Jpn. J. Appl. Phys. Vol. 42 (2003) pp. 7259–7264
Part 1, No. 12, December 2003

Slurry Utilization Efficiency Studies in Chemical Mechanical Planarization

Ara PHILIPOSSIAN and Erin MITCHELL

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

(Received March 4, 2003; accepted July 15, 2003; published December 10, 2003)

The residence time distribution of slurry in the pad-wafer interface was experimentally determined and used to calculate the slurry utilization efficiency (η) of the chemical mechanical planarization (CMP) process. Slurry utilization efficiency represents the percentage of slurry that actually participates in the polish by entering the region bounded between the wafer and the pad. Results show that η ranges from 2 to 22%, depending on operating conditions such as applied wafer pressure, relative pad wafer velocity, slurry flow rate and pad surface texture (i.e. type of pad grooving).
[DOI: 10.1143/JJAP.42.7259]

KEYWORDS: chemical mechanical planarization (CMP) slurry utilization efficiency mean residence time coefficient of friction (COF) slurry flow rate pad surface texture

1. Introduction

Literature states the importance of the presence of slurry in the wafer-pad region,[1–3] however, little has been published on the subject. For certain processes, slurry cost represents almost half of the total cost of ownership of the entire chemical mechanical planarization (CMP) operation,[4] and the notion that a significant portion of this slurry may be wasted lends importance to understanding and optimizing this process parameter. Residence Time Distribution (RTD) analysis for various slurry flow rates has shown that not all of the slurry introduced onto the pad actually participates in the polish process.[5] By employing classical reactor design theory and RTD techniques, this study provides insight into optimum consumables and operating parameters for maximizing slurry utilization efficiency and minimizing slurry use.

2. Apparatus

A fully automated scaled version of a Speedfam-IPEC 472 polisher was constructed for this study as shown in Fig. 1.

Table I shows the appropriate scaling factors for each parameter as well as the numerical comparison of the scaled polisher's typical values with that of the Speedfam-IPEC 472.

The slurry's kinematic viscosity and the fluid film thickness between the pad and the wafer were assumed to be the same for the two systems, therefore, Reynolds number was used to scale the platen and wafer speeds (i.e. the relative pad-wafer velocity in the scaled model was matched to that of the full-scale model). Furthermore, the scaled polisher's platen-to-wafer diameter ratio and slurry flow rate normalized by the platen area corresponded to the values for the full-scale polisher. For wafer pressure, ranges typically found on an industrial polisher were applied to the scaled apparatus. The main body of the apparatus consists of a Struers Rotopol-35 polisher with a variable speed platen. A drill press modified to include a DC motor for variable head rotation, provides motion and down force to the wafer during polish. To apply a given load to the wafer, a traverse with a weighted carriage, is mounted atop the drill press. The slurry is injected onto the center of the pad. A conditioning system is mounted to the polisher, which can be used either in-situ or ex-situ. The polisher resides on a sliding table that measures the frictional force between the wafer and the pad.

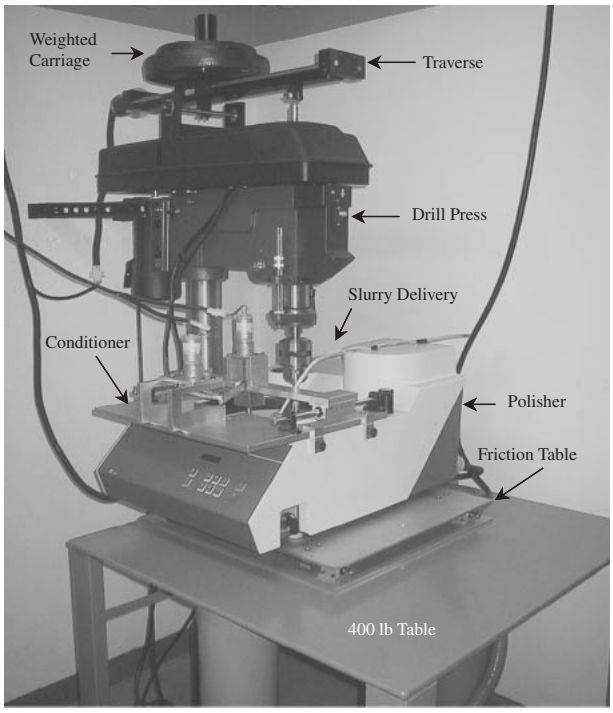

Fig. 1. Scaled polisher showing the diamond disc conditioner, the drill press and the traverse assembly.

