Dispersion Number Studies in CMP of Interlayer Dielectric Films

Ara Philibossian* and Erin Mitchell

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

Determining the factors that cause flow nonidealities during chemical mechanical planarization (CMP) is critical for controlling and optimizing the process. This study explores aspects of the fluid dynamics of CMP on interlayer dielectric films. The residence time distribution of slurry under the wafer was experimentally determined and used to calculate the dispersion number of the fluid in the wafer-pad region based on a dispersion model for nonideal reactors. Furthermore, lubrication theory was employed to explain trends in flow behavior as operating conditions were varied. The results indicate that at low wafer pressure and high relative pad-wafer velocity, the slurry flow exhibits nearly ideal plug flow behavior. As pressure increases and velocity decreases, flow begins to deviate from ideal behavior and the slurry becomes increasingly more mixed beneath the wafer. These phenomena were found to be the result of variable slurry film thickness between the pad and the wafer, as measured by changes in the coefficient of friction between the pad and the wafer.

© 2003 The Electrochemical Society. [DOI: 10.1149/1.1627352] All rights reserved.


Understanding the fluid dynamics, specifically the extent of flow nonidealities during chemical mechanical polishing (CMP) is essential for maintaining a stable and predictable process especially as film thicknesses continue to decrease and wafer sizes continue to increase in accordance with the International Technology Roadmap for Semiconductors (ITRS). Previous research indicates that slurry distribution under the wafer significantly influences the process. Despite a wide consensus that slurry transport is a critical parameter, limited experimental research has been performed on slurry flow under the wafer. The purpose of this study is to develop a greater understanding of fluid dynamics using lubrication theory and the residence time distribution (RTD) technique to determine the vessel dispersion number at a variety of operating parameters. By employing classical RTD techniques coupled with well-defined vessel dispersion models for nonideal reactors, this study quantifies the extent of axial dispersion (and therefore flow nonideality) as functions of slurry flow rate, wafer pressure, and pad-wafer velocity.

The dispersion model is used to describe nonideal reactors, where the axial dispersion is superimposed on the plug flow of a fluid. The dispersion number is defined as

\[
\text{Dispersion number} = \frac{D}{v \times L}
\]  

In the above equation, \( D \) is the axial dispersion coefficient, \( v \) is the average superficial velocity, and \( L \) is a characteristic length. The dispersion number is a measure of the ratio of the rate of transport by diffusion and the rate of transport by convection. The deviation from plug flow in the reactor is apparent from the magnitude of the dispersion number. As the dispersion number approaches zero, dispersion is considered to be negligible, and the behavior of the system is said to approach that of a plug flow reactor (PFR). As the dispersion number approaches infinity, there is a large degree of dispersion, and the behavior of the reactor approaches perfectly mixed flow as in a continuously stirred tank reactor (CSTR).

The degree of dispersion is a critical factor in CMP because it indicates the degree of nonideality in the flow system. It is desirable to operate under conditions that create the most ideal system (whether a PFR or CSTR) since this creates a greater degree of predictability and control. Because it is virtually impossible to attain an ideal CSTR (dispersion number \( \rightarrow \infty \)), it is much more realistic to strive for ideal PFR behavior. The overall goal of this study is to investigate the fluid dynamics and flow nonidealities of the ILD CMP process to optimize operating conditions such that predictable and robust planarization processes can be developed.

Apparatus

A scaled version of a Speedfam-IPEC 472 polisher was constructed for this study. 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, the 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). 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 (Fig. 1). The polisher was designed after the system developed at Tufts University through an elaborate technology exchange program with the University of Arizona. The main body of the apparatus consists of a Steuers 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 the polish. To apply a given load to the wafer, a traverse with a weighted carriage is mounted atop the drill press. Slurry is injected onto the center of the pad. A conditioning system is mounted on the polisher, which can be used either in situ or ex situ.

