Journal of Engineering Mathematics **50:** 1–24, 2004.

A theory of pad conditioning for chemical-mechanical polishing

LEONARD J. BORUCKI , THOMAS WITELSKI[1], COLIN PLEASE[2] , PETER R. KRAMER[3] and DONALD SCHWENDEMAN[4]

Formerly Motorola Inc, Tempe, Arizona 85284 U.S.A.; E-mail: LenBorucki@intelligentplanar.com; [1]*Department of Mathematics, Duke University, Durham, North Carolina 27708-0320, U.S.A.;* [2]*Faculty of Mathematical Studies, University of Southampton, Southampton, England SO17 1BJ;* [3]*Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, New York 12180, U.S.A.;* [4]*Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, New York 12180, U.S.A.*

Received 22 april 2003; accepted in revised form 16 December 2003

Abstract. Statistical models are presented to describe the evolution of the surface roughness of polishing pads during the pad-conditioning process in chemical-mechanical polishing. The models describe the evolution of the surface-height probability-density function of solid pads during fixed height or fixed cut-rate conditioning. An integral equation is derived for the effect of conditioning on a foamed pad in terms of a model for a solid pad. The models that combine wear and conditioning are then discussed for both solid and foamed pads. Models include the dependence of the surface roughness on the shape and density of the cutting tips used in the conditioner and on other operating parameters. Good agreement is found between the model, Monte Carlo simulations and with experimental data.

Key words: abrasive wear, chemical-mechanical polishing, surface roughness

1. Introduction

Chemical-mechanical polishing (CMP) is a procedure for reducing height variations in material layers that have been deposited on silicon-wafer surfaces. Planarization considerably simplifies the task of building multilayer integrated-circuit structures while staying within the depth-of-focus limitations of photolithography tools. First applied to microelectronics by IBM in the early 1980s, CMP is now one of the fastest growing and most essential processing techniques in the electronics industry. In this article, we will introduce and investigate a model for a little-studied but important component of a CMP tool that is used to maintain process stability and uniformity.

In one large class of single-wafer CMP tools, the silicon wafer being processed is pressed against a rotating polishing pad; see Figure 1. During polishing, a chemically-reactive slurry containing abrasive particles is delivered on the pad ahead of the wafer. Pad-surface asperities that are tall enough to touch the wafer trap abrasive particles and drag them across the surface. This abrasive action, combined with chemical properties of the slurry, is responsible for the removal of material from the surface of the wafer. The same abrasive action, however, also produces wear and degradation of the polishing-pad surface. Our focus here is the modelling of the process used to maintain the surface of the pad. In the *conditioning* process, a diamond-embedded rotating disk is swept back and forth radially across the pad either during polishing (*in situ* conditioning) or between wafers (*ex situ* conditioning). The diamonds create microscopic cuts or furrows on the pad surface that continually uncover new pad material and maintain roughness.

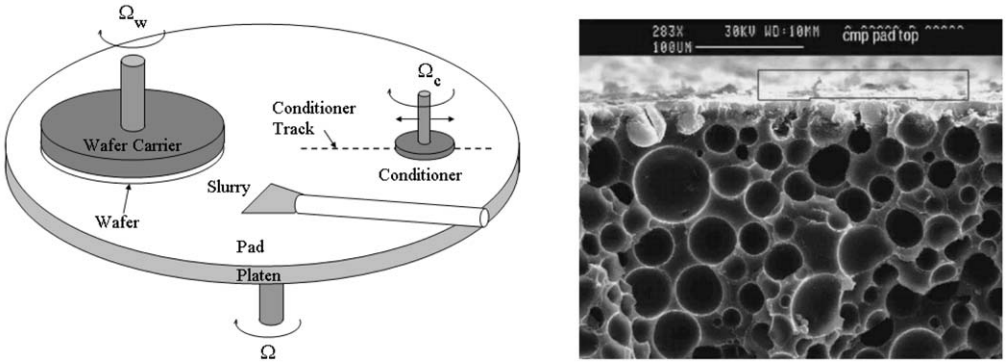

Figure 1. (Left) Schematic of a typical single-wafer rotary CMP tool. (Right) Scanning Electron Micrograph cross-section of a used, conditioned void-filled polyurethane polishing pad. Surface asperities can be seen at the top of the image. The scale bar at the top center is 100 microns (0·1 mm) long. Voids average about 30 microns in diameter and occupy about 60% of a planar cross-section.(Data by Letitia Malina, Motorola)

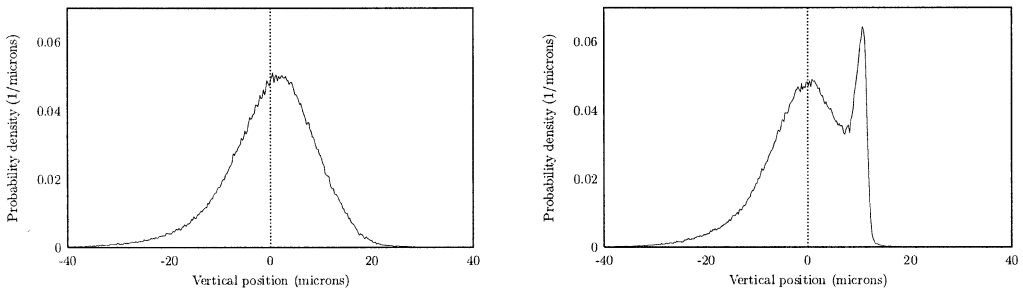

Figure 2. Surface height PDFs from a newly conditioned pad (left) and for the same pad after polishing wafers for 31 minutes without conditioning. (right) [1]. The surface heights are given relative to the mean. The left side of each distribution is determined by the pad void structure while the right side is influenced by conditioning and wear. The secondary peak in the right-hand PDF indicates that many high asperities have been abraded away and the surface has become more planar.

The details of the pad-surface roughness depend partly on the pad structure and partly on wear and conditioning. Polishing pads typically consist of relatively soft materials such as polyurethane. They are 1–2 mm thick and may be half a meter or more in diameter. Variations in the pad-surface height relative to the mean are on the order of tens of microns and their distribution is not generally Gaussian. The distribution of heights above the mean is controlled by pad conditioning and surface-wear processes. If the pad is manufactured from a foam, the form of the distribution far below the mean is determined by the void structure of the pad material. The physical structure of a foamed pad can be locally very non-uniform. A magnified cross-section of a common type of closed cell, void-filled polyurethane polishing pad is shown in Figure 1. The voids in this pad average 30 microns in diameter. In a planar cross-section, they occupy about 60% of the total area. The surface-height probability-density function (PDF) can be measured by optical interferometry. A typical measured PDF is shown in Figure 2 [1].

In the absence of conditioning, asperities that periodically come into contact with the wafer as the pad rotates are abraded by the slurry particles. This results in the formation of a secondary peak in the right-hand tail of the PDF that moves toward the primary peak and grows with time (Figure 2) [1]. The appearance of the secondary peak is associated with a

decline in the average polish rate. One explanation for this is that surface wear increases the actual contact area, thereby lowering the average real pressure at a fixed load. The reduction in the real contact pressure is then responsible for the decline in the polish rate.

In [2], it was shown that the evolution of the surface height PDF $\phi(z,t)$ due to slurry abrasion can be modeled by a Hamilton-Jacobi-like conservation equation,

$$\frac{\partial \phi}{\partial t} = \frac{4c_a E^* \sqrt{\kappa_s}}{3\pi} \frac{\partial}{\partial z} \left(H(z-d(t)) \sqrt{z-d(t)}\, \phi \right), \tag{1.1}$$

where H is the Heaviside step function, $z = d(t)$ is the vertical position of the wafer being polished, E^* is the pad Young's modulus divided by one minus the square of the Poisson ratio and κ_s is the mean asperity tip curvature. The parameter c_a is a wear coefficient that depends on the slurry characteristics and is proportional to the sliding speed between the pad and the wafer as described below. This equation produces a change in the PDF ϕ due to abrasion of asperities having height $z > d$ under the wafer. The variable t in (1.1) may be thought of as measuring the cumulative wear time. In Appendix A we derive modifications needed to this equation in order to combine it with the present theory.

Equation (1.1) applies to solid (void-free) pads for any amount of wear and to foamed pads if the wear effects are limited to a thin layer, on the order of the void length scale. In [2], Equation (1.1) was derived by assuming that the rate of decrease of asperity height is proportional to the product of the sliding speed and asperity tip contact pressure and by applying a conservation-law argument to balance the rate at which asperity heights enter an interval $(z, z+\Delta z))$ with the rate at which wear reduces the heights below z. When Equation (1.1) is considered in the (illustrative but unrealistic) case where the polishing surface stays at a fixed height above the initial surface of a solid pad, then it can be solved exactly using the method of characteristics,

$$\phi(z,t) = \begin{cases} \phi_0(z) & z < d, \\ \frac{\mathcal{W}}{2\sqrt{z-d}} \left(t + \frac{2}{\mathcal{W}} \sqrt{z-d} \right) \phi_0 \left(d + \frac{1}{4} \mathcal{W} (t + \frac{2}{\mathcal{W}} \sqrt{z-d})^2 \right) & z > d, \end{cases} \tag{1.2}$$

where $\phi_0(z)$ is the surface height PDF at the beginning of the wear process and $\mathcal{W} = 4c_a E^* \sqrt{\kappa_s}/(3\pi)$. This solution develops an integrable singularity at $z = d$ representing the portion of the pad-surface that has been worn smooth by the wafer. After a sufficiently long but finite time, this singularity converges to a δ-distribution with its amplitude approaching a limiting value given by the fraction of the pad-surface originally protruding above the wafer height d, $\int_d^\infty \phi_0(z)\, \mathrm{d}z$. In the more realistic case, when the vertical position of the wafer decreases with time as a result of load balance, the secondary peak may or may not involve a singularity (See Figure 14). It is known experimentally that, if conditioning is then resumed on a worn pad, the secondary peak in Figure 2b moves to the right, decreases in magnitude, and disappears. We will later solve the modified version of (1.1) numerically in this case and compare it with the data in Figure 2b (See Figure 15b). One objective of this paper is to extend the ideas in [2] to explain this behavior by accounting for not only wear due to polishing but also to consider simultaneously the conditioning process.

