Title
Chemical mechanical polishing: asperity-based computational modeling at the feature and die scales

Author
Salazar, Brian

Publication Date
2021

Peer reviewed|Thesis/dissertation

Chemical mechanical polishing: asperity-based computational modeling at the feature and die scales

by

Brian Salazar

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering- Mechanical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Hayden K Taylor, Chair
Professor Grace X Gu
Professor Ronald Gronsky

Fall 2021

Chemical mechanical polishing: asperity-based computational modeling at the feature and die scales

Abstract

Chemical mechanical polishing: asperity-based computational modeling at the feature and die scales

by

Brian Salazar

Doctor of Philosophy in Engineering- Mechanical Engineering

University of California, Berkeley

Professor Hayden K Taylor, Chair

Semiconductor manufacturing is the process used to create electronic computer chips. Advancing semiconductor process technology enables the design of new computer chips with lower power requirements and faster speeds. For our society, this enables expanded availability of communications, higher levels of defense, more intelligent modes of transportation, and higher capacity consumer electronics.

Chemical mechanical polishing (CMP) is a crucial step in semiconductor manufacturing, as it enables the manufacture of ever-smaller integrated circuits that would not be possible without it. One significant CMP application is the shallow trench isolation (STI) process, where CMP is used to planarize oxide; this oxide subsequently forms dielectric isolation between transistors. While critical and ubiquitous in semiconductor manufacturing, CMP presents many open engineering problems, including the need to reduce chip defects and maintain high efficiency. For STI in particular, the relevant efficiency metric is the planarization efficiency, which evaluates the normalized difference in material removal rate within desired and undesired areas. Many components synergize to make an efficient CMP process, including the polishing pad, the polish material (wafer), the conditioning disk, and the abrasive slurry.

This thesis focuses on the polishing pad's role in polishing. As recent advances in additive manufacturing enable a much larger space of possible pad designs, this thesis provides a framework for modeling the efficiency of a pad design, which may guide optimal pad design selection. At present, polishing pad selection requires significant trial and error, which involves testing multiple pad designs (at a high cost) until the CMP tool user identifies a suitable pad design. The CMP tool user repeats this trial and error process for each new chip design since each has its polish requirements. A computational prediction tool that replaces this trial and error process would be beneficial for reducing the development time.

However, computational modeling of this process is a highly complex task, as the process contains billions of polygons that would be impossible to represent computationally. I present a framework for simulating the polish between a given chip layout and pad design. In particular, this work enables the simulation of polishes using segmented pads, consisting of a higher elastic modulus contacting material embedded in a lower elastic modulus material. The simulation of a multimaterial CMP pad in such detail is novel. The computational method uses a contact wear model and well-established tribology and contact mechanics techniques. In CMP, the pad's surface texturing plays a dominant role, as only a few asperities that protrude furthest from the pad are contacting the polish material. This thesis shows the use of microscopy techniques to scan an entire, used CMP pad surface. The computational method then uses that experimental scan as input to contact simulations.

Before simulating die polishing, we must understand how features with a given size, shape, and height behave when polished by a given pad. This thesis first presents a feature scale model, which predicts a pad's planarization efficiency when polishing features with known size, shape, and height. This feature scale model operates by simulating the solid contact between the pad and features across the range of relative positions and rotations and subsequently averaging the behavior across those positions and rotations. This information feeds into the die-scale model, which predicts the polish behavior for arrangements of various features.

As a result of this modeling work, we can look forward to application-specific CMP pad development. Until now, the design of a CMP pad relied heavily upon trial and error at high experimental cost; this model now informs that design process for the first time.

To my family, who immigrated to the U.S. so that their children may have access to education.

Contents

List of Figures

List of Tables

Acknowledgments

Over my past six years as a graduate student, I have met the most amazing people, and I am deeply appreciative to each and everyone of them. Perhaps my greatest joys emerged from teaching Engineering 27/29 for 8 semesters (where I cumulatively had 350+ students in my sections) and the Pre-Engineering Program's Maker Design Studio (PREP MDS) for 4 summers (where I cumulatively met an additional 350 students). While my role may have been "instructor," I assure you that it was I who learned something new from each interaction with these wonderful and bright UC Berkeley students.

I am thankful to all the wonderful undergraduate researchers who have worked alongside me over the years. To Nicole Panditi, Levi Seidel, Christina Ong, Kai Yun, Amanda Hernandez, Sean Sullivan, Jeffrey Spitz, Jie Qiu, Matthew Immesoete, Ricardo Avila, Edith LLontop, Sharjeel Laeeq, Hams Laeeq, Alberto Ibarra, Michael Hong, Ayush Sehgal, Lizzette Corrales, Claudia Renero Lopez, Hannah Dam and Zane Schemmer: thank you.

It has been a joy to work alongside other graduate student members in the Design for Nanomanufacturing group, including Dr. Hossein Heidari, Artos Cen, Dr. Brett Kelly, Dr. Hannah Gramling, Michael Loy, Nicole Luk, Alvin Li, Dr. Kristyn Kadala, Dr. Kyle Chung, and Joe Toombs. I'm extremely grateful to my closest collaborators from the civil engineering department, Dr. Parham Aghdasi, Dr. Ian Williams, and Dr. Yi Shao, and it has been a pleasure to meet with them on a near-weekly basis over the past years.

From my time at Appled Materials, Inc., I would like to thank Dr. Rajeev Bajaj, Daniel Redfield, Ashwin Chockalingam, Mayu Yamamura, Venkat Hariharan, Jason Fung, Raghava Kakireddy, Aniruddh Khanna, and Shiyan Jayanath for helpful discussions, and for all the CMP knowledge that you instilled upon me when I was a young graduate student.

I would like to extend my deepest gratitude to my dissertation committee chair and advisor Professor Hayden Taylor, and my success would not have been possible without him. I will always cherish our time teaching together, conversing about semiconductor manufacturing, and dreaming up new concrete reinforcement technologies. I am pretty sure the number of emails exchanged between us numbers in the thousands, and it takes an exceedingly special mentor like Hayden to have the dedication and energy to reply to those email and then still meet with me weekly. Thank you for providing me with another arrow in my quiver.

There are a few additional faculty members who have helped make my Berkeley experience so special. To my dissertation and qualifying exam committee members Professor Ronald Gronsky and Professor Grace Gu, I would like to express my deepest appreciation for all your guidance. To Professor Sara McMains and Liwei Lin who also participated in my qualifying exam committee, thank you for taking the time to learn about my work, for your advice, and for challenging me. I am extremely grateful to Professor Scott Moura, who took a chance and allowed me to give back to PREP (a program of which we were both participants at one point). Thank you for helping me develop my instructional skills. I cannot begin to express my thanks to Professor Claudia Ostertag, who has taught me so much and has been an exceedingly supportive mentor.

The mechanical engineering technical staff are superb in every way, Mike Neufer, Alex Jordan, Daniel Paragas, Scott McCormick, Dennis Lee, Katherine Hom, and Jacob Gallego astound me in their dedication and wisdom.

To Sofia Arevalo, thank you for being an amazing partner, for being my most trustworthy climbing belay partner, for many late night study sessions in Etcheverry Hall, for venturing into the Yosemite and Sequoia wildernesses to camp with me, and, the cherry on top, for many helpful discussions on nanoindentation of polymeric surfaces. And thank you for helping me raise the most adorable dog Jack, who loves nothing more than cuddling on the couch, and is the best companion that a computational researcher could ask for.

This research used the Savio computational cluster provided by Berkeley Research Computing at the University of California, Berkeley. This work also used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Specifically, it used the Comet system at the San Diego Supercomputing Center through the HPC@UC program.

Chapter 1

Introduction

1.1 Chemical mechanical polishing: background

Chemical mechanical polishing (CMP; also known as chemical mechanical planarization) is ubiquitous in semiconductor manufacturing. When an integrated circuit (IC)—such as your computer's processor—needs to be fabricated or when a micro-electro-mechanical system (MEMS) device—such as the accelerometer and gyroscope in your smartphone—needs to be produced, the CMP process is employed. Fabrication of those consumer products likely requires multiple CMP steps, making CMP a critical and frequently employed process. Because the deposition of thin films onto patterned semiconductor substrates is non-planar, CMP enables the fabrication of advanced IC chips and MEMS devices. That non-uniform film thickness complicates subsequent steps in the long sequence of semiconductor processes. Hence, CMP's role is to selectively polish the thin film to make the resultant film more planar and ready for subsequent process steps.

CMP tools, such as the one described in Section A.2, are used to perform the CMP process, and tool makers consistently release new tools. There is typically a high cost associated with running a CMP tool due to high equipment, consumable material, and cleanroom operating costs. As a result, tool users aim for low process times (also known as high-throughput). Figure 1.1 contains a schematic of a typical CMP tool. The polish material is a silicon wafer that has undergone at least one thin film deposition. The schematic shows a CMP process in which a wafer and pad rotate about their respective centers, with their rotations being non-concentric. Additionally, the wafer sweeps back and forth along the pad in one direction, which I will describe as horizontal. There are other CMP forms, but this dissertation focuses on this specific form since the experiments employed this particular setup.