The main differences between this polisher and the one developed at Tufts University are as follows: the polisher at Tufts University employs an optical technique known as dual emission laser induced fluorescence (DELIF) to quantify the extent of flow nonidealities associated with the CMP process. The technique is based on two high-resolution, spatially aligned cameras that measure the fluorescence of the slurry beneath the wafer. Since the cameras must be able to see under the wafer, the traditional wafer carrier, and retaining ring assembly, along with the backing film and the actual wafer, are replaced with a thick transparent glass disk having a nominal diameter of 75 mm. In the case of the University of Arizona polisher, flow nonidealities are quantified by measuring the coefficient of friction in the wafer-pad region in response to the introduction of slurries containing varying amounts of silica abrasives (see the section on Theory and Experimental Techniques). To measure the frictional force between the pad and the wafer during polishing, a sliding table is placed beneath the polisher. The table consists of a bottom plate and a top plate that the polisher is set upon. As the wafer (in this case having a nominal diameter of 100 mm) and the pad are engaged, the top plate slides with respect to the bottom plate.
in only one direction due to friction between the pad and wafer. The degree of sliding can then be quantified by coupling the two plates to a load cell. The load cell is attached to a strain gauge amplifier, which sends a voltage to a data acquisition board. The apparatus is calibrated to report the force associated with a particular voltage reading.

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 the RTD, can be obtained 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. The apparatus is described in greater detail elsewhere by Philipossian et al.

Theory and Experimental Technique

In the case of CMP, the reactor can be assumed to act as a closed vessel, since there is a change in flow pattern at the boundaries and the movement of fluid is assumed to be plug flow outside of the reactor. For closed vessels, the variance is a function of the dispersion number as follows:

$$\sigma^2 = \frac{2D^2}{vL} - \left(\frac{D}{vL}\right)^2 (1 - e^{-vL/d})$$

[2]

In this equation $\sigma^2$ is the variance of the residence time distribution curve (E curve) and $\tau$ is the mean residence time (MRT) of the fluid in the reactor. These parameters can be calculated directly from the E curve using the following relationships:

$$\sigma^2 = \int_0^\infty (t - \tau)^2 e^{-\tau} d\tau$$

[3]

$$\tau = \int_0^\infty t e^{-\tau} d\tau$$

[4]

A new technique, based on the measurement of COF, was developed 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. The technique relies 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). 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 to obtain the E curve.

Prior to all experiments, the polishing pad was subjected to a 30 min ex situ conditioning process followed by a 5 min break-in with a dummy wafer using the same slurry as for the experiments. For each experiment the system was first allowed to reach steady state using a slurry with a particular abrasive concentration, causing the old slurry to be replaced and allowing the system to reach a new steady state. Throughout this entire process, the COF of the system was measured once a second, allowing the system to reach a new steady state. These tests were carried out using two different sets of consumables and conditions, which will be referred to as phase I and phase II. The experimental conditions for the two phases are summarized in Tables II and III. The fumed silica slurry used for phase I experi-