Table I. Scaling factors used in constructing the scaled polisher.

Parameter	Scaling Factor	Speedfam-IPEC 472	Scaled Polisher
Down Pressure	1	4 psi	4 psi
Platen Speed	Reynolds Number	Relative pad-wafer velocity of 0.5 m per second (30 rpm)	Relative pad-wafer velocity of 0.5 m per second (55 rpm)
Platen Diameter/ Wafer Diameter	D_{platen}/D_{wafer}	51 cm/20 cm	31 cm/10 cm
Slurry Flow Rate	Platen Surface Area	215 cc per minute	80 cc per minute

All polishing parameters are computer controlled and monitored. In addition, the computer synchronizes the friction table to the polishing process so that real-time friction data—crucial for determining RTD—can be ob-

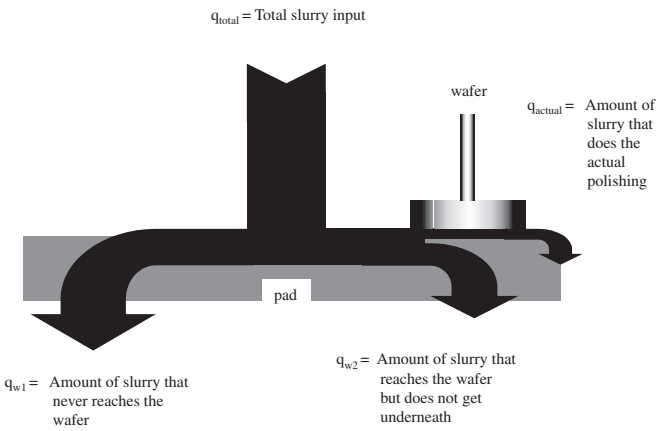

Fig. 2. Various paths the fluid can take during CMP.

tained during polishing. For any given run, the coefficient of friction (COF) is determined by dividing the shear force divided by the normal force applied to the wafer. Philipossian *et al.* have described the experimental apparatus in detail elsewhere.[5)]

3. Theory

The assumption that only a small percentage of slurry introduced to the system actually reaches the wafer and participates in the polish is illustrated in Fig. 2.[5)] Based on this, slurry utilization efficiency (η) can be expressed as:

$$\eta = \frac{q_{\text{actual}}}{q_{\text{total}}} \times 100\% \tag{1}$$

In eq. (1), q_{actual} is the slurry flow rate beneath the wafer, and q_{total} is the total flow rate of the slurry introduced onto the pad. While q_{total} is a known entity for a given process, the value of q_{actual} must be found experimentally. For a closed flow system, the flow rate in the reactor is equal to the ratio of reactor volume (V) to the mean residence time (τ) of the fluid:

$$q_{\text{actual}} = \frac{V}{\tau} \tag{2}$$

The above equation is based on the premise that there be no dead-zones or stagnation regions within the reactor (i.e. the entire reactor is essentially an active flow system). Calculations performed by Philipossian *et al.* have shown the Stokes number to be around 0.0001 for ten percent silica slurry comprised of 100-nm particles.[6)] This indicates that the experimentally determined values of mean residence time for the slurry are representative of both the fluid and the suspended solids. Moreover, numerical Euler-Euler granular multi-phase slurry flow simulations[6)] have shown that, for grooves as deep as 0.6 mm (which are twice as deep as the grooves studied here), the abrasive particle distribution non-uniformity across the entire depth of the groove is less than 9 percent for particles ranging from 100-nm to 1 micron in diameter. This indicates that more than 91% of the slurry is effectively getting purged in and out of the pad groove regions and is actively contributing to the overall fluid dynamics of the process, thus justifying the application of eq. (2).