The setup consists of a wafer loaded into a wafer carrier (also called the polishing head). Silicon wafers have standard sizes, with a 300 mm diameter being the typical size used in semiconductor fabrication plants. A retaining ring surrounds the wafer edge and controls the pressure at the wafer edge. The wafer carrier rotates and translates horizontally at

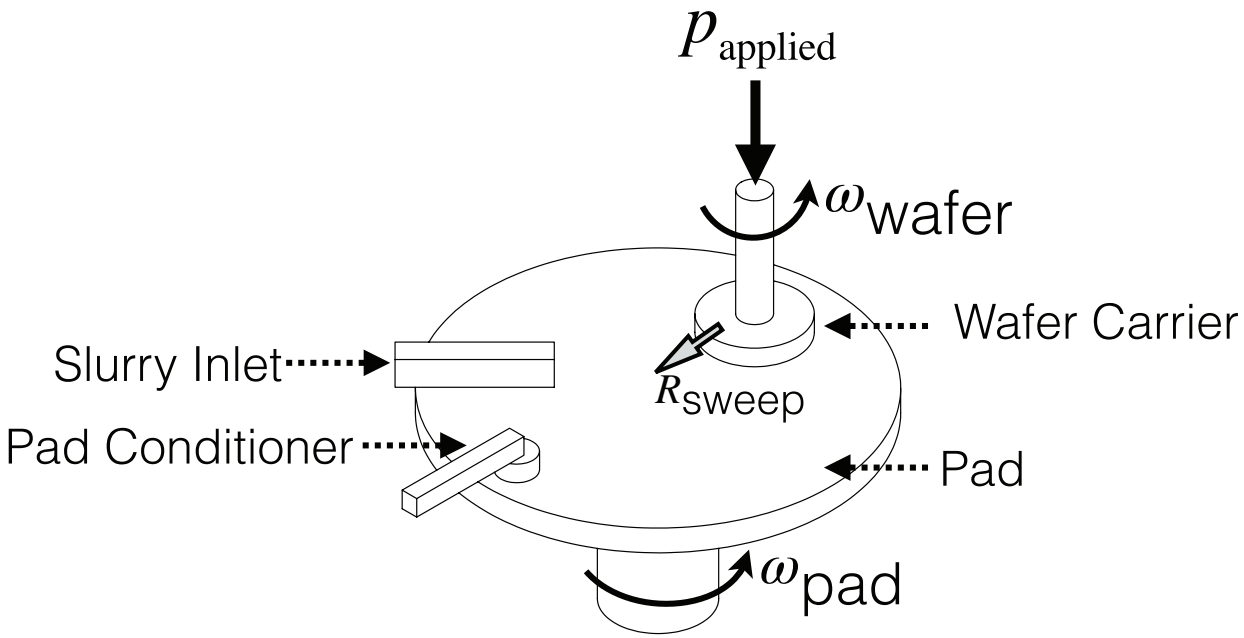


Figure 1.1: Schematic of a CMP tool, showing the location of the wafer carrier, pad, pad conditioner, and slurry source. The pad's rotational motion, the wafer's rotational and sweeping motions, and the applied pressure between the wafer and pad are illustrated.

prescribed speeds while a set down pressure is applied. Wafer carriers contain sensors that monitor the applied pressure on the wafer. Advanced wafer carriers may have a membrane between the wafer and retaining ring. This membrane allows for finer control of the pressure distribution: by creating zones that are a set radial distance from the wafer center and through the retaining ring, the average pressure within these zones may be adjusted to achieve higher uniformity.

At the interface between the wafer and pad is a slurry, which contains abrasive particles. A common abrasive particle choice is cerium oxide (ceria), which has a diameter on the order of tens of nanometers. Slurry has multiple roles: to provide the particles that abrade into the film and provide chemical modification and weakening of the film. Since slurry constantly flows out the system (as it flows off the pad edge), it is also constantly supplied into the system through an inlet.

This dissertation focuses on one particular CMP component: the pad. Its purpose is to provide the pressure required to remove material from the wafer. The pad typically has a diameter of about 76 cm and is made from a polymeric material. The pad is onto the platen, and it provides the rotational motion.

The fourth CMP component is the conditioning disk (also known as the pad conditioner or diamond disk), which imposes a certain roughness on the pad by mechanical abrasion. The disk contains multiple, small diamond protrusions; fine disk designs contain smaller, closer

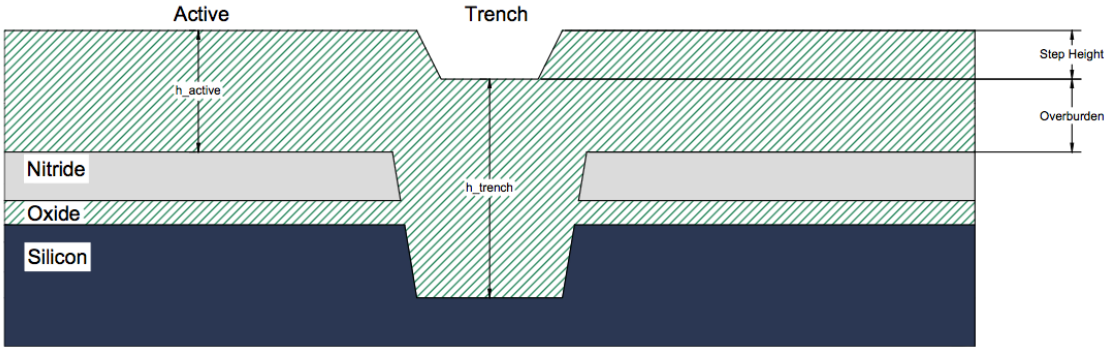
packed diamonds than in coarse disk designs. During conditioning, the conditioner presses onto the pad with a prescribed downforce, which causes the tallest diamond protrusions to abrade the pad [1]. Before any wafers are polished, the pad will undergo a break-in step, wherein it may undergo conditioning for around an hour. The conditioning allows the pad to attain a specific and uniform roughness characteristic. As the pad polishes a wafer, the pad itself will flatten (known as glazing) and wear, which can be detrimental to the polishing uniformity [2], [3]. Hence, the pad needs to be conditioned throughout its life so that its roughness can stay consistent over time. There are two conditioning options: ex-situ and in-situ. In in-situ conditioning, the conditioning happens while a wafer is being polished. The drawback is that a pad particulate that the conditioner has dislodged may contact the wafer and create a defect. In ex-situ conditioning, the conditioning happens in-between wafer polishes and not while a wafer is being polished. The drawback is that more time is required to process one wafer. During conditioning, both the pad and conditioning disk experience wear. Kim investigated this wear mechanism and found that the polish performance may vary over the pad's lifetime [4].

There are many forms of CMP, categorized into metal CMP and dielectric CMP, where the naming refers to the polish material. Two common metal CMP applications are the damascene and dual-damascene processes, where copper (a conductor) is polished. Two common dielectric CMP applications are interlayer dielectric (ILD) CMP and shallow trench isolation (STI) CMP. In ILD CMP, a dielectric material located between metal interconnects is polished. This dissertation focuses on the shallow trench isolation process, where silicon dioxide (also called oxide, an insulator) is polished. Section 1.2 further describes this process. The modeling principles discussed in this work are readily adaptable to include other forms of CMP.

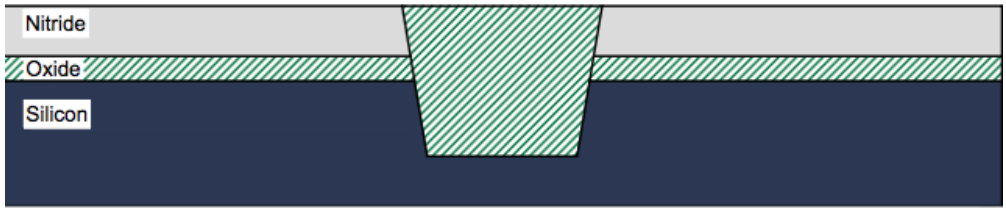
1.2 The shallow trench isolation (STI) process

The shallow trench isolation process is a multi-step process used for fabricating transistors and involves deposition, etching, patterning, and polishing steps. Figure 1.2a is a schematic that shows the wafer before the CMP step. The silicon substrate contains an etched feature. Above the silicon is a thin pad oxide layer, followed by a nitride layer and another thicker oxide layer. The schematic depicts the active and trench oxide regions, and a step between the two regions is visible. CMP's goal is to remove the active oxide and to flatten this step and, as shown in Figure 1.2b. The nitride layer serves as a stop layer and protects the oxide layer beneath it [5]. However, the CMP process is inherently non-ideal due to factors described in Section 1.6. Those factors mean that some features will polish faster than other features. CMP users must ensure complete active oxide clearing across the wafer for subsequent nitride stripping). To accomplish complete active oxide clearing, CMP users may decide to overpolish the wafer, leading to the pad potentially removing more trench oxide than desired. Figure 1.2c illustrates this undesired trench oxide removal (called dishing). Additionally, while the nitride layer has a higher resistance to polish than the oxide, it

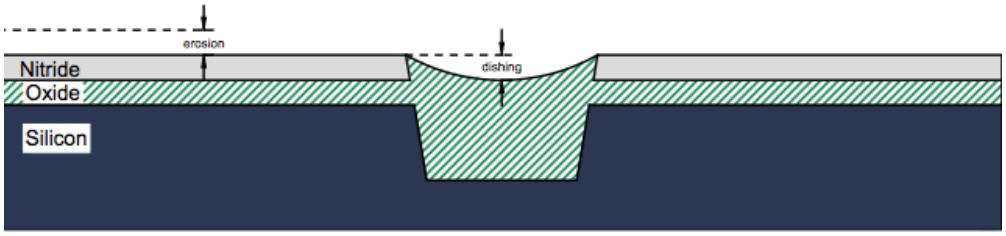
may slowly polish, leading to erosion; significant nitride erosion may lead to damage in the underlying active silicon [6].



(a) Prior to polishing



(b) Following an ideal polish



(c) Following a non-ideal polish

Figure 1.2: Schematic of a single feature on the wafer (in cross-section) undergoing the CMP step of the STI process. (a) shows the feature's geometry prior to polishing. (b) shows the feature's geometry following an ideal polish, including complete active oxide removal and no dishing. (c) shows the feature's geometry following a non-idea, realistic polish, including dishing and potential nitride erosion.

This dissertation focuses on models that can assist CMP users to attain that first process goal: to remove the active oxide as quickly as possible and reduce dishing. There exist published works that assist with the second goal of predicting the nitride erosion profile [7]. Future work may expand this model towards more accurate nitride erosion predictions.

1.2.1 Polish metrics

There are several quantitative metrics for determining the quality of a CMP process. For STI CMP, we define the planarization efficiency as

$$P.E. = \frac{R_{\text{active}} - R_{\text{trench}}}{R_{\text{active}}} \tag{1.1}$$

where R_{active} is the material removal rate in the active region (units of length per time) and R_{trench} is the material removal rate in the trench region.

The presence of the step between active and trench regions means that contacting asperities will compress to a lesser extent in trench regions than in active regions, leading to $R_{\text{active}} > R_{\text{trench}}$. It then follows that the planarization efficiency may have any value between 0 and 1. A CMP process with a planarization efficiency of 1 is ideal, and a planarization efficiency of 0 corresponds to having no selectivity between the trench and active oxide regions. Note that the planarization efficiency varies with respect to time and position on the wafer. The planarization efficiency's temporal dependence arises from the decrease in feature step heights as the polish proceeds. The smaller feature heights have a higher likelihood that a pad asperity will contact the trench and experience higher trench pressure since they may compress pad asperities to a greater extent than a feature with a larger step height. The planarization efficiency's spatial dependence arises from multiple factors, including the pressure distribution, slurry distribution, and temperature distribution.

