

STOCHASTIC MODELS FOR CHEMICAL-MECHANICAL POLISHING

JOSÉ RAMÍREZ,* *Cornell University*

B. RIDER,** *Duke University*

Abstract

Chemical-mechanical polishing is an essential processing technique in the electronics industry for planarizing materials that have been deposited on silicon wafer surfaces. Recently, various statistical models have been introduced to describe this process. The picture is one of surfaces evolving in time according to a Markovian law. To date these models have been studied mostly on the level of simulation. The purpose of this note is to put some of the phenomena seen in simulations on firm analytical ground and to formulate open probabilistic questions related to these problems.

Keywords: Surface roughness; Markov processes; Queueing.

1. Introduction

We study a Markov process of surfaces which has as its motivating application a fast growing and essential processing technique in the electronics industry known as chemical-mechanical polishing (CMP). Roughly put, CMP is a procedure for removing materials that have been deposited on silicon wafer surfaces and is known to greatly simplify the task of building multilayer integrated surface structures. While the mechanisms underlying CMP have been qualitatively described (from an engineering perspective there is even the text [12]), most aspects of this important process had not modeled mathematically until recently in [2] and [3]. Our work is based on a model introduced in the latter paper.

In the class of CMP tools under investigation, the silicon wafer is held against a rotating polishing pad (Figure 1). During polishing, a chemically-reactive slurry is sprayed onto the pad ahead of the wafer while a diamond-covered rotating plate called a conditioner is pressed into the pad, the purpose being to maintain the roughness of the pad. The action of the slurry combined with dragging the roughened pad across the surface is understood to be responsible for the removal of material from the wafer. The goal of [3], continued here, is to well understand the conditioning process. Figure 2 (left) shows an example of the height distribution of a newly conditioned pad. In the absence of conditioning, the height distribution forms an additional peak to the right which then grows in time. This corresponds to the pad surface possessing expanding planar regions, which in turn decreases the average contact pressure with the wafer and so also the average polishing rate. It is therefore of great interest in industry to

* Postal address: Department of Mathematics, 310 Mallot Hall, Cornell University, Ithaca, NY 14853. Email: ramirez@math.cornell.edu

** Postal address: Department of Mathematics, Duke University, Durham, NC 27708. Email: rider@math.duke.edu

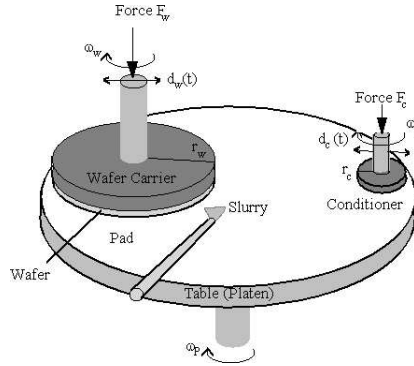


FIGURE 1: Schematic of CMP tool [3].

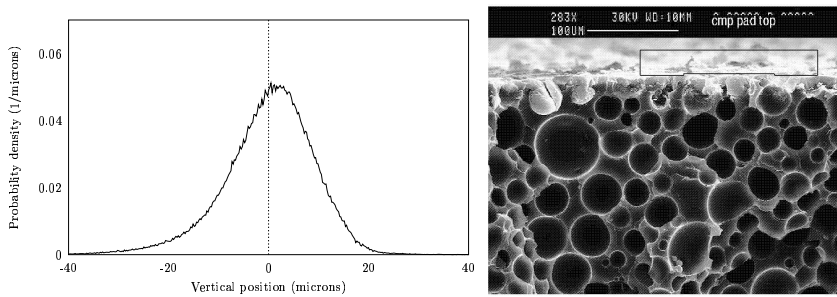


FIGURE 2: Left: Empirical probability density function of one-point surface height for a pad after 20 minutes of conditioning ([8], [2]). The picture is shifted to produce zero mean. Right: Scanning Electron Micrograph cross section of a conditioned void-filled polyurethane pad. The scale bar at the top indicates 100 microns. Voids average about 30 microns in diameter and occupy about 50% of any cross section. (Data by Letitia Malina, Motorola [2])

have an effective model for the evolution of this distribution along with other measures of pad roughness due to conditioning.

One notable feature of the height distribution is the asymmetry of the tails. In modeling this or any statistic stemming from the conditioning process, one needs take into account that the pads themselves are typically made from soft materials such as polyurethane which are locally far from uniform. A magnified cross-section of a common pad material is depicted in Figure 2 (right); it is clearly not solid but “foamed”. The darker circles correspond to voids which occupy roughly 35% of pad volume and 50% of any cross section. We will see that this void structure is largely responsible for the left (or lower) tail of the distribution.

Let us now describe the stochastic model for pad conditioning introduced by the authors of [3]. Understandably, it rests on a number of simplifications of the physical system. First, time is discretized, and one projects down to a two-dimensional cross section of the surface. In cross section, the diamond array of the conditioning plate is replaced by a saw-toothed function lying above the pad. Now, the diamonds are engineered to be of uniform size and shape and are distributed infirmly over the conditioner. Based on this and the relative size and rotation speed of the conditioning

plate to that of the pad, [3] argues that one may restrict attention to the effect of a single diamond swept uniformly at random over a fixed segment of pad in each time step; the picture then extended periodically. Lastly, the manner in which the conditioner moves into the pad must be considered. In actuality there is a type of feedback mechanism: a balance of forces controls the speed at which the conditioning plate is pushed into the pad so that material is removed evenly across the surface. The model ignores this dependence in favor of a rate of descent that is solely a function of time. The result of these comments is the following Markovian law of surface development: with $z = H(x, n)$ the surface height at time n and position x along the pad,

$$H(x, n + 1) = \min \left\{ H(x, n), V(x - X_n) - c(n) \right\}. \quad (1.1)$$

in which X_1, X_2, \dots is a sequence of independent identically distributed Uniform random variables on $[0, 1]$, V describes the geometry of the diamond (below $V(x) = \text{const}|x|$), and the (non-negative) $c(n)$ models the movement of the plate into pad. The evolution (1.1) is subject to a planar initial condition, $H(\cdot, 0) = \text{const}$, and is taken to be periodic at its boundary: H is a map from the circle to a lower half plane (the pad is extended indefinitely below).

Of course, (1.1) only describes a conditioning plate's descent into a solid pad. In [3], physical arguments along with a continuum in time limit lead to Hamilton-Jacobi like equations for the development of a suitable approximation of the probability distribution of H at a given point. This was carried out for stationary ($c(n) \equiv 0$) and constant rate ($c(n) = c \times n$) conditioning. Simulations are then run and shown to support the use of these PDE approximations. The effect of void-filled pads is then added after-wards in a heuristic fashion.

