A Contact-Mechanics-Based Model for General Rough Pads in Chemical Mechanical Polishing Processes

Chunyang Feng, Changhao Yan, Jun Tao, Xuan Zeng, and Wei Cai

Abstract: In this paper, a general rough-pad model is proposed for the chemical mechanical polishing (CMP) process. The proposed rough-pad model has several advantages over existing models. First, general height distribution functions and autocorrelation functions are used to describe the pad surface, which are easier to obtain than pad asperity height and curvature distributions in existing models. Second, the spectral representation technique and nonlinear transformation method used in the model allow rough-pad surfaces with general pad surface height distributions and autocorrelation functions. Thus, no assumption is made on the surface geometry and statistics of the pad. A conjugate-gradient iteration scheme combined with the fast Fourier transform technique is used to solve the resulting wafer–pad rough-contact problems to fully take into account the bulk deformation of the pad and the interactions among neighboring asperities. Model predictions are in good agreement with the experimental data in the existing literature. Based on the proposed model, the effects of CMP process parameters and underlying pattern geometries on dishing and erosion can be evaluated. The proposed model may also be used as a CMP pad design tool for improving dishing and erosion.

© 2009 The Electrochemical Society. [DOI: 10.1149/1.3133238] All rights reserved.

Manuscript submitted January 16, 2009; revised manuscript received April 14, 2009. Published May 20, 2009.

Chemical mechanical polishing (CMP) is widely used as the primary planarizing technique for the fabrication of integrated circuits. During the CMP process, a patterned wafer attached to a rotating carrier is pressed face down onto a rotating pad flooded with a polishing slurry containing chemical compounds and abrasive particles. Despite its apparent simplicity, the material-removal mechanism of CMP is very complicated indeed. Many believe that the chemicals in the slurry weakens the wafer film surface, and the weakened film is then removed by the mechanical interaction among the wafer, the polishing pad, and the abrasive particles in the slurry.

Despite being a predominant planarizing technique, CMP is known to suffer from undesired pattern-dependent problems, such as metal dishing and oxide erosion, the magnitudes of which are dependent on feature sizes including both linewidth and space. Metal dishing is the difference between the height of the oxide in the spaces and that of the metal in the trenches. Oxide erosion is defined as the difference between the oxide thickness before and after CMP. Figure 1 illustrates metal dishing and oxide erosion in a copper CMP. These two phenomena have a significant impact on circuit performance and yield loss. Thus, it is important to develop models that cannot only predict the evolution of the wafer profile during CMP process, but can also be used to assist in optimizing the CMP process to improve dishing and erosion.

As the polishing pads play a very important role in material removal, it is important to relate the pad properties to the topographical results predicted by the CMP models. Over the past decades, many CMP pad models have been developed as reviewed in Ref. 2, in which an extensive overview of the CMP process and modeling is presented. The existing CMP pad models may be classified into two categories: The smooth- and rough-pad models. In the smooth-pad models, the polishing pad is assumed to be perfectly smooth and metal dishing is mainly caused by the pad bending.3,4 In Ref. 3–5, the pad is assumed to be continuous elastic and contact mechanical models are presented. In Ref. 6, the pad behaves like linear springs and an analytical model for dishing and erosion is proposed. However, the assumption of the smooth-contact surface is not realistic, because the real pad surface is rough and the silicon wafer is mainly supported by pad asperities. Therefore, it is crucial to include the effects of pad surface roughness in CMP modeling. Several CMP rough-pad models adopt the Greenwood–Williamson (GW) approach5 to model the pad–wafer rough contact.6,8 In these models, the pad–wafer asperity contact is modeled based on the GW model with different asperity height distributions, such as exponential distribution,8 Gaussian distribution,8 and periodic distribution.11

Although the GW model7 has been widely accepted in CMP process modeling and can predict important trends in the effects of pad surface properties on dishing and erosion, its shortcomings are also well known and limit its usefulness. For example, in the GW model, the asperity height distribution, asperity curvature, and density are used to represent the pad surface roughness. However, the concept of asperity is still an unresolved problem and needs to be investigated further.12 Reference 13 shows that the former definition of asperity, “peaks” on a surface profile, is quite wrong and suggests that the earlier Archard concept9,10 that roughness consists of “protuberances on protuberances on protuberances” should be used.12 However, the limitation of Archard’s model of a rough surface is that it cannot easily be related to surface measurements. Although there are some other definitions of pad surface asperity,12 it is a known fact that relating the measured pad surface height distribution with a given sampling rate to pad asperity height distribution needs more comprehensive work.12 Thus, the difficulty in directly measuring or evaluating the model parameters limits its usefulness. The asperity geometries in the GW model are oversimplified and the Hertzian contact theory is applied for a single asperity deformation. Furthermore, the interactions between neighboring asperities are neglected, which limits the accuracy of the model. Because the pad surface roughness is strongly affected by both the inherent pad microstructure and surface conditioning, generalizations about asperity statistics have limited utility.13 The wafer surfaces evolved during CMP process are not usually ideally flat surfaces (with recesses and bumps). Thus, the assumption of Hertzian