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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Scaling factor</th>
<th>Speedfam-IPEC 472</th>
<th>Scaled polisher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Down pressure</td>
<td>1</td>
<td>4 psi</td>
<td>4 psi</td>
</tr>
<tr>
<td>Platen speed</td>
<td>Reynolds number</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5 m/s (30 rpm)</td>
<td>0.5 m/s (55 rpm)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51 cm/15 cm</td>
<td>31 cm²·10 cm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>175 cm³/min</td>
<td>65 cm³/min</td>
</tr>
<tr>
<td>Platen diameter/wafer diameter</td>
<td>$D_{platen}/D_{wafer}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slurry flow rate</td>
<td>Platen surface area</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Figure 1. Image of scaled polisher showing the diamond disk conditioner, the drill press, and the traverse assembly.](image-url)
Table III. Phase II experimental conditions.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wafer pressure</td>
<td>2, 4, and 6 psi</td>
</tr>
<tr>
<td>Platen and wafer speed</td>
<td>40, 80, 120, and 160 RPM</td>
</tr>
<tr>
<td>Relative pad-wafer velocity</td>
<td>0.31, 0.62, 0.93, and 1.24 m/s</td>
</tr>
<tr>
<td>Conditioner rotation speed</td>
<td>30 rpm</td>
</tr>
<tr>
<td>Conditioner oscillation speed</td>
<td>20 rpm</td>
</tr>
<tr>
<td>Slurry flow rate</td>
<td>60 cm$^3$/min</td>
</tr>
<tr>
<td>Initial slurry type</td>
<td>Ultrapure water</td>
</tr>
<tr>
<td>Final slurry type</td>
<td>du Pont Air Products Syton OXK</td>
</tr>
<tr>
<td>Pad type</td>
<td>Rodel IC-1000 K-Groove</td>
</tr>
<tr>
<td>Wafer</td>
<td>4 in. thermally grown silicon dioxide</td>
</tr>
</tbody>
</table>

There are advantages and disadvantages to using the COF method described above to determine MRT and dispersion number as opposed to the previously mentioned DELIF technique developed at Tufts University. The COF method does not require any additional equipment other than the sliding friction table to measure COF. The DELIF method, on the other hand, requires two spatially aligned high-resolution digital cameras and the accompanying acquisition hardware and software. This apparatus is relatively expensive and considerably more difficult to operate since the cameras must be focused and the images matched to within one pixel. In addition, the DELIF technique requires the room to be completely dark to avoid any interference with the fluorescence emitted from the dyes. The greatest advantage of the COF method is that actual polishing conditions are mimicked during MRT and dispersion number measurements. An actual silicon or silicon dioxide wafer is held in a carrier complete with a retaining ring and backing film to hold the wafer in place, as in an actual industrial ILD polishing process. The DELIF method, on the other hand, does not allow for use of a silicon wafer, carrier, or retaining ring. Since the camera must be able to acquire images from under the wafer, a transparent glass disk gimbaled directly to a central rod is used instead of the wafer assembly. During actual polishing, a retaining ring surrounds the wafer, creating a discontinuity in the area within which the slurry flows. In addition, glass wafers tend to start off having either a concave or a convex shape, with the shape changing during polishing. Such changes in shape can lead to hysteresis, which in turn can cause uneven load distribution, possibly altering the entrainment of slurry. DELIF only allows analysis of a small window under the wafer, which may or may not provide a good representation of the entire wafer. The COF data since the method with which MRT and dispersion number are 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. Figure 6, showing the COF response to the fluid inputs, emphasizes the nonlinear relationship between the COF and solids content. This relationship necessitated the construction of a concentration response in order to produce an F curve. This was achieved by solving for solid concentrations from each COF value on the COF vs. time plot using the polynomial fit previously found to correlate the two values. This process yielded a plot of slurry abrasive concentration as a function of time. An example of a plot of slurry solids concentration over time, obtained from combining the data found in Fig. 5 and Fig. 6, is shown in Fig. 7. From the concentration response, F and E curves were calculated as for phase I experiments. These procedures are explained in further detail elsewhere.

The dispersion number was calculated from the E curve for all combinations of conditions listed in Tables II and III using Eq. 2 through 4.

Figure 2. COF as a function of PL-4217 silica concentration.

Figure 3. Actual response (COF vs. time) corresponding to 2.5 wt % slurry displacing 25 wt % slurry (PL-4217) during a polish process.
method measures an average under the entire wafer. The DELIF method also requires that the pads be dyed black so as not to interfere with fluorescence intensity. Dyed pads may have different mechanical properties compared to conventional pads. The COF method uses conventional undyed pads. Given the above information, it is possible that in the DELIF technique, some or all of the deviations from actual polishing conditions can affect the determination of the MRT and dispersion number. Therefore, the COF method is a valuable alternative, since the MRT is measured under the same conditions as the actual process.