To make progress on the mathematical modeling of the conditioning process, we will employ several simplifying assumptions. First, we explain some of the nuances in the application of Equation (1.1) that are related to the actual details of the geometry in Figure 1 and derive a simpler time-averaged version of this equation. See Appendix A for the detailed analysis. Second, under appropriate conditions on the relative rotation rates of the pad and the conditioner, it is shown in Appendix B that the sweeping motion of a circular conditioner

yields a uniform density of furrows cut over a large portion of the polishing pad. This analysis is used to replace the circular conditioner by a simpler one-dimensional "bar conditioner" used in our mathematical model. The diamonds on the bar conditioner are assumed to create furrows in the pad by cleanly cutting the pad material. This makes it possible to avoid inclusion of a mechanical theory to describe furrow formation by a combination of plastic deformation (in which material is simply pushed aside) and cutting as described, for example, in [3] or [4]. Given these simplifications, we use the bar conditioner to develop a model for surface-roughness development on solid pads for fixed height conditioning and for constant cut-rate conditioning. A key relation involving the intrinsic roughness of a foamed pad is then introduced that allows us to extend the solid-pad-conditioning models to foamed pads. Finally, we develop a model that combines wear and conditioning on solid pads and discuss the corresponding model for foamed pads. The latter extends Equation (1.1) so that it applies for any amount of wear of a foamed pad.

2. The basic conditioning problem

We consider a two-dimensional radial cross-section of the pad that encounters the equivalent conditioner once per pad rotation. The cutting surface of this tool has an array of identical diamond tips, which for simplicity will be assumed to be triangular with opening angle α, with mean separation distance ℓ given by Equation (B3); see Figure 3. The theory can be easily generalized to cutting tips with more general shapes. While in a real conditioner the diamond tips protrude a finite distance, we will assume that the protrusion is large enough that we can neglect the presence of the conditioner plate. Compared to the length scales for the conditioning process set by the diamond tips, the pad dimensions are very large, so the pad will be assumed to be semi-infinite in the vertical direction. The lateral direction is divided into intervals of length ℓ, each containing one diamond tip situated at a uniform random location within its interval during each cut. In our model, the conditioning process will be statistically homogenous along the radial direction so that the statistical structure of the pad does not depend on the radial position.

We start by modeling conditioning without wafer wear, the *ex situ* case. The goal is to describe the evolution of the surface of the polishing pad in which the circular conditioner has been replaced by the equivalent one-dimensional "bar" conditioner (see Appendix B). The key simplifying assumption is that, given the position of the conditioner, the ability of any diamond to cut the pad material is not dependent on the shape of the pad-surface it encounters. Pad material that has been cut away is assumed to then be washed away by the slurry rather than being redeposited elsewhere.

We introduce a coordinate system based on the initial state of the pad. We use x to describe the horizontal distance from a fixed point on the pad radial cross-section. Vertical positions will be measured relative to the initial position of the pad-surface, $z = 0$. For solid pads, the initial pad-surface will be assumed to be smooth. For pads made from "foamed" materials, it will be assumed that a horizontal planar cut has been made at $z = 0$ and that the portion of the pad above $z = 0$ has all been removed. When plotting distributions it is common engineering practice to plot surface-height measurements relative to the average height as in Figure 2. For the purposes of the subsequent analysis, however, it is more useful to define the distributions relative to the z-coordinate defined here, which is fixed in the laboratory reference frame.

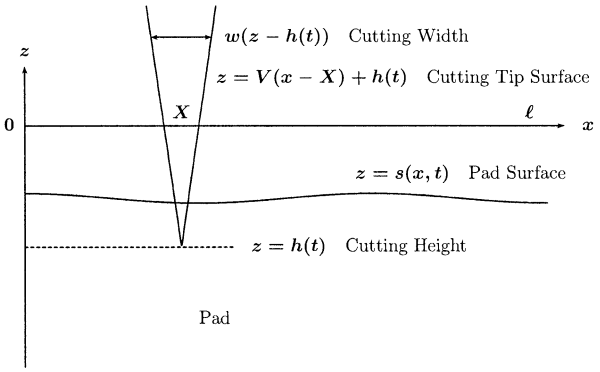

Figure 3. The coordinate system and terminology used in the paper. A cutting tip on the conditioner face is shown displaced horizontally by a random amount X and vertically by $h(t)$.

The profile of each of the diamond cutting tips will thus be represented by the function

$$V(x) = \nu|x|, \qquad -\ell/2 < x < \ell/2, \tag{2.1}$$

where $\nu = \cot(\alpha/2)$. We will assume that variations in the pad-surface are always smaller than $\nu\ell/2$, so that the conditioner only cuts part of the surface on any pass. In our analysis, we will make use of the width

$$w(z) = 2z/\nu, \qquad z \geq 0. \tag{2.2}$$

of the cutting tip at a vertical distance z above its apex.

The vertical position of a solid pad-surface can be expressed as a single-valued function $z = s(x, t)$; we will generalize our model to foamed pads later in Section 5. We model the evolution of a small fixed region on the pad cross-section as it interacts with the conditioner on successive rotations of the pad. If the rotation rate of the platen is Ω radians/sec, these encounters occur at discrete times t_n, where

$$t_n = 2\pi n/\Omega. \tag{2.3}$$

We describe conditioning as a discrete random process in which the surface of the pad is modified as a result of each encounter with the conditioner, $s(x, t_n) \rightarrow s(x, t_{n+1})$. Hence, within any interval associated to a single diamond cutter (say the interval $[0, \ell]$), a single pass of the conditioner results in the transformation

$$s(x, t_{n+1}) = \min\left(s(x, t_n), V(x - X_n) + h(t_n) \right), \tag{2.4}$$

where $h(t_n)$ gives the vertical position for the cutting tip and X_n is a uniformly-distributed random variable on $[0, \ell]$ that gives the horizontal position of the tip on the n^{th} pass of the conditioner. In order for our model to respect the statistical homogeneity of the true conditioning process (and not behave differently near the edges of the interval $[0, \ell]$), we interpret (2.4) with periodic boundary conditions so that a diamond cutter centered near the right edge of the interval will also cut a bit near the left edge of the same interval (rather than the adjacent interval belonging to the next diamond cutter). A Monte Carlo model similar to the one employed here is described in [5]; however, we go beyond [5] by also developing analytic models that are specific to the pad conditioning problem.

We assume that the vertical position of the conditioner, as measured by the position of the cutting tips, $z = h(t)$, is a prescribed function of time that is specified as part of the conditioning process. We will consider two simple cases for positioning the conditioner: (i) when $h(t) = h_0$ is a constant height, and (ii) when the height $h(t)$ decreases at the cut rate c. In these simplified cases, the assumption that $h(t)$ is not influenced by pad asperities allows us to consider the steps in (2.4) to be independent.

We will describe the structure of the surface in terms of the function $q(z, t)$, defined as the fraction of the surface which remains above height z after time t. More precisely, we think here of sampling the surface structure over a radial interval of length L, and defining $q(z, t)$ as the measure of the set of x values in this interval for which $z > s(x, t)$, divided by L. We expect universal behavior for $q(z, t)$ (no dependence on L nor significant variations from experiment to experiment) when L is chosen large enough to contain sufficiently many independent cuts so as to adequately sample the statistics of the spatial structure. All of our model calculations will make predictions for the universal statistics of the spatial structure, which will be relevant for $L \gg \ell$ (with $L < r_p$) at all times or for $L = \ell$ after enough time so that each interval of this length contains many (independent) cuts (see Appendix B).

Because the cutting process is statistically homogenous along the radial direction, we can equivalently write

$$q(z, t) = \text{Prob}(z < s(x, t)) \tag{2.5}$$

for any choice of horizontal position x, where the probability refers to the randomness in the cutting process. It is also useful to characterize the surface in terms of the probability that the height of a point on the surface does not exceed z,

$$p(z, t) = \text{Prob}(z \geq s(x, t)) = 1 - q(z, t), \tag{2.6}$$

and the function

$$\phi(z, t) = \frac{\partial p(z, t)}{\partial z} = -\frac{\partial q(z, t)}{\partial z}. \tag{2.7}$$

The function $p(z, t)$ is the cumulative density function (CDF) for surface height, $q(z, t)$ is the complementary cumulative density function (CCDF), and $\phi(z, t)$ is the surface-height probability-density function (PDF).

To describe the progress of the conditioning process, we will track the average surface height over a sample radial interval of length L:

$$\bar{s}(t_n) = \frac{1}{L}\int_0^L s(x, t_n)\,\mathrm{d}x = \int_{h(t_n)}^0 z\phi(z, t_n)\,\mathrm{d}z, \tag{2.8}$$

and the RMS (root-mean-square) roughness, as measured by the standard deviation, $\sigma(t_n)$,

$$\sigma^2(t_n) = \frac{1}{L}\int_0^L [s(x, t_n) - \bar{s}(t_n)]^2\,\mathrm{d}x = \int_{h(t_n)}^0 [z - \bar{s}(t_n)]^2\phi(z, t_n)\,\mathrm{d}z. \tag{2.9}$$

3. A fixed-height conditioner on a solid pad

To start the analysis of the model, consider the simplest case where the conditioner is maintained at a constant height, $h(t) = h_0$ (where $h_0 < 0$), and the pad is solid. A typical sequence

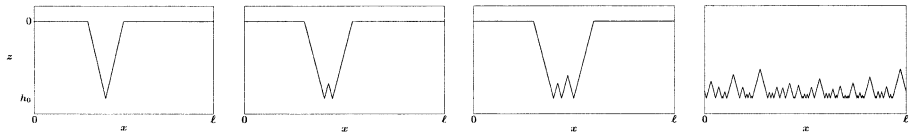

Figure 4. A typical sequence of solid pad-surface profiles after successive steps of the discrete conditioning process (2.4), for $n = 1, 2, 3$, and $n = 50$ with an initially flat pad-surface.