Figure 1. (Color online) Definition of dishing and erosion.
contacts in the GW model is not valid. The model of Ref. 9 predicts a false finite metal line dishing in the limit of zero linewidth. Therefore, a true topographical description and a rigorous treatment of CMP physics are critical to the analysis of CMP pads.16

Other than the GW approach7 using a statistical description of surfaces and pressures, the direct numerical solution of rough-contact problems for surfaces collected from measured profiles or computer-generated ones exhibits a more promising approach with the advent of computer technology and fast numerical techniques.17 Terrell et al.18,19 proposed a voxel-based wear model in which the pad–wafer rough-contact problem is numerically solved based on contact mechanics. Although the model can predict a similar trend to the experimental data, the roughness characterization of the pad is too simple. It is well known that both height distribution functions and spatial functions are needed to describe the texture of a rough surface. However, only the height parameter is used to generate the rough surfaces in Ref. 18 and 19, and the spatial information is ignored. Autocorrelation function and autocorrelation length have been widely used in surface-related studies to provide spatial information in addition to height parameters. For example, the effects of autocorrelation on friction behavior is analyzed in Ref. 20. Moreover, only Gaussian pad topography is considered in Ref. 18 and 19. However, previous studies show that the surface height distribution of the pad may deviate significantly from normality.15 Therefore, it is necessary to calculate the evolution of the wafer profile for non-Gaussian pad topography.

To address the issues raised above, the present work focuses on developing a general rough-pad model with measurable model parameters that can accurately evaluate the effects of pad properties on the evolution of the wafer surface and is amenable for use in pad design and process optimization. Here, the model uses the spectral representation technique21 and nonlinear transformation method22 to numerically generate the pad rough surfaces with given surface height probability density functions (pdf) and autocorrelation functions. Then, a conjugate gradient (CG)-based iteration scheme23 combined with the fast Fourier transform (FFT) technique is applied to solve the wafer–pad rough-contact problem. Finally, the obtained contact-pressure distribution is used to determine the local removal rate of the underlying patterns. The model does not make any assumption on the pad asperity shape and the pad surface height distribution, and fully takes into account the bulk deformation of the pad and the interactions among neighboring asperities. In this model, the pad surfaces are characterized by two functions: The height pdf and autocorrelation function, which are easily obtained by experimental measurements of the pad surfaces, for example, using light interferometry.24 Using the model, the effects of other CMP process parameters, the applied down pressure, the selectivity of the slurry, and the underlying pattern geometries on dishing and erosion can also be examined.

The rest of the paper is organized as follows. Background describing polishing pad surface morphology is reviewed in the pad surface morphology section. The methods for the pad rough surface modeling and characterization are proposed in the Numerical gen-

Figure 2. SEM image of a conditioned pad.26

eration of pad rough surfaces section. Formulation of the wafer–pad rough-contact problem and the numerical methods used to solve this problem are presented in the Wafer–pad rough contact model section. The overall simulation procedure is summarized in the Simulation methodology section. Model calibration procedure and numerical experiments are provided in the Model Calibration and Numerical Experiments sections, respectively. Finally, conclusions are drawn.

Model Description

**Pad surface morphology.**— Generally, a pad is made of polyurethane. The surface is covered with 30–50% pore density (diameter 40–60 μm), and each pore is separated with wall structures (asperities of width 10–50 μm). Figure 2 shows the scanning electron microscope (SEM) image of a conditioned pad. The microscale features of asperities and pores can be seen clearly in the SEM image. The surface roughness is strongly affected by both the inherent pad microstructure and surface conditioning before and during polishing. Various metrology methods have been proposed to characterize pad surfaces.15,25 Figure 3 shows representative pad surface data from a conditioned pad and their height distribution.26 It shows a Gaussian or Gaussian-type distribution with some distortion. As stated in Ref. 15, the pad surface height probability distribution function can usually be described by one of the Pearson family of distributions. The autocorrelation function can also be easily obtained from the pad surface data. As stated in Ref. 24, the conditioned polishing pad usually exhibits isotropic roughness.

From the above discussions, we can see that the shape of the pad asperities and the surface height distribution vary with the pad in-

Figure 3. (Color online) A conditioned pad surface representative line scan and pad height probability distribution.26
herent properties and conditioning parameters. So, to evaluate the effect of the pad surface properties on polishing results, a numerical method is needed to simulate the pad surfaces given arbitrary pad surface statistics.