As for the disadvantages, the COF technique results in an F curve with sustained oscillatory behavior due to inherent stick-slip phenomena (as seen in Fig. 3) while the DELIF method produces a higher signal-to-noise ratio. The DELIF technique uses one type of slurry tagged with different dyes, while the COF technique requires a change in slurry solids concentration to produce the desired tracer output. As a result, the calculated MRT corresponds to the mean time to replace the old slurry with the new slurry in the reactor, not the mean time during which a fluid element remains in the reactor. However, the MRT induced by a change in fluid is an adequate

Figure 4. Residence time distribution (E curve) corresponding to COF response.

Figure 5. Polynomial fit of experimental data of the COF as a function of Syton OX-K silica concentration.

Figure 6. Actual response (COF vs. time) corresponding to 20 wt % Syton-OXK slurry displacing water during a polish process.

Figure 7. Concentration response (silica content vs. time) corresponding to 20 wt % Syton-OXK slurry displacing water during a polish process.
approximation to the MRT under steady-state conditions, because properties such as density and viscosity are similar in both fluids. Based on the reasons cited above, the advantages of the COF method for determining MRT are considered to outweigh its disadvantages, thus warranting a comprehensive study of the technique.

Results and Discussion

The fact that total variability in measuring the MRT is about ±10%, coupled with the relationship between the MRT and dispersion number defined by Eq. 2, the uncertainty in the calculated value of the dispersion number may be as high as ±40%. The dispersion number for phase I experiments is depicted in Fig. 8. The slurry flow rate has a minor impact on the dispersion number under the range of conditions tested. Furthermore, there is negligible dispersion, and the reactor is near plug flow at all conditions, except in the case of high pressure and low velocity. Under these conditions the reactor deviates somewhat from ideal plug flow behavior, but since the dispersion number is small compared to infinity, the system is still far from a perfectly mixed system.

To better understand these trends, the dispersion number was plotted as a function of the average COF corresponding to the operating conditions used for each run. The average COF was found by averaging the COF of the system when exposed to the initial slurry input and the COF reached during the final slurry input. The dispersion number as a function of average COF is illustrated in Fig. 9, which shows that at conditions of low COF, the dispersion number is close to zero, thus indicating plug flow. At a certain value of COF (between 0.25 and 0.30), the dispersion number begins to rise and the reactor begins to deviate from plug flow.

Phase II experiments (Fig. 10) give the dispersion number as a function of wafer pressure at four different relative pad-wafer velocities. The data shows a similar trend as for phase I experiments. At 2 psi, the dispersion number is small regardless of the pad-wafer relative velocity. At higher pressures (4 and 6 psi), the dispersion number is low at high velocities (0.93 and 1.24 m/s), but increases significantly at lower velocities (0.31 and 0.62 m/s). Data taken at 6 psi displays a more dramatic rise than at 4 psi when velocity is decreased, indicating that the most axial dispersion occurs at high pressures and low pad-wafer relative velocities. The dispersion number as a function of the average COF is shown in Fig. 11. In this case, the average COF is the mean of the COF at 0, 3, 6, 12, and 20% abrasive concentrations. These trends are consistent with phase I data. The dispersion number is nearly zero at a low COF and then at a certain threshold COF, it begins to rise with the COF. The threshold COF values are similar for both phase I and phase II indicating that as a first approximation, the pad and slurry type have minimal effects on the dispersion number.