In this paper, we first derive exact expressions for the (one-point) height distribution, $P(H(x, n) \leq z)$, for stationary and constant rate conditioning under the assumption of a solid pad. In the stationary case which occupies Section 2 we can do more. We consider the *distribution of heights*, $\text{meas}\{x : H(x, n) \leq z\}$, and can describe various moments along with fluctuation formulas for long time. For constant rate conditioning we compute the one-point function for large enough n and then give asymptotic formula for the mean and variance of H ; this is Section 3. Section 4 builds on the result of Section 3 in order to describe the more important case of constant rate conditioning of foamed pads. We model the pad by placing independent identically distributed voids at each point of a two-dimensional Poisson random field.¹ In the presence of this random environment the height distribution is described in terms of a convolution of the solid-pad height distribution and one describing that of the typical void size. Interestingly, this analysis touches on a classical problem of Queueing Theory. The result is a model which captures the main qualitative features of the empirical distribution pictured above. In a last section we describe some related problems for future work.

2. Stationary Conditioning of a Solid Pad

As in [3], we start with the idealization of stationary conditioning: $c(n) = 0$. In this warm-up model we are able to obtain exact formulas for several measures of the

1. This model in itself is known as the Poisson blob or Boolean model of continuum percolation [9].

surface roughness. Also following [3], $V(x)$ is chosen to be a symmetric wedge of constant slope $1/\varepsilon$. Things are normalized so that the plate sits directly above the (solid) pad at height = 1, the minimum of the diamond extending down to height = 0. The evolution of the pad surface (1.1) then takes the form

$$H_\varepsilon^0(x, n+1) = \min\left\{H_\varepsilon^0(x, n), \frac{1}{\varepsilon}|x - X_n|\right\}, \quad (2.1)$$

where $H_\varepsilon^0(x, 0) \equiv 1$ (the superscript 0 indicating stationarity).

We consider the distribution of heights,

$$\mathcal{D}_{n,\varepsilon}(z) \equiv \text{meas}\left\{x : H_\varepsilon^0(x, n) \leq z\right\}, \quad (2.2)$$

obtaining both moment and fluctuation formulas for this statistic. Note that by stationarity (in x) $E[\mathcal{D}_{n,\varepsilon}(z)] = P(H_\varepsilon^0(\cdot, n) \leq z)$, recovering the one-point function. The key to our expressions is that the surface at time n may be reconstructed from the position of the cuts. That is, all statistics of interest may be derived from the distribution of the Order Statistics of n uniform random variables on the interval $[0, 1]$.

Theorem 2.1. *For $H_\varepsilon^0(\cdot, n)$ the random surface at time n there are the following description of the height distribution $\mathcal{D}_{n,\varepsilon}$. With X^1, X^2, \dots, X^n the Order Statistics of the first n cut positions:*

$$\mathcal{D}_{n,\varepsilon}(z) = \sum_{k=1}^{n-1} (X^{k+1} - X^k) \wedge (2z\varepsilon) + \left((1 - X^n + X^1) \wedge 2z\varepsilon\right). \quad (2.3)$$

This expression allows explicit computation of the mean and variance variance of the height distribution:

$$E\left[\mathcal{D}_{n,\varepsilon}(z)\right] = P\left(H_\varepsilon^0(x, n) \leq z\right) = 1 - (1 - 2z\varepsilon)^n, \quad (2.4)$$

and, as long as $\varepsilon < 1/4$,

$$\text{Var}\left[\mathcal{D}_{n,\varepsilon}(z)\right] = \frac{2}{(n+1)}(1 - 2z\varepsilon)^{n+1} + \frac{(n-1)}{(n+1)}(1 - 4z\varepsilon)^{n+1} - (1 - 2z\varepsilon)^{2n} \quad (2.5)$$

It follows that $\mathcal{D}_{n,\varepsilon}(z) \rightarrow 1$ with probability one as $n \uparrow \infty$ for any fixed $z > 0$.

Remark. It is interesting to note that the arguments in [3] led to an expression quite close to ours for the mean of $\mathcal{D}_{n,\varepsilon}(z)$.

Remark. The surface energy, $\mathcal{E}_{n,\varepsilon} = \int_0^1 |(H_\varepsilon^0)'(x, n)|^2 dx$, provides another measure of roughness. This object can also be described in terms of the Order Statistics of the cut positions:

$$\mathcal{E}_{n,\varepsilon} = \frac{1}{\varepsilon^2} \sum_{k=1}^{n-1} (X^{k+1} - X^k) \wedge (2\varepsilon) + \frac{1}{\varepsilon^2} \left((1 - X^n + X^1) \wedge 2\varepsilon\right). \quad (2.6)$$

An easy consequence is that $\mathcal{E}_{n,\varepsilon}$ converges to $1/\varepsilon^2$ with probability one as $n \uparrow \infty$. Clearly from (2.6) one may compute various statistics as in (2.4) and (2.5).

Theorem 2.1 shows that the surface falls below any pre-arranged level at exponential speed with fluctuations also exponentially decaying. A better way to view the asymptotics is to consider small ε and $n \simeq \alpha/\varepsilon$ for $\alpha > 0$ and notice that

$$\lim_{\varepsilon \downarrow 0} E \left[\text{meas} \left\{ x : H_\varepsilon^0(x, \frac{\alpha}{\varepsilon}) \leq z \right\} \right] = \lim_{\varepsilon \downarrow 0} P \left(H_\varepsilon^0(\cdot, \frac{\alpha}{\varepsilon}) \leq z \right) = 1 - e^{-2z\alpha}. \quad (2.7)$$

This is to be compared to the exponential decay of the height distribution reported in simulations ([3]), and demonstrates that optimal conditioning times should be inversely proportional to the cutting slope, yielding a non-trivial (and predictable) distribution of heights.

The observation (2.7) is really an old result of Smirnov [10] who shows that the sum of the indicators of the events $\{X^{k+1} - X^k \geq x/n\}$ tends to $\exp(-x)$ in probability. However, having computed the additional piece of information that $1/\varepsilon = n$ times the variance (2.5) is of order one, suggests that, in this simple model, $\mathcal{D}_{n,\varepsilon}$ has a central limit theorem. The result is the following:

Theorem 2.2. *With a fixed ε , $z < 1$ and $n \uparrow \infty$, the difference between 1 and $\mathcal{D}_{n,\varepsilon}(z)$ when multiplied by n tends to an exponential random variable of mean one. More informative, in the limit of small ε with time of order ε^{-1} , one finds the following convergence to a Gaussian distribution:*

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \left[\text{meas} \left\{ x : H_\varepsilon^0(x, \lceil \frac{\alpha}{\varepsilon} \rceil) \leq z \right\} - (1 - e^{-2z\alpha}) \right] = \mathcal{N}(0, \phi(\alpha, z)) \quad (2.8)$$

in which $\phi(\alpha, z) = 2\alpha^{-1}e^{-2z\alpha}(1 - (1 + 2z\alpha)e^{-2z\alpha})$.