Semiconductor manufacturers also seek to quantify the amount of variation in the oxide thickness, both for a specific die and across the entire wafer. The within-die-non-uniformity (WIDNU) is the range in oxide thickness across a specific die, and the within-wafer-non-uniformity (WIWNU) is the range in oxide thickness across the entire wafer. Ideal polish processes would have WIDNU and WIWNU values approaching zero. Additionally, the CMP process may introduce defects onto the wafer, such as embedded particles, scratches caused by slurry particles, and scratches caused by conditioning disk diamond debris [8]–[10]. A cleaning process is performed after CMP to remove some of these defects, but not all defects can be corrected. The defectivity is often expressed as the number of defects measured within a given area or as a count across the entire wafer, and the type of defect may also subdivide it.

Since CMP tool time is costly, CMP users seek to process wafers as quickly as possible. To quantify this, we evaluate the process' blanket removal rate. The blanket removal rate is the change in polish film thickness divided by the polish time for an unpatterned wafer, typically averaged across many locations on the wafer. The blanket removal rate depends on many process parameters, including the pad, wafer film material [11], slurry, applied pressure, rotational speeds, and conditioning disk.

1.3 Challenges for CMP processes

When a wafer needs to be polished, a CMP user must select various CMP process parameters, and costly trial-and-error experiments guide these parameter selections. CMP intimately couples the conditioning disk, slurry, pad, retaining ring, and wafer; changing any one of these consumables is likely to alter the polish behavior [12]. Furthermore, each of those consumables has additional process parameters associated with it, such as the conditioning disk's down-force and sweeping motion and the slurry's flow rate. Zhao discusses how the various process input parameters synergize to create multiple distributions across the wafer (for temperature, sliding distance, slurry, and pressure), which then determine the removal rate distribution [13]. A particularly non-obvious example of how multiple input parameters may interact is in regards to the wafer rotational speed to pad rotational speeds ratio and the wafer horizontal speed to wafer rotational speed ratio, which Zhao showed to have drastic effects on how uniformly a slurry particle slides across the wafer [14]. By developing a kinematic analysis of a single particle trajectory across the wafer surface, Zhao found that some combinations of those two process ratios result in some wafer regions where a particle did not contact as much as the remaining regions [14].

One of the CMP user's goals is to minimize the slurry mean residence time (so that fresh slurry may be incorporated) and maximize the slurry utilization efficiency (to minimize slurry wastage). Mu experimented with grooved pads and found that the mean residence time increased with the pad groove width and decreased with the applied pressure [15]. Additionally, Mu found that the slurry utilization efficiency increased with the pad groove width only up to a certain point before plateauing and did not change with the applied pressure.

The conditioning disk selection and the path that the conditioning disk takes as it sweeps across the pad are of great importance. CMP tools typically allow the user to specify the disk's velocity across multiple zones, with each zone specified as a certain radial distance from the pad center [16]–[18]. Lee investigated sweeping profile effects and determined that this choice affects the wear behavior across the pad, which may subsequently affect the polish behavior [19]. Ban developed a computational simulation of pad conditioning and used it to study the tilt angle's effect on the pad surface texture, where the tilt angle is the angle between the conditioning disk and pad [20]. Ban's simulations showed that for a zero tilt angle, the sweeping profile held a dominant effect on the pad's nonuniformity [20]. However, for a non-zero tilt angle, the tilt angle held a dominant effect on the pad's nonuniformity, compared to the sweeping profile [20].

Khanna determined that conditioning disk diamonds that are too sharp lead to an unstable CMP performance, and the conditioning disks themselves require a break-in period so they may reach a stable regime [21]. Khanna additionally discussed the coupling between pad design and conditioning disk selection and experimentally determined that a porous pad achieves the best uniformity with a fine conditioner. In contrast, a solid pad achieves the best uniformity with a comparatively moderately-coarse conditioner [22]. Khanna's experiments on conditioning disk parameter selection show that optimizing the pad's surface roughness

can lead to improvements in the material removal rate and planarization efficiency while also minimizing overpolishing effects and reducing the slurry flow rate requirements [23]. Additionally, Khanna found that optimizing both the surface roughness and the pad surface hardness improved the planarization efficiency and defectivity compared to a standard pad.

The process temperature is a parameter that a CMP user cannot specify but is crucial to determining and understanding the polish behavior, as elevated temperatures may affect the pad's mechanical properties and the slurry's chemical reactivity. Matsunaga found that an increase in the polish temperature results in a higher removal rate [24]. Khanna's experiments showed that the polish temperature might increase substantially, especially with long polish times, which leads to a decrease in the pad's storage modulus, and results in a higher removal rate and a lower planarization efficiency [25]. Similarly, Suzuki showed that pads may exhibit nonlinear viscoelastic behavior [26]. Viscoelastic pad behavior may affect the polish since pad asperities may more readily reach into trench regions as the loading time and operating temperature increase. A similar effect is that the pad—and especially soft pads—may shrink over time due to moisture loss [27]. That shrinkage results in non-uniform pad conditioning which may affect polishing performance [27].

As the semiconductor industry advances towards more sophisticated chip designs, with smaller feature sizes and more layers, CMP needs to advance with the goals of achieving a higher planarization efficiency, a higher removal rate, and a lower defectivity. Significant research efforts aim to improve all aspects of CMP, including both the CMP tools and consumables. Conditioning disk designers frequently introduce new disks with more sophisticated diamond designs, such as described in [28]. Additionally, the development of new slurries with new abrasive particles and chemical compositions progresses, as described in [12], [29]. Finally, pad development progresses, as described in Section 1.4.

As the number of available CMP consumables is consistently increasing, CMP tool users need guidance to select the optimal consumable combination. This work aims to guide CMP pad selection, which may significantly reduce the amount of trial-and-error required to optimize CMP for a particular wafer design.

1.4 Novel CMP Pads

CMP pads are composed of polymeric materials, which allows for sufficiently high contact pressures to generate material removal while allowing sufficient pad compliance and conformation to the wafer. Cook classified pads into four categories: I. pads based on impregnated felts, II. synthetic suede, III. closed-cell polyurethane foams, and IV. solid grooved polymers [30], [31]. Pad class III is the most commonly used, with pad class IV being the second most commonly used. This work discusses pads of both Class III and Class IV.

The use of a solid (non-porous) pad allows for attaining a higher planarization efficiency, while the incorporation of porosity allows the pad properties to be tuned (towards a lower effective modulus), which may result in lower defectivity at the cost of a lower planarization efficiency.

Pads contain both microtexture (imposed onto the pad by the disk conditioner) and macrotexture (determined by the pad groove designs); Li studied how that texturing may facilitate slurry transport [32]. When generating a surface microtexture on a solid pad, that texture is a function of only the conditioning disk, whereas when using a porous pad, that texture is a function of both the conditioning disk and the porosity size and distribution. Pads may be grooved, which means that grooves are cut into the polymer; these grooves may serve as pad compliance mechanisms and as slurry reservoirs. As the slurry is dispensed into a CMP process, there are two important metrics to evaluate: the slurry mean residence time is the amount of time that the slurry remains in the pad-wafer interface, and the slurry utilization efficiency is the volumetric flow rate ratio of slurry that contributed to polishing to the total amount [15].

Perhaps the most studied pad is DuPont's IC1000$^{\text{TM}}$ pad, composed of a porous polyurethane, is classified as a hard pad, and is available in various groove designs. IC1000$^{\text{TM}}$ is used as a top pad, then stacked and adhered onto a sub-pad. DuPont's Suba$^{\text{TM}}$ IV is the sub-pad commonly used with IC1000$^{\text{TM}}$; Suba$^{\text{TM}}$ IV is highly compressible, which allows the top pad to better conform to the wafer and hence improves wafer uniformity.

Modern manufacturing processes, such as 3D printing, allow CMP pads of a much wider range of geometries and material combinations to be fabricated [33]–[35]. The question is now: if CMP pads can be additively manufactured, how do we decide on optimal pad designs? This work aims to provide a CMP prediction tool that accounts for the pad's role. Our prediction tool may subsequently allow the pad design to be optimized for polishing a given die design.

1.4.1 Segmented CMP pads

The pads discussed in this work are known as segmented pads, which are monolithic and not composed of multiple different pads. These pads still contain distinct top and sub-pad regions, but the fabrication process allows the pad to be formed as one. The top pad consists of high modulus segments, which are the portions that contact the wafer. This segmentation allows the pad to achieve a degree of local planarity (across the width of the segment) while allowing longer range conformation across the wafer. In between the segments are grooves, which may serve as slurry reservoirs and affect slurry transport. A pad without any segments is likely to possess both high nonuniformity and low material removal rates since it cannot conform to the wafer, and the slurry cannot reach into the pad-wafer interface. The subpad is designed to enable the segments to flex and conform to the wafer, and it does not directly contact the wafer.

1.5 Material removal mechanisms

The mechanisms by which material removal occurs are not entirely understood. It is still a subject of debate whether pad asperities directly contact the wafer and maintain the entire

applied load or if the slurry carries some or even all of that load. Further, when supposing that the pad asperities are contacting the wafer, what is the nature of that contact and what is occurring at the interface. In particular, what percentage of material removal is caused by pad asperities directly contacting the wafer and what percentage is caused by slurry particles that are trapped between the pad asperities and the wafer?

1.5.1 Lubrication regimes and the Stribeck curve

For two contacting, solid, lubricated bodies, there are three possible lubrication regimes: boundary lubrication, mixed lubrication, and full film lubrication [36]. Other writers may refer to these three regimes by other names, such as contact lubrication, mixed lubrication, and hydrodynamic lubrication / hydroplaning. In boundary lubrication, the solid bodies are so close together that the pressure is carried by contacting asperities. In full film lubrication, there exists a lubricant film that prevents the two bodies from directly contacting. That lubricant film has a thickness and maintains a hydrodynamic pressure, and it is important to note that the lubrication film discussed here is distinct from the thin film that is being polished by CMP. There are two sub-regimes within full film lubrication: elastohydrodynamic lubrication (EHL) and hydrodynamic lubrication. For a system in a full film lubrication regime, if the deformation within at least one of the bodies is on the same order of magnitude as the film thickness, then it is considered as EHL; otherwise, if the solid deformation is negligible compared to the film thickness, it is considered as hydrodynamic lubrication. Finally, the mixed lubrication regime is a transition between boundary lubrication and full film lubrication: the pressure is carried by both contacting asperities (in some regions) and a hydrodynamic lubrication film (in the remaining regions).

To help identify the lubrication regime between two solid bodies, the Stribeck curve may be generated [37]. The Stribeck curve plots the coefficient of friction against the Hershey number on logarithmic axes. The Hershey number is the dynamic viscosity ($\mathrm{Pa} \cdot \mathrm{s}$) times the rotational speed $\left(\frac{1}{\mathrm{s}}\right)$ divided by the applied pressure (Pa).