Based on the above, in the case of CMP, the reactor volume can therefore be approximated by the slurry holding capacity of the pad as it contacts the wafer. The latter can be defined as:

$$V = \pi r^2 \times \delta_{\text{eff}} \tag{3}$$

In eq. (3), r represents wafer radius and δ_{eff} is the effective slurry thickness height based on a model proposed by Philipossian and Olsen:[7)]

$$\delta_{\text{eff}} = Ra \times \alpha + (1 - \alpha)\delta_{\text{groove}} \tag{4}$$

In eq. (4), Ra represents average pad surface roughness as measured by a stylus profilometer. The parameter α is the ratio of the surface area of 'up-features' of a pad to the surface area of a flat pad, and δ_{groove} refers to the depth of the perforations or the grooves of the pad.

Generally speaking, to begin to understand and predict the fluid dynamics in a reactor, it is useful to determine how long elements of a fluid reside in the vessel. Introducing a stimulus in the form of a tracer, and measuring the response can be used to determine the distribution of residence times of the flowing fluid. Elements of fluid take different routes through the vessel so they require different lengths of time to exit the reactor. The distribution of these times for the fluid leaving the vessel is referred to as the residence time distribution (RTD). This represents the exit ages of individual fluid elements.

The RTD technique involves an abrupt introduction of a tracer into a reactor operating at steady-state. The tracer may be introduced into the system in the form of a step of fluid. Constructing an E-curve and calculating the value for τ requires a series of steps beginning with the introduction of a tracer input. With no tracer initially present, a step input of tracer of concentration (C_0) is imposed on the fluid stream entering the vessel. A time record of the concentration of tracer in the exit stream (C) is measured. The results are normalized so that time is zero ($t_0 = 0$) at the moment the tracer is introduced. The concentration of tracer in the exit stream is compiled as C/C_0, such that the resulting curve rises from 0 to 1 over time. This plot is known as the F-curve. The E-curve is related to the F-curve via the following equation:[8)]

$$\frac{dF}{dt} = Edt \tag{5}$$

where Edt represents the fraction of the fluid leaving the reactor with an age between t and $t+\Delta t$. The mean of the E-curve, which represents τ, is found from:

$$\tau = \int_0^\infty t \times Edt \tag{6}$$

In the above equation t represents elapsed time. For a given process, once the E-curve is known, the corresponding slurry utilization efficiency can be calculated by rearranging the above equations to obtain eq. (7) below:

$$\eta = \left[\frac{100\pi r^2[\alpha Ra + (1 - \alpha)\delta_{\text{groove}}]}{q_{\text{total}}}\right] \times \left[\frac{1}{\int_0^t t \times Edt}\right] \tag{7}$$

Table II. Experimental conditions.

Parameter	Phase I	Phase II
Wafer Pressure (PSI)	2 and 4	2, 4 and 6
Pad-wafer velocity (m/s)	0.31 and 0.62	0.31, 0.62, 0.93 and 1.24
Slurry flow rate (cc/min)	40, 60 and 80	60
'Initial' slurry type	Fujimi PL-4217 fumed silica slurry at 25% solids	Water
'Final' slurry type	Fujimi PL-4217 fumed silica slurry at 2.5% solids	DuPont Air Products Syton OXK colloidal silica slurry at 20% solids
Pad type	Freudenberg FX-9 Perforated	Rodel IC-1000 K-Grooved
Wafer type	100-mm Bare Silicon	100-mm Thermal Silicon Dioxide