In the proofs of Theorems 2.1 and 2.2 we make repeated use of the following fact.

Lemma 1. *Let X^1, \dots, X^n denote the Order Statistics of X_1, \dots, X_n , n independent uniform $[0, 1]$ random variables. Then for the corresponding spacings $\{X^k - X^{k-1}\}$ in the bulk (here we consider k from 2 to n) one has that*

$$P(X^k - X^{k-1} \in da) = n(1-a)^{n-1}da \quad (2.9)$$

and

$$P(X^k - X^{k-1} \in da, X^j - X^{j-1} \in db) = n(n-1)(1-a-b)^{n-2}dad b \quad (2.10)$$

for any $j \neq k$, $0 \leq a, b \leq 1$ and $a + b \leq 1$ where necessary. For the edge variable we have $P(1 - X^n + X^1 \in da) = n(n-1)a(1-a)^{n-2}$ and

$$P(1 - X^1 - X^n \in da, X^k - X^{k-1} \in db) = n(n-1)(n-2)(1-a-b)^{n-3}dad b$$

for the edge-bulk correlations.

Proof. Recall (see [4] pg. 127 for example) the equivalence in law between (X^1, X^2, \dots, X^n) and $(T_1/T_{n+1}, T_2/T_{n+1}, \dots, T_n/T_{n+1})$ in which $T_k = \tau_1 + \tau_2 + \dots + \tau_k$ with $\{\tau_k\}$ independent exponentials of mean one. Then (2.9) follows from

$$P(X^k - X^{k-1} \leq a) = P(\tau_1 \leq aT_{n+1}) = \int_0^{\frac{a}{1-a}} \frac{t^{n-1}e^{-t-s}}{(n-1)!} ds dt = 1 - (1-a)^n.$$

The proof of (2.10) and the expressions involving the edge variable are similar in spirit; their verification is not reported.

Proof of Theorem 2.1. The evaluation of $\mathcal{D}_{n,\varepsilon}(z)$ can be made in terms their Order Statistics $X^1, X^2 \dots$ of the centers of the first n cuts is made evident by writing

$$\begin{aligned} \text{meas}\{x : H_\varepsilon^0(x, n) \leq z\} &= \int_0^1 \mathbf{1}_{\{H_\varepsilon^0(x, n) \leq z\}}(x) dx \\ &= \sum_{k=1}^n \left\{ \int_{X^k}^{X^{k+1}} \mathbf{1}_{\{H_\varepsilon^0(x, n) \leq z\}}(x) dx \right\} \end{aligned} \quad (2.11)$$

where we take $X^{n+1} = X^1$. Focus on a bulk term ($1 \leq k \leq n-1$), things being the same. For the k -th integral in (2.11) there two cases depending on whether $|X^{k+1} - X^k| \leq 2\varepsilon$ or not. In the former case, the surface between X^k and X^{k+1} forms an isosceles triangle with peak $\varepsilon(X^{k+1} - X^k)/2$ at the midpoint $(X^k + X^{k+1})/2$. In the latter it is a trapezoid: H_ε^0 growing linearly from X^k to $X^k + \varepsilon$, $H_\varepsilon^0 \equiv 1$ on $[X^k + \varepsilon, X^{k+1} - \varepsilon]$, and then returning to 0 at $x = X^{k+1}$. Therefore, for $0 \leq z < 1$, the portion of the interval $[X^k, X^{k+1}]$ with $H_\varepsilon^0 \leq z$ is either the entire interval when $|X^{k+1} - X^k| \leq 2z\varepsilon$ or it is just the subintervals $[X^k, X^k + \varepsilon z]$ and $[X^{k+1} - \varepsilon z, X^{k+1}]$. In symbols that is:

$$\int_{X^k}^{X^{k+1}} \mathbf{1}_{\{H_\varepsilon^0(\cdot, n) \leq z\}}(x) dx = (X^{k+1} - X^k) \wedge (2z\varepsilon),$$

proving the representation (2.3).

Next, with the aid of Lemma 1, obtaining the mean and variance of $\mathcal{D}_{n,\varepsilon}(z)$ are straight forward (though tedious) computations. We record the results. For the expectation, one needs that for any positive constant $c < 1$,

$$E[(X^k - X^{k-1}) \wedge c] = n \int_0^1 (s \wedge c)(1-s)^{n-1} ds = \frac{1}{(n+1)} - \frac{(1-c)^{n+1}}{(n+1)},$$

and similarly

$$E[(1 - X^n + X^1) \wedge c] = n(n-1) \int_0^1 (s \wedge c)s(1-s)^{n-1} ds = \frac{2}{(n+1)} - \frac{(1-c)^n}{(n+1)}(1-c+cn)$$

which explains (2.4).

As to the variance, we restrict to $c < 1/2$ and compute separately for each spacing. First for the diagonal terms: in the bulk we have

$$E\left[\left\{(X^k - X^{k-1}) \wedge c\right\}^2\right] = \frac{2}{(n+1)(n+2)} - \frac{2(1-c)^{n+1}}{(n+1)(n+2)}(1+c+cn) \quad (2.12)$$

with

$$E\left[\left\{(1 - X^n + X^1) \wedge c\right\}^2\right] = \frac{6}{(n+1)(n+2)} - \frac{(1-c)^n}{(n+1)(n+2)}(6 - 2c^2 + 6cn + 2c^2n^2)$$

at the edge. For the covariances of the individual spacings there is

$$\begin{aligned} & E\left[\{(X^k - X^{k-1}) \wedge c\}\{(X^j - X^{j-1}) \wedge c\}\right] \\ &= n(n-1) \int_0^1 \int_0^1 (s \wedge c)(t \wedge c)(1-s-t)^{n-2} ds dt \\ &= \frac{1}{(n+1)(n+2)} - \frac{2(1-c)^{n+2}}{(n+1)(n+2)} + \frac{(1-2c)^{n+2}}{(n+1)(n+2)} \end{aligned}$$

for the bulk-bulk and

$$\begin{aligned} & E\left[\{(1 - X^n + X^1) \wedge c\}\{(X^k - X^{k-1}) \wedge (2z\varepsilon)\}\right] \tag{2.13} \\ &= n(n-1)(n-2) \int_0^1 \int_0^1 (s \wedge c)(t \wedge c)(1-s-t)^{n-3} s ds dt \\ &= \frac{2}{(n+1)(n+2)} - \frac{(1-c)^{n+1}}{(n+1)(n+2)} (4-2c+nc) + \frac{(1-2c)^{n+1}}{(n+1)(n+2)} (2-2c+nc) \end{aligned}$$

for the edge-bulk terms. Putting together the displays between (2.12) and (2.13) in the appropriate fashion with $c = 2z\varepsilon$ yields (2.5). The proof is complete.

