Hydrodynamics of Slurry Flow in Chemical Mechanical Polishing

A Review

Elon J. Terrell and C. Fred Higgs III[z]

Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213-3890, USA

Chemical mechanical polishing (CMP) is a process that is commonly used to planarize wafer surfaces during fabrication. Although the complex interactions between the wafer, pad, and slurry make the CMP process difficult to predict, it has been postulated that the motion of the slurry fluid at the wafer–pad interface has an important effect on the wafer surface wear distribution. This paper thus serves as a review of past studies of the hydrodynamics of slurry flow during chemical mechanical polishing. The reviewed studies include theoretical and numerical models as well as experimental measurements.

Chemical mechanical polishing (CMP) is a manufacturing process that is used to planarize the surfaces of small-scale devices such as integrated circuits and hard disk read/write heads during fabrication. CMP has emerged as a critical fabrication step due to the demand for faster and more complex small-scale devices with multilevel interconnects. During CMP, the wafer containing the device is mounted face-down onto a rotating carrier and pressed against a rotating polishing pad that is flooded with chemically reactive slurry containing abrasive nanoparticles, as shown in Fig. 1. The mechanical and chemical interactions between the wafer, pad, and slurry cause the surface of the wafer to wear to atomically smooth levels. Although CMP is widely used in industry, much of the physics behind CMP is not known because of the complex phenomena at the wafer–pad interface. These complexities include (*i*) the slurry flowfield and film thickness distribution in the wafer–pad interface, (*ii*) the material wear effects caused by interactions between contacting wafer and pad asperities, (*iii*) the effects of wafer and pad surface roughness on the slurry flowfield, (*iv*) the material wear effects caused by the nanoparticles, and (*v*) the effect of the nanoparticles on the rheology of the slurry. The lack of detailed knowledge of these effects has reduced CMP optimization into a mostly empirical process. Thus, CMP modeling and experimentation has become critical for understanding CMP and minimizing the amount of trial-and-error schemes that are currently necessary for CMP optimization.

A number of experimental and theoretical studies have been conducted in order to analyze different aspects of the CMP process. A great deal of CMP research involves analysis of the material removal rate (MRR). A generalized expression for the wafer surface material removal rate is given by Preston's wear equation, as follows

$$MRR = \frac{kPU}{H} \qquad [1]$$

where *k* is the nondimensional Preston's wear coefficient, *P* is the wafer downforce, *V* is the relative velocity of the wafer–pad interface, and *H* is the hardness of the wafer surface. Preston's wear equation has commonly been used as an approximation for global MRR.

A number of more sophisticated wafer surface wear models have been developed to account for various physical phenomena that take place during CMP. Nanz and Camilletti[1] provided a critical review of CMP models up until 1995. Several studies have taken a contact mechanics approach toward CMP analysis, assuming that the wafer and pad surfaces are in direct sliding contact during the CMP process. Additional studies have assumed that the wafer and pad are

z E-mail: higgs@andrew.cmu.edu

in partial contact and have used a hybrid contact mechanics/fluid mechanics approach toward analyzing CMP. Finally, a set of studies have analyzed the CMP process solely using fluid mechanics, assuming that the wafer and pad surfaces are completely separated by slurry. From their review, Nanz and Camilletti have indicated the importance of the slurry flowfield to the CMP process as well as the need for more in-depth understanding of slurry flow at the wafer–pad interface. Therefore this paper serves as a review of past studies in slurry hydrodynamics during CMP. These studies include film thickness and hydrodynamic pressure modeling, numerical fluid flow modeling, and experimental investigations.

CMP Hydrodynamic Modeling

In the modeling studies that are discussed in this paper, slurry hydrodynamic analysis was used to find expressions for a number of parameters, including the slurry pressure field, film thickness distribution, and shear rate.