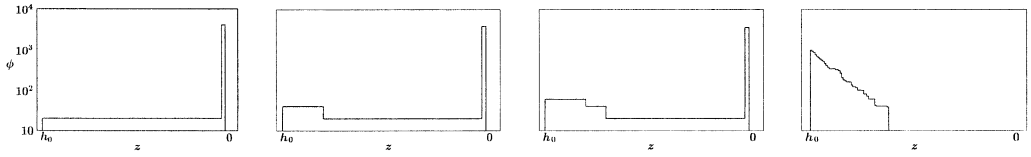

Figure 5. Pad surface height distributions generated by (2.4) corresponding to the sequence shown in Figure 4.

of pad-surfaces generated by (2.4) was calculated using a Monte Carlo simulation; see Figure 4. In this figure, the pad-surface was initially smooth, $z = s(x, 0) = 0$, corresponding to a surface distribution $\phi(z, 0) = \delta(z)$. The surface-height distribution $\phi(z, t_n)$ for a single experiment over a sampling interval $L = \ell$ at corresponding steps is shown in Figure 5. We observe that the conditioning process gradually converts the initial smooth surface to another nearly worn-smooth surface at $z = h_0$ with a truncated PDF on $h_0 \le z \le 0$. Figure 5 shows the amplitude of the δ-distribution corresponding to the initially smooth surface ($z = 0$) gradually decreases as the distribution of the surface shifts toward $z = h_0$. A plot of the roughness integral (2.9) confirms that the roughness initially grows quickly, but decreases for long times; see Figure 7.

The surface-height distribution for this example is not universal over the sampling interval $L = \ell$ for small n, and will look somewhat different from experiment to experiment. However, for even moderately large n, the spatial statistics over the interval $L = \ell$ are well described by the universal formulas for the conditioning process which we now develop.

It is possible to study this process analytically by considering the shape of the pad-surface at each step and then accounting for all possible scenarios for the positions where the successive steps occur. However, this is algebraically tedious and impractical after a few steps. Therefore we consider a different approach.

The central idea is to consider the probability $q(z, t_n)$ that after n cuts, the surface height at a given point x remains above a certain level z. Before the first cut, we assume the pad is uniform for all $z \le 0$ with $s(x, 0) = 0$, so $q(z, 0) = H(-z)$. At later times, $q(z, t)$ will only be modified for vertical positions $h_0 \le z \le 0$ cut by the conditioner, and we focus on this range of values for the moment. At each successive step, the surface height will be cut if and only if the center of the horizontal position of the cutting tip falls within a distance $\frac{1}{2}w(z-h_0) = (z-h_0)/\nu$ of the given point x (taking periodic boundary conditions). This event happens with probability $w(z-h_0)/\ell = 2(z-h_0)/(\nu\ell)$. We consider only the case where the conditioner scrapes a very small part of the surface per revolution, so that the nondimensional parameter

$$\lambda = \frac{2h_0}{\nu\ell}, \qquad (\lambda \le 0) \tag{3.1}$$

satisfies $|\lambda| \ll 1$.

The location of successive cuts is independent of all previous events. Consequently, the probability of survival of the surface at position x above a height z can after n cuts be modeled

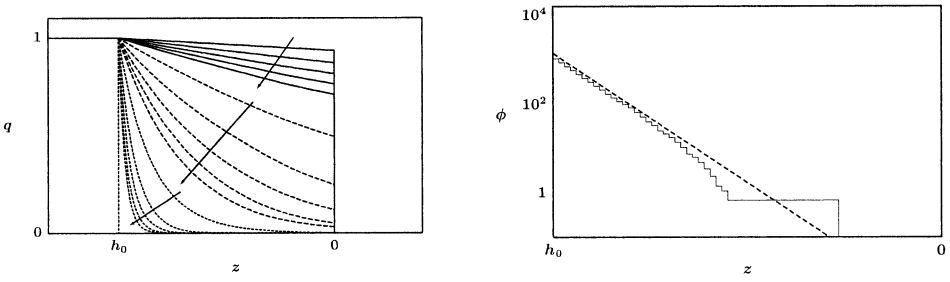

Figure 6. (left) Development of the CCDF $q(z,t)$ (3.8) corresponding to the simulation in Figure 4 for $n = 10, 20, ...50, 100, 200, ..., 500, 1000, 2000, ...5000$. (right) Comparison of the PDF produced by the Monte Carlo simulation for $\nu = 0{\cdot}5$, $\ell = 1$, $h_0 = -0.01$ at $n = 2000$, with the long-time asymptotic PDF (3.6).

as the probability of n identically, independently distributed failures for the cutting tip to come close enough to x. Generalizing our result to apply also for arbitrary z, we therefore achieve the formula

$$q(z,t_n) = H(-z)\left(1 - \frac{1}{\ell}w(z-h_0)H(z-h_0)\right)^n. \tag{3.2}$$

It then follows that the cumulative distribution function is

$$p(z,t_n) = 1 - H(-z)\left(1 - \frac{2}{\nu\ell}(z-h_0)H(z-h_0)\right)^n, \tag{3.3}$$

where we have used (2.2) for $w(z-h_0)$. By differentiating $p(z,t_n)$ with respect to z, we obtain the PDF,

$$\phi(z,t_n) = H(-z)H(z-h_0)\left(1 - \frac{2}{\nu\ell}(z-h_0)\right)^{n-1}\left[\frac{2n}{\nu\ell} + \left(1 + \frac{2h_0}{\nu\ell}\right)\delta(z)\right]. \tag{3.4}$$

After many steps in the process (2.4), the influence of the initial condition, $\phi(z,0) = \delta(z)$, becomes negligible (Fig. 5a,b,c) and the long-time PDF can be approximated by

$$\phi(z,t_n) = \frac{2n}{\nu\ell}\left(1 - \frac{2}{\nu\ell}(z-h_0)\right)^{n-1} + O(n^{-1}), \qquad h_0 \le z \le 0, \qquad n \to \infty, \tag{3.5}$$

see Figure 5d. Since the conditioner is assumed to cut once per pad revolution and the pad revolves at a rate Ω we can interpret this for continuous times as

$$\phi(z,t) \sim \frac{\Omega t}{\pi\nu\ell}\exp\left[\frac{\Omega}{2\pi}\log\left(1 - \frac{2}{\nu\ell}(z-h_0)\right)t\right], \qquad t \to \infty. \tag{3.6}$$

We can simplify this expression further using the assumption $|\lambda| \ll 1$ to yield an exponential distribution in z for $h_0 \le z \le 0$,

$$\phi(z,t) \sim \frac{\Omega t}{\pi\nu\ell}\exp\left[-\frac{\Omega}{\pi\nu\ell}(z-h_0)\,t\right], \qquad t \to \infty. \tag{3.7}$$

In Figure 6a we show the evolution of the complementary cumulative density function (CCDF) $q(z,t)$, defined by

$$q(z,t) = 1 - \int_{h_0}^{z}\phi(z',t)\,\mathrm{d}z'. \tag{3.8}$$

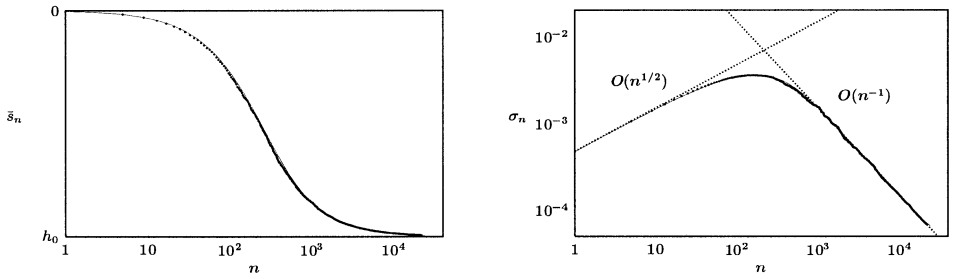

Figure 7. The average surface height (2.8) and RMS surface roughness (2.9) for the conditioning process in Figure 4. Initially, the roughness increases rapidly, but for long times, the roughness decreases as the amplitude scale of surface variations decrease. The dots shows the direct results of a Monte-Carlo simulation of (2.4), calculated using the spatial integrals in (2.8, 2.9). The almost indistinguishable solid curves are the calculated expected values from the probabilistic model (3.4).

for the sequence of conditioning steps considered in Figure 4. This shows how the distribution with the step due to the initial smooth surface transforms to become a step describing a new smooth surface. We note that the CCDF can be interpreted as the density of the remaining pad material at a given height z at time t.

While the expression (3.4) is more complicated than the limiting distribution (3.7), it has significant value since it describes the conditioning process at all times. From (2.8), we can use (3.4) to obtain the average surface height as,

$$\bar{s}(t_n) = \frac{2n}{\nu\ell}\int_{h_0}^{0} z\left(1 - \frac{2}{\nu\ell}[z - h_0]\right)^{n-1} \mathrm{d}z. \tag{3.9}$$

It is convenient to non-dimensionalize this expression in terms of the parameter λ from (3.1):

$$\bar{s}(t_n) = nh_0\lambda\int_{1}^{0} y(1 - \lambda[y-1])^{n-1}\,\mathrm{d}y = \frac{h_0}{(n+1)\lambda}\left((n+1)\lambda + 1 - (1+\lambda)^{n+1}\right). \tag{3.10}$$

For $|\lambda| \ll 1$, this expression may be simplified to

$$\bar{s}(t_n) = \frac{h_0}{n\lambda}(1 + n\lambda - \mathrm{e}^{n\lambda})(1 + O(|\lambda|^{1/2})), \tag{3.11}$$

which is uniformly accurate for all n. Hence in the following limiting cases the average surface height is given by

$$\bar{s}(t_n) \sim \begin{cases} -\frac{1}{2}h_0\lambda n, & n \ll 1/|\lambda|, \\ h_0 + h_0/(\lambda n), & n \gg 1/|\lambda|. \end{cases} \tag{3.12}$$

For short times, the surface height decreases linearly; for long times, it converges to the cutting height $z = h_0$. (See Figure 7a).