Proof of Theorem 2.2. Noting the exchangeability of $X^{k+1} - X^k$, one could employ general techniques ([14], for example). However, we find it easiest to do things by hand, making use of the equivalence in law noted in the proof of Lemma 1.

For the first statement we may write: with $\{\tau_k\}$ independent exponentials of mean one and T_k their running sum,

$$\begin{aligned} \lim_{n \uparrow \infty} n \left\{ 1 - \text{meas} \{x : H_\varepsilon^0(x, n) \leq z\} \right\} &= \lim_{n \uparrow \infty} n \left[1 - \sum_{k=1}^{n-1} (X^{k+1} - X^k) \wedge (2\varepsilon z) \right] \\ &= \lim_{n \uparrow \infty} \frac{n}{T_{n+1}} \left[\tau_{n+1} - \sum_{k=1}^n [\tau_k - \tau_k \wedge (2\varepsilon z T_{n+1})] \right] \end{aligned}$$

with the second equality holding in distribution. Now the ratio n/T_{n+1} tends to one almost surely, and as $\tau_{n+1} \sim \tau$ for any n , the claim follows because

$$E \left[\sum_{k=1}^n [\tau_k - \tau_k \wedge (2\varepsilon z T_{n+1})] \right]^2 \leq n^2 E[\tau_1^2] P\left(\tau_1 \geq 2z\varepsilon \sum_{k=2}^n \tau_k\right) \rightarrow 0$$

by a large deviation estimate.

The proof of (2.8) follows the same format. The desired sum of terms in the Order Statistics is replaced by

$$\begin{aligned} & \frac{1}{\sqrt{\varepsilon} T_{\lceil \alpha/\varepsilon \rceil + 1}} \left[\sum_{k=1}^{\lceil \alpha/\varepsilon \rceil} \tau_k \wedge (2z\varepsilon T_{\lceil \alpha/\varepsilon \rceil + 1}) - (1 - e^{-2z\alpha}) T_{\lceil \alpha/\varepsilon \rceil + 1} \right] \tag{2.14} \\ &= \frac{1}{\varepsilon T_{\lceil \alpha/\varepsilon \rceil + 1}} \times \left[\sqrt{\varepsilon} \left(\sum_{k=1}^{\lceil \alpha/\varepsilon \rceil} \tau_k \wedge 2z\alpha - (1 - e^{-2z\alpha}) T_{\lceil \alpha/\varepsilon \rceil + 1} \right) \right. \\ & \quad \left. + \sqrt{\varepsilon} \left(\sum_{k=1}^{\lceil \alpha/\varepsilon \rceil} \tau_k \wedge (2z\varepsilon T_{\lceil \alpha/\varepsilon \rceil + 1}) - \tau_k \wedge 2z\alpha \right) \right]. \end{aligned}$$

Similar to before $\varepsilon T_{\lceil \frac{\alpha}{\varepsilon} \rceil + 1} \rightarrow \alpha$ with probability one as $\varepsilon \downarrow 0$ and so may be considered constant from the start. Next the computation $E[\tau_1 \wedge \gamma - \tau_1(1 - e^{-\gamma})]^2 = 2e^{-\gamma}(1 - (1 + \gamma)e^{-\gamma})$ and the central limit theorem combine to show that for the first term,

$$\begin{aligned} & \sqrt{\varepsilon} \left[\sum_{k=1}^{\lceil \frac{\alpha}{\varepsilon} \rceil} (\tau_k \wedge 2z\alpha) - (1 - e^{-2z\alpha}) T_{\lceil \frac{\alpha}{\varepsilon} \rceil + 1} \right] \\ &= \sqrt{\varepsilon} \sum_{k=1}^{\lceil \frac{\alpha}{\varepsilon} \rceil} \left(\tau_k \wedge 2z\alpha - \tau_k(1 - e^{-2z\alpha}) \right) + \sqrt{\varepsilon} T_{\lceil \frac{\alpha}{\varepsilon} \rceil + 1} \rightarrow \mathcal{N}\left(0, \alpha\phi(\alpha, z)\right) \end{aligned}$$

in distribution as $\varepsilon \downarrow 0$. The proof is finished by overestimating the variance of the remaining sum in (2.14) by an object tending to zero. The details are straight-forward.

3. Moving Plate Conditioning

Next we consider $c(n) = c \times n$, which describes the motion of the conditioner into the (still solid) pad at fixed velocity $0 < c \ll 1$. To distinguish from the previous section this process is denoted by

$$H_\varepsilon^c(x, n+1) = \min \left\{ H_\varepsilon^c(x, n), \frac{1}{\varepsilon} |x - X_n| - cn \right\}. \quad (3.1)$$

At time 0 we fix $H_\varepsilon^c(\cdot, 0) = 0$ and note that H can reach arbitrarily low depths. Also, the diamond is taken to extend indefinitely (it is a semi-infinite wedge of slope $1/\varepsilon$). We will make essential use of this assumption below, though admittedly that it has serious failings in terms of a model. The real diamonds are of fixed height with the flat part of the plate simply pushing (not cutting) into the pad when in contact. The hope is that the infinite wedge is a reasonable approximation at the tip, and gets better at smaller velocities.

The surface can no longer be reconstructed from knowing the positions of the cuts: for that both the position and the time of occurrence would be required. Instead, we observe that the process is stationary in time once that parameter is larger than a constant n^* defined below. In this stationary regime we compute explicitly the law of the one-point surface height. The result is the following:

Theorem 3.1. *The law of the random variable $H_\varepsilon^c(x, n)$ is independent of $x \in [0, 1]$ and also of time (n) once that value is large enough. In particular, for $n \geq \frac{1}{2c\varepsilon} = n^*$, we have the equivalence in law*

$$H_\varepsilon^c(x, n) = -nc + \left\lfloor \frac{1}{2c\varepsilon} \right\rfloor c + \min \left\{ \mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_{\lfloor \frac{1}{2c\varepsilon} \rfloor} \right\} \quad (3.2)$$

in which the \mathcal{U}_k are independent and uniformly distributed on the interval $[-kc, -kc + \frac{1}{2\varepsilon}]$.