Numerical generation of pad rough surfaces.— A rough pad surface with a given height statistic was generated at each time step to make contact with the wafer surface, similar to Ref. 18. As in Ref. 9, to provide a unified structure and reduce the computational effort, the underlying patterned wafer was represented as a periodic structure. Thus, the generated pad rough surface should also be periodic with the same period to eliminate the edge effect.

Gaussian random surface generation.— If the pad surface height distribution is assumed to be Gaussian,21,27 Gaussian surfaces with zero mean and a given autocorrelation function \( R \) are generated using a spectral representation method.21 The spectral representation method expands the stochastic field as a sum of trigonometric functions with random phase angles. The amplitudes are deterministic and depend only on the power spectrum of the stochastic field.21 The method to generate the one-dimensional (1D) and two-dimensional (2D) random surfaces is succinctly reviewed here, and for a detailed method description, the readers can refer to Ref. 21 and 27.

The spectral representation of the 1D rough surface \( f(x) \) is given by the following expression21

\[
f(x) = \sqrt{2} \sum_{n=0}^{N-1} A_n \cos(\omega_n x + \phi_n)
\]

Equation 1 can asymptotically lead to a Gaussian stochastic field as \( N \to \infty \) due to the central limit theorem. The amplitude \( A_n \) is the deterministic amplitude which is dependent on the power spectrum of \( f(x) \), \( \omega_n \) is the discretized frequency, and \( \phi_0, \phi_1, \ldots, \phi_{N-1} \) are independent random phase angles uniformly distributed in \([0, 2\pi]\). These parameters are defined as

\[
A_n = \sqrt{2S(\omega_n)\Delta \omega} \quad n = 0, 1, 2, \ldots, N - 1
\]

\[
\omega_n = n\Delta \omega \quad n = 0, 1, 2, \ldots, N - 1
\]

\[
\Delta \omega = \frac{\omega_0}{N}
\]

and \( A_0 = 0 \) or \( S(\omega_0) = 0 \) \[5\]

In Eq. 2, \( S \) is the power spectral density function of \( f(x) \). In Eq. 4, \( \omega_0 \) represents a cutoff frequency beyond which the power spectral density function may be assumed to be zero.

The power spectral density function \( S \) is related to the autocorrelation function \( R \) based on the Wiener–Khintchine theory

\[
S(\omega) = \frac{1}{2\pi} \int R(\tau) e^{-i\omega \tau} d\tau
\]

It is easy to see that the simulated rough surface \( f(x) \) given by Eq. 1 is periodic with period

\[
L = 2\pi/\Delta \omega
\]

The generation of rough surfaces through Eq. 1 may be computationally expensive. To leverage the FFT technique, Eq. 1 can be alternatively written as

\[
f^{(i)}(p \Delta x) = \text{Re} \left\{ \sum_{n=0}^{M-1} B_n \exp(i(n\Delta \omega)(p \Delta x)) \right\}
\]

\[
p = 0, 1, 2, \ldots, M - 1
\]

where \( \text{Re} \) is the real part, \( M \) is the number of FFT points, and \( B_n \) is given by

\[
B_n = \sqrt{2A_n} e^{i\phi_n} \quad n = 0, 1, 2, \ldots, M - 1
\]

The discretized random surface realization \( f^{(i)}(p \Delta x) \) in Eq. 8 is periodic with period \( L \). Hence, \( \Delta x \) and \( \Delta \omega \) are related as follows

\[
M \Delta x = L = \frac{2\pi}{\Delta \omega}
\]

With the aid of Eq. 10, Eq. 8 can be rewritten as

\[
f^{(i)}(p \Delta x) = \text{Re} \left\{ \sum_{n=0}^{M-1} B_n e^{i(2\pi np/M)} \right\} \quad p = 0, 1, 2, \ldots, M - 1
\]

Now the FFT technique can be readily applied on Eq. 11 to accelerate the computation.

The procedure of generation of 2D rough surfaces is similar to the above 1D case. The 2D rough surface \( f(x_1, x_2) \) can be generated by the following expression21

\[
f(x_1, x_2) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \left[ A^{(1)}_{n_1 n_2} \cos(\omega_1 x_1 + \omega_2 x_2 + \phi_{n_1 n_2}^{(1)}) + A^{(2)}_{n_1 n_2} \cos(\omega_1 x_1 - \omega_2 x_2 + \phi_{n_1 n_2}^{(2)}) \right]
\]

where

\[
A^{(1)}_{n_1 n_2} = \sqrt{2S(\omega_{1n_1} - \omega_{2n_2})\Delta \omega_1 \Delta \omega_2}
\]