Similarly, we can obtain the RMS roughness as

$$\sigma^2(t_n) = \bar{s}^2\,(1+\lambda)^n + nh_0^2\lambda\int_{1}^{0}\left[y - \frac{\bar{s}}{h_0}\right]^2(1 - \lambda[y-1])^{n-1}\,\mathrm{d}y, \tag{3.13}$$

with the following uniformly valid approximation for small $|\lambda|$:

$$\sigma^2(t_n) = \frac{h_0^2}{\lambda^2 n^2}\left(1 + 2n\lambda\mathrm{e}^{n\lambda} - \mathrm{e}^{2n\lambda}\right)(1 + O(|\lambda|^{1/2})). \tag{3.14}$$

In the limiting cases, we find

$$\sigma^2(t_n) \sim \begin{cases} -\frac{1}{3}h_0^2\lambda n, & n \ll 1/|\lambda|, \\ h_0^2/(\lambda n)^2, & n \gg 1/|\lambda|. \end{cases} \tag{3.15}$$

Hence for short times the surface roughness increases while for much longer times the roughness decreases as the pad approaches the surface $z = h_0$. This implies that there exists an intermediate time when the surface roughness is maximized. From (3.15), this time is approximately

$$n_{\max} \approx \frac{3^{1/3}}{|\lambda|}, \tag{3.16}$$

see Figure 7b. This analysis suggests that there is an optimal conditioning time as with insufficient conditioning the surface is relatively smooth and planar (and therefore produces low actual contact pressures) while for long times the surface also approaches a planar limit. We have not attempted here to derive measures of the expected mechanical polishing rate. Further discussion of the relevant rough surface contact mechanics can be found in [8–11].

4. A moving conditioner on a solid pad

Having gained insight into the behavior of the static conditioner, we now extend these ideas to account for a conditioner that moves in a deterministic manner into the pad with $h(t_n)$ as a given function. Again we consider the material at a particular height z and look at a single step of the process. The fraction of pad material remaining at vertical position z after the $(n+1)^{\text{st}}$ step in the conditioning process, when the diamond tips are at position $z = h(t_{n+1})$, is

$$q(z, t_{n+1}) = \left(1 - \frac{2}{\nu\ell}\Big(z - h(t_{n+1})\Big)\, H(z - h(t_{n+1}))\right) q(z, t_n), \tag{4.1}$$

given that the state at the previous step is described by $q(z, t_n)$. This Equation for the transition on the $(n+1)^{\text{st}}$ step can be rewritten as

$$\frac{q(z, t_{n+1}) - q(z, t_n)}{t_{n+1} - t_n} = -\frac{\Omega}{\pi\nu\ell}\Big(z - h(t_{n+1})\Big)\, H(z - h(t_{n+1}))\, q(z, t_n), \tag{4.2}$$

where we have introduced $t_{n+1} - t_n = 2\pi/\Omega$ using (2.3). To describe the process on time scales much larger than $1/\Omega$, we can approximate the difference quotient on the left with a time derivative of q to yield a differential Equation for $q(z, t)$,

$$\frac{\partial q}{\partial t} = -\frac{\Omega}{\pi\nu\ell}\big(z - h(t)\big)\, H(z - h(t))\, q. \tag{4.3}$$

This is the basic Equation that governs the changes to the distribution created by a conditioner with triangular cutting tips. The problem is fully specified if the parameters describing the structure of the conditioner, (Ω, ℓ, ν), the motion of conditioner, $h(t)$, and the initial structure of the pad-surface, $q(z, 0)$, are known. We note that, if the conditioner is fixed, that is $h(t) = h_0$, (4.3) yields the same exponential controlling factor as given by the analysis in the previous section, (3.7). For general cutting tips, with shape defined by $w(z)$, the corresponding PDE is

$$\frac{\partial q}{\partial t} = -\frac{\Omega}{2\pi\ell} w(z - h(t))\, H(z - h(t))\, q. \tag{4.4}$$

The above generalizion will be used later to compare the theory with data – see Figure 15.

Some additional insight into the behavior of the conditioner can be found by considering the simple case where the conditioner is taken to move into the pad at a constant speed c,

$$h(t) = h_0 - ct. \tag{4.5}$$

Such a situation would occur if a constant force were applied to the conditioner and initial transients had subsided. The density distribution would then obtain a steady shape that would also move down at constant speed c. We now seek the structure of this steady distribution.

Adopting the moving coordinate Z with $Z = z - h(t)$ and with $Q(Z,t) = q(z,t)$ the Equation becomes

$$\frac{\partial Q}{\partial t} + c\frac{\partial Q}{\partial Z} = -\frac{\Omega}{\pi \nu \ell} Z H(Z) Q. \tag{4.6}$$

Neglecting initial transient behavior, this PDE reduces to an ordinary differential Equation for the long-time steady profile traveling wave solution $Q(Z)$,

$$c\frac{\mathrm{d}Q}{\mathrm{d}Z} = -\frac{\Omega}{\pi \nu \ell} Z H(Z) Q. \tag{4.7}$$

We also have the condition that $Q(0) = 1$ to represent the fact that the pad is solid ahead of the moving conditioner, for $Z < 0$. Hence we conclude that

$$Q(Z) = \exp\left(-\frac{\Omega}{2\pi c \nu \ell} Z^2\right) \qquad 0 \le Z \le -h(t). \tag{4.8}$$

Hence the CCDF is Gaussian. We then obtain the surface height density distribution from (2.7) as

$$\phi(z,t) = \frac{\frac{\Omega}{\pi c \nu \ell}(z - h_0 + ct)}{1 - \exp\left(-\frac{\Omega}{2\pi c \nu \ell}(h_0 - ct)^2\right)} \exp\left(-\frac{\Omega}{2\pi c \nu \ell}\Big(z - h_0 + ct\Big)^2\right), \tag{4.9}$$

for $h_0 - ct \le z \le 0$. This is a linear distribution near the cutting height $z = h(t)$, while for $z \gg h(t)$ it is nearly Gaussian. This distribution is not comparable with (3.7) since the limit of a fixed conditioner, $c \to 0$, is a singular limit of (4.9).

The above argument is easily extended to apply to general functions $w(z)$ for the diamond width. The steady state CCDF, which we simply state here, is

$$Q(Z) = \exp\left(-\frac{\Omega}{2\pi c \ell}\int_0^Z w(\zeta)\,\mathrm{d}\zeta\right). \tag{4.10}$$

This generalization will be used in Section 5 to model the freshly conditioned pad PDF in Figure 2.

We continue for the moment with triangular diamond shapes. Setting $\tilde{h}_0 = h_0 - ct$, and $z = \tilde{h}_0 y$, we can describe the distribution in terms of a dimensionless parameter, Λ, representing the ratio of the pad revolution speed Ω to the cutting speed c:

$$\Lambda = \frac{\tilde{h}_0^2 \Omega}{2\pi c \nu \ell}. \tag{4.11}$$

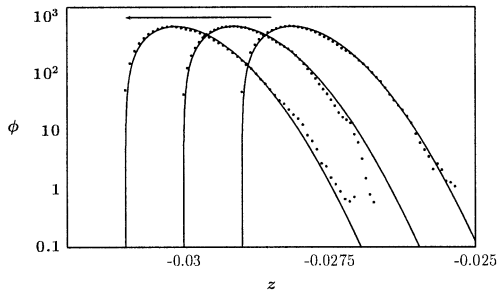

Figure 8. Snapshots of the surface height PDF at equally spaced times. Results from Monte Carlo simulations are shown by dots, the solid curves are the prediction from (4.9).

Note that Λ is dependent on time through $\tilde{h}_0$ and that $\Lambda \to \infty$ as $t \to \infty$. Then the average pad height is

$$\bar{s}(t) = \frac{2\Lambda\tilde{h}_0}{1 - \mathrm{e}^{-\Lambda}} \int_1^0 y(y-1) \exp\left(-\Lambda(y-1)^2\right) \mathrm{d}y. \tag{4.12}$$

This integral can be evaluated exactly to yield

$$\bar{s}(t) = \left(\frac{1}{1 - \mathrm{e}^{-\Lambda}}\right)\left(1 - \frac{\sqrt{\pi}\,\mathrm{erf}(\sqrt{\Lambda})}{2\sqrt{\Lambda}}\right)\Big(h_0 - ct\Big), \tag{4.13}$$

the asymptotics of this expression for $t \to \infty$ yield that

$$\bar{s}(t) \sim \left(h_0 - \sqrt{\frac{c\nu\ell\pi^2}{2\Omega}}\right) - ct \qquad t \to \infty. \tag{4.14}$$

That is, the mean of the surface height lags behind the cutting height $h(t)$ by a known amount related to the cutting speed and other conditioner parameters.

Similarly, the integral for the variance (2.9) can also be evaluated exactly to yield

$$\begin{aligned}\left(\frac{\sigma(t)}{h_0 - ct}\right)^2 \;=\; & \frac{1}{\Lambda}\left(\frac{1 - \mathrm{e}^{-\Lambda}[1+\Lambda]}{1 - \mathrm{e}^{-\Lambda}}\right) + \sqrt{\frac{\pi}{\Lambda}}\,\frac{\mathrm{erf}(\sqrt{\Lambda})}{1 - \mathrm{e}^{-\Lambda}} \\ & \left(\mathrm{e}^{-\Lambda} - \frac{1}{4}\sqrt{\frac{\pi}{\Lambda}}\,\mathrm{erf}(\sqrt{\Lambda})[1 + \mathrm{e}^{-\Lambda}]\right)\end{aligned} \tag{4.15}$$

and similarly using asymptotics for $t \to \infty$, we find the limiting RMS roughness behaves as

$$\sigma^2(t) \sim \frac{(4-\pi)\pi c\nu\ell}{2\Omega} \qquad t \to \infty. \tag{4.16}$$

Figures 8 and 9 compare these analytical predictions with results from Monte Carlo computer simulations of the conditioning process. Figure 8 shows excellent agreement of the PDF with (4.9) and Figure 9 shows rapid convergence to the predicted asymptotic results as time increases.