The surface at time n is then supported on $[-nc, -nc + \frac{1}{2\varepsilon}]$ and is described by the distribution function: with $0 \leq z \leq 1$,

$$\begin{aligned} P\left(H_\varepsilon^c(x, n) \leq -nc + \frac{1}{2\varepsilon}z\right) &= E\left[\text{meas}\left\{x \in [0, 1] : H_\varepsilon^c(x, n)(x) \leq -nc + \frac{1}{2\varepsilon}z\right\}\right] \\ &= 1 - \prod_{k=0}^{\lfloor \frac{z}{2c\varepsilon} \rfloor} \left(2\varepsilon kc + (1 - z)\right). \end{aligned} \quad (3.3)$$

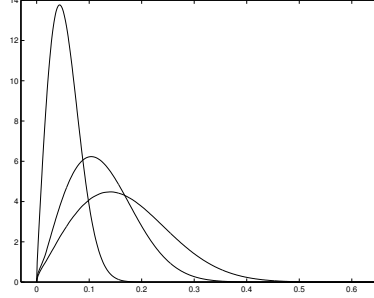


FIGURE 3: Probability density of the normalized random variable $2\varepsilon(H_\varepsilon^c(\cdot, n) + nc)$ where the parameter $(2c\varepsilon)^{-1}$ equals 500, 100, and 50 from left to right.

Note the normalization is made so that $z = 0$ corresponds to the deepest part of the surface.

Remark. The energy for this model is easily seen to be constant $= 1/\varepsilon^2$ for $n \geq n^*$.

Figure 3 depicts the density identified in Theorem 3.1 for various values of the parameter $c \times \varepsilon$; it is quadratic at its peak and vanishes linearly at $z = 0$. While the upper tail appears to well reflect the observations, we see that the assumption of a solid pad misses even the roughest properties of the lower tail. The figure puts attention on a natural asymptotic regime in which velocity is small and/or the slope of the wedge is large ($c\varepsilon \ll 1$). In this regime the normalized height distribution is well approximated as in

$$P(2\varepsilon(nc + H_\varepsilon^c(z, n)) \in dz) / dz = 1 - \exp\left[\sum_{k=0}^{\lfloor \frac{z}{2c\varepsilon} \rfloor} \log(2c\varepsilon k + (1-z))\right] \quad (3.4)$$

$$\simeq 1 - \exp\left[\frac{z}{2c\varepsilon} \int_0^1 \log(zx + (1-z)) dx\right] = 1 - \exp\left[-\frac{1}{2c\varepsilon}(z + (1-z)\log(1-z))\right].$$

This expression may be used to approximate moments of the surface height. For example, when (3.4) is valid, one has that

$$E[H_\varepsilon^c(\cdot, n)] = -nc + \frac{1}{2\varepsilon} \int_0^1 \prod_{k=0}^{\lfloor \frac{z}{2c\varepsilon} \rfloor} (2c\varepsilon k + (1-z)) dz \quad (3.5)$$

$$\simeq -nc + \frac{1}{2\varepsilon} \int_0^1 e^{[-\frac{1}{2c\varepsilon}(z+(1-z)\log(1-z))]} dz \simeq -nc + \frac{1}{2} \sqrt{\frac{c\pi}{\varepsilon}}$$

after a Laplace integral approximation. An expression for the variance,

$$Var[H_\varepsilon^c(\cdot, n)] \simeq \frac{1}{e} \sqrt{\frac{c\pi}{\varepsilon^3}} \quad (3.6)$$

is arrived at by a similar procedure.

Proof of Theorem 3.1. That past a certain point the law of the surface is stationary in time is one advantage of having the cutting wedge extend indefinitely. With n large

enough, the surface at time n is completely determined by the last $\frac{1}{2c\varepsilon}$ steps or cuts of the diamond array. Since these cuts are independent and of the same distribution, the stationarity follows. Simply observe that, for any k , $\min_x \{H_\varepsilon^c(x, k)\} = -kc$ and, after an additional $n^* = \lceil \frac{1}{2c\varepsilon} \rceil$ steps, $\max_x \{H_\varepsilon^c(x, k + n^*)\} = -kc$. In other words, in time n^* the current surface is erased and so does not figure into the statistics of the process for all future times.

Now focus on the height at some fixed $x \in [0, 1]$ and consider times n larger than n^* . Each time there is a new cut, the surface at x may or may not be affected depending on whether the height at x is greater than or less than the potential new height created by this cut. The latter in turn is a simple function of the distance between the center of the cut relative to x . (We use the word potential to indicate that that this would indeed be the new height provided there had been some surface at x to cut away.) The fortunate fact is that because surface refreshes itself so-to-speak in a finite number of steps, we do not need to keep track of whether a given cut is felt or not at x (we say real or potential) in order to describe the distribution of the surface at large times.

Not only is the value of H_ε^c at x at time $n - n^*$ replaced by a new height at time n caused by the intervening cuts, which particular cut is irrelevant: the height at time n is simply that due the cut of minimum height (maximal depth) achieved among the last n^* cuts. That is,

$$H_\varepsilon^c(x, n) = -nc + \frac{1}{2c} + \min_{n - \frac{1}{2c\varepsilon} \leq k \leq n} \left\{ \text{potential new height at time } k \right\}.$$

Computing the distribution of the later is not difficult. With X_k the uniformly distributed position of the k -th cut we have the following equalities in distribution:

$$-kc + \frac{1}{\varepsilon} |x - X_k| = -kc + \frac{1}{\varepsilon} |X_k| = \text{Uniform} \left[-kc, -kc + \lceil \frac{1}{2c\varepsilon} \rceil \right]$$

in which the first equality is due to the fact that we are working on the circle. This completes the derivation of (3.2), denoting by \mathcal{U}_k the random variable which is uniformly distributed on $[-kc, -kc + n^*]$.

The expected measure of that part of the surface below a fixed level now follows from the rotation invariance along with the representation (3.2) of the surface as the minimum of independent variables. With $H_\varepsilon^c(x, n) \in [-nc, -nc + \frac{1}{2\varepsilon}]$, we make the normalization in terms of $0 \leq z \leq 1$:

$$\begin{aligned} E \left[\text{meas} \left\{ x : H_\varepsilon^c(x, n) \leq nc - \frac{1}{2\varepsilon} z \right\} \right] &= \int_0^1 E \left[\mathbf{1}_{\{H_\varepsilon^c(x, n) \leq -nc + \frac{1}{2\varepsilon} z\}} \right] dx \\ &= P \left[H_\varepsilon^c(\cdot, n) \leq -nc + \frac{1}{2\varepsilon} z \right] = P \left[\min \{ \mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_{\frac{1}{2\varepsilon}} \} \leq \frac{1}{2\varepsilon} (z - 1) \right] \\ &= 1 - \prod_{k=1}^{\frac{1}{2\varepsilon}} P \left[\mathcal{U}_k \geq \frac{1}{2\varepsilon} (z - 1) \right] = 1 - \prod_{\ell=0}^{\lceil \frac{z}{2\varepsilon} \rceil} (1 - z + 2\varepsilon c \ell). \end{aligned}$$

The last line is due to the independence among the \mathcal{U}_k 's. The proof is finished.