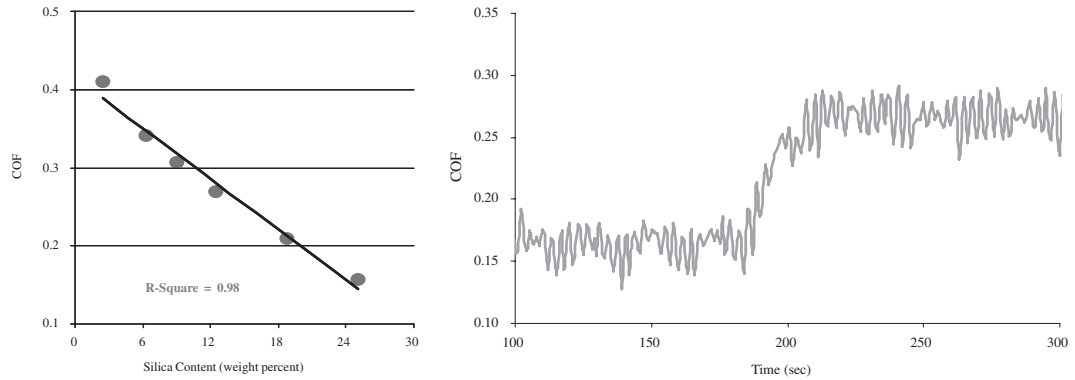

Fig. 3. Relationship between COF and silica concentration (a) and the resulting F-curve (b).

4. Experimental Technique

In this study, a new technique is developed to measure the E-curve in the wafer-pad interface as a function of key operating parameters. The technique relies on the change in shape of the F-curve to an instantaneous disturbance within the system caused by the sudden replacement of water with slurry.[9] The RTD method takes 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 can then be differentiated [see eq. (5)] to obtain the E-curve.

Prior to all experiments, the pad was subjected to 30 min of ex-situ conditioning followed by a 5-min break-in with a dummy wafer. The slurry used for conditioning and break-in was the same one used for performing the RTD experiments. For each experiment, the system was first allowed to reach steady state using slurry with a particular abrasive concentration. The slurry was then switched instantaneously to one with a significantly different abrasive concentration, causing the old slurry to be replaced and allowing the system to reach a new steady state. Throughout this entire process, COF was measured 1000 times a second. By normalizing the COF versus time response curve, an F-curve was produced, from which the E-curve was then constructed and used to measure the mean residence time.

Two different sets of experiments were conducted. In this study, they are referred to as Phase I and Phase II. Table II summarizes the experimental conditions. The fumed silica slurry used for Phase I experiments yielded a linear relationship between COF and silica abrasives concentration as seen in Fig. 3(a). This enabled the F-curve to be constructed directly from the COF response, a sample of which is shown in Fig. 3(b). The corresponding E-curve is depicted in Fig. 4.

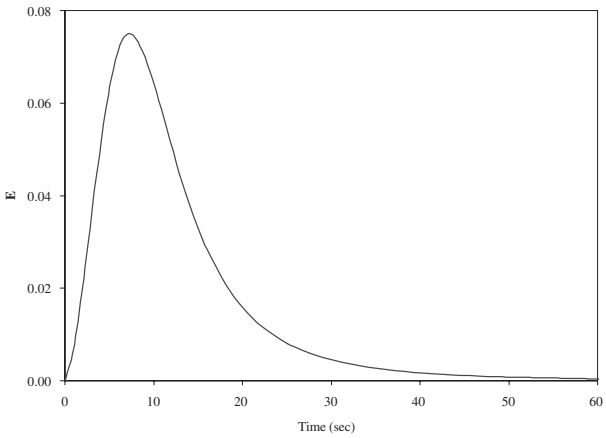

Fig. 4. E-Curve corresponding to data sets shown in Fig. 3.

The colloidal silica slurry employed in Phase II gave a non-linear relationship between COF and silica abrasives concentration [Fig. 5(a)]. The COF response to the fluid input highlights the nonlinear relationship between COF and solids content as seen in Fig. 5(b). The dashed line in Fig. 5(b) represents the point in time where the 'initial' slurry began to be replaced by the 'final' slurry. The above-mentioned non-linear relationship necessitated the construction of a concentration response in order to produce an F-