As previously noted, the total variability in the dispersion number is approximately ±40%. However, the specific value of the dispersion number is not as important for the work presented here, compared to the relative magnitudes of the dispersion number among the various conditions analyzed. Results from both phase I and phase II indicate that the dispersion number is small when either the pressure is low or the velocity is high. This indicates that the CMP reactor approaches ideal plug flow behavior at all conditions, except at the combination of high pressure and low velocity. Furthermore, at low COF, the dispersion number is not a function of COF as the value remains constant and near zero. Under conditions of high COF, however, the dispersion number becomes an increasing function with COF.
The above phenomenon is explained by recognizing that as the COF increases, slurry film thickness decreases. Therefore, at a low COF a mostly continuous fluid layer exists between the pad and the wafer. Under these conditions, convective transport is much greater than diffusive transport since there are few obstacles to disrupt flow. Slight decreases in the thickness of the fluid layer induced by a change in operating conditions, i.e., wafer pressure or relative pad-wafer velocity, cause the COF to increase but do not substantially affect the fluid flow. As the fluid thickness continues to decrease, at some point the fluid transport mechanism begins to change as the fluid layer becomes less continuous. At some point, as indicated by the threshold COF, the discontinuity of the fluid layer creates a reduction in convective flow due to an increased 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 caused by pad asperities acting as baffles against the flow. The threshold COF indicates the point where the film thickness is small enough such that diffusive transport begins to dominate over the convective transport.

COF and slurry film thickness are both system outputs, meaning they cannot be controlled directly. Instead, these polishing attributes have been correlated to operating conditions such as slurry flow rate, wafer pressure, and relative pad wafer velocity. Decreasing pressure and increasing velocity were both found to increase slurry film thickness and decrease COF. For this reason, these parameters also had an impact on the dispersion number of the system. Low pressure and high velocity caused the film thickness to be high, COF to be low, and therefore the dispersion number to be near zero (ideal plug flow). At high pressure and low velocity, the opposite was true and the dispersion number deviated from ideal behavior.

The range of dispersion numbers reported in this study are more or less similar to those described by Coppeta using the DELIF technique. Both methods show that pad grooving does not impact the dispersion number, and that increasing the relative pad-wafer velocity and slurry flow rate both act to reduce axial mixing. The main discrepancy between these results and Coppeta’s is that Coppeta did not find wafer pressure to have any significant effect on the dispersion number, whereas, the data reported in this study indicated that the dispersion number is highly dependent on wafer pressure. This disagreement most likely stemmed from the fact that Coppeta’s experiments were carried out with 75 mm concave and convex continuous glass plates without the use of a carrier film and a retaining ring. Such systems have been shown by Scarfo to have different properties in terms of angle of attack, slurry entrainment, as well as local and global pressure nonuniformities, including suction.

The dispersion number is an important parameter for understanding the fluid dynamics of CMP. It indicates the degree of mixing that occurs within the reactor volume. Ideal flow is preferred for any system because it results in a more predictable process. An ideal plug-flow reactor is a more realistic goal to strive for because the ideal CSTR is unattainable, as infinite mixing can never be reached. The PFR is also favored because it causes unwanted polish products to be swept out of the reactor more quickly than the CSTR, which allows products to be mixed with new reactants and remain under the wafer for several rotations. Furthermore, it is of great importance to operate at conditions that yield COF values that are well below the threshold value. At the threshold, the dispersion number shifts from a steady value near zero to an increasing function with COF. For the conditions examined in this study, the threshold COF appears to be between 0.25 and 0.30. From a process control perspective, operating near or above
this threshold COF could potentially cause large deviations in the process when operating conditions fluctuate only slightly.

Conclusions

The dispersion number of the fluid entrained beneath the wafer during CMP was calculated using the residence time distribution technique. Results indicate that the fluid exhibits near ideal plug flow behavior under operating conditions of low wafer pressure and high relative pad-wafer velocity due to the presence of a substantial fluid layer indicated by low COF. When operating at a low COF, the dispersion number is constant as operating conditions change. At a threshold COF value (between 0.25 and 0.3), increases in the COF cause increases in axial dispersion due to high tortuosity of the flow. It is desirable to operate in the region of low COF since the flow through the system is nearly ideal and the Dispersion Number is stable with respect to minor changes in operating conditions or disturbances to the system.

Acknowledgments

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

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

References