4. Conditioning of Foamed Pads

Finally we come to the problem of constant rate conditioning in the presence of void-filled (foamed) pads. This process will simply be denoted $H^*(\cdot, n)$ though it depends

on c and ε in a similar manner as before. In fact, $H^*(\cdot, n)$ is described as the process $H_\varepsilon^c(\cdot, n)$ descending into a random environment which, importantly, is independent of its motion.

We model this random environment by centering a circular void of independent identically distributed radius at each point of a Poisson random field in the plane. (The idea being that the pad in the mechanical setup is a sample of a much larger piece of statistical homogeneous material, see Figure 2.) While this model itself is well studied under the name of the Poisson blob or Boolean model of Continuum Percolation [9], we have not seen the particular statistic needed here discussed directly in that (or any other) literature. Of interest to us in understanding the effect of such a Poisson cloud of voids on the one-point height distribution of H^* is the process induced along any fixed line in the plane by its intersection with the set of voids.

We denote the input parameters as ν , the Poisson rate, and \mathfrak{r} , the individual void radius with density $P(\mathfrak{r} \in dr) = f(r)dr$. It is assumed that the latter is compactly supported, $f(r) = 0$ for $r > R_0$; the pad cross section indicating there are no arbitrarily large voids. Next, let us fix a vertical line Z corresponding to a given horizontal coordinate x at which we want to measure $H^*(x, n)$. This line is only influenced by the voids in the strip $[x - R_0, x + R_0]$. Now, the (normal) projection of every point lying in this vertical strip onto the center line Z is also a Poisson point process with the new rate $2\nu R_0$. Of course, not all these points are seen in the sense that not every void will intersect Z . That is, this process must be thinned, leaving us with a rate

$$\lambda = 2\nu R_0 P(\text{given void intersects } Z) = 2\nu \int_0^{R_0} \int_0^h f(r) dr dh. \quad (4.1)$$

(Conditioning on a point in one half of the strip being at a fixed vertical coordinate, its horizontal position is uniformly distributed on $[0, R_0]$). Centered at each remaining point on Z (after thinning) is a hole of length \mathfrak{L} caused by the intersecting void. The law of \mathfrak{L} is easily computed. With h the distance from Z to the corresponding Poisson point and given that $\mathfrak{r} \leq h$ (the void indeed intersects), one has $\mathfrak{L} = 2\sqrt{\mathfrak{r}^2 - h^2}$ and so also

$$P(\mathfrak{L} = \ell) = \frac{\ell}{\int_0^{R_0} r f(r) dr} \int_{\ell/2}^{R_0} \frac{f(r) dr}{\sqrt{r^2 - \ell^2/4}} \quad (4.2)$$

in which the notation indicates a density in $\ell \in [0, 2R_0]$. The description is then as follows. Along any vertical line in the pad we have a Poisson point process of rate λ augmented by independent holes with law (4.2) centered at each such point. We will view this process restricted to a half-line $[0, \infty)^2$ but observe that as defined it is in stationarity. We denote by P_S the corresponding law with E_S its expectation. As a final remark, it is most convenient for the present application that it is the same to consider the holes as initiated at the individual Poisson points (instead of being centered as was natural from the construction):

Lemma 2. *The process P_S is equivalent in law to that of having independent random holes of distribution \mathfrak{L} initiating at each point of a Poisson process of rate λ .*

2. The process of cuts *descends* into the pad seen as occupying the negative half-line. Still, it is convenient to keep hole lengths and other distances along the line as positive quantities.

The proof of this fact is deferred to the end of the section. Continuing, we note that the individual holes under P_S may of course overlap forming (again independent and identically distributed) composite holes of arbitrary length. It is this composite length that is important here. More to the point, denote by \mathfrak{D} the distance from any given point along Z to bottom of the hole that point lies in. This definition is made noting that \mathfrak{D} will certainly have a positive probability of being zero (there being a positive probability the general point does not find a hole). With the distribution of \mathfrak{D} in hand, the distribution of $H^*(x, n)$ is easily seen to agree with $H_\varepsilon^c(x, n)$ of the previous section when that value does not happen to coincide with a hole, and, if it does, it is simply the sum of H_ε^c and the distance to the bottom of the that hole.

Now the determination of the law of \mathfrak{D} happens to be a problem in Queueing Theory, this length being connected to the busy periods of the $M/G/\infty$ queue, and we may follow ideas from [13] and [11]. The result is:

Theorem 4.1. *Bring in $\mathfrak{m} = E[\mathfrak{L}]$, the mean of the individual hole length (4.2), as well as the distribution function*

$$G(z) = (1 - e^{-\lambda \int_0^z P(\mathfrak{L} \geq u) du}) / (1 - e^{-\lambda \mathfrak{m}})$$

with g the corresponding density. We have

$$P_S(\mathfrak{D} \in dz) = e^{-\lambda \mathfrak{m}} \delta_0(dz) + (1 - e^{-\lambda \mathfrak{m}}) \sum_{k=1}^{\infty} (e^{-\lambda \mathfrak{m}} (1 - e^{-\lambda \mathfrak{m}})^{k-1}) g^{(k)}(z) dz \quad (4.3)$$

where $g^{(k)}$ denotes the k -fold convolution and $e^{-\lambda \mathfrak{m}}$ identifies P_S (no hole at a given point). The law of $H^*(x, n)$ is now determined by $H^*(x, n) = H_\varepsilon^c(x, n) - \mathfrak{D}$.

Remark. Recall that from the results of the previous section, Theorem 4.1 provides a closed form of the density of $H^*(x, n)$ for all sufficiently large n only.

The structure of the distribution (4.3) allows for simple expressions for moments of H^* at a given point in terms parameters coloring the model. For example,

$$E[H^*(x, n)] = E[H_\varepsilon^c(x, n)] - (e^{\lambda \mathfrak{m}} - 1) \int_0^\infty xg(x) dx$$

and

$$Var[H^*(x, n)] = Var[H_\varepsilon^c(x, n)] + (e^{\lambda \mathfrak{m}} - 1) \left[\int_0^\infty x^2 g(x) dx - \left(\int_0^\infty xg(x) dx \right)^2 \right].$$

More important is the description of the tail of the distribution (4.3) which we see to be exponential. This is detailed in the following.