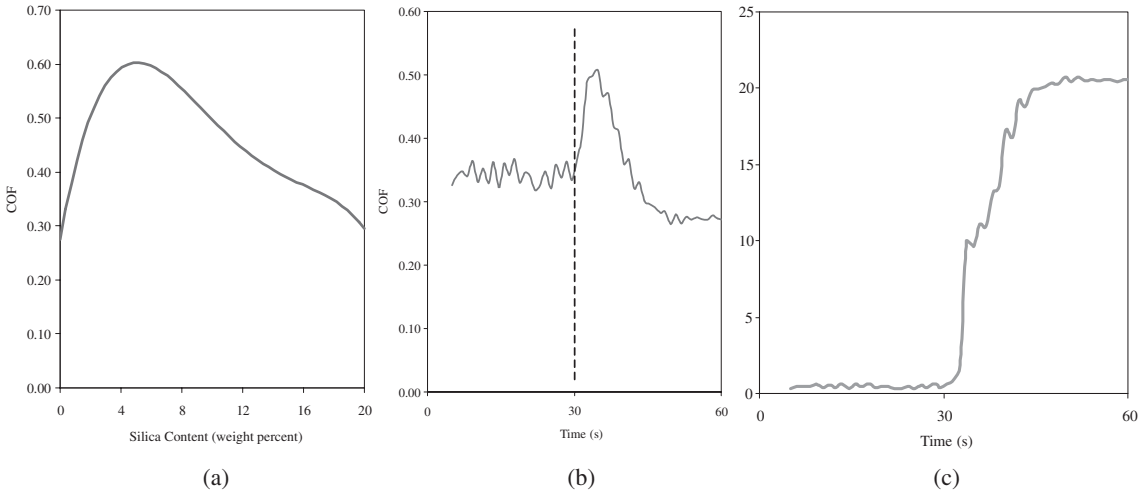

Fig. 5. COF as a function of Syton-OXK colloidal silica slurry solids concentration (a); actual response (COF versus time) corresponding to 20% weight percent Syton-OXK slurry displacing water during a polish process (b); corresponding silica concentration response (c).

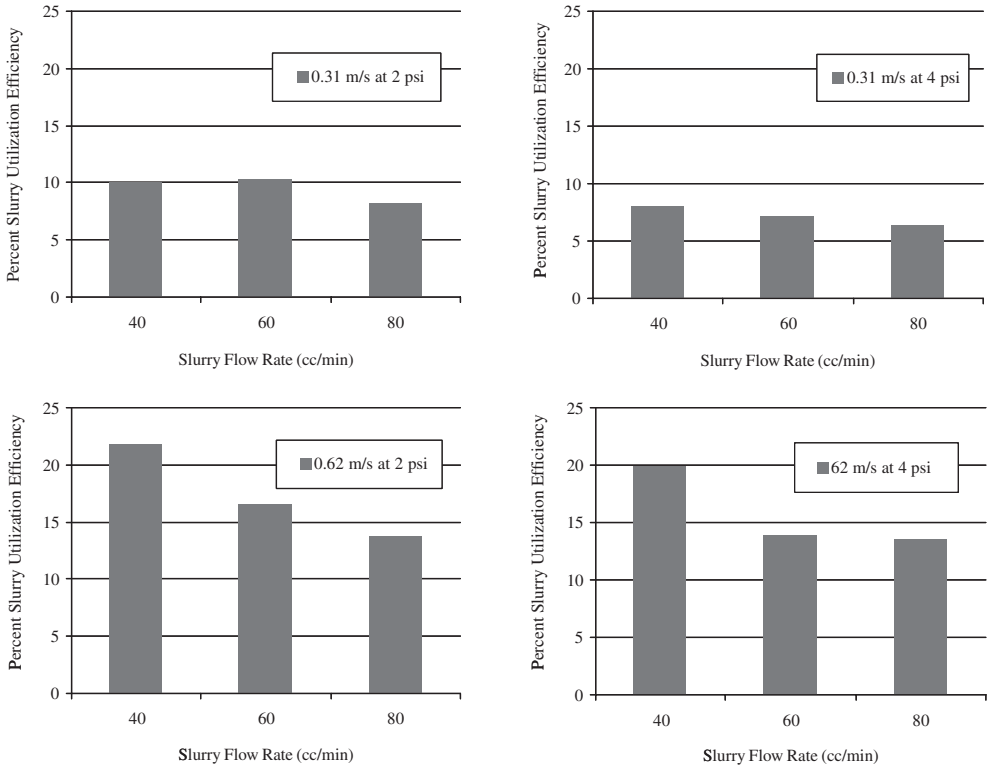

Fig. 6. Slurry utilization efficiency as a function of pressure, velocity and slurry flow rate for Phase I studies.

curve. This was achieved by solving for solids concentration from each measured COF value on the COF-versus-time plot using the trend shown in Fig. 5(a). This process yielded a plot of slurry abrasive concentration as a function of silica concentration as seen in Fig. 5(c).