Slurry Film Thickness and Hydrodynamic Pressure Modeling

Assuming that a thin slurry film separates the wafer and pad surfaces during CMP, the film thickness and pressure distribution of the slurry can be related using the Reynolds equation, shown in 1D form as follows

$$\frac{d}{dx}\left(h^3\frac{dp}{dx}\right) = 6\mu U\frac{dh}{dx} \qquad [2]$$

where *p* is the hydrodynamic pressure, *h* is the local film thickness, μ is the dynamic viscosity of the slurry, *U* is the relative velocity of the bottom surface, and *x* is the downstream distance. Analysis of the slurry pressure distribution is of importance in CMP hydrodynamic studies because it gives insight into how much the wafer and pad surfaces are being pushed away from each other (positive pressure) or sucked toward each other (negative pressure) along the length of the interface.

A study by Sundararajan et al.[2] involved the derivation of 2D wafer-scale lubrication and mass-transport models for an assumed hydrodynamic slurry interface during CMP. They started the model using the one-dimensional, steady-state Reynolds equation (Eq. 2), assuming that the film thickness had a convex shape due to bending of the wafer. It is important to note the difference between a convex and concave wafer according to CMP terminology. As shown in Fig. 2, a wafer is termed convex if it is bent toward the pad in the middle and concave if it is bent away from the pad in the middle. Certain constant parameters in the film thickness expression were solved by assuming two constraints: (*i*) that the integral of the pressure distribution across the length of the wafer is equal to the applied load, and (*ii*) that the movement of the forces around the center of the wafer is zero. After solving for the film thickness and pressure distributions, the authors calculated the slurry velocity distribution
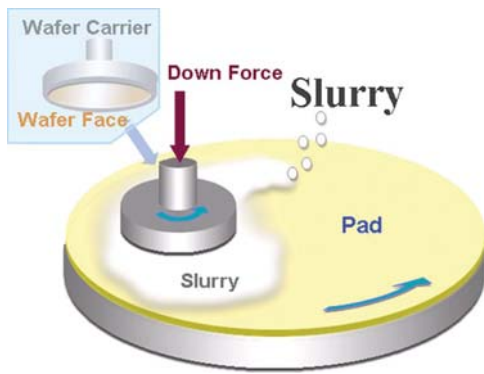
Figure 1. Diagram of the CMP process.

Figure 3. Diagram of 3D pad–surface model from from Nishioka et al.[3] Adapted with permission, © 1999 IEEE.

across the wafer using the expression for Couette/Poiseuelle flow, given as follows

$$u(x,y) = U\left(1 - \frac{y}{h}\right) - \frac{dp}{dx}\frac{h^2}{2\mu}\frac{y}{h}\left(1 - \frac{y}{h}\right) \qquad [3]$$

where U is the velocity of the pad surface, h is the film thickness, and p is the hydrodynamic pressure. From the results of their model, the authors found that certain conditions caused the slurry flow to have an unstable separation region. The results of their lubrication study were combined with mass-transport theory in order to predict the material removal rate distribution over the surface of the wafer.

Nishioka et al.[3] presented an analytical model for the slurry film thickness and wafer–pad coefficient of friction during CMP, accounting for pad surface roughness. The pad was modeled as a moving 3D sinusoidal surface, while the wafer was modeled as a flat, stationary surface. The expression for the pad surface is given as follows

$$h(x,y) = h_0 + R_P \cos\frac{2\pi x}{\lambda_x}\cos\frac{2\pi y}{\lambda_y} \qquad [4]$$

where h_0 is the mean line of the pad, R_P is the peak roughness, and λ_X and λ_Y are the wavelengths of the pad in the x and y directions, respectively. A diagram of the resultant sinusoidal surface is shown in Fig. 3. The parameters needed to define the sinusoidal surface (such as wavelength and roughness amplitude) were determined by performing roughness analysis on an actual pad. From this model, expressions for the mean hydrodynamic pressure and mean shear stress were derived as functions of the minimum film thickness. Their prediction of the pressure variation with film thickness is shown in Fig. 4. In order to validate their model, they compared the predicted coefficient of friction of their model to the measured coefficient of friction from experimental CMP tests.