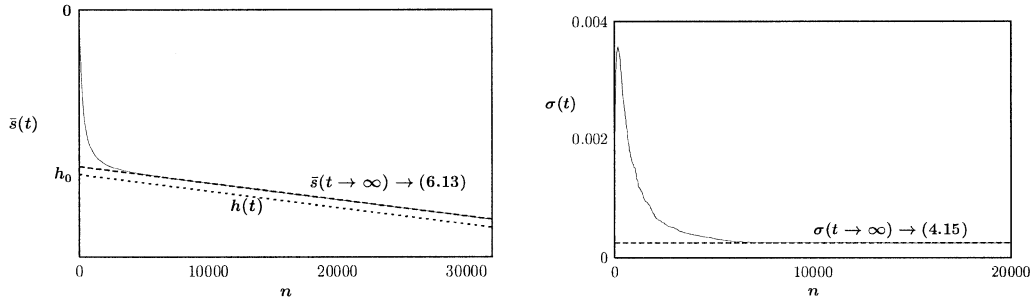

Figure 9. Comparison of the theory and Monte Carlo simulations for the mean value and variance of the surface height with a moving conditioner: the mean lags behind the cutting height by a constant amount (left) and for long time the variance approaches a constant given by (4.16) (right).

5. Conditioning of foamed pads

There are two important differences between a solid pad and a foamed pad as shown in Figure 1. The first is that due to the presence of spherical voids, the surface of a foamed pad is generally not a single-valued function $z = s(x, t_n)$, see Figure 10a. However, we do not need to model the surface in its full generality. Instead, we are interested only in that part of the full surface that can be measured by line-of-sight optical interferometry. Using this simplification, the visible surface is single-valued but may be discontinuous where it drops into an exposed void (Figure 10a).

The second difference is that the distribution of void sizes in the pad material adds other length-scales into the conditioning problem. For a foamed pad, a period interval of length ℓ may show significant non-uniformity in the spatial properties of the pad as compared with other randomly selected slices of length ℓ, and these "quenched" nonuniformities will influence the statistics of the surface sample even after many cuts. We will assume that these non-uniformities are statistically homogeneous in the sense that taken over a large number of sample cross-sections of length ℓ, or over a large number of experiments, their statistics do not depend on position. In numerical experiments we need to consider the conditioning process over a much larger sampling interval L in order to observe the universal statistics of the surface height. For convenience, we take $L = N\ell$ with N a large integer.

The Monte Carlo simulation of conditioning of foamed pads involves a discrete representation of the pad volume, rather than just its surface (as was sufficient for a solid pad). We discretize a pad cross-section of length L and sufficient depth into a uniform Cartesian grid. For each cell in the grid, the simulation will keep track of whether the pad is solid there, $\rho_{i,j} = 1$, or whether the cell is in a void, $\rho_{i,j} = 0$, where the indexing corresponds to the position $x_i = i\Delta x$, $z_j = j\Delta z$. To construct the simulated foamed pad (see Figure 10b), we start from a solid pad, with $\rho_{i,j} = 1$ at each cell. Void center locations and void radii are then generated one at a time from a statistical model and $\rho_{i,j}$ is set to 0 at cells that lie inside a void. Figure 10b shows a small section of a simulated foamed pad that has been generated by this process. The full simulation has length $L = 200\ell$ and contains over 100, 000 disjoint circular voids.

The pad-surface $z = s(x_i, t_n)$ is found by scanning each column of $\rho_{i,j}$ from the top down until the first occurrence of a $\rho = 1$, that is,

$$s(x_i) \equiv \max_j \{z_j | \rho_{i,j} = 1\}. \tag{5.1}$$

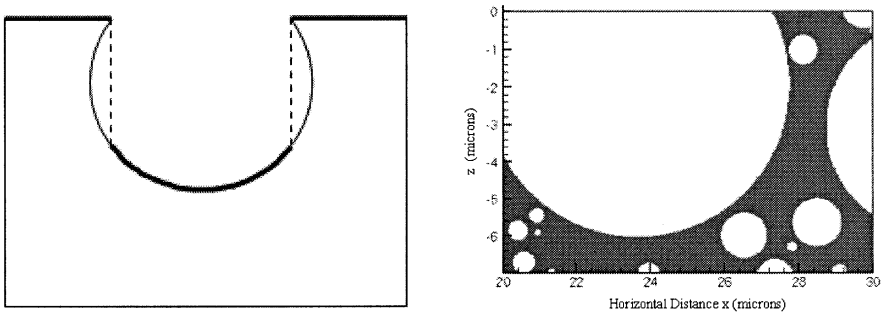

Figure 10. (left) The line-of-sight surface (heavy curve) on a foamed pad may be discontinuous. (right) A small section of a simulated 20,000 micron wide foamed pad model.

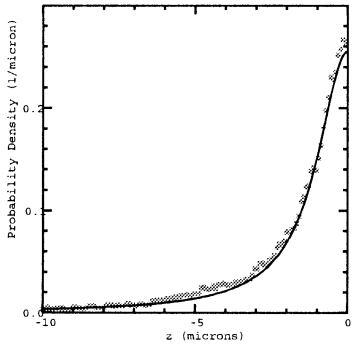

Figure 11. Intrinsic PDF (circles) of the line-of-sight surface of the foamed pad model in Fig. 10b. The solid curve is the best least squares fit for $z < 0$ to the Monte Carlo simulation surface data with the function $\Phi(z) = 1/(a_0 + a_2 z^2) + a_3\delta(z)$. (The delta function at $z = 0$ is not shown in the PDF.)

The conditioning process is then done using a moving array of N equally-spaced cutting tips with a randomly generated shift X_n, as before (see Figure 12a). Pad material is removed by zeroing $\rho_{i,j}$ in all cells above the cutting surface, that is

$$\rho_{i,j}(t_{n+1}) = \begin{cases} \rho_{i,j}(t_n) & \text{if } z_j < C(x_i, t_n), \\ 0 & \text{if } z_j \geq C(x_i, t_n), \end{cases} \tag{5.2}$$

where the conditioner cutting surface is

$$C(x, t_n) = h(t_n) + \sum_{k=1}^{N} V(k\ell - X_n). \tag{5.3}$$

Now consider the statistics of the surface of a foamed pad prior to cutting. When a large sample of pad is sliced open with a planar cut, the exposed surface has some "intrinsic" or natural surface height CCDF $\mathcal{Q}(z)$. For the case of a solid pad, $\mathcal{Q}(z)$ is just the Heaviside function $H(-z)$, but for a foam, $\mathcal{Q}(z)$ is determined by the density of voids, the distribution of void sizes and the fraction of the pad that lies exactly on the cut plane. We consider only foamed pads that are statistically homogeneous in the sense that the intrinsic CCDF is independent of the location and orientation of the cut. While it should sometimes be possible to derive an analytic model for $\mathcal{Q}(z)$ from the method used to construct the voids, we will instead approximate $\mathcal{Q}(z)$ using a numerical fit to Monte Carlo data (Figure 11).

To understand what is happening for conditioning of a foamed pad, imagine that a solid pad is simultaneously being cut in an identical manner as shown in Figure 12a. Let the surface

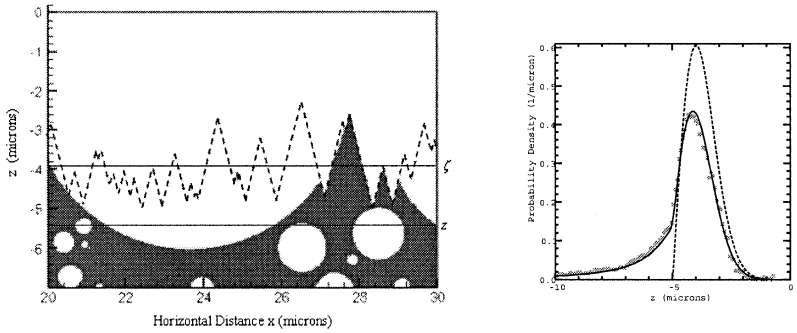

Figure 12. (left) The section of the pad model in Figure 10b after 1000 conditioning passes. Also shown (dashed curve) is the corresponding surface of a solid pad that has received an identical treatment. The CCDF for the foamed pad can be derived from that of the solid pad and the intrinsic CCDF $\mathcal{Q}(z)$. The PDF of the solid pad provides information about the density of points near height ζ on the dashed curve while the intrinsic CCDF of the foamed pad determines the probability of finding points on the foamed surface between ζ and z. (right) Monte Carlo (circles) and analytic (solid) PDFs of the pad model in Figure 12. Also shown (dashed) is the analytic PDF for the identically conditioned solid pad. The transition at about $z = -5$ μm marks the depth of the conditioner tips.

of the foamed pad be $s_f(x,t)$ (f for "foamed") and that of the virtual solid pad be $s(x,t)$. Then for any fixed x, we can write $s_f(x,t) = s(x,t) + U$, where U is the (negative) vertical displacement of the void lying immediately below the virtual solid pad-surface. According to our void model assumptions, U is a random variable which is independent of $s(x,t)$ and has $\mathcal{Q}(z)$ as its CCDF. Consequently, the CCDF $q_f(z,t)$ for the conditioned foamed surface height is just given by the convolution of the CCDF $q(z,t)$ for the virtual solid pad-surface height and the probability density $\Phi(z) = -\mathrm{d}\mathcal{Q}/\mathrm{d}z$ for the intrinsic surface height of the foamed pad [12, pp. 143–148]:

$$q_f(z,t) = \int_z^0 q(\zeta,t)\Phi(z-\zeta)\,\mathrm{d}\zeta, \tag{5.4}$$

As an example, suppose that the surface of the virtual solid pad moves from $z = 0$ to $z = h$ by any process, such as wear, constant height conditioning or slicing. Then the steady-state CCDF for the solid pad is $H(h-z)$ and

$$q_f(z) = \int_z^0 H(h-\zeta)\Phi(z-\zeta)\,\mathrm{d}\zeta = \int_z^h \Phi(z-\zeta)\,\mathrm{d}\zeta = \mathcal{Q}(z-h). \tag{5.5}$$

Thus, the final foamed pad CCDF is just a translation of the initial CCDF.

In Figure 12b we compare the Monte Carlo PDF for the pad in Figure 12a after 1000 passes with the analytic model obtained from the above convolution. The parameters for the conditioner cutting tips and the cutting speed are the same as in the previous comparisons.