The reasons behind the linear and non-linear relationships between COF and abrasive concentration for the fumed and colloidal slurries, respectively, are not well understood. This, however, does not compromise the integrity of the data since the method with which τ 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.

5. Results and Discussion

Figure 6 shows the slurry utilization efficiency as a function of flow rate, wafer pressure and relative pad-wafer velocity for Phase I experiments. Results corresponding to slurry utilization efficiency as a function of wafer pressure and relative pad-wafer velocity for Phase II experiments are shown in Fig. 7. In general, the results indicate that η ranges from 2 to 22 percent depending on the choice of process settings and consumables. The total error in calculating η in each plot is approximately $\pm 10\%$.

Figure 6 indicates a general decrease in slurry utilization

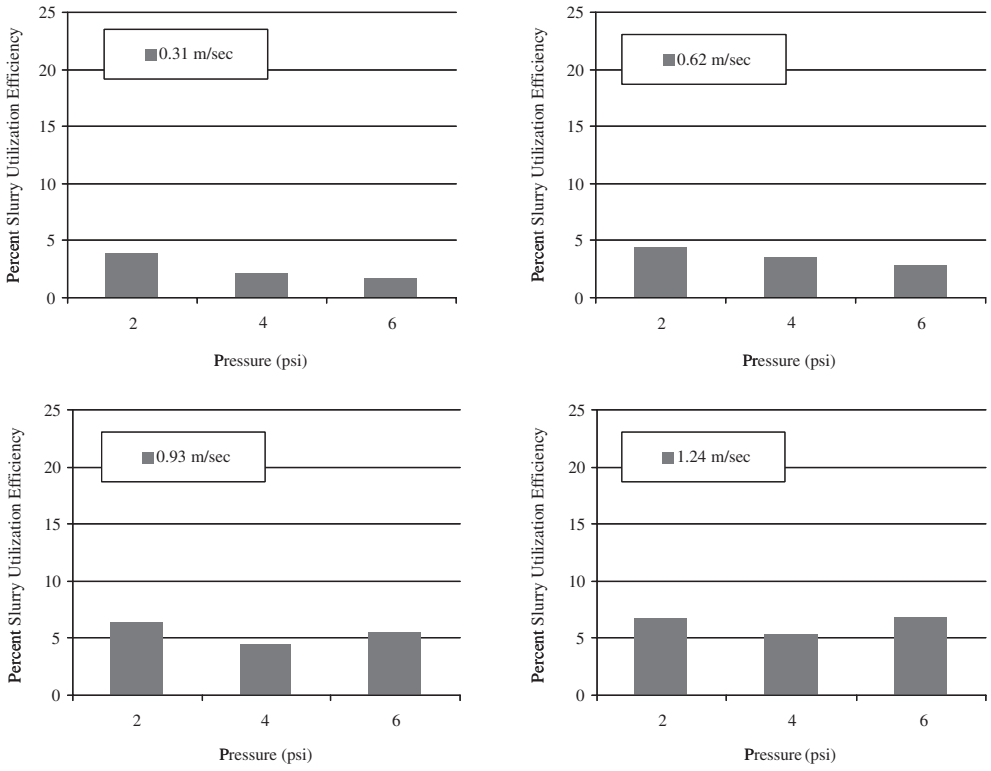

Fig. 7. Slurry utilization efficiency as a function of pressure, velocity and slurry flow rate for Phase II studies.

efficiency with increasing flow rate. On the average, a 2-fold reduction in slurry flow rates results in a roughly 25% increase in η. This trend becomes more pronounced at the higher pad-wafer speed. While the cause (or causes) for these findings are still under investigation, the observed trend highlights the dual environmental and cost-of-ownership (COO) benefits gained by reducing flow rate. It should be noted that the above trend is in stark contrast to the common (but unsubstantiated) belief that increasing flow rate increases slurry utilization, and underscores the unfortunate current situation regarding high the COO associated with various slurries.