\[
A^{(2)}_{n_1 n_2} = \sqrt{2S(\omega_{1n_1} + \omega_{2n_2})\Delta \omega_1 \Delta \omega_2}
\]

\[
\omega_{1n_1} = n_1 \Delta \omega_1 \quad \omega_{2n_2} = n_2 \Delta \omega_2
\]

\[
\Delta \omega_1 = \frac{\omega_0}{N_1} \quad \Delta \omega_2 = \frac{\omega_0}{N_2}
\]

and

\[
S(0, \omega_{2n_2}) = S(\omega_{1n_1}, 0) = 0 \quad n_1 = 1, 2, \ldots, N_1 - 1 \]

\[
n_2 = 0, 1, 2, \ldots, N_2 - 1
\]

In the above equations, \( S \) is the power spectral density function, and \( \omega_{1n_1} \) and \( \omega_{2n_2} \) represent the cutoff wavenumbers corresponding to the \( x_1 \) and \( x_2 \) axes in the space domain. The term \( \phi_{n_1 n_2}^{(1)} \) and \( \phi_{n_1 n_2}^{(2)} \) are both independent random phase angles uniformly distributed in \([0, 2\pi]\).

Similar to the 1D case, Eq. 12 can be reformulated as follows to use the FFT technique

\[
f^{(i)}(p_1 \Delta x_1, p_2 \Delta x_2) = \text{Re} \left\{ \sum_{n_1=0}^{M_1-1} \sum_{n_2=0}^{M_2-1} B^{(1)}_{n_1 n_2} \exp \left( \frac{2\pi i n_1 p_1}{M_1} + \frac{2\pi i n_2 p_2}{M_2} \right) \right\}
\]

\[
+ B^{(2)}_{n_1 n_2} \exp \left( \frac{2\pi i n_1 p_1}{M_1} - \frac{2\pi i n_2 p_2}{M_2} \right)
\]

where

\[
p_1 = 0, 1, 2, \ldots, M_1 - 1 \quad p_2 = 0, 1, 2, \ldots, M_2 - 1
\]

\[
B^{(1)}_{n_1 n_2} = \sqrt{2A^{(1)}_{n_1 n_2}} \exp(i\phi^{(1)}_{n_1 n_2})
\]

\[
B^{(2)}_{n_1 n_2} = \sqrt{2A^{(2)}_{n_1 n_2}} \exp(i\phi^{(2)}_{n_1 n_2})
\]

Figure 4 shows two 1D surface realizations with different correlation lengths, which have the autocorrelation function and corresponding power spectral density function given by

\[
R(\tau) = \sigma^2 e^{-\tau/\eta}
\]
Non-Gaussian random surface generation.— In the aforementioned section, a technique has been described to generate the pad rough surfaces with a Gaussian height distribution. For the generation of non-Gaussian surfaces with a given height distribution function and an autocorrelation function, the nonlinear transformation method in Ref. 22 will be used.

Let \( X(\bar{\tau}) , \bar{\tau} \in R^m \) be the target non-Gaussian random surface and \( F \) and \( R \) be the cumulative distribution function (CDF) and autocorrelation function of \( X(\bar{\tau}) \). The nonlinear transformation method generates \( X(\bar{\tau}) \) by a nonlinear transformation \( g \) of a standard Gaussian process \( Y(\bar{\tau}) \), \( \bar{\tau} \in R^m \)

\[
X(\bar{\tau}) = g[Y(\bar{\tau})]
\]

The nonlinear transformation algorithm consists of the following four steps. First, the nonlinear transformation function \( g \) is taken as

\[
g(y) = F^{-1}(\Phi(y))
\]

where \( \Phi(y) \) is the standard Gaussian CDF. Second, to generate realizations of the standard Gaussian process \( Y(\bar{\tau}) \), the autocorrelation function of \( Y(\bar{\tau}) \) should be obtained. Let \( \rho(\bar{\tau}) , \bar{\tau} \in R^m \) be the autocorrelation function of \( Y(\bar{\tau}) \). According to Ref. 22, \( \rho(\bar{\tau}) \) must satisfy the following equation

\[
\sigma^2 R(\bar{\tau}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y)g(z)\Phi[y, z; \rho(\bar{\tau})]dydz
\]

where \( R(\bar{\tau}) \) is the autocorrelation function of \( X(\bar{\tau}) \), \( \Phi( , , ) \) is the 2D standard Gaussian density, and \( \sigma \) is the standard derivation of the target non-Gaussian field. Third, after the autocorrelation function \( \rho(\bar{\tau}) \) of \( Y(\bar{\tau}) \) is obtained, realizations of the standard Gaussian process \( Y(\bar{\tau}) \) can be generated using the spectral representation method described before. Finally, realizations of the non-Gaussian random surface \( X(\bar{\tau}) \) can be calculated from the realizations of the Gaussian process \( Y(\bar{\tau}) \) and the nonlinear transformation in Eq. 26.