Studies by Shan et al.[4,5] have been used to analyze the slurry hydrodynamic pressure distribution across the wafer during CMP. Their analysis was conducted using a combination of mathematical modeling and validation experiments. The modeling aspect of their study involved the use of the one-dimensional Reynolds equation taken over a constant-pad-velocity line as shown in Fig. 5. The
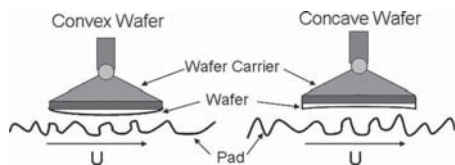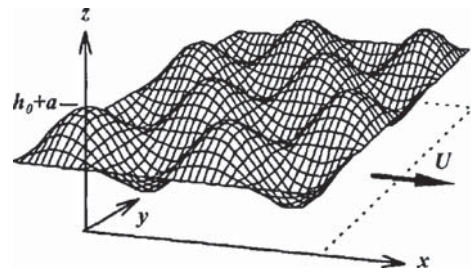
wafer surface was assumed to be fixed for simplification purposes. The slurry film thickness was estimated by assuming a contact stress distribution across the wafer and then solving for film thickness using the Greenwood and Williamson contact stress model.[6] This resulted in a film thickness distribution that was smallest at the edges of the wafer and largest in the middle. From the film thickness distribution the authors used a finite-differencing algorithm on the 1D Reynolds equation (Eq. 2) to determine the predicted pressure distribution across the wafer. Their analysis showed that the slurry pressure distribution is subambient at certain locations along the length of the wafer, implying that the wafer is "sucked down" at those locations. Validation experiments were conducted using a commercial benchtop polisher that was fitted with preconditioned polishing pads. The pressure distribution was measured by outfitting the simulated wafer surface with a series of pressure taps whose data was acquired using an electronic pressure transducer. For these experiments, water was substituted for slurry as the interfacial fluid in order to prevent the possibility of the particles interfering with the pressure taps. The results of the validation experimentation also showed a region of subambient pressure and corresponded well with the predicted results.

A study by Higgs III et al.[7] expanded on the studies by Shan et al.[4,5] by performing a 2D analysis of the entire wafer instead of a line of constant pad radius. This study, like the previous study, involved a combination of experimental pressure measurements and mathematical modeling. The mathematical model was created using the polar form of the Reynolds Equation, given as follows

$$\frac{\partial}{\partial r}\left(rh^3\frac{\partial p}{\partial r}\right) + \frac{1}{r}\frac{\partial}{\partial\theta}\left(h^3\frac{\partial p}{\partial\theta}\right) = 6\mu(r\omega)\frac{\partial h}{\partial\theta} \qquad [5]$$

where r and θ are radial and tangential coordinates along the wafer–pad interface, respectively, and ω is the rotational speed of the pad. The film thickness, h, was found by assuming a given contact stress distribution and then finding the attack angle of the wafer by balancing forces and moments about the pivot point. As a result of this study, the authors found yet again that a significant portion of the pressure distribution was subambient, as shown in Fig. 6. The validation experiments were conducted using a tabletop polisher using
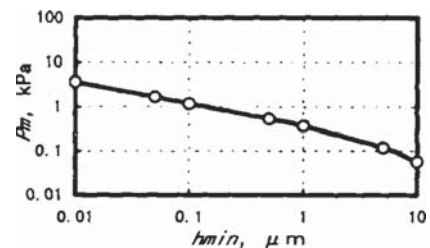

Figure 2. Diagram showing the difference between a convex and a concave wafer.