Previous researchers have performed experiments to generate Stribeck curves of particular CMP processes [38]–[41]. Liang found that their copper CMP process likely operates under boundary lubrication [42]. Kasai found that their wafer may not have been held level during polishing but instead angled toward its leading-edge, leading to a non-uniform lubrication film thickness [43]. Doy presented a CMP process where the lubrication regime depended on the pad groove design; using a positive spiral design resulted in a mixed lubrication regime while using a negative spiral design resulted in boundary lubrication [44]. Recently, Borucki developed a theory where the lubrication film is envisioned to cycle between providing lift and suction forces on the wafer at high frequencies, and the lubrication film carries a higher percentage of the load when the wafer is in lift mode [45]. While there is no consensus within the CMP community on which lubrication regime a CMP process experiences, and it is possible that different CMP processes will undergo different lubrication regimes, the boundary lubrication regime is more stable and results in the higher removal rates that

CMP users often desire. As such, we will ignore any lubrication film thickness effects and develop our model for boundary lubrication.

1.5.2 Two-body vs three-body contact

As the pad is pressed against the wafer, there are two different types of contact: there is two-body contact in which the pad asperities directly contact the wafer, and there is three-body contact in which abrasive particles are embedded into the interface between contacting pad asperities and the wafer. The presence of abrasive particles at the contact interfaces leads to higher local pressures than if they were not present; the abrasive particles may be thought of as high hardness, low contact area indenters that plow through the chemically-weakened oxide.

Previous researchers have studied the abrasive particles' role in polishing. Bozkaya states that the majority of material removal is caused by three-body contacts [46], and Kim envisions polishing as a long sequence of particles producing fine scratches on the wafer [47]. Sampurno found that the material removal rate tends to increase linearly with the median particle size [48] and Srinivasan provided a detailed overview of chemical mechanisms that facilitate material removal [5]. Recently, Jeon performed experiments that showed that the particle material choice might alter the material removal rate's relationship with the contact area ratio [49]. In Jeon's experiments, increasing the contact area ratio resulted in a higher material removal rate when using ceria particles but resulted in a lower material removal rate when using silica particles [49]. Jeon posited that this behavior change was due to the particle's bond strength. Ceria forms a stronger bond with the oxide film than silica does, so the ceria particles were more likely to penetrate the film, whereas the silica particles were more likely to roll across the oxide instead of penetrating [49]. When using ceria, increasing the contact area ratio led to more contacting particles, and hence higher removal rates [49]. This is in contrast to the silica particles, for which increasing the contact area ratio led to lower pressures on individual particles, thereby increasing the probability of rolling instead of penetrating and leading to a lower material removal rate [49]. While this dissertation focuses on two-body contact, the role of abrasive particles could be incorporated later.

Knowing the contact area ratio between the pad and wafer is important. The contact area ratio is the ratio between the real contact area and the apparent (nominal) contact area. The contact area ratio depends on many CMP process parameters, including the pad design. A two-body contact scenario with a contact area ratio close to 1 would produce much lower contact pressures than a scenario with a contact area ratio close to 0. Section 1.9 provides a review of previous experimental efforts to measure the contact area ratio in CMP processes. This work proceeds by assuming that CMP's contact area ratio is indeed much smaller than 1% so that material removal occurs due to high contact pressures between pad asperities and weakened oxide.

1.6 Length scales at which to study CMP

CMP modeling is an interesting and complicated problem to study because of the multiple length scales at which polish occurs. Pads have a diameter of about 760 mm, and they polish wafers with a diameter of 300 mm. The wafers contain tessellations of dies, which will later be diced out of the wafer to become individual computer chips. Those dies contain feature arrangements of multiple sizes and shapes. Then, it is pad asperities and slurry abrasives which cause material removal.

Figure 1.3 shows the four relevant length scales at which CMP variation occurs, as discussed in this dissertation, and other researchers may categorize the length scales slightly differently [50]. The individual particle scale is on the order of 10 nm–100 nm, and the particle size and concentration are contributing parameters [48], as is the pad asperity's shape [51], [52]. At the feature-scale, which is on the order of 10 nm–1 mm, the feature's size, shape, and height are important. At the die-scale (on the order of 10 mm), the die's design—namely the pattern density arrangement—may affect the polish. Namely, a given feature's polishing may be affected by its neighbors. The final length-scale is the wafer scale, which is on the order of 300 mm. The wafer's average pattern density may contribute at this scale, as wafers with a larger feature area will distribute the pressure over that larger area, leading to lower local pressures and hence lower removal rates. Additionally, wafer edge effects are well-documented [53]–[55], as features near the edge of the wafer may experience higher local pressures and hence higher removal rates compared to features near the wafer center.

The accumulation of effects across these four relevant length scales leads to non-ideal polishes, and accounting for all of these contributors is a complex task. Additionally, because of this multi-scale nature, it would be computationally infeasible to use more traditional modeling methods, such as the finite element method, because the number of required elements would be too large. This dissertation focuses on two length-scales: Chapter 3 discusses feature-scale modeling and Chapter 4 discusses die-scale modeling. Future work could incorporate the particle and wafer scales.

1.7 Tribology and contact mechanics: previous work

As this dissertation relies heavily on and advances existing tribology and contact mechanics techniques, it is important to contextualize this work within the existing literature.

1.7.1 Single and multi-asperity models

Classical contact mechanics often deals with the stress distribution that develops in a smooth, elastic half-space when it is loaded by a smooth indenter, and Johnson [56] provides an overview of those methods. In particular, the classic Hertz theory of contact assumes that the materials behave elastically, the contact area ratio is small, each body is an elastic half-space, and the surfaces are frictionless. The finite element method has been used to simulate

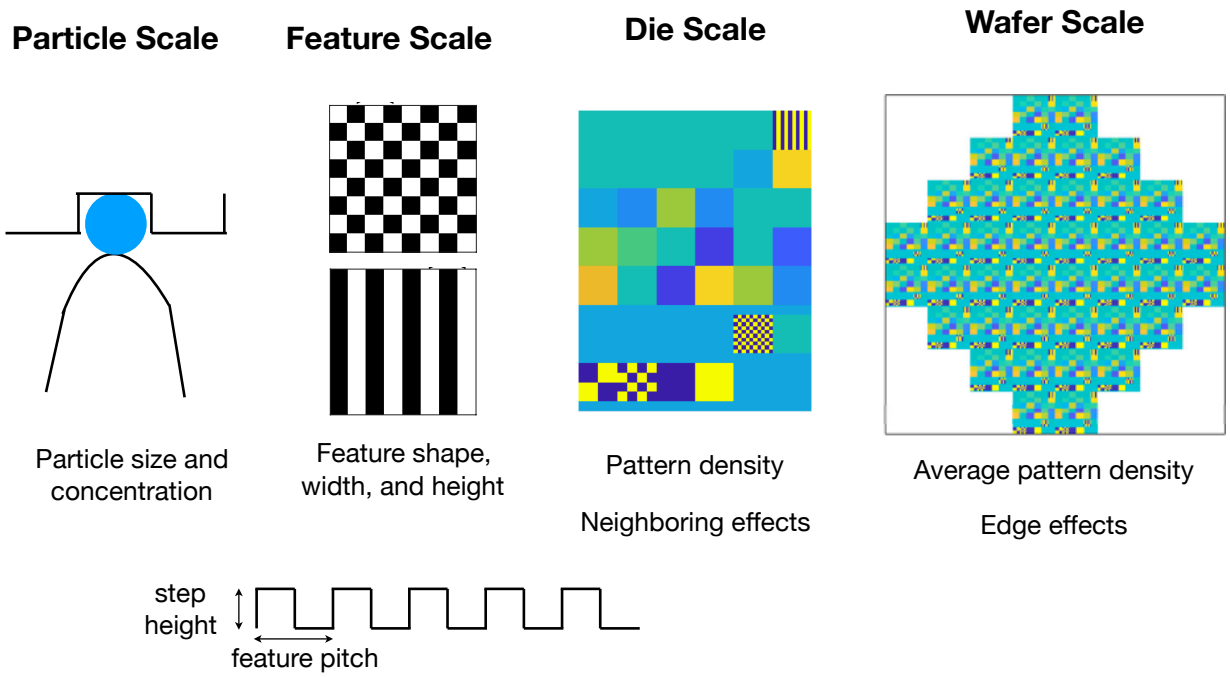


Figure 1.3: The various length-scales at which polishing non-uniformity occurs. At the scale of an individual particle, the particle size and concentration may affect the contact pressure. At the feature scale, the feature shape, size, and height may affect the feature's planarization efficiency. A die's layout, and especially its pattern density distribution, may lead to non-uniform polishing at the die scale. Finally, at the wafer scale, the wafer's average pattern density may affect the average removal rate, and the polishing behavior near the wafer edge may differ from that near the wafer center.

the contact between a rigid, single asperity (indenter) and an elastic, perfectly-plastic layer on an elastic, perfectly-plastic substrate, but the finite element method is limited to small simulation domains, due to computational cost [57].

Real contact scenarios contain surfaces where multiple asperities are contacting, and not where only a single asperity is contacting, and the solution procedure for multi-asperity contact problems often uses results from single-asperity contact problems [58]–[60]. Greenwood and Williamson developed the often-cited theory for the contact of two rough, nominally-flat surfaces; Greenwood and Williamson [61] is one of the early works where a multi-asperity model is developed. Critically, Greenwood and Williamson employ Hertzian contact theory and assume that the contacting asperities deform independently of each other, which drastically simplifies the contact problem and may not be realistic for CMP. Additionally, the Greenwood and Williamson theory allows for a small amount of plastic deformation, which occurs when an asperity's contact pressure exceeds a certain value, and Greenwood and Williamson [61] argue that the elastic theory may still be employed as long as the plastic

contact zone is small in comparison to the elastic contact zone. They implement their theory for the cases when the asperity heights have either an exponential or Gaussian distribution and use these assumptions to predict the contact area. Komvopoulos used a 2D finite element method to simulate the contact between a rigid surface with multiple asperities and an elastic half-space, and found that closely-spaced, contacting asperities may influence each other [62].

1.7.2 Models that incorporate measured surfaces

The use of measured roughness profiles directly in simulations is a significant milestone for contact mechanics theory, as the measured profiles may more closely match reality than theoretical distributions. Webster analyzed the contact between two rough surfaces, in 2D, and incorporated 2D roughness which was measured from a real surfaces [63]. Cole expanded that analysis for layered surfaces, by investigating the contact between a flat, elastic layer on an elastic substrate and a rough, elastic indenter, in 2D, also with 2D measured roughness [64].