To show the accuracy in generating non-Gaussian surfaces, we considered non-Gaussian surfaces with autocorrelation function in Eq. 22 and a Gaussian-type distribution with asymmetric variance

\[
pdf(x) = \begin{cases} 
\frac{2}{\sqrt{2\pi}(\sigma_1 + \sigma_2)} e^{-\frac{(x - m_0)^2}{2(\sigma_1 + \sigma_2)^2}} & x < m_0 \\
\frac{2}{\sqrt{2\pi}(\sigma_1 + \sigma_2)} e^{-\frac{(x - m_0)^2}{2(\sigma_1 + \sigma_2)^2}} & x \geq m_0
\end{cases}
\]

Figure 6 shows the comparison of the target pdf in Eq. 29 with...
Figure 7. (Color online) Comparison of average autocorrelation function of 100 surface realizations generated by the nonlinear transform method with given autocorrelation function in Eq. 22, where $\eta = 30 \mu m$.

The average pdf of 100 generated non-Gaussian surface realizations. Figure 7 shows the comparison of the expected autocorrelation function in Eq. 22 with the average autocorrelation function of 100 generated non-Gaussian surface realizations. We can see that a good approximation of the target pdf and autocorrelation function can be obtained.

Wafer-pad rough contact model. — Pad–wafer rough contact is assumed to be point contact, which is illustrated in Fig. 8. The pad is assumed to be linearly elastic, and the wafer is assumed to be purely rigid. It is further assumed as in Ref. 9 that the pressure in the slurry is constant at the length scale under consideration and hence does not contribute to variations in pad deflection.

After the rough pad surface was generated at each time step using the methods described in the last subsection, the contact pressure distribution between the wafer and the pad was determined based on the contact-mechanics theory. The three-dimensional (3D) formula to describe the pad deformation $u(x,y)$ due to the contact pressure $p(x,y)$ is

$$u(x,y) = \int_A G(x-x',y-y')p(x',y')dx'dy'$$  \[30\]

where $G(x,y) = 1/(\pi E\sigma^2)(x^2 + y^2)\eta$ is the Green’s function with a local half-space of the wafer and $E^* = E/(1-v^2)$ is the equivalent Young’s modulus of the pad, where $E$ and $v$ is the Young’s modulus and Poisson’s ratio of the pad, respectively. The surface gap between wafer and pad, $h(x,y)$, after deformation can be written as

$$h(x,y) = h_i(x,y) + u(x,y)$$  \[31\]

where $h_i(x,y)$ is the initial interfacial gap and $c$ is the rigid-body approach, as shown in Fig. 8. Because the contact pressure $p(x,y)$ cannot be tensile and the surface gap should approach zero in the real contact area denoted by $A_c$, the following constraints should also hold

$$h(x,y) = 0 \quad p(x,y) > 0 \quad (x,y) \in A_c$$  \[32\]

Therefore, the 3D rough-contact problem can be modeled by Eq. 30 subjected to boundary constraints in Eq. 31-33.

As stated in Ref. 3, the above formulation in Eq. 30 expresses the displacement at one point as a function of pressure over the entire contact area. Therefore the surface displacement is nonlocal and the interactions between asperities are automatically taken into account.

The formulation can also be modified into a 2D case. The plane-strain deformation, $u(x)$, under a periodic pressure distribution is given by the following equation

$$\frac{da(x)}{ds} = \frac{2}{LE^*} \int_{-L/2}^{L/2} \rho(s)\cot L \int_{-L/2}^{L/2} \frac{\pi(x-s)}{L} ds$$  \[34\]

where $L$ is the period of the structure. After integrating Eq. 34, the contact problem for a 2D periodic structure can be modeled as

$$u(x) - C = -\frac{2}{\pi E^*} \int_{-L/2}^{L/2} \frac{\rho(s)\ln|\sin L|}{L} ds$$  \[35\]

where C is an integration constant.

To solve the rough-contact problem in Eq. 30-33, the approaches proposed in Ref. 29 are applied in this paper. As stated in Ref. 29, the above contact problem can be formulated as a quadratic optimization problem with linear inequality constraints, and a CG-based method can be used to solve it iteratively with fast convergence. The information of real contact area, contact-pressure distribution, and interfacial gap can be determined simultaneously. Because the underlaying patterns are represented as periodic structures, for the 3D case, the FFT-based technique can be readily utilized here in combination with the iterative scheme to accelerate matrix-vector multiplication instead of using a direct matrix method.