Figure 4. Variation of mean pressure with minimum film thickness from Nishioka et al.[3] Adapted with permission, © 1999 IEEE.
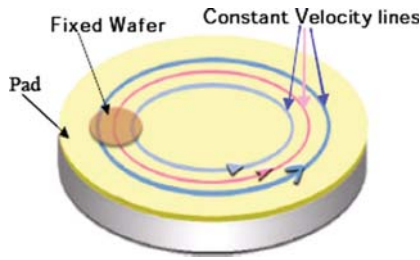
Figure 5. Diagram of constant pad velocity line that was analyzed in the 1D CMP hydrodynamic studies.

water as a substitute for slurry in a method similar to that of Shan et al.[4,5] The predicted results matched up well with experimental data. More works related to the modeling of subambient hydrodynamic slurry pressure can be found in the literature[8-12] but are omitted from this paper for the sake of brevity.

Cho et al.[13] presented a 2D mathematical model which predicts the slurry film thickness and pressure distribution during CMP. For this analysis the authors used the polar form of the Reynolds equation. They assumed that both the wafer and pad surfaces were completely flat, although the wafer was allowed to tilt from its center pivot point in order to balance forces and moments. The authors also assumed that the slurry was a particle-free, incompressible, Newtonian fluid for their analysis. By coupling the tilt of the wafer (film thickness distribution) with the balance of forces and moments on the wafer (pressure distribution), the authors were able to solve for both using the numerical Newton–Raphson method.

Thakurta et al.[14] developed a model for slurry film thickness and velocity distribution during CMP and compared the predicted results with the results of experiment. Their model, outlined in Fig. 7, accounted for the porosity and deflection of the pad surface. The theoretical model was created by first assuming a parabolic shape for the convex wafer as follows

$$h(x,y) = h_0 + S_X\left(\frac{x}{a}\right) + S_Y\left(\frac{y}{a}\right) + \delta_0\left(\frac{x^2 + y^2}{a^2}\right) \qquad [6]$$

where h_0 is the centerline height of the wafer, δ_0 is the wafer dome height, a is the radius of the wafer, and S_X and S_Y are the horizontal and vertical slopes of the wafer due to its equilibrium angle on the gimbal. Additionally, the pad surface topography, given by $s(x,y)$, accounted for elastic deformation of the pad, which was specified to be directly proportional to the slurry hydrodynamic pressure. The pad porosity was also accounted for by assuming that a certain amount of slurry seeps into the pad depending on the hydrodynamic pressure distribution. Lubrication approximations were then used to simplify the polar Navier–Stokes equations, assuming the slurry to

be a Newtonian, particle-free fluid, which provided the following governing equation

$$-12\mu\nabla \cdot [(h_w - s)^3\nabla p] + \frac{(h_w - s)}{2}\nabla \cdot (\vec{U}_w + \vec{U}_p)$$

$$+ \frac{(\nabla h_w + \nabla s)}{2} \cdot (\vec{U}_p - \vec{U}_w) + V_w - V_p = 0 \qquad [7]$$

where μ is the dynamic viscosity of the fluid, h_w and s are the wafer and pad surface topographies, respectively, p is the hydrodynamic pressure, and U_w and U_p are the velocities of the wafer and pad, respectively. The governing equation was solved using an iterative finite difference scheme by being subjected to a balance of forces and moments around the pivot point. From this analysis, the authors were able to calculate the slurry pressure field, flowfield, and film-thickness distribution. The pressure field from their study is shown in Fig. 8, showing concentric isobars that are greater than ambient everywhere in the flowfield. Figure 9 shows vector plots of the interfacial flowfield at different vertical "slice" locations inside the wafer–pad gap. As Fig. 9 shows, the slurry flowfield appears to closely follow the motion of the pad near the pad ($z^* = 0$), then transitions into following the motion of the wafer as vertical location of interest increases. In the vertical location directly next to the wafer ($z^* = 1$), the slurry flow approximately follows the motion as the wafer.