Following that work, many researchers have analyzed layered, rough surfaces in 3D; most notably Nogi and Kato analyzed the frictionless contact between layered, elastic bodies and a rough, rigid indenter where the indenter's topography was represented as the sum of a smooth sphere and 3D roughness measured from a nominally flat, real surface [65]. Nogi and Kato's work is also significant for employing the conjugate gradient method (CGM) to iteratively solve the inverse problem (i.e. to determine the pressure distribution required to impose a certain displacement) and, within each of those CGM iterations, using the fast Fourier transform to determine the surface displacements given a pressure distribution. The Nogi and Kato method also uses an iterative approach to determine the contact region, which was previously described by Johnson [56]. The boundary element method approach taken by Nogi and Kato, where the contacting layer's pressures are calculated only at its contacting boundary and not throughout the entire body, is significantly less computationally expensive than a finite element method-based approach.

Peng and Bhushan (and later Cai and Bhushan) built upon Nogi and Kato's work, but instead of using a CGM solver, Peng and Bhushan used a quadratic programming approach for solving the inverse problem, which may yield better results in contact scenarios with large contact area ratios [66], [67]. Though Peng and Bhushan presented results using artificially generated surfaces, they also stated that the method applies when using measured surfaces. While previous research modeled surfaces as elastic–perfectly plastic, Chen et al. and later Wang et al. developed contact mechanics theories for an elastic–plastic surface, and presented results using measured roughness profiles [68], [69].

Other models take experimental asperity scans and fit characteristic parameters to form a representative surface, such as a sinusoidal surface [68] or a fractal surface [70], [71] for use in contact simulations. Yan developed the theory for contact between 3D fractal surfaces and an elastic–plastic layered media [72].

1.7.3 Recent advances and applications

More recent advances in contact mechanics focus on incorporating the role of adhesion at the contacting interfaces [73], [74] and on developing the boundary element method for the mixed lubrication regime (both solid–solid contact and hydrostatic fluid carry the load) [75]. Additional studies on the contact between nominally flat surfaces investigate the role of dynamic sliding effects [76], the behavior of oblique, contacting asperities [77], [78], and the importance of asperity spacing and total number of asperities when one of the surfaces is regularly patterned [79]. Contact mechanics-based approaches, like the ones discussed here, have been used to model a variety of manufacturing processes, including the micro-embossing of thermoplastic layers [80], nanoimprint lithography [81], and metal imprinting [82].

1.8 Previous CMP models

CMP modeling is a rich field, and there exist hundreds of previous modeling attempts in the literature. The modeling goals vary from researcher to researcher but the understanding gained from previous models is invaluable.

1.8.1 Models at a single length scale

A wide body of previous CMP modeling research exists, and specifically, many models focus on CMP at one particular length scale. Stine, Ouma, et al. [83]–[85] significantly advanced the understanding of CMP's pattern dependence. Stine et al. presented an analytical model for the removal rate as a function of the die's pattern density which, while limited in its predictive ability, is a remarkably simple and thus useful model [83], [84]. Ouma et al. expanded upon that work by introducing the concept of the effective pattern density, and using the effective pattern density instead of the local pattern density within the model [85]. The effective pattern density is computed by convolving the local pattern density with a filter, where that filter has a characteristic length equal to the planarization length. The planarization length is the distance over which the pad is able to completely conform to the polish material's topography [8], [85]. Ouma et al. showed a method for extracting the planarization length from experiments using a characterization pattern layout, and their experiments showed the planarization length to be several millimeters [85].

Many existing models are based on Greenwood and Williamson's theory, including at the asperity scale [86], [87] or at the feature scale [88], [89]. Vlassak developed one such Greenwood and Williamson-based model at the feature scale, and discussed that the model is limited to only features with size greater than the asperity size [90]. Seok used finite-element software to determine the stiffness of asperities of various sizes, which are assumed hemispherical in shape [91]. Boning provided an understanding of a pad asperity's modulus; by using a Greenwood and Williamson–based approach and modeling the asperity heights as having an exponential distribution, Boning developed a relationship between the contact area ratio and the pad asperity modulus [92]. Laursen et al. presented a 2D feature-scale

model which predicts the polish film thickness as a function of the polish time, by calculating the contact pressures between the pad and film [93].

1.8.2 Models which integrate fluid and solid mechanics

There also exist CMP models that incorporate the effects of both solid contact and fluid mechanics, at the feature-scale [94] or at the wafer-scale [95]–[97]. Wang developed a theoretical feature-scale model using smoothed particle hydrodynamics [98]. By envisioning a single asperity with a given shape (namely, a Gaussian shape), Wang found that the fluid film pressure depends on the feature shape. Notably, the feature-scale model analyzed various asperity to feature configurations, by allowing the asperity to sweep horizontally across. However, the approach assumed the pad asperity to be rigid, and pad deformation may significantly affect the pressure distribution.

Significant insight may be extracted from multi-scale CMP models. Kim created an asperity- and wafer-scale model, with a modified version of Greenwood and Williamson's theory where the asperity height distribution is a Gaussian distribution that is clipped to have a maximum asperity height [99]. By solving for both the solid–solid contact stresses and the fluid pressures using finite element analysis, Kim found that the contact pressure may vary significantly across the wafer.

1.8.3 Models which account for the spatial dependence of applied pressure

Suzuki developed a finite-element analysis-based wafer-scale model for CMP processes which use an airbag-type wafer carrier [55]. By simulating the contact stresses between the wafer/retaining ring and pad, Suzuki found that the wafer edge's contact pressures vary significantly from the wafer center's contact pressures. When comparing against experimentally measured material removal rates across the wafer edge, Suzuki found that the simulated contact pressures did not directly correlate with the material removal rates, and posited that the coefficient in Preston's equation (discussed in Section 2.2) varies with respect to position, since the abrasive particle distribution may be uneven. Chen also developed a finite-element analysis-based wafer-scale model, and Chen's analysis allows for multiple membrane zone pressures within the wafer carrier [100]. Chen modeled both flat and bowed wafers, and found that the zone pressures could be adjusted to reduce the size of the wafer edge affected zone for both types of wafers. Wang similarly developed a finite-element analysis-based wafer-scale model, which allows for multiple zone pressures, and found strong agreement between simulations and experimentally-measured material removal rates [101], [102]. Zhao expanded upon that work, and found that the zone pressure selections may lead to wafer bending, which affects the fluid film pressure [54].

1.8.4 CMP models that incorporate measured roughness

A limited number of CMP models directly used measured surface roughness. Terrell et al. used a voxel-based simulation approach, and while the model incorporated wafer topography measured using profilometry, it also used a statistically-generated pad topography and was implemented for only small domains [94]. Similarly, in [103], Terrell et al. scanned the substrate for use in simulations, though they also assumed the pad was flat. Gray modeled the static contact between pad asperities (taken from a confocal microscopy scan of a small pad portion) and a blanket wafer, which allowed for contact area ratio estimation [104]. Yu incorporated a measured pad topography to calculate the contact area ratio between the pad asperities and blanket wafer and the contact area ratio's interactions with the pad design, but did not allow for the plastic behavior of asperities [105].

1.8.5 Understanding the contact temperature

Borucki developed a 3D fluid flow and thermal model for copper CMP, which also incorporates a Greenwood-Williamson model [106]. He found that contacting asperities may experience a higher temperature than previously understood, with simulations suggesting that contacting asperity summits reached as high as 180 °C while the pad temperature (at a spot near the wafer leading edge) remained closer to 40 °C.

1.8.6 Understanding the mechanics of grooved pads

Baisie et al. developed finite element analysis-based models for pad conditioning [107], [108]. By assuming linear-elastic, isotropic pad material properties, modeling the pad porosity as a reduction in the pad's effective modulus, and assuming that the contact interface between the pad and conditioner is perfectly smooth, Baisie et al. provided insights into how conditioning occurs. In particular, Baisie et al. discovered that as the simulated conditioner contacts the pad, it is imposing a tilting motion onto the pad segment, and it is the softer sub-pad deformation which allows the segment to remain closer to planar [108].

1.8.7 Asperity-based models for the polishing of patterned wafers

It is worth noting that many of the preceding works involved the study unpatterned wafers and/or flat contact scenarios. We will now devote our attention exclusively to asperity-based contacts with patterned wafers. Feng developed a feature-scale model, where the pad is envisioned as a linear-elastic half-space, the pad texturing is modeled using either a Gaussian or non-Gaussian distribution, and the asperity distribution is characterized using pad scans [109]. By using a Green's function approach to predict dishing in a metal CMP process, the interactions between neighboring asperities are not neglected and the model is able to predict increasing dishing amounts as the feature size increases.

Xie [110] and Fan [111], [112] both developed models that assume asperities to behave elastically, and to have negligible width. Thus, their models have no asperity size dependence, and an asperity could just as easily reach into a small trench as it could to a large trench. Xie's pattern density step height (PDSH) model is particularly notable, as it predicts the active and trench pressures as a function of only the feature's step height, by assuming an exponential asperity height distribution and the mean asperity height λ [110]. Meanwhile, Xu used an exponential asperity height distribution to determine the amount of feature dishing as a function of the feature size in a metal CMP process [89]. The height distribution is modified so that there is an asperity-size dependence, through the use of a fitting parameter. While the use of statistical asperity height distributions may be insightful, scans for pads used in this dissertation do not show the asperities in these experiments to have such a mathematically-represented distribution. It is critically important to accurately capture the asperity profile, as they determine the real contact area, and thus the contact pressures.

1.8.8 Multi-scale CMP models

No review of CMP modeling would be complete without discussing the influential work by Luo and Dornfeld, who developed theories at multiple scales, including a 3D feature-scale model that simulates evolution, a wafer-scale model (for both the boundary and hydrodynamic lubrication modes), and a feature to die-scale model [113]. The multi-scale model operates by envisioning pads to have a bulk layer with given stiffness and a uniform, flat asperity layer with given thickness and stiffness. Since that asperity layer may have lower stiffness than the bulk layer, the model is better able to capture dishing and feature size dependence. The model procedure involves multiple steps, including using a 2D boundary element model to determine the feature's effective asperity layer stiffness, which would lead to realistic feature size dependence. Then, that effective asperity layer stiffness is used in a 1D topography evolution model, where the pad mechanics are now envisioned as springs and dashpots and not with a boundary element approach. This last point is why it was crucially important to use the 2D model; if the asperities are envisioned as independent springs, there is no feature-size dependence, but the use of an effective asperity layer stiffness whose value depends on the feature size (a larger feature corresponds to a lower asperity layer stiffness) does allow for feature size dependence. Within this 1D model, at the feature-scale, the contact between the wafer and the asperity layer is computed and at the die-scale, the contact between the wafer and the pad bulk layer is computed so that topography evolution may be predicted.