After the contact-pressure distribution is obtained, the local material-removal rate (MRR) can be calculated by applying Preston’s law of a linear relationship between removal rate and pressure

$$MRR(x,y) = \frac{MRR_p}{P_t} p(x,y)$$  \[36\]

where MRR(x,y) is the MRR at the point (x,y), MRR_p is the MRR under the applied nominal pressure, and $P_t$ is the applied nominal pressure. So the local topography changes can be directly linked to the removal rate by

$$\frac{\delta S(x,y,t)}{\partial t} = -MRR(x,y)$$  \[37\]

where $S(x,y,t)$ is the wafer surface height at the point $(x,y)$ in time $t$.

Although the linear relationship between removal rate and local pressure is used in the present formulation, it is not essential to the model and any removal rate law in which the removal rate is a function of the local pressure could be used instead.

Simulation methodology. — A flowchart of the overall simulation procedure is shown in Fig. 9. First, the simulator reads in the initial surface topography for the underlying patterns and the CMP process-related parameters including the applied nominal pressure, the simulation time, and the MRR under the applied nominal pressure. In the meantime, the pad parameters consisting of the equiva-
The main peak of the histogram is taken as the average height level of the oxide surfaces, whereas the smaller peak indicates the average height level of the bottom of the dished metal features. From Fig. 10, the average dishing can be directly read off as the distance between the two peak maxima. According to the erosion definition, the average erosion can be obtained as the difference between the initial dielectric height and the main peak maxima. The procedure described above is very flexible in dishing and erosion calculation. It can either be applied to a single line structure to obtain the local averaged dishing and erosion, or to the whole computational domain to obtain the global averaged dishing and erosion.

Model Calibration

Before using the model to predict the wafer surface evolution during CMP process, the model parameter values including both the process-related parameters and the pad parameters need to be given. However, if the independently measured data are directly used for some of the model parameter values, the model cannot predict the real CMP process accurately. This is because the model in this paper only focuses on the contact mechanical aspect of the CMP process, while other important mechanical interactions in CMP, namely, fluid mechanics and particle dynamics, are not involved. Thus, the incompleteness of the contact mechanical model with respect to the real CMP process leads to the discrepancy of the values of the model parameters from the real measured ones.

The model parameter values can be obtained through model calibration. Calibrating the model for a given CMP process is an optimization process. The optimization criterion is to minimize the root-mean-square (rms) error between the model simulation results and the measured data by finding optimized model parameters. The necessary inputs to the calibration procedure include the measured experimental data of the polished wafer topography and the initial model parameter values. The measured topography information can be dishing, erosion, or the surface heights of the metal lines and space regions. The model parameters consist of the applied nominal pressure, the MRRs, the equivalent Young’s modulus of the pad, the standard derivation, and correlation length of the pad surface height. The model parameters should be divided into two groups. One includes the parameters that need to be extracted through the calibration procedure and the other one is left unchanged. The overall calibration procedure operates in iterations. During each iteration, the simulation in Fig. 9 is called, then the simulation results are compared with the silicon data. The subset of the parameter values is optimized such that the error between the simulation results and the measured data has decreased. The optimized parameters will be used in the subsequent iterations. The procedure terminates when the rms error between the simulation results and the measured data is below the given threshold. Once the model parameters have been calibrated, the model can be used to predict dishing and erosion on any pattern structures polished under the same process conditions as those used in the model calibration.

Numerical Experiments

In this section, a series of experiments are conducted to verify the proposed general rough pad model and investigate the effects of pad rough surface parameters on dishing and erosion as well as the underlying trench filtering effect using the proposed model. The simulation time step is 0.001 s in all the following experiments. All of the simulation results in this section are based on model parameters of the applied nominal pressure $P = 28$ kPa and a selectivity of 3, if not mentioned otherwise. The Gaussian pad surface height distribution is used from the first to sixth subsections and the non-Gaussian distribution is used in the last subsection. The correlation function is taken to be as in Eq. 22. Furthermore, for simplicity, the surface topography of the underlying structure is initially assumed flat, if not mentioned otherwise.

Convergence of discretization in space.—To numerically solve the system equation in Eq. 30 and 35, both the wafer and pad sur-
faces need to be discretized into mesh grids. In this paper the piece-wise constant-collocation scheme was used, where the pressure $p$ and deformation $u$ are both taken as constants in each mesh grid. To check the influence of the choice of the number of mesh grids in the simulation domain on the computed polishing results, the variations of the metal dishing with different mesh resolutions have been calculated for the purpose of comparison. As shown in Fig. 11, the dishing results converge at a grid of 40 mesh points. Further refinement of mesh dimension beyond the value of 40 yields relatively little changes.