Jeng and Tsai[15] presented a CMP model which combines hydrodynamic lubrication theory with granular flow analysis in order to account for the motion of the slurry with abrasive nanoparticles. The hydrodynamic aspect of this model was based on the macroscopic Navier–Stokes equations, while the granular flow aspect of this model was based on microscopic molecular theory from a separate study.[16] These two approaches were combined and simplified into a set of governing equations which described particle-fluid motion. Their resultant model predicted that the material removal rate increases proportionately with particle size. They later expanded upon their previous model by accounting for pad roughness effects using a combination of flow factors from separate studies.[17-20] The studies by Jeng and Tsai focused a significant amount of attention on the effect of the abrasive particles in CMP, which is an important aspect of CMP that is often neglected in literature. However, their studies assumed that the slurry was completely composed of particles, which is an exaggerated assumption because slurry is composed primarily of fluid and contains only a trace amount (3–5 wt %) of particles.[21,22]

Chen and Fang[23] presented a mathematical model which predicts the slurry film thickness and pressure distribution during CMP. The wafer was assumed to have a convex shape while the pad was assumed to be completely flat and horizontal. The resultant pressure distribution was found by deriving the polar form of the Reynolds equation and solving it by expanding the pressure distribution into
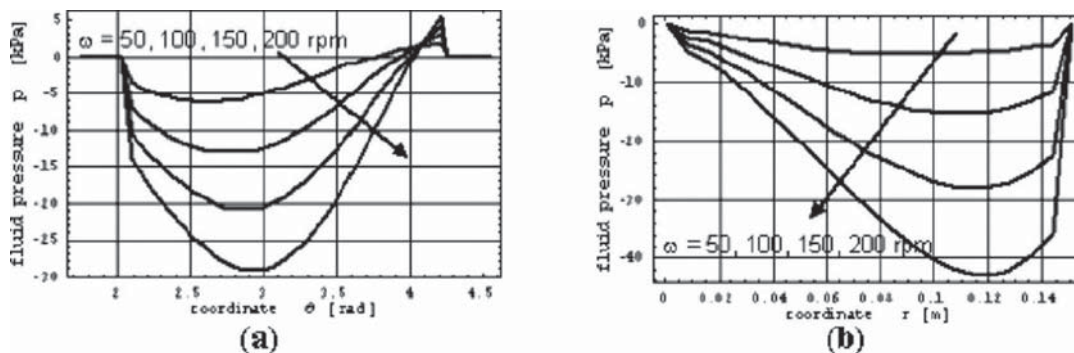

Figure 6. Predicted 2D fluid pressure from Higgs III et al.[7]: (a) tangential and (b) radial.
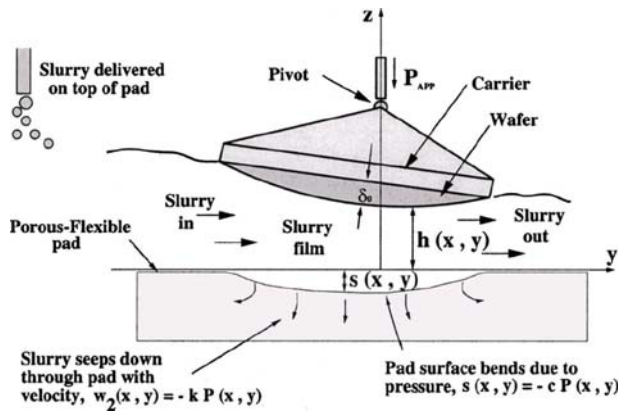
Figure 7. Diagram of modeling domain used in Thakurta et al.[14] Reproduced by permission of The Electrochemical Society, Inc.

Stoum–Liouville eigenfunctions. As a result of their work they found that the pressure distribution takes on a half-parabolic shape across the radial direction from the center of the wafer. All predicted pressures from this study were greater than atmospheric.