1.9 Existing contact area ratio estimations

1.9.1 Static methods for measuring the contact area ratio

Many researchers have attempted to experimentally measure the contact area ratio under dry, static conditions [114]–[122]. This typically involves pressing a transparent material (sapphire or quartz) onto the pad and scanning through the transparent material with a confocal microscope. Sun used this technique and found a linear relationship between the contact area and the applied pressure, with the contact area ratio under a 2 psi applied pressure being 0.026% [115]. Isobe measured contact area ratios between 0.1% and 0.5% under a 35 kPa (5 psi) applied pressure [119]. Jiao found that the contact ratio is highly dependent on the pad surface temperature [118]. By placing a heating element below the pad, and calibrating a model for the pad surface temperature, Jiao found that the contact area ratio increased from 0.029% at 25 °C to 0.092% at 45 °C. Elmufdi measured contact area ratios between 1.0% and 3.6% for a 3 psi polish pressure process with various pads [120]. Elmufdi's subsequent work found that the contact area ratio is highly dependent on the conditioning disk choice and conditioning time, with the measured contact area ratios ranging from 0.07% to 13.40% (for a 4 psi polish pressure process and various pad options) [121].

1.9.2 Dynamic methods for measuring the contact area ratio

It is difficult to accurately measure the contact area ratio from experiments, since the presence of slurry, velocity, and temperature all affect the pad deformation, but these variables are difficult to simulate using the previously described static contact area measurements. Instead, researchers have created dynamic contact area measurement apparatuses.

Uneda developed a method for in-situ contact area measurements, where the pad contacts against an image rotation prism in the presence of slurry. [123], [124]. When light was incident upon the top of the prism, the prism bottom completely reflected the light at locations where there was no contact and diffusely reflected light at locations where there was contact. These images were then captured using a camera, and the contact area ratio was found to range between .3% and 1.1%, with a 10 kPa (1.45 psi) applied pressure.

Gray conducted dynamic contact ratio measurements on grooved pads, using Dual Emission Laser Induced Fluorescence [104]. In these experiments, the pad was pressed and rotated against a transparent window, and two dyes were incorporated into the slurry. The light emission through the transparent window provided insight into the fluid film thickness, with higher light intensity corresponding to a larger film thickness. The contact area ratio was extracted from the fluid film thickness, and the contact ratio ranged between 0.1% and 4% for a 1.7 psi process. Gray additionally discussed possible limitations associated with performing these dynamic measurements on grooved pads, as the slurry may accumulate at the segment edges, thereby obscuring the measurements, and recognized that noise in the measurement process prevented contact ratio measurements below 0.1%.

Jiang used a simulation-based approach to estimate the contact area [125]. By scanning a small pad portion with a confocal microscope, Jiang imported that texture into a finite element analysis package, applied a nonlinear pad material model, and imposed contact with a rigid flat plane. By simulating the loading and unloading, Jiang was able to measure the amount of plasticity that developed in contacting asperities, and found that plasticity developed not only in the contacting asperity tips but also in the surrounding areas, suggesting that asperities do not behave elastically and independently.

1.9.3 Implications of previous contact area ratio measurements on CMP fundamental understanding

The preceding discussion illuminated the wide range in contact area ratio estimates, with estimates spanning from as low as 0.026% to as large as 13.40%. When interpreting these previous measurements, it is important to recognize that each experiment was performed under different conditions—pad, slurry, conditioner, substrate material, applied pressure, etc.—and those conditions directly affect the contact area ratio. Additionally, each methodology contains some amount of measurement uncertainty, and the associated resolution of imaging might prevent measurement of small contact area ratios ($< 0.1\%$). These two factors make cross-comparison exceedingly difficult but it is still possible to draw generalizations about where CMP typically operates. Most previous work found that CMP processes typically operate with a contact area ratio around or below 1%, and potentially multiple orders of magnitude below 1%, and only a minority of studies found higher values. This operating zone with a contact area ratio below 1% may be desired for high removal rate processes, since the sparse contact leads to high pressures and slurry access (though it may have drawbacks for defectivity). For processes where the contact area ratio is below 1%, it is the tallest few pad asperities that dominate the contact and models that assume flat pads are unrealistic.

An important limitation of all the works discussed in this section is the use of non-patterned (i.e. blanket) wafers. While using blanket wafers to characterize a CMP process and to estimate the contact ratio is very insightful, there is not yet enough understanding on how the polish behavior translates from blanket wafers to patterned wafers.

1.10 The need for CMP modeling

CMP is an exceedingly complex process, with decisions on the consumable set and tool parameters being intimately coupled with the design that is to be polished. The industry provides hundreds of options for just one consumable, and that consumable choice has a domino effect on the remaining consumable choices. At present, many CMP process decisions are driven by trial and error (such as in Wang [126]), which can be time-consuming and costly. At the same time, CMP is a ubiquitous process and is used for many steps in the fabrication of the technology that drives our world, so reducing trial and error could lower production costs, increase factory throughput, and enable fabrication of more advanced chips.

Because the pad-wafer interface is so small, and consists of opaque materials that are rotating at high velocities, we cannot directly measure what happens at the interface. We can, however, use models to extract insights on what occurs at the interface, and use those insights to engineer more efficient processes.

Additionally, CMP users may have conflicting process goals, as there is often a tradeoff between planarization efficiency and defectivity, a tradeoff between the removal rate and defectivity, and another tradeoff between the WIDNU and the WIWNU. Simulations may assist in finding optimal process parameters and in designing new process consumables, which can enable advanced semiconductor designs.

One particular modeling thrust that would benefit IC designers is the ability to integrate CMP models into computer-aided design software, such as presented in [127]. If IC designers possessed the ability to quickly predict how a particular layout would behave in CMP, they could detect non-uniformity hotspots and take actions to reduce those impacts. One method for this is to use dummy-fill (i.e. incorporate additional features into the layout which do not contribute to the final chip's performance but aid in maintaining polishing uniformity). The decisions on where to incorporate dummy fill may be driven by modeling efforts [128].

CMP is filled with unanswered, fundamental questions which modeling efforts may attempt to answer. Table 1.1 contains a listing of open questions, and this list is by no means comprehensive.

Table 1.1: Listing of various fundamental questions which CMP models may aim to answer.

With respect to conditioning
- What is the optimal surface texturing to impart to a pad surface for a high removal rate and planarization efficiency yet low defectivity?
- How consistent does the pad surface need to remain as polishing progresses, to reduce wafer-to-wafer non-uniformity?
- What is the optimal disk sweeping profile for minimal variation across the pad?

With respect to pad design
- What is the optimal pad segment size and modulus for a high removal rate and planarization efficiency yet low defectivity?
- What is the optimal pad groove width and modulus for a high removal rate and planarization efficiency yet low defectivity?
- What is the optimal subpad thickness and modulus for a high removal rate and planarization efficiency yet low defectivity?
- How are the pad mechanics affected by the temperature?

With respect to feature-scale considerations
- How does a feature's planarization efficiency depend on the pad's segment size?
- How does a feature's size, shape, and height affect its planarization efficiency?
- How can the amount of dishing within a feature be predicted / minimized?

With respect to die-scale considerations
- How can a designer optimize their layout design for minimal within-die-non-uniformity?
- How can a process' planarization length be determined?
- How can optimal dummy fill patterns be determined?
- Can die topography be predicted sufficiently fast for computer-aided design applications?
- Within a die, where are the non-uniformity hotspots?

With respect to wafer-scale considerations
- How can a process achieve balance between the WIDNU and the WIWNU?
- Can we predict the polish time required for all features across the wafer to polish?
- How do the process parameters (such as the applied pressure, wafer's rotational speed and the pad's rotational speed) affect non-uniformity?
- How can the optimal head membrane zone pressures be selected?
- How can the size of the affected zone at the wafer edge be reduced?
- How do the retaining ring size and positioning affect the polishing near the wafer edge?

With respect to the process in general
- What is the material removal law?
- How can the optimal slurry chemistry, particle size, and particle distribution be selected?
- Is the process operating in a stable lubrication regime or in an unstable regime?

1.11 The need for a better CMP model

Recent advances in manufacturing have enabled new and almost infinite pad designs, and it is these new manufacturing capabilities that have ignited the need for this modeling work. Pad manufacturers seek pad designs that achieve high planarization efficiencies for a wide variety of features, low defectivities, and high conformity across the wafer. They have many fundamental questions, including those listed in Table 1.1.

Despite all the insightful modeling efforts described in Section 1.8, additional modeling work is still required. In particular, a better CMP model requires

- An understanding of how neighboring asperities interact, and not a representation of pad asperities that behave as independent springs.
- Realistic pad textures, based on pad scans, and not flat pad surfaces or purely theoretical asperity distributions.
- A thorough representation of the pad mechanical behavior as a composite material, including the presence of segments and multiple constituent materials, and not a lumping of the pad behavior into a layer with an effective modulus.
- Analysis for patterned wafer, and not solely blanket wafers whose polish behavior may differ from that of patterned wafers.
- An understanding of multi-scale polish behaviors, as they may be more accurate and/or computationally efficient than models which represent only one length scale.

The work presented here, and previously in [129], [130], is the first model which attempts to incorporate these five requirements.

1.12 Thesis structure

Chapter 1 has reviewed existing understanding of the CMP process, summarized previous modeling efforts, and motivated the need for a more sophisticated model. Chapter 2 provides a contact mechanics based, 3D framework for predicting the pressure distribution and material removal between the wafer and pad. The contact mode is assumed to be solid-solid, and the boundary element approach described accounts for interactions between neighboring asperity contacts. The pad design is intimately studied, so that the pad segment, subpad matrix, and the pad asperity deformations are all accounted for and based on the constituent material properties. The wafer features have topography, and are represented as rigid compared to the polymeric pad. Importantly, the pad texturing is accounted for, by scanning used pads and incorporating the measured pad asperities within the wafer topography.

The contact mechanics framework described in Chapter 2 is then employed at the feature scale, as described in Chapter 3. The feature-scale model predicts the planarization

efficiency of a feature with given size, height, and shape, and those predictions may then be used in the die-scale model described in Chapter 4. That chapter also contains useful insights gained from the model on how layout affects polish behavior. Chapter 5 contains some concluding remarks and provides suggestions on future modeling work. Appendix A provides information on the experimental measurements used to acquire data that inform this physically-based model.