The model parameters for this experiment are as follows:

1. elastic modulus: 29 MPa
2. Poisson’s ratio: 0.2
3. standard derivation: 0.5 μm
4. correlation length: 30 μm

Comparison with the smooth pad model.—To check the convergence of the model with respect to the standard derivation $\sigma$ of the pad surface height, the dishing results of the proposed model with $\sigma$ approaching zero ($\sigma = 0.005 \mu$m) and Chekina’s smooth-pad model were compared. As shown in Fig. 12, our model results are in good agreement with the results of Chekina’s smooth-pad model.

The model parameters for this experiment are as follows:

1. elastic modulus: 29 MPa
2. Poisson’s ratio: 0.2
3. standard derivation: 0.005 μm
4. correlation length: 30 μm

Model calibration.—In this subsection, two calibration experiments are investigated and discussed. The pad surface statistic parameters of the model are calibrated by fitting to the experimental data, and the other model parameters are obtained based on previous literature. The model prediction and the experimental data for metal dishing are compared.

The first set of experimental results used are from Park et al. The pad used in this experiment is believed to be similar to an IC1000 pad based on several publications. The top pad of the IC1000 pad is considered to be the same as that of the IC1400 pad. Therefore, the elastic modulus and Poisson’s ratio are taken as 29 MPa and 0.2, respectively. It is reported that the selectivity, the
removal rate under the nominal pressure of the metal to the dielectric, is much smaller if the pattern effect is taken into account. Therefore, the patterned selectivity cannot be directly obtained from a blanket removal rate. Based on the results of Ref. 34, the selectivity is taken as 3. The calibration process follows the procedure described in the Model Calibration section. The estimated values of normalized steady-state dishing are plotted in Fig. 13 in comparison with experimental data.

As shown in the figure, our model agrees well with the experimental results. Especially, our model can capture well the break point phenomena, where the dishing increases with higher pattern density and reaches its maximum in the 60–70% density range, and suddenly decreases beyond that point. The good agreement demonstrates the predictive capability of our model.

The model parameters for this experiment are as follows:

Specified parameters:
1. elastic modulus: 29 MPa
2. Poisson’s ratio: 0.2
3. applied nominal pressure: 28 kPa
4. selectivity: 3

Calibrated parameters:
1. standard derivation: 0.07 μm
2. Correlation length: 20 μm

Another set of experimental data used in the paper are from Saxena et al.4 The experiments were conducted at an applied pressure of 28 kPa with a potassium hydroxide based slurry as well as an IC1400 pad. Furthermore, the pattern density was fixed at 50% for all experiments. The pattern wafers were polished using a two-step polishing procedure, 13 min with the first-step slurry, followed by a 2 min polishing with the second-step slurry. The steady-state dishing data point was chosen to obtain the calibrated parameters in the model.

Figure 14 shows the comparison of the predicted results of our model with the experimental data from Saxena et al.4 From it, we can see that our model was able to predict the dishing with reasonable accuracy.

The model parameters for this experiment are as follows:

Specified parameters:
1. elastic modulus: 29 MPa
2. Poisson’s ratio: 0.2
3. applied nominal pressure: 28 kPa
4. selectivity: 3

Calibrated parameters:
1. standard derivation: 0.07 μm
2. Correlation length: 20 μm

Figure 15. (Color online) Evolution of the profile and the dishing of a metal line as a function of polish time. The linewidth is 20 μm and the line space is 80 μm.

Figure 16. (Color online) Evolution of the wafer surface for a 3D case.

Figure 17. (Color online) Metal dishing as a function of linewidth. The pattern density used here is 50%.
The average pad asperity radius was around 10 μm. In this experiment, the increase in the steady-state dishing with respect to the model, the correlation length of the fixed pattern density. In this experiment, the correlation length of the amount of steady-state dishing as a function of trench width with a material. The simulation topographies after this section are as follows:

The second experiment was a 3D case. In this experiment, the underlying pattern was a raised square feature with the initial step height of 0.5 μm. The whole structure was coated with the same material. The simulation topographies after \( t = 10 \) and 50 s of polishing are shown in Fig. 16. The corner rounding phenomenon can also be seen in this 2D case.

The model parameters used for both of the two experiments in this section are as follows:

1. applied nominal pressure: 6 psi
2. elastic modulus: 300 MPa
3. Poisson’s ratio: 0.4
4. standard derivation: 5 μm
5. correlation length: 30 μm

**Effects of underlying trench sizes.**— Figure 17 shows the amount of steady-state dishing as a function of trench width with a fixed pattern density. In this experiment, the correlation length of the pad surface was set to be 10 μm, so as stated in the previous section, the average pad asperity radius was around 10 μm. According to the model, the increase in the steady-state dishing with respect to linewidth can be divided into two regimes. When the linewidth is smaller than the average asperity size, only the asperity with smaller size can contribute to the pressure on the trench bottom and large asperities are filtered. When the linewidth is bigger than 20 μm, it is wide enough for the pad asperities to reach the trench bottom.