Slurry shear rate modeling.— It is possible that the shear rate of the slurry is of great importance to CMP hydrodynamics. Runnels and Eyman[24] have postulated that the wafer surface wear rate is directly proportional to the shear rate of the slurry in the hydrodynamic lubrication regime according to the following equation

$$MRR = K\sigma\tau \qquad [8]$$

where *MRR* is the material removal rate, *K* is the Preston coefficient, σ is the normal stress, and τ is the shear stress of the slurry fluid.

Equation 8 is proposed only for hydrodynamic lubrication. If the minimum film thickness is on the order of or less than the average surface roughness, then the effect of solid contact must be taken into account.[14] Several solid contact CMP models are currently available in literature, if solid–solid contact is to be assumed. Details of the solid contact models are outside the scope of this paper.

A study by Sohn et al.[25] involved the derivation of expressions for the shear rate for slurry flow at the wafer–pad interface. Assuming that the slurry behaved as a particle-free, Newtonian fluid, the authors used a simplified version of the incompressible Navier–Stokes equations to model the flowfield. For this analysis it was
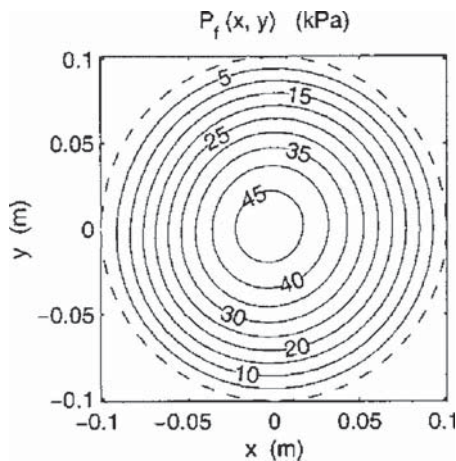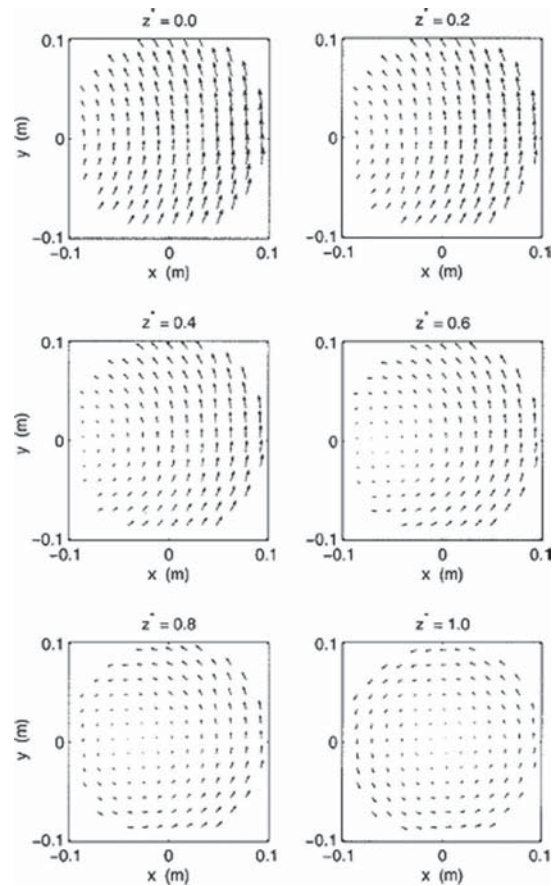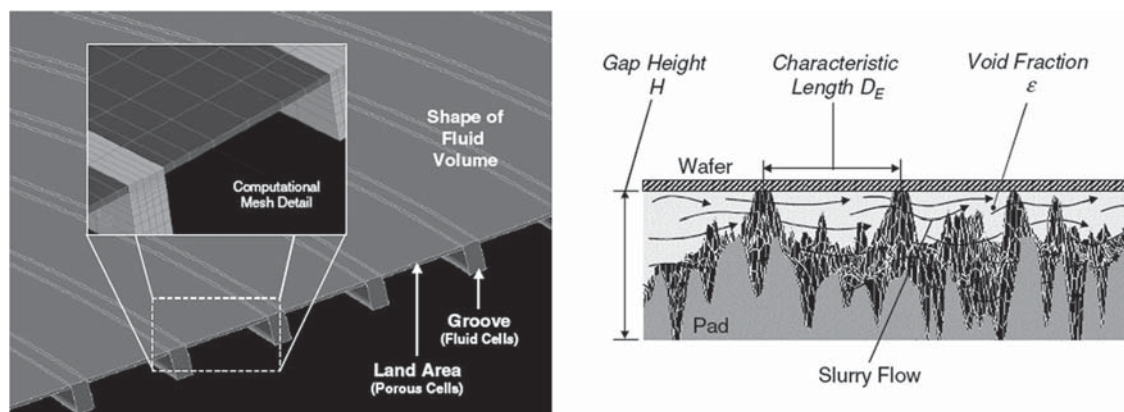