Corollary 4.1. *The random variable \mathfrak{D} conditioned on being strictly positive is stochastically dominated above and below by random variables X_a and X_b whose laws are given by*

$$P(X_{a,b} \in dx) / dx = e^{-\lambda_{a,b} \mathfrak{m}_{a,b}} e^{-\lambda_{a,b} x} \sum_{k=1}^{\infty} \lambda_{a,b}^k P_k(x / \mathfrak{m}_{a,b}) \quad (4.4)$$

in which P_k is the density of the sum of k independent $[0, 1]$ -uniform variables and the constants are: $\lambda_a = \lambda$ and $\mathfrak{m} = 2R_0$, $\lambda_b = \lambda P(\tau \geq r)$ and $\mathfrak{m} = \sqrt{2}r$ for any $r \leq R_0$. The law(s) (4.4) exhibit an exponential tail: for $x \gg 1$, $x^{-1} \log P(X_{a,b} \geq x) \simeq -c < 0$ with $c = \lambda_{a,b}(1 + z^*)$ in which z^* is the non-zero real root of $z + 1 = e^{\kappa z}$ with $\kappa = \lambda_{a,b} \mathfrak{m}_{a,b}$.

Note that one is free to optimize over r in the above in order to produce the best possible lower estimate. The main point is that it has been demonstrated how a simple model of foamed pads can produce a left tail which is qualitatively similar to that of the empirical law pictured in the Introduction. In fact, a numerical fit of the empirical tail carried out in [3] indicates that it exhibits exponential decay.

Proof of Theorem 4.1. Let us denote by f the density of interest. To begin the analysis, we must consider a non-stationary version of our process P_S of random holes along the vertical line. In particular we take the same statistical description on $[0, \infty)$ with the condition that there is no hole at $z = 0$. We denote by P_0 , E_0 and the like the measure, expectation and other quantities computed under this 0-boundary condition. Then, according to [13] (p. 210),

$$\begin{aligned} M_0(z) &= E_0 \left[\# \text{ holes with initial point in } [0, z] \right] \\ &= \lambda \int_0^z P_0(u) du \equiv \lambda \int_0^z P_0(\text{no hole} \in du) = \int_0^z e^{-\lambda \int_0^u P(\mathcal{L} \geq v) dv} du. \end{aligned} \quad (4.5)$$

On the other hand

$$M_0(z) = (\tau)(z) + (\tau \circ R)(z) + (\tau \circ R \circ R)(z) + \dots \quad (4.6)$$

in which $\tau(z)$ is the exponential distribution of rate λ and R is the distribution of the distance between the start of two consecutive holes.

Back in our stationary regime these same quantities take the form

$$P_S(z) = \lim_{z \uparrow \infty} P_0(z) = e^{-\lambda m} \text{ and } M_S(z) = \int_0^z P_S(u) du = \lambda z e^{-\lambda m}, \quad (4.7)$$

in which we have identified the probability of the conditioner not finding a hole. And, similar to (4.6) there is the expression

$$M_S(z) = (\tau \circ f)(z) + (\tau \circ R \circ f)(z) + (\tau \circ R \circ R \circ f)(z) + \dots$$

This may then be combined with (4.6) and (4.5) to produce $M_S(z) = (M_0 \circ f)(z)$ and so also

$$\begin{aligned} \widehat{f}(s) \equiv \int_0^\infty e^{-sx} f(x) dx &= \left[s e^{\lambda m} \int_0^\infty e^{-sx - \lambda \int_0^x P(\mathcal{L} \geq u) du} dx \right]^{-1} \\ &= \frac{e^{-\lambda m}}{1 - (1 - e^{-\lambda m}) s \widehat{G}(s)}. \end{aligned} \quad (4.8)$$

The distribution $G(z)$ is as in the statement and and, of course, $s\widehat{G}(s) = \widehat{g}(s)$ with $g = G'$. Written another way this is

$$\widehat{f}(s) = e^{-\lambda m} + (1 - e^{-\lambda m}) \sum_{k=1}^{\infty} e^{-\lambda m} (1 - e^{-\lambda m})^{k-1} [\widehat{g}(s)]^k,$$

which may be inverted to complete the proof.

Proof of Corollary 4.1. The domination is done by coupling to an environment of square holes of deterministic size. To go above, clearly H^* is only made larger by replacing each random circular void by a square void of side $2R_0 = m_a$. On the other hand, removing any voids will make D and so H^* smaller. Let us remove all voids of radius smaller than some $r > 0$ which reduces the rate figuring in P_S to $\lambda P(\tau \geq r) = \lambda_b$. Things are made smaller still by then replacing each remaining void with the inscribed square of side $\sqrt{2}r = m_b$. In this way we have determined the parameters figuring into versions of the P_S induced by fixed square voids.

The rest of the analysis is the same and so the form of the law (4.4) follows from (4.3) upon noting that now: $P(\mathcal{L} > u) = 1$ for $u \leq m_{a,b}$ and 0 otherwise. It is then the case

$$g(z) = g_{a,b}(z) = \lambda_{a,b}(1 - e^{\lambda_{a,b}m_{a,b}})^{-1}e^{-\lambda_{a,b}z},$$

and (4.4) comes out with a little algebra.

As for the estimate of rate of exponential decay, this is seen from the Laplace transform of \mathfrak{D} in the present setting. Forgetting the (a,b) -subscripts, we consider P_S with rate λ and fixed individual hole length, $\mathfrak{L} \equiv m$. The transform (4.8) is then

$$\widehat{f}(s) = e^{-\lambda m} \frac{s + \lambda}{s + \lambda e^{-m(s+\lambda)}}, \quad (4.9)$$

and the decay of the corresponding law coincides with the abscissa of convergence. That is, one has decay of type e^{-cx} for $-c$ the right-most pole of (4.9). Note that the singularity $s = -\lambda$ is removable and that there is just one other real root of $s + \lambda e^{m(s+\lambda)} = 0$. It is the latter that determines the decay.

Changing variables $z = -s/\lambda - 1$, puts attention on the left-most solution (besides $z = 0$) of $z + 1 = e^{\kappa z}$ for $\kappa = \lambda m$. Take $\kappa > 1$, the opposite case being treated similarly. There is a real root z^* lying in $(-1, 0)$. To show there are no other roots of smaller real part, we apply Rouché's Theorem ([1], p.153) to $f(z) = e^{\kappa z} - (z + 1)$ and $g(z) = z + 1$ along a contour \mathcal{C}_R defined as follows. Pick a real x such that $z^* < x < 0$ and let \mathcal{C}_R consist of the vertical line segment $\Re(z) = x$, $|\Im(z)| \leq R$ which is then attached to the half-circle of radius R extending into left half-plane. It is required to show that $|e^{\kappa z}| = |f + g| < |g| = |z + 1|$ along \mathcal{C}_R , then by taking $R \uparrow \infty$ it follows that there is just the one zero of f in $\Re(z) < x$. On the vertical segment, we have that $e^{\kappa x} < x + 1$ by design, and this inequality maintains up and down the line as $|e^{\kappa z}|$ is constant there while $|z + 1|$ grows. The necessary bound for the half-circle is even easier. The proof is finished.