In Figure 13, we show the best fit to the newly conditioned foam pad data in Figure 2a that can be obtained from Equation (5.4) using an intrinsic PDF of the form $\Phi(z) = a_0 \exp(a_1 z)H(-z) + a_2\delta(z)$, the general steady-state solid pad model for a moving conditioner (4.10) and a power law $w(z) = b_0 z^{b_1}$ for the diamond shape. The optimal Φ is also shown. The strength $a_2 \approx 0{\cdot}4$ of the delta function in Φ was taken to correspond to the fraction of non-void area in a planar cross-section of a pad sample. For any value of a_1, a_0 was then chosen to make the function integrate to 1. This intrinsic PDF therefore has one free parameter. Values

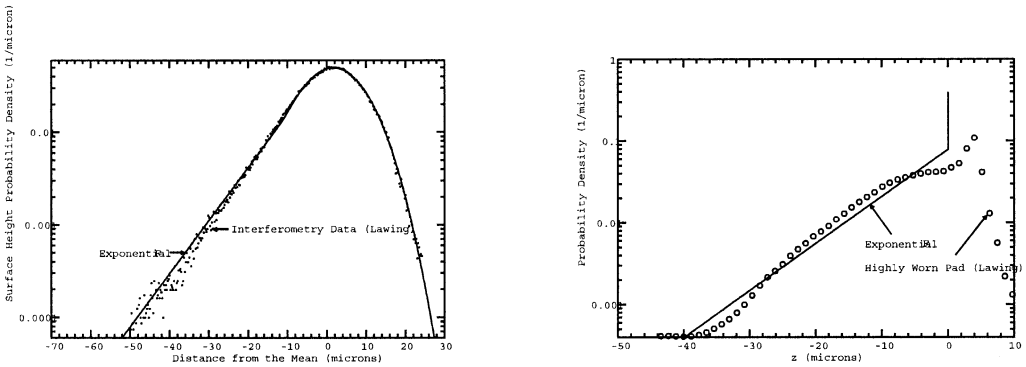

Figure 13. Left: Best fit of Equation (5.4) to the newly conditioned pad data in Figure 2a using an exponential intrinsic PDF with a delta function at $z = 0$ and a power law diamond shape. Right: Optimized intrinsic PDF and a measured PDF from a highly worn pad from [1].

for the diamond density and the conditioner and pad radii were estimated or taken from [1]. These imply an equivalent bar conditioner diamond spacing of $\ell = 16\mu$m. While not all of the model parameters can be determined from the data in [1], the shape exponent b_1 and the composite parameter $a = (\Omega/\ell)(b_0/c)$ can be uniquely extracted. Given the rotation rate Ω and an estimate of the diamond width w_0 at the matrix from [1], it is then possible to relate the cut rate c to the diamond protrusion d via $d = (w_0/b_0)^{1/b_1} = (w_0(\Omega/\ell)/(ca))^{1/b_1}$. A protrusion of about 400 μm above a 200 μm diameter base in the matrix is predicted to correspond to a cut rate of about 10 μm per minute, consistent with the characterization in [1] that the conditioner that produced Figure 2 has a high cut rate.

6. Simultaneous conditioning and wear

Having derived a model for the conditioner acting on the pad by itself, we now consider how to extend the model to account for wear created during wafer polishing. This is the *in situ* case. First, we rewrite the wear equation (A4) from appendix 7 in terms of the CCDF rather than the PDF. In the context of the simplified conditioning model used in the previous sections, we then present a combined PDE that describes both processes in the case of a moving conditioner and a solid pad and we find a traveling wave solution. Finally, we discuss a wear and conditioning model for foamed pads.

6.1. SOLID PADS

To begin, we note that Equation (A4) (and equivalently (1.1)) version (A.4) in the appendix) is easily integrated with respect to z to cast it in terms of the CCDF q. Assuming that $\phi(z,t)$ decays quickly enough as $z \to \infty$ so that $\sqrt{z - h(t)}\,\phi(z,t) \to 0$ at any time t, we obtain

$$\frac{\partial q}{\partial t} = \mathcal{W}(r) H(z-d)\sqrt{z-d}\,\frac{\partial q}{\partial z}, \tag{6.1}$$

where the function $\mathcal{W}(r)$ is described in Appendix A. This Equation describes wear for solid pads for all times t.

To combine conditioning and wear, we observe that when the bar conditioner is used, these two processes are both linear in q and occur sequentially on each rotation (Figure 17).

The combined effect on q over an integral number of periods can therefore be modeled by adding the right-hand sides of Equations (4.3) and (6.1). To simplify notation, we define a "conditioning function" as

$$C = \frac{\Omega}{\pi \nu \ell}(z - h)H(z - h), \tag{6.2}$$

and a "wear function" as

$$W = \mathcal{W}(r)\sqrt{z - d}\, H(z - d). \tag{6.3}$$

The basic PDE for the combined wear and conditioning model for a moving bar conditioner is then

$$\frac{\partial q}{\partial t} = -Cq + W\frac{\partial q}{\partial z}. \tag{6.4}$$

We expect this Equation to approximate the actual discrete sequential processes well on time scales that are much larger than a half-sweep of the conditioner.

To gain more insight into the behavior of the combined model, we again examine the steady solution for a constant cutting speed c. Let $Z = z - h(t)$ and $Q(Z, t) = q(z, t)$ as before and denoting the difference between the wafer and conditioner depths by $D(t) = d(t) - h(t)$, Equation (6.4) becomes

$$\frac{\partial Q}{\partial t} - \frac{\mathrm{d}h}{\mathrm{d}t}\frac{\partial Q}{\partial Z} = -\frac{\Omega}{\pi \nu \ell} Z H(Z) Q + \mathcal{W} H(Z - D)\sqrt{Z - D}\, \frac{\partial Q}{\partial Z}. \tag{6.5}$$

In applications, $D(t) > 0$. At steady state, $D(t)$ is a constant D. If $h(t) = h_0 - ct$, then the steady-profile traveling wave solution satisfies the linear ODE

$$c\frac{\mathrm{d}Q}{\mathrm{d}Z} = -\frac{\Omega}{\pi \nu \ell} Z H(Z) Q + \mathcal{W} H(Z - D)\sqrt{Z - D}\, \frac{\mathrm{d}Q}{\mathrm{d}Z}. \tag{6.6}$$

After gathering $\mathrm{d}Q/\mathrm{d}Z$ terms, this ODE is easily solved. When $0 < Z < D$, the solution is the Gaussian function obtained earlier for steady conditioning of a solid pad at constant speed since there is no wear below $Z = D$. Hence,

$$Q(Z) = \exp\left(-\frac{\Omega}{2\pi c \nu \ell} Z^2\right) \qquad 0 < Z < D \tag{6.7}$$

while for the remainder of the region there is wear and

$$Q(Z) = \exp\left(-\frac{\Omega}{2\pi c \nu \ell} D^2\right) \exp\left(-\frac{\Omega}{\pi \nu \ell}\int_D^Z \frac{\zeta\, \mathrm{d}\zeta}{c - \mathcal{W}\sqrt{\zeta - D}}\right) \qquad Z > D. \tag{6.8}$$

The steady state solution given by Equations (6.7) and (6.8) can be seen to meet the necessary continuity requirements at $Z = D$ and the boundary condition $Q = 1$ at $Z = 0$ that corresponds to having only solid pad below the cutting tips. Example solutions are shown in Figure 14.

The structure of the tail of the PDF is affected by wear and can be determined by examining Equation (6.8). It can easily be seen that the integrand in Equation (6.8) has a singularity at $\zeta = D + (c/\mathcal{W})^2$. As the singularity is approached from the left, $Q(Z)$ approaches zero. This implies that the relevant part of the solution lies entirely to the left of the singularity. The wear

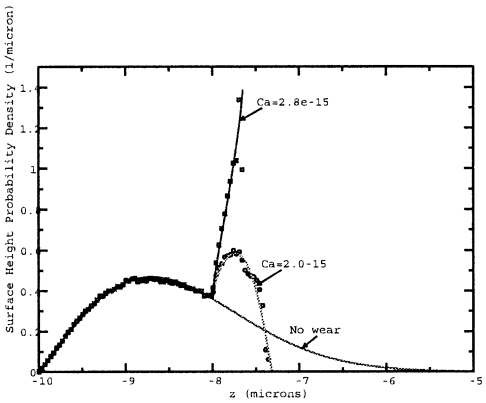

Figure 14. PDFs $\phi = -\mathrm{d}Q/\mathrm{d}z$ (solid curves) produced by the solid pad steady state conditioning and wear model for various values of the wear parameter $\mathcal{W}$ (in units of $\sqrt{\mu m}$/min). The separation between the cutting tips and the wafer surface is 2 μm. Monte Carlo results (circles and squares) are shown for comparison.

function $\mathcal{W}$ can be varied by changing C_a (using a different slurry) or E^* (using a different pad material). We find that the PDF may fail to have a secondary peak if $\mathcal{W}$ is sufficiently small, that it has a single interior peak between D and the singularity for larger $\mathcal{W}$, and that above some threshold it increases monotonically between D and the singularity. Numerical results are shown in Figure 14 as a function of the choice of wear parameter $\mathcal{W}$ for a wear-time ratio of $\eta = 0{\cdot}2124$ at equal pad and wafer rotation rates. The numerical solutions are also compared with Monte Carlo simulations, which are discrete and incorporate asperity wear directly in the form of Archard's law, as described in [2]. The wear parameter is larger than would be realistic during actual *in situ* polishing in order to clearly show the secondary wear peak.

6.2. FOAMED PADS

Wear and conditioning of a foamed pad can be modeled by combining the governing PDE (6.4) for a virtual solid pad with the basic convolution (5.4) for the foamed pad CCDF,

$$q_f(z,t) = \int_z^0 q(\zeta,t)\Phi(z-\zeta)\,\mathrm{d}\zeta. \tag{6.9}$$

An initial condition is required for q; for a cleanly sliced pad, it is $q(z,0) = H(-z)$. More generally, a Laplace transform can be applied to Equation (6.9) to obtain an initial q from any given initial q_f.