Figure 9. Slurry flowfield at different vertical locations inside the wafer–pad gap from Thakurta et al.[14] Reproduced by permission of The Electrochemical Society, Inc.

assumed that both the wafer and pad surfaces were rotating, no-slip walls. Additionally, the wafer and the pad were assumed to be perfectly parallel to each other, which resulted in a constant slurry hydrodynamic pressure. The authors simplified the Navier–Stokes equations by assuming that the Reynolds number and aspect ratio are both negligibly small. From this analysis the authors were able to derive closed-form expressions for the slurry velocity field and shear rate. They found that when the wafer rotational speed is greater than the speed of the pad, the shear rate is greatest at the edges of the wafer. However, when the wafer rotational speed is the same as that of the pad, the shear rate is uniform throughout the wafer.

Discussion of hydrodynamic modeling studies.— The discrepancy between the predicted results of each of these models appears to be rooted in the different assumptions of the wafer and pad surface geometries. The studies by Sundararajan et al.,[2] Thakurta et al.,[14] Jeng and Tsai,[15,17] and Chen and Fang[23] assumed that the wafer surface was slightly convex, which resulted in a pressure distribution that was greater than ambient everywhere in the domain. In contrast, the studies by Shan et al.[4,5] and Higgs III et al.[7] incorporated contact stress models into their analysis and thus ended with a film thickness that was smallest at the edges of the wafer and largest in the middle. As a result, the former group of authors predicted a pressure distribution that was greater than ambient everywhere in the domain, while the latter group predicted a region of subambient pressure. In order to determine the correct pressure distribution, one must analyze the film thickness distribution and account for wafer

Figure 8. Pressure distribution under rotating wafer from Thakurta et al.[14] Reproduced by permission of The Electrochemical Society, Inc.

Figure 10. Slurry flow domain and diagram of a real wafer–pad interface from Muldowney[28] using Fluent 6.1, a commercial CFD solver. Reprinted with permission from Materials Research Society.

bending and pad deflection. Numerous studies have been conducted which analyze the deflection of the wafer and pad during CMP, although a detailed discussion of these studies is omitted from this paper for the sake of brevity.

Additionally, we must take note of some of the simplifications that are used in these models. The assumption of the slurry as a Newtonian fluid is the most widely used simplification in each of these models, with the exception of the studies by Jeng and Tsai,[15,17] who assumed that the slurry was composed completely of small "granular" particles.

Numerical Studies in CMP Hydrodynamics

A few studies have been conducted which use computational fluid dynamics (CFD) to analyze the slurry flowfield in CMP. CFD solvers are advantageous for this analysis due to their ability to input complex flow domains and solve transport equations for multiphase flow and chemical reactions.