Chapter 2

Modeling framework

2.1 Modeling goals

This chapter presents a framework for analyzing the solid–solid contact between a CMP pad and a patterned wafer. The framework is used in Chapter 3 to calculate the planarization efficiency of a feature with a given size, height, and shape. It is subsequently used in Chapter 4 to simulate the topography evolution of an entire die. Section 1.11 expands upon the desired goals for this model. Vital to this work is the incorporation of real asperity profiles, taken by scanning used pads, as described in Section 2.7.

2.2 Preston's equation

One of the most difficult tasks in computational mechanics is understanding how contact pressures translates to material removal. There is a special class of processes, known as Prestonian processes, for which the following linear relationship holds:

$$\bar{R} = k \cdot \bar{v} \cdot \bar{P}$$

where $\bar{R}$ is the average material removal rate $\left(\frac{\mathrm{m}}{\mathrm{s}}\right)$, k is a process constant $\left(\frac{1}{\mathrm{Pa}}\right)$, $\bar{v}$ is the average relative rotational velocity between the pad and the wafer $\left(\frac{\mathrm{m}}{\mathrm{s}}\right)$, and $\bar{P}$ is the average contact pressure (Pa).

Whether a CMP process is well approximated as Prestonian depends on many factors, including the slurry choice and polish material. For the slurry, the abrasive material, abrasive size distribution, the chemistry, the flow rate, and the transport mechanisms may all affect whether the process is Prestonian. Additionally, not all substrate materials exhibit a Prestonian relationship. Preston conducted experiments on glass (one of whose main constituent is silicon dioxide) where he showed that glass exhibits a Prestonian relationship [131]. To determine whether a process is Prestonian, process developers must run multiple blanket wafers, with wafers tested at various values for the relative velocity and average down pressure. Using pre- and post-polish average oxide thickness measurements to compute the material

removal rate, process developers may plot $\bar{R}$ against $\bar{v} \cdot \bar{P}$; a linear relationship here suggests a Prestonian process and a non-linear relationship suggests a non-Prestonian process. This same process may be used to fit the material removal law, to obtain a phenomenological model.

Researchers have suggested alternative process laws [132]–[134]. Runnels points out that Preston's equation is limited in that it does not account for large changes in polish behavior that would occur when switching between lubrication regimes, and proposes using the shear stress—instead of the average velocity—in the material removal rate equation [135], [136]. Maury found that a modified Prestonian law, consisting of a piece-wise linear relationship between the material removal rate and average relative velocity times applied pressure, agreed with their oxide polishing experiments [133]:

$$\bar{R} = \begin{cases} k_1 \cdot \bar{v} \cdot \bar{P} & \bar{v} \cdot \bar{P} \leq \left(\bar{v} \cdot \bar{P}\right)_{\text{transition}} \\ \bar{R}_0 + k_2 \cdot \bar{v} \cdot \bar{P} & \bar{v} \cdot \bar{P} > \left(\bar{v} \cdot \bar{P}\right)_{\text{transition}} \end{cases}$$

where $k_1 > k_2$ are process constants. Additional research, summarized in Castillo-Mejia [137], has studied a generalized Prestonian equation:

$$\bar{R} = k \cdot \bar{v}^{\alpha_1} \cdot \bar{P}^{\alpha_2}$$

where α_1 and α_2 are process constants. As described in [137], various researchers have attempted to fit values of α_1 and α_2 for their experiments, with some results suggesting that the material removal rate has a sublinear dependence on both the average pressure and average relative velocity. For example, Hocheng found that values of $\alpha_1 = 0.5$ and $\alpha_2 = 0.5$ had reasonable agreement with experiments [134].

Additional rate laws with increasing complexity have been proposed, as summarized in Luo [113]. In this work, I proceed by assuming that all processes are locally Prestonian, as described in [137], so that

$$R(x, y) = k \cdot v(x, y) \cdot P(x, y) \tag{2.1}$$

where $R(x, y)$ is the material removal rate, k is a process constant, $v(x, y)$ is the relative rotational velocity between the pad and the substrate that is being polished (i.e. the wafer), and $P(x, y)$ is the contact pressure. By assuming that k is a constant, this model implicitly assumes that the temperature and slurry distributions are constant across the wafer. Lee proposed a version of this locally Prestonian rate law which raises the normalized contact pressures to an exponent and raises the normalized relative velocities to another exponent [138], [139]. Using finite element analysis simulations to predict the contact pressures, Lee found that raising the normalized contact pressure to a power of 3.77 led to the best agreement with experiments. However, these results are limited in that the method assumes smooth contact between the pad and wafer, and the presence of asperities may significantly affect the contact pressures.

By examining the literature, it is clear that the material removal law is very specific to the particular process. Since one of the modeling goals is to use as simple a model as

possible, the model presented here proceeds by using the locally Prestonian model described in Equation 2.1 and the methodology allows for alternative rate laws (as demonstrated in Section 4.3.6).

2.3 The planarization efficiency for Prestonian processes

One of the modeling goals is to predict a process' planarization efficiency. I make use of the planarization efficiency definition given in Equation 1.1. For a Prestonian material, I can then express the planarization efficiency as

$$P.E. = 1 - \frac{P_{\text{trench}}}{P_{\text{active}}} \tag{2.2}$$

Since P_{trench} and P_{active} depend on both time and space, the planarization efficiency also varies with the polish time and wafer position.

2.4 Matrix representation of pressure and topography variables

In this simulation, the variables that are functions of position are implemented as two-dimensional matrices. For example, the pressure distribution $P(x, y)$ and the pad topography $w(x, y)$ are both implemented as matrices on a corresponding grid with grid size $p \times p$. The notation $P[i, j]$ and $w[i, j]$ means to select the pressure value and the pad topography value at the i-th row and j-th column of the corresponding matrices. The matrices have size $M \times N$. This dissertation will use both the continuous representations and the discretized representations of variables interchangeably.

2.5 Parametrizing the pad design

The pad design space is quite expansive, and I narrow the design space to consist of segmented pads, as shown in Figure 2.1. Much of the analysis focuses on concentric-ring, grooved pads where the segment material is of much higher modulus than the subpad material. Chapter 4 also discusses grooved pads with a non-concentric ring design, and where the segment material is not very different from the subpad material.

2.6 Decomposing the pad deformation

I take a contact mechanics approach and analyze the solid–solid contact between the pad segments and wafer features. I assume three dominant modes of pad deflection: due to pad

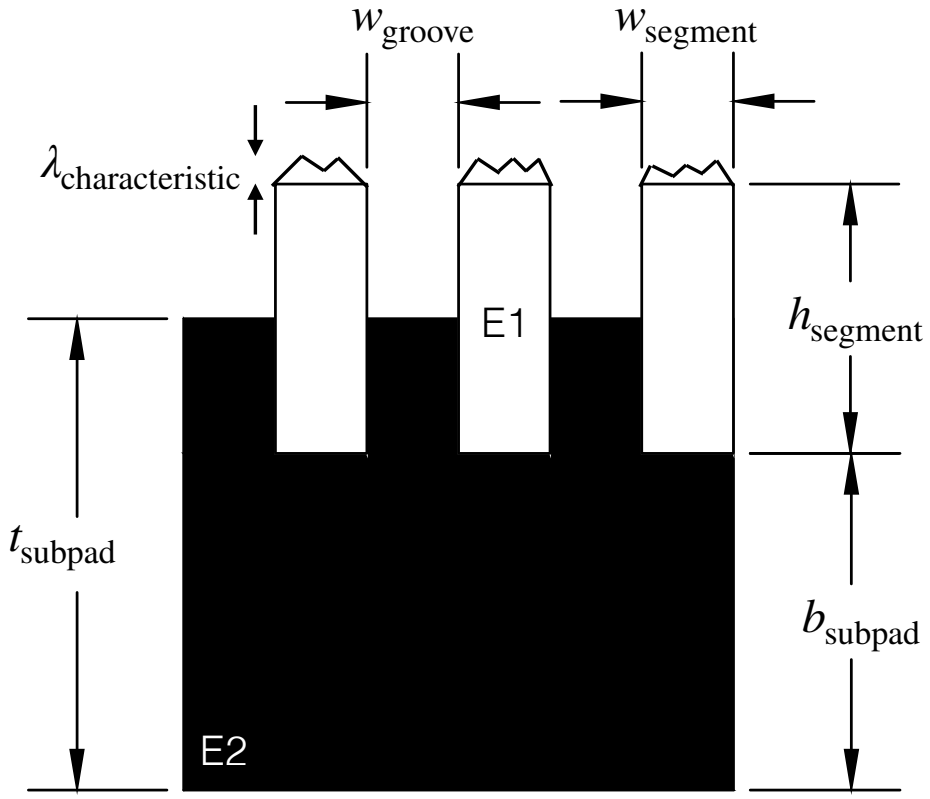


Figure 2.1: Diagram showing the parametrization of a segmented pad (shown in cross-section). The pad consists of segments embedded into a subpad. The variables are further described in Section 2.1.

surface elasticity, due to bulk downward displacement of a pad segment, and due to bulk rotation of a pad segment. I use surface response functions to determine the resultant pad surface topography when given a pressure distribution:

$$w = C_{\text{surface}} * P + C_{\text{pressure}} * \bar{P}_{\text{segment}} + C_{\text{moment}} * P_{\text{moment}} \tag{2.3}$$

where $*$ represents the convolution operation and C's represent the surface responses. P is the pressure distribution represented as a matrix. $\bar{P}_{\text{segment}}$ is the average pressure on a segment. P_{moment} is the effective moment on a segment. For computational efficiency, the convolutions are performed using fast Fourier transforms on MATLAB `gpuArray` structures.

2.6.1 Pad surface elasticity

I use Nogi and Kato's approach to determine the elastic response of the ring material's surface when a unit pressure is applied to the central element [65]. The response function is shown in Figure 2.2.

Table 2.1: Table of parameters that describe a pad design, and typically ranges for those parameters. The variable names correspond with those shown in Figure 2.1.