As stated in a previous section, directly applying the GW model in CMP modeling will lead to false results especially when feature size approaches zero. It can be seen in Fig. 17 that our model can clearly predict zero dishing in the limit of zero linewidth, which further proves that our model is more realistic.

The model parameters used in this section are as follows:

1. elastic modulus: 29 MPa
2. Poisson’s ratio: 0.2
3. standard derivation: 0.1–6 μm
4. correlation length: 10 μm

**Effects of topographical characteristics of Gaussian surfaces.**— Figure 18 illustrates the effect of the standard derivation \( \sigma \) of the pad surface height on dishing and erosion. Figure 18a clearly shows that the metal dishing increases with respect to \( \sigma \). Figure 18b shows the erosion as a function of \( \sigma \). The effect of \( \sigma \) on erosion is not significant and erosion decreases slightly with respect to \( \sigma \). This phenomenon is because as \( \sigma \) increases, a bigger fraction of the pressure is supported by the recessed metal area, which increases the removal rate of the metal, but decreases that of the oxide, as also observed in Ref. 9.

The model parameters used are as follows:

1. elastic modulus: 29 MPa
2. Poisson’s ratio: 0.2
3. standard derivation: 0.1–6 μm
4. correlation length: 30 μm

Figure 19 illustrates the effect of correlation length of the pad surface height on dishing and erosion. Figure 19a clearly shows that the steady-state metal dishing increases with increasing correlation...
length up to a certain value beyond which it decreases dramatically. This phenomenon is because as the correlation length is larger than half of the linewidth, decreasing the correlation length means the pad asperities can reach the bottom area more easily and thus the larger the amount of dishing. When the correlation length is smaller than half of the linewidth, the pad asperity size is so small compared to the linewidth that they can reach the recessed area freely. So, the smaller the amount of deformation of asperities on the top area of the underlying pattern structure, the more difficult for the asperities above the recessed area of the underlying pattern structure to touch the recessed area. According to Hertzian theory, the smaller the asperity radius, the smaller the amount of deformation of the asperity under a given pressure. Therefore, when the correlation length is smaller than half of the linewidth, as the correlation length decreases, the dishing decreases.

Figure 19b shows the erosion as a function of the correlation length. The opposite trend is observed. This phenomenon is because as the dishing increases, a bigger fraction of the pressure is supported by the recessed metal area. While as the total pressure is a constant, the pressure on the oxide space decreases. So the erosion decreases. As a result, the trend of erosion is opposite to that of dishing.

The model parameters used are as follows:

1. elastic modulus: 29 MPa
2. Poisson’s ratio: 0.2
3. standard derivation: 0.1 μm
4. correlation length: 5–35 μm

Effects of non-Gaussian pad surfaces.— Finally, the analysis is carried out with non-Gaussian pad surfaces. As stated in the previous section, the pad surface height probability distribution function is often not Gaussian but can usually be described by a Gaussian-
type distribution with some distortion. Here, the pdf given in Eq. 29 was used to examine the effect of the asymmetry of the pdf on polishing results.

Four different height distributions are used for comparison, as shown in Fig. 20. The corresponding metal dishing and pattern profiles are shown in Fig. 21 and 22, respectively.

As illustrated by these figures, the asymmetric distributions have great influences on polishing results, and asymmetric distributions tend to have more sharp corners of surface profile than symmetric ones.

Conclusion

In this paper, a general rough-pad model was proposed for the CMP process. The model supports arbitrary height distribution functions and autocorrelation functions to characterize the rough-pad surface, which can be easily obtained from measurements. The resulting wafer–pad rough-contact problem was solved by a CG-based iteration method. The present model can be used as a tool to optimize the CMP process. In future work, obtaining experimental data on pad surface height and wafer surface height pre- and post-CMP can help to further test and calibrate the model predictions.

Acknowledgments

This research was partly supported by the National Major Science and Technology special project no. 2006ZX01035-001-06 of China during the 11th five-year plan period, by NSFC research project no. 60676018 and no. 60806013, by the National Basic Research Program of China under grant no. 2005CB321701, by the doctoral program foundation of the Ministry of Education of China (no. 200802460068), by the International Science and Technology Cooperation program foundation of Shanghai (no. 08510700100), and by the program for Outstanding Academic Leader of Shanghai. W.C. was partly supported by the U.S. National Science Foundation (grant no. CCF0727791) and the U.S. Department of Energy (grant no. DEFG0205ER25678).

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

References