Proof of Lemma 2. This is accomplished by identifying the generating functions. First, build up P_S by discretizing. Let $\lambda_k = \lambda P(\mathcal{L} = k/N)1/N$ with $\ell_k = k/N$ for k running between 1 and $\lfloor 2R_0N \rfloor$. Consider the k -th process of Poisson points of rate λ_k with symmetric holes of (fixed) length ℓ_k about each point. P_S may be obtained by the superposition of these processes followed by the continuum or $N \uparrow \infty$ limit.

Next let μ_k denote the counting measure of points for the k -th process. The support of the measure

$$\eta_k(dx) = \int_{-\infty}^{\infty} 1_{[-\ell_k/2, \ell_k/2]}(x - y) \mu_k(dy)$$

describes the process of holes created at this level. Its generating function is

$$E \left[\exp \left\{ \int_{-\infty}^{\infty} f d\eta_k \right\} \right] = \exp \left\{ \lambda_k \int_{-\infty}^{\infty} \left(e^{\int_{-\ell_k/2}^{\ell_k/2} f(x+y) dy} - 1 \right) dx \right\}$$

in which f is any reasonable, compactly supported test function. Therefore, with $\tilde{\eta} = \lim_{N \uparrow \infty} \sum \eta_k$ the totality of holes under P_S we find that

$$E_S \left[\exp \left\{ \int_{-\infty}^{\infty} f d\tilde{\eta} \right\} \right] = \exp \left\{ \lambda \int_0^{\infty} P(\mathcal{L} \in d\ell) \int_{-\infty}^{\infty} \left(e^{\int_{-\ell/2}^{\ell/2} f(x+y) dy} - 1 \right) dx \right\}. \quad (4.10)$$

Now with the integration in x on the right of (4.10) extending over the line, a simple change of variables will shift the domain of the inner-most integral to $[0, \ell]$. But that is just what one would have found by carrying out the entire procedure, beginning with the holes initiating rather than centered at each Poisson point.

5. Comments and Future Directions

The ideal would be that our final model (Section 4) is simple enough (i.e., one can make rather direct estimates and some exact computations) while possessing enough freedom in its parameters to be tuned in line with the real statistics. We have made no attempt at carry out a true statistical comparison of this model to real data, but hope that such a test might be undertaken by practitioners. On this note we mention that even the statistics of the pad itself poses a possibly interesting problem. The Poisson cloud of voids considered here was taken out of sheer convenience (it allows us to compute) though it did reproduce qualitatively features of the height distribution indicated by numerics.

Next, recall that the goal along the way was to describe the evolution of the roughness of the pad. The engineering literature contains several studies of surface roughness (see [5], [6], and [7] for example), typically assuming an a priori statistical description of the heights (for example, multi-variate Gaussian). With this, measures of roughness are built out of curvature statistics, numbers of local maxima/minima, and the like. The present work marks a beginning for CMP in that we have computed the one-point function for the height, but certainly any real measure of surface deviation from planar should include spatial correlations or other higher order statistics. It would therefore be interesting to extend results of Section 3 to encompass the types of statics computable in the stationary model. (In fact, one reason we spent time on this simpler model was to contrast how much more one can compute here than in the more important constant rate case.) Of equal or greater interest would be multi-point correlations of the type $P(H_\varepsilon^c(x_0, n) \in dz_0, H_\varepsilon^c(x_1, n) \in dz_1)$, after which one can consider the same in the presence of voids.

Also, one obvious gap in the above is a description of the transient behavior of the fixed rate conditioning problem. It may be that optimal roughness is obtained prior to stationarity. In another direction, one might consider a non-constant velocity model in which the current rate depends on the current surface statistics. This is more in line with the actually process in which the conditioner descends into the pad due to a balance of forces.

Finally, we mention that the real process is three dimensional. It seems that it would not be too hard to extend the above in a natural way to that setting.

Acknowledgements

Many thanks to Tom Witelski for introducing us this problem. Thanks to both Tom and L. Borucki for ensuing conversations, access to their work [3] as it developed, and

the use of Figures 1 and 2 drawn from that paper. In addition, the second author gratefully recognizes the support of grant NSF-9983320.

References

- [1] Ahlfors, L. (1979) *Complex Analysis* (3rd edition), McGraw-Hill, Inc.
- [2] Borucki, L.J. (2002). Mathematical modeling of polishing rate decay in chemical-mechanical polishing. *Journal of Engineering Mathematics* **43/2-4**, 105-114.
- [3] Borucki, L.J., Please, C, Witelski, T, Kramer, P., Schwendeman, D. (2003). A theory of pad conditioning for chemical-mechanical polishing, *submitted to Journal of Engineering Mathematics*.
- [4] Durrett, R. (1991). *Probability Theory and Examples*. Waddworth and Brooks/Cole, Belmont, Ca.
- [5] Greenwood, J.A. (1984) A unified theory of surface roughness, *Proceedings of the Royal Society of London A* **393**, 133-157.
- [6] Greenwood, J.A., Williamson, J.B.P. (1996). Contact of nominally flat surfaces. *Proceedings of the Royal Society of London A* **295**, 300-319.
- [7] Greenwood, J.A., Wu, J.J. (2001). Surface roughness and contact: an apology *Meccanica* **36**, 617-630.
- [8] Lawling, A.S. (2002). Polish rate, pad surface morphology and pad conditioning in oxide chemical mechanical polishing *Proc. of the 5th Int. Symposium on Chemical Mechanical Polishing, Philadelphia: ECS PV202-1*, 45-60.
- [9] Meester, R., Roy, R. (1996). *Continuum Percolation*. Cambridge University Press.
- [10] Smirnov, N.V. (1949) Limit distributions for the terms of a variational series. *AMS Transl Series* **1**, **67**.
- [11] Stadje, W. (1985) The busy period of the queueing system $M/G/\infty$, *Journal Applied Prob*, **22**, 697-704.
- [12] Steigerwald, J.M., Murarka, S.P., Gutman, R.J. (1997) *Chemical Mechanical Planarization of Microelectronic Materials*. New York: John Wiley and Sons, Inc.
- [13] Takacs, L (1962). *Introduction to the Theory of Queues* Oxford University Press.
- [14] Taylor, R., Daffer, P., Patterson, R. (1985). *Limit Theorems for Sums of Exchangeable Random Variables*, Rowman & Allanheld, Publishers.