One of the first CMP numerical studies was conducted by Runnels and Eyman,[24] who created a numerical model to predict the slurry film thickness and hydrodynamic pressure. Assuming that hydrodynamic lubrication takes place between the wafer and pad surfaces, they imposed a sample slurry flowfield domain into a numerical code and solved it using a Galerkin finite element scheme. Both the wafer and pad surfaces were modeled as being rigid, smooth, no-slip walls. The pad was assumed to be flat while the wafer was designed to be convex with a specified radius of curvature. The wafer and pad walls were bounded by an additional wall which joined the two. This bounding wall was given a stress-free boundary condition in order to allow fluid to enter and exit the domain freely. The slurry film thickness was found by balancing the hydrodynamic forces with the applied load and pivoting the wafer such that the moment around the pivot point was zero. From the results of these simulations, the authors were able to find the amount of load that can be supported by the wafer as well as the minimum film thickness of slurry between the wafer and the pad.

Fu and Chou[26] used CFX-3D, a commercial numerical solver, to solve for the slurry flowfield in a CMP domain between the wafer and the pad. For their simulation, both the wafer and pad were modeled as being perfectly rigid, flat, and smooth, while the slurry was modeled as being a Newtonian, incompressible, particle-free fluid. The wafer–pad gap was fixed at a given input value, either 20 or 40 μm. Both the wafer and pad walls were modeled as having no-slip boundary conditions, while the remaining surfaces were modeled as stress-free boundaries in order to allow the slurry to freely enter and exit the computational domain. The resultant slurry shear stress distribution from the simulation was used to estimate the material removal rate from the wafer surface.

Yao et al.[27] used the software package Fidap, a commercial numerical solver, to model the slurry flow pattern between two moving surfaces at different locations at the wafer/pad interface. They chose not to model the entire wafer/pad domain but rather modeled various geometries along the tangent of the wafer in order to conserve computational resources. Each of the geometries was square and featured different scales of roughness. The boundary conditions for each of the geometries were dependent on the rotational movements of the pad and wafer. The slurry flow was modeled using the incompressible Navier–Stokes equations, assuming that the slurry exhibited nonNewtonian behavior due to the abrasive nanoparticles. They assumed that the material removal rate was directly proportional to the slurry shear stress as postulated in Runnels and Eyman, and then used Fidap's time interval updating capability to change the slurry film thickness over time based on the predicted wafer surface wear. Using this method they were able to determine the amount of material removal that occurs to the surface roughness after a given amount of polishing time. They were also able to derive an empirical model for the instantaneous polish rate with respect to time.

Muldowney[28] used the commercial numerical solver Fluent 6.1 to model the slurry velocity field, thermal field, and chemical reactions between the pad and wafer during CMP. The CFD domain in this study is shown in Fig. 10. For this simulation, the wafer was modeled as being flat and smooth, while the pad was modeled as having a rough topography in the form of a series of concentric circular grooves that are separated by circular concentric "asperities." The gap between the "asperities" and the wafer surface was modeled as being porous in order to account for the porous flow of slurry through the asperities. The pad and the wafer surfaces were modeled as no-slip/no-penetration boundaries, while the slurry was assumed to be a Newtonian fluid. From this analysis, the author was able to show the resultant velocity profile of the slurry at the wafer–pad interface. Figure 11 shows the predicted velocity profile, which appears to have a stagnation/backflow region that is comparable to the midgap ($z^* = 0.6$) velocity field predicted by Thakurta et al.[14] (Fig. 9).

Rogers et al.[29] used a combination of numerical modeling and experimental testing to analyze the flow of slurry during CMP. For their numerical study they used Fluent, a commercial fluid flow solver, to analyze the flow of both the slurry and the surrounding air outside the wafer–pad interface. Their 2D flow domain consisted of the linearly moving pad surface, the surrounding air/slurry volume, and the fixed wafer surface, which was represented as a rigid punch. The motion of the air and slurry were analyzed using Fluent's volume of fluid (VOF) solver, which modeled the air and slurry as two immiscible fluids. Both the wafer and pad walls were modeled as no-slip boundaries, while the side walls were modeled as cyclic

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.