At first glance, the structural information contained in Φ would appear to play no role in the time evolution of a foamed pad CCDF; *i.e.*, Equation (6.4) for q could be integrated to the desired time without reference to Φ after which a single application of Equation (6.9) would provide q_f. This is correct if $h(t)$ and $d(t)$ are externally imposed. However, in a real CMP tool, these quantities are always a result of load balancing involving feedback from the structure of q_f. In the theory of surface contact given by Greenwood and Williamson [9], the contact pressure at pad radius r within the wafer bounds is

$$P(r,t) = -\frac{4\eta_s E^*}{3\sqrt{\kappa_s}} \int_{d(t)}^{\infty} (\zeta - d(t))^{3/2} \frac{\partial q_f(r,\zeta,t)}{\partial \zeta}\,\mathrm{d}\zeta, \tag{6.10}$$

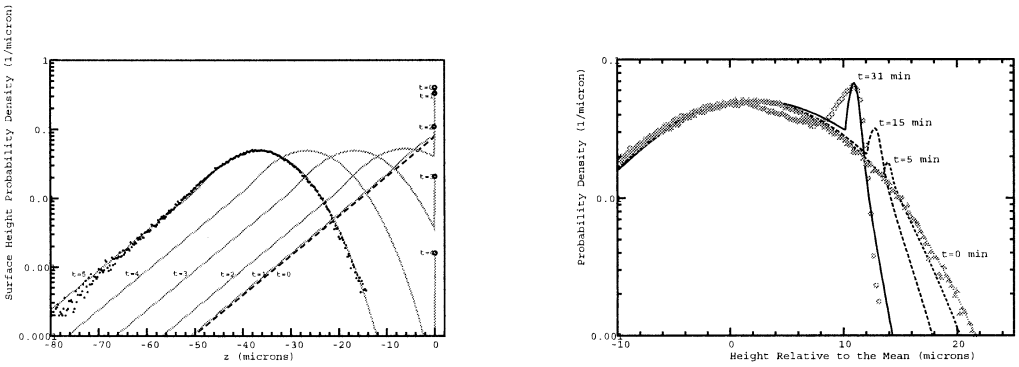

Figure 15. Left: Hypothetical evolution of an exponential intrinsic PDF with a delta-distribution surface component into the freshly conditioned pad data [1] in Figure 2a. Snapshots are shown at one minute time intervals. Right: Comparison of the subsequent wear of the surface height distribution on the left in the absence of conditioning (solid curve) with the data in Figure 2b.

where the radial dependence of q_f has been shown explicitly. Load balance is imposed by integrating Equation (6.10) over the wafer surface to get the total force and adjusting $d(t)$ so that the total force equals the applied load. Thus, pad structural information enters equation (6.4) via $d(t)$ in a load-balanced application.

As an example, we consider a foam with intrinsic PDF $\Phi(z) = a_0 e^{a_1 z} H(-z) + a_2 \delta(z)$ as described in Section 5 and calculate the transient evolution of ϕ at constant cut rate that may have led to the freshly conditioned pad data in Figure 2a. Numerical results are shown in Figure 15a. The simulation suggests that the pad-surface reached steady state in 4–5 minutes of conditioning.

Figure 15b shows a load-balanced simulation of the subsequent abrasive wear of the PDF in Figure 15a into the measured PDF in Figure 2b in the absence of any further conditioning. Two parameters, the wear function factor C_a and the mean asperity tip curvature κ_s (which also affects load balance), were optimized to obtain agreement with the secondary peak in Figure 2b. The location of the secondary peak and the shape of the tail produced by the model agree with the data. The observable differences may be attributable to the simplicity of the intrinsic PDF used and to factors such as platen run-out that are not included in the model.

7. Summary and conclusions

We have developed a simplified theory of conditioning and wear in chemical-mechanical polishing. We have shown that when the conditioner rotates sufficiently slowly, a circular conditioner that executes a constant speed radial sweep can be replaced by an equivalent bar conditioner that has the same cumulative effect on the pad within the region where the wafer is polished. In the context of the bar conditioner, we have then explicitly derived models for the evolution of the surface height distribution of a solid pad during both fixed height and fixed cut-rate conditioning. For foamed pads, we have derived a very fundamental result that relates the foamed pad CCDF to that of an identically conditionened virtual solid pad via a convolution involving the foamed pad intrinsic PDF. The results of all of these models agree with the corresponding Monte Carlo simulations. For combined conditioning and wear of solid pads we obtained a single PDE that also is in agreement with Monte Carlo simulations. Finally, conditioning and wear of a foamed pad is modeled by combining the virtual solid pad

evolution Equation with the fundamental convolution that relates the solid and foamed pad CCDFs.

A theory of conditioning and wear should be connected with a theory of wafer polishing. The overall theory would make it possible to understand in principle the relationship between the design and operation of the conditioner, the structure of the pad, and the measures of polishing performance that are of final importance in the industrial applications of CMP.

Acknowledgements

We gratefully acknowledge the 18th annual workshop on Mathematical Problems in Industry (MPI) at Rensselaer Polytechnic Institute for bringing about this joint collaborative work. The workshop and the participation of L. Borucki were partially supported by the NSF under RPI's VIGRE program in the Mathematical Sciences. Collaboration by T. Witelski was supported by a fellowship from the Alfred P. Sloan Foundation.

Appendix A. CMP tool operation and geometry

We describe in detail practices that are typical of the operation of a rotary CMP tool (see Figure 1), the geometry of these tools, and the related simplifying assumptions that we will adopt. Further information about CMP processing can be found in [6, Chapters 1,2]. The key objectives in this section are to replace Equation (1.1) by an Equation that produces the same change in ϕ after each full pad rotation without involving the details about the intermittent applicability of Equation (1.1) and to explicitly include information about the wafer geometry and relative sliding speed.

A.1. Description of the wear process

The wafer is a circular disk with radius r_w that rotates with constant angular speed Ω_w in the same direction as the pad (which we take here as clockwise). In practice, Ω_w is usually close to the angular speed of the pad, Ω, since maintaining equal speeds promotes uniform polishing. For simplicity we may assume that the center of the wafer is fixed at position $(-c_w, 0)$ relative to the center of the pad, with $c_w > r_w$, see Figure 16.

Within the annular ring $c_w - r_w \le r \le c_w + r_w$ the wear on the pad is not uniform for two reasons; (i) the sliding speed of a point on the pad relative to the wafer may depend on r, and (ii) the physical extent of the wafer surface (and hence the relative time duration of the wear per pad revolution) depends on r. For any fixed value of r within the ring, the relative amount of wear per revolution is given by the angle subtended by the part of the path C passing under the wafer at that radius divided by 2π; we call this the *wear-time ratio*,

$$\eta(r) = \frac{1}{\pi}\cos^{-1}\left(\frac{c_w^2 + r^2 - r_w^2}{2rc_w}\right), \qquad |r - c_w| \le r_w. \tag{A1}$$

The effective influence of $\eta(r)$ is to describe the reduction in the rate of wear due to the finite-sized wafer as compared with uniform wear over the entire annular ring. Since the fraction of time $\eta(r)$ that a small part of the pad spends under the wafer depends on the distance from the center of the pad, a pad with an initially spatially uniform surface height PDF will evolve into one for which the PDF is nonuniform.

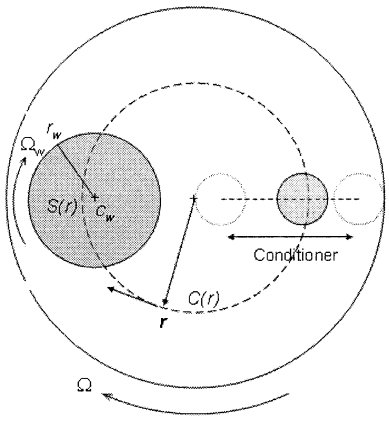

Figure 16. Top view of a CMP tool geometry. A small region of pad at the end of the vector **r** of length r travels along a circular path, encountering the wafer once per period. It will also encounter the circular conditioner during a few of the periods in each outward or inward radial conditioning sweep.

As mentioned above, the factor c_a in the wear Equation is proportional to the relative sliding speed V_s between the pad and the wafer; thus we may write $c_a = C_a V_s$. During an encounter with the wafer, the relative velocity at a point on the pad initially at (r, θ_0) and currently at $(r, \theta(t))$ depends on both the pad and wafer rotation rates and is given by

$$\mathbf{V}_s = -(\Omega_w - \Omega) r \sin\theta(t)\, \mathbf{i} + [(\Omega_w - \Omega) r \cos\theta(t) + \Omega_w c_w]\, \mathbf{j} \tag{A2}$$

where $\theta(t) = \theta_0 - \Omega t$. In the special case when $\Omega_w = \Omega$, the sliding speed is $V_s = \Omega c_w$, but otherwise the speed is a function of r and t, $V_s(r, t) = |\mathbf{V}_s|$. For use in the wear Equation (1.1) we want a measure of the sliding speed that captures the radial dependence of the wear rate after large numbers of pad revolutions while averaging out the details of the variation in rate within each period. We note that since all points at radius r see the same sliding velocity history, we take $\theta_0 = 0$ and use first-order averaging [7, pp. 262–270] to obtain

$$\begin{aligned} \bar{V}_s(r) &= \frac{\Omega}{2\pi\eta(r)} \int_{\pi(1-\eta(r))/\Omega}^{\pi(1+\eta(r))/\Omega} V_s(r, t)\, \mathrm{d}t \\ &= \Omega_w c_w + \frac{r^2(\Omega - \Omega_w)}{c_w} \frac{\sin\pi\eta(r)}{\pi\eta(r)} + O\left(\Omega_w c_w \left(\frac{(\Omega - \Omega_w) r^2}{\Omega_w c_w^2}\right)^2\right). \end{aligned} \tag{A3}$$

We will use $\bar{V}_s(r)$ in (1.1) to describe the evolution of the PDF $\phi(z, t)$ with a parametric dependence on r. This replacement of the temporally fluctuating sliding speed by its average is reasonable provided the time scale of the fluctuations of the sliding velocity, $r_w/(c_w\Omega)$, is much smaller than the time scale over which the surface height of the pad evolves. In practice, the ratio of these time scales does not exceed $1/40$.