Figures 6 and 7 both show significant increases in efficiency with increasing relative pad-wafer velocity. This is due to the fact that a faster moving pad helps carry a greater percentage of the slurry under the wafer since the fluid elements close to the surface of the pad are traveling at the same velocity as the pad. In the case of Phase I experiment, a 2-fold increase in velocity causes a similar increase in η. However in the case of Phase II experiments, a 4-fold increase in velocity is required in order to increase utilization by a factor of 2. This difference is most likely due to the differences in the slurry retention capabilities of Perforated pads compared to K-Grooved pads. This will be discussed in greater detail later in this section.

Regarding the effect of pressure, there is a slight increase in slurry utilization efficiency with decreasing pressure in all cases except for high relative pad-wafer velocities. It is well known that low pressures cause an increase in the slurry film thickness between the wafer and the pad, thus reducing the tortuosity of the system and allowing slightly higher amounts of slurry to penetrate the wafer-pad interface.[10,11] The fact that the effect of pressure on slurry utilization efficiency is not pronounced is consistent with the slight increases in the fluid film thickness (i.e. 10 to 20 microns) with decreasing pressure.

The most dramatic difference in the data contained in Figs. 6 and 7 is the fact that, on the average, the slurry utilization efficiencies reported for Phase I are far greater than those reported in Phase II. For the four conditions corresponding to both sets of experiments (i.e. 0.31 and 0.62 meters per second, and 2 and 4 PSI), the average slurry utilization efficiency for the FX-9 Perforated pad is 11.75% (Phase I), while the value for IC-1000 K-Grooved pad is 3.75% (Phase II). This corresponds to a 3.2 times difference in the values of slurry utilization efficiency.

In fact, the slurry holding capacity in the space bounded between the wafer and the grooves in the Perforated pad is 1.8×10^{-6} cubic centimeters compared to 5.8×10^{-7} cubic centimeters for the for the K-grooved pad. The ratio of the two volumes is 3.1. This is quite close to the ratio of their respective slurry utilization efficiencies, which indicates that the greater the amount of slurry trapped in the grooves, the less the amount of waste from the slurry flowing off the side of the pad before reaching the wafer. This relationship can be beneficial in designing pads and surface textures with optimal slurry retention and therefore slurry utilization attributes.

6. Conclusion

As stated previously, slurry represents nearly 50% of the total CMP cost of ownership. The cost of slurry is high compared to other factors involved in operating the CMP module, and therefore the motivation for reducing slurry use is great. Using the residence time distribution technique, this study has shown that slurry utilization efficiency can be

measured experimentally using classical reactor design principles. Results show that η ranges from 2 to 22%, depending on operating conditions such as applied wafer pressure, relative pad wafer velocity, slurry flow rate and pad surface texture (i.e. type of pad grooving). Careful selection of operating conditions, including minimizing flow rate, and opting for pad surface textures which increase slurry retention are relatively simple options which have been explained and scientifically rationalized in this study.

Acknowledgements

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

1) D. Stein, D. Hetherington, M. Dugger and T. Stout: J. Electron. Mater. **25** (1996) 1623.
2) K. Kim, S. Moon and H. Jeong: Electrochem. Soc. Proc. **99** (1999) 402.
3) A. Sikder, F. Giglio, J. Wood, A. Kumar and M. Anthony: J. Electron. Mater. **30** (2001) 1520.
4) K. Holland: Micro **20** (2002) 52.
5) A. Philipossian and E. Mitchell: Micro **20** (2002) 85.
6) A. Philipossian, L. Racz, J. Lu and C. Rogers: Technology Requirements for Advanced CMP Applications. Proc. CMP Tech. Meeting of Jpn. Soc. Precision Eng., San Francisco, CA, 2000.
7) A. Philipossian and S. Olsen: Proc. 8th Int. CMP-MIC Conf., Marina Del Rey, CA, 2003, p. 25.
8) O. Levenspiel: *Chemical Reaction Engineering* (John Wiley & Sons, New York, 1972).
9) G. Froment and K. Bischoff: *Chemical Reactor Analysis and Design* (John Wiley & Sons, New York, 1979).
10) J. Lu: M. S. Thesis, Tufts University, Massachusetts USA, 2001.
11) E. Mitchell: M. S. Thesis, University of Arizona, Arizona, USA, 2002.