We now combine the above considerations to obtain a version of Equation (1.1) that applies over the entire pad rotational period, not just during the periods of wafer contact. By considering the change in ϕ on a timescale comparable with a pad rotation, we can use the wear-time ratio $\eta(r)$ to write Equation (1.1) for the effective wear per pad rotation,

$$\frac{\partial\phi}{\partial t} = \mathcal{W}(r) \frac{\partial}{\partial z} \left(H(z - d(t)) \sqrt{z - d(t)}\, \phi \right), \tag{A4}$$

where, for brevity of notation, we have defined a "wear function",

$$\mathcal{W}(r) = \eta(r) C_a \bar{V}_s(r) \frac{4E^* \sqrt{\kappa_s}}{3\pi} \tag{A5}$$

(In the MKS system, $\mathcal{W}$ has units of $m^{1/2}/sec$). Because at a given radius r, the relation between the elapsed time t and the wear time τ is $t = \tau/\eta(r)$, (A4) produces the same change in ϕ over any integral number of pad rotations as Equation (1.1) without explicitly resolving the intermittency of the wear process.

A.2. Description of the conditioning process

Like the wafer, the conditioner is a rotating disk pressed onto the pad-surface. The conditioning disk has a much smaller radius r_c and is periodically swept over a fixed radial track. We will assume that the conditioning sweep occurs along the positive x-axis with constant translational speed v_c and constant rotational speed Ω_c between $x = r_c$ and $x = r_p - r_c$, where r_p is the pad radius (Figure 16). As the conditioner traverses the track, it gradually removes pad material. In a real tool, the translational speed of the conditioner is often controlled so that the removal is as uniform as possible over the conditioning track. That is, the local mean surface height is close to constant across the pad and during the conditioning process it decreases at a constant rate c, called the *cut rate*. While variable sweep speeds can be incorporated into our approach, the essential features are illustrated at constant speed.

Unlike the wafer, the points and sharp edges on the diamonds on the conditioner face cut the pad rather than abrading it. The essential difference is that abrasion is a gradual process in which the removal rate depends on the applied load, the contact time and the geometry of the abraded surface (*i.e.*, asperity heights) as in Equation (A4) while in cutting the amount of material removed depends strongly on the shape of the cutting object. Thus, equation (A4) cannot be used to describe conditioning.

In the next section, we consider the influence of some of the conditioner parameters and construct an equivalent conditioning tool with a simplified geometry.

Appendix B. Simplification of the conditioner

For the purpose of studying CMP, it is not necessary to resolve all of the fine-scale details of the conditioning process, so long as their cumulative net effects can be understood. Consequently, we find that the average effect of the actual circular conditioning tool can be modeled by a very much simplified equivalent tool.

Conditioning disks are usually 0·025–0·05 m in radius and are covered with diamond cutting tips that some manufacturers distribute randomly and uniformly over the disk. Selected according to their size, shape and other qualities, the diamonds are embedded in a tough matrix from which they may have a carefully controlled protrusion. When the diamonds are uniformly distributed, the density of grooves or furrows cut into the pad is independent of the sliding direction of the disk. Furthermore, when the conditioner and pad rotate in the same direction and the rotation rate of the conditioner is sufficiently slow relative to that of the pad, each diamond cuts a non self-intersecting furrow during each pad rotational period. This happens when the conditioner and pad rotate at the same rate, for example. Under the assumptions of uniformity and slow rotation, the density of furrows is independent of the conditioner rotation rate, so we may analyze the density using a non-rotating disk. Furthermore, if the

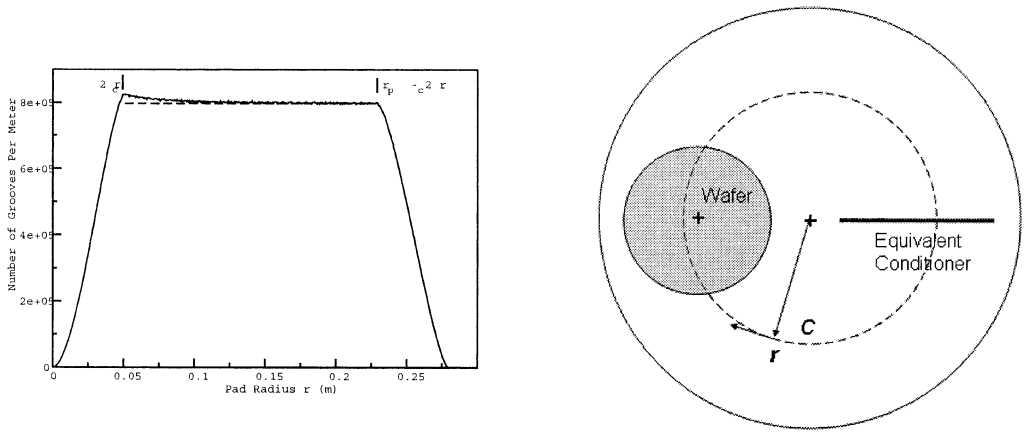

Figure 17. (left) Comparison of the furrow density as a function of pad radius r from a direct simulation of groving by a rotating conditioner with uniformly distributed diamonds with the analytic approximation (dashed line) derived in the paper. (right) For the model of *in situ* conditioning, the circular conditioner is replaced by the equivalent bar conditioner. A small portion at **r** of a radial cross-section encounters both the wafer and conditioner once per period.

sweep speed is sufficiently slow relative to Ω_c, then the shape of the cut that each diamond makes across a fixed radial line on the pad will be close to that of the diamond.

With these assumptions, we can estimate the average density of furrows cut in an annulus between r and $r + \Delta r$, where r and $r + \Delta r$ are both between $2r_c$ and $r_p - 2r_c$. At any constant sweep speed, all of the diamonds on the conditioner eventually pass through this annulus during a half-sweep. If Δr is chosen so that $\Delta r = 2\pi v_c/\Omega$, then each diamond will make exactly one new cut through any radial cross-section of the annulus with angular label θ_0. The total number of cuts by the end of a half-sweep will therefore be the total number of diamonds on the disk, $N_D = \rho_D \pi r_c^2$, for this particular annulus and for any θ_0, where ρ_D is the density of diamonds on the disk face. The average density of new cuts in the annulus (the number per unit length of a radial cross-section) is then $N_D/\Delta r$ by the end of the sweep.

We further simplify the conditioning model by distributing the new cuts in the annulus evenly between the pad rotations comprising the half-sweep. Since the time for the half-sweep is $(r_p - 2r_c)/v_c$ or

$$N_s = \frac{(r_p - 2r_c)/v_c}{2\pi/\Omega} = (r_p - 2r_c)/\Delta r \tag{B1}$$

rotational periods, the average density of new cuts introduced per rotation is

$$\frac{N_D/\Delta r}{N_s} = \frac{N_D/\Delta r}{(r_p - 2r_c)/\Delta r} = \frac{N_D}{r_p - 2r_c}. \tag{B2}$$

The average spacing between new cuts per rotation is consequently

$$\ell = \frac{r_p - 2r_c}{\rho_D \pi r_c^2}. \tag{B3}$$

The average spacing of new cuts per turn depends on the total number of diamonds on the disk and the length of the half-sweep but not on r or on the sweep speed v_c as long as the latter is sufficiently slow. As seen in Figure 17, which compares this approximation with a

direct simulation of grooving during a half-sweep, the estimate is very good between $r = 2r_c$ and $r = r_p - 2r_c$. These are the approximate limits within which the wafer is polished. In the simulation, the tracks on the pad of a large number of uniformly distributed diamonds were explicitly calculated over the half-sweep using a rotating conditioner.

The above formula informs us that the circular conditioner may be simplified by remounting the $\rho_D \pi r_c^2$ diamonds on its face in a line on a bar-shaped tool of length $r_p - 2r_c$. The diamonds on this bar, sometimes referred to here as the *equivalent* or *idealized linear* conditioner, are arranged with a mean spacing of ℓ and each execute independent random radial displacements as the pad turns. On successive rotations of the pad, a fixed radial cross-section will then then encounter the bar conditioner with the diamonds at a uniformly distributed random displacements. The bar conditioner is equivalent to the circular conditioner in the sense that by the end of one half-sweep, it produces the same density of new cuts between $2r_c$ and $r_p - 2r_c$ as the circular conditioner produces. However, the equivalent conditioner operates on a sample of the pad at radius r at every turn of the pad rather than for just for a fraction of the half-sweep (Figure 17).

References

1. A. S. Lawing, Polish rate, pad-surface morphology and pad conditioning in oxide chemical mechanical polishing. In: S. Seal, R. L. Opila, C. Reidseme Simpson, K. Sundaram, H. Huff and I. I. Suni (eds.), *Proc. of the Fifth International Symposium on Chemical Mechanical Polishing*, Philadelphia, PA, May 12–17, 2002. The Electrochemical Society Proc. Vol. PV2002-1 (2002) 46–60.
2. L. Borucki, Mathematical modeling of polish rate decay in chemical-mechanical polishing. *J. Engng. Math.* 43 (2002) 105–114.
3. B. Bhushan, *Introduction to Tribology*. New York: John Wiley and Sons, Inc. (2002) 732pp.
4. J. A. Williams and Y. Xie, The generation of wear surfaces by the interaction of parallel grooves. *Wear* 155 (1992) 363–379.
5. S. Jacobson, P. Wallen and S. Hogmark, Correlation between groove size, wear rate and topography of abraded surfaces. *Wear* 115 (1987) 83–93.
6. J.M. Steigerwald, S.P. Murarka and R.J. Gutman, *Chemical Mechanical Planarization of Microelectronic Materials*. New York: John Wiley and Sons, Inc. (1997) 324pp.
7. J. A. Murdock, *Perturbations: Theory and Methods*, volume 27 of *Classics in Applied Mathematics*. Philadelphia: Society for Industrial and Applied Mathematics (SIAM) (1999) 509pp.
8. J. A. Greenwood, A unified theory of surface roughness. *Proc. R. Soc. London A* 393 (1984) 133–157.
9. J. A. Greenwood and J. B. P. Williamson, Contact of nominally flat surfaces. *Proc. R. Soc. London A* 295 (1966) 300–319.
10. J. A. Greenwood and J. J. Wu, Surface roughness and contact: an apology. *Meccanica* 36 (2001) 617–630.
11. N. Yu and A. A. Polycarpou, Contact of rough surfaces with asymmetric distribution of asperity heights. *J. Tribology* 124 (2002) 367–376.
12. W. Feller, *An Introduction to Probability Theory and its Applications*, Vol II, 2^{nd} ed. New York: John Wiley & Sons Inc. (1971) 669pp.