Parameter	**Range**	**Meaning**
E_1	400–4000 MPa	Pad segment material's Young's modulus
E_2	1–100 MPa	Sub-pad material's Young's modulus
w_{segment}	0.2–3 mm	Segment width
w_{groove}	0.2–3 mm	Groove width
h_{segment}	0.2–1 mm	Segment height
b_{subpad}	0.2–2 mm	Height from pad bottom to the pad segments' bottom
t_{subpad}	0.5–2.5 mm	Sub-pad thickness
$\lambda_{\text{characteristic}}$	10–100 µm	Characteristic asperity height
E_{surface}	1000–10000 MPa	Pad segment material's surface modulus

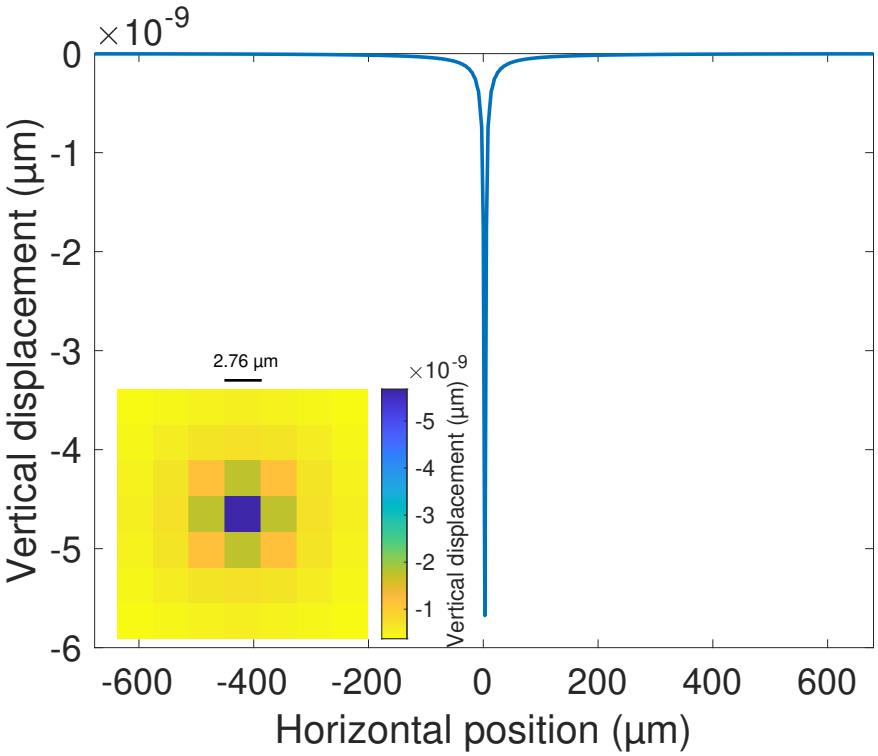


Figure 2.2: The surface response function C_{surface} representing the elasticity of the pad surface (for Pad A). Inset shows the top view of the response, zoomed onto the loading location.

I treat the pad segments as elastic half-spaces with no surface roughness, which contact a rough topography. That topography is composed of the wafer feature topography and the measured pad surface topography. Note that the wafer feature topography is represented as

locally flat within active/trench regions, as the local roughness of the wafer is much smaller than the pad's roughness.

2.6.2 Pad segment bulk behavior

The pad segment bulk behavior is decomposed as containing both bulk downward displacement and bulk rotation of the entire segment. Using the Abaqus 6.14 finite-element package, I simulated the mechanical behavior of the pads. Since CMP pads are large (762 mm diameter) compared to an individual pad segment, and polish happens sufficiently far from the center of the pad so that the ring curvature is low, I use a plane-strain approach as shown in Figure 2.3, for computational efficiency.

I seek to understand how the pad deforms when one of its segments is loaded with either a unit downward pressure or a unit moment. The number of pad segments within the simulation domain is variable, as further described in Section 3.4, and the method described here required that the number of pad segments be a power of two. The loaded segment is then selected as the (floor $\left(\frac{\text{\# of pad segments in the domain}}{2}\right) + 1$)th segment.

In the example with 2 pad segments shown in Figure 2.3, I simulate two unit cells of the pad segments, with the second segment surface set as the loading location. The left and right edges of the simulation domain are given periodic boundary conditions, using the method discussed in Wu [140]. The bottom of the pad is set to be fixed, which represents the steel platen in the physical system. Since the pad deflections are small, I employ a linear-elastic material behavior, which requires only the two materials' Young's moduli and Poisson's ratios. I used plane strain square elements, with an element size of 10 µm.

The unit response functions are then the vertical displacement along the top of the pads, and, for this example, are shown in Figure 2.4. Since the contact mechanics model presented in this dissertation may use a resolution different from the 10 µm element size at which Abaqus produced the unit response functions, these unit response functions need to be sampled at the proper resolution.

2.6.3 Convolution and fast Fourier transforms

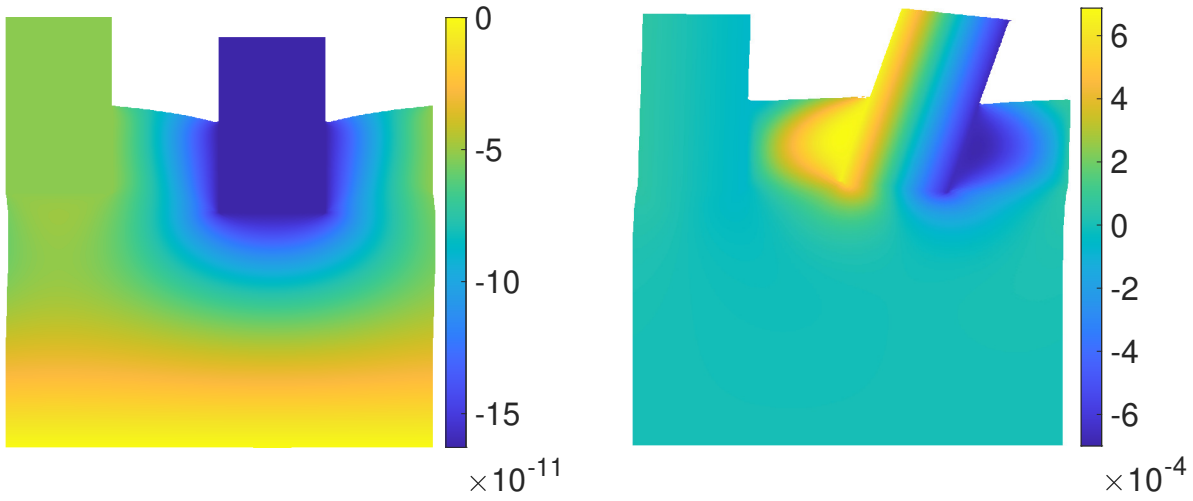
Given the large number of convolution operations that need to be performed in this modeling framework, the computational speed of that operation is critical. For a general response $\hat{w}$ that is the convolution between a kernel C and pressure P, such that

$$\hat{w} = C * P, \tag{2.4}$$

the convolution may be rewritten in Fourier space:

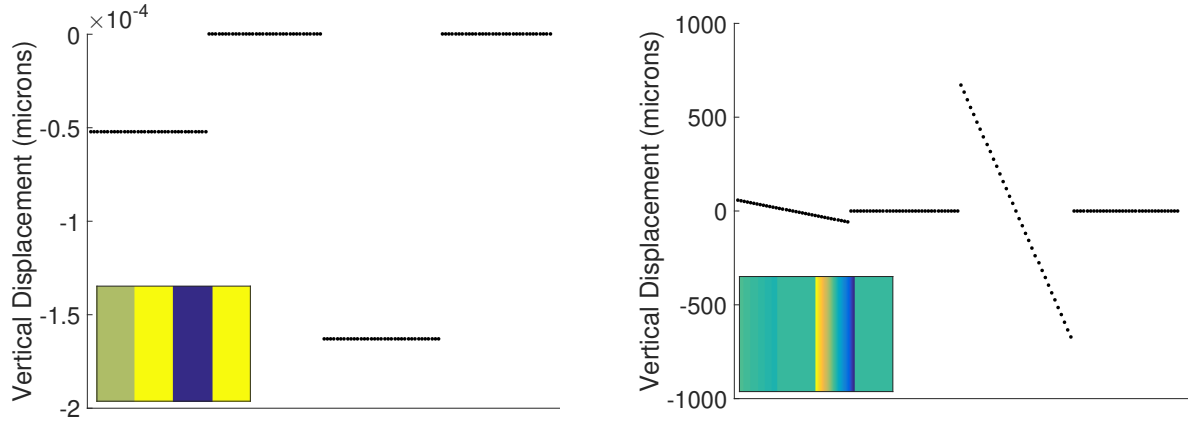
$$\texttt{fft2}(\hat{w}) = \texttt{fft2}(C) \odot \texttt{fft2}(P) \tag{2.5}$$

where `fft2` is the two-dimensional fast Fourier transform (since the variables presented here are implemented as two-dimensional matrices). The multiplication represented by $\odot$ is a



(a) Unit displacement (b) Unit rotation

Figure 2.3: Vertical displacement (in meters) from plane-strain finite element simulations of the second segment being loaded with a unit downward displacement or unit moment (for Pad A). The deformation is scaled by a factor of 10^6 for (a) and by 0.05 for (b).



(a) Unit displacement (b) Unit rotation

Figure 2.4: The bulk unit response functions, which are the vertical displacements along the top of the pad segments from the previous finite element simulations. The pad grooves are demarcated from the pad segments, as the displacement in the grooves is set to zero (since the subpad's displacement within the grooves is not relevant for wafer contact). The insets contain top views of the pad surface displacements and the main plots show horizontal cross-sections through the top views (for Pad A).

component-wise multiplication, and hence a computationally intensive matrix-multiplication is not required. By taking the two-dimensional, inverse fast Fourier transform, I have

$$\hat{w} = \texttt{ifft2}\left(\texttt{fft2}(C) \odot \texttt{fft2}(P)\right) \tag{2.6}$$

This approach implicitly assumes periodic boundary conditions for P, C, and $\hat{w}$.

2.6.4 Determining the Pressure Distribution

I follow the approach used by Taylor [80] for microembossing and decompose the pressure distribution into two kinds:

1. The pressure distribution P_e that would cause unit downward displacement of all points in the contact set
2. The pressure distribution P_f that would cause filling of the feature trenches by points within the contact set

To find P_e, I use MATLAB's `bicgstab` function to find the pressure in Equation 2.3 when $w = [w_e] = [-1]$. Similarly, to find P_f, I use MATLAB's `bicgstab` function to find the pressure in Equation 2.3 when

$$w = [w_f] = \begin{cases} 0 & \text{if pad is contacting active} \\ h_{\text{step}} & \text{if pad is contacting trench} \end{cases} \tag{2.7}$$

The resultant pressure distribution is

$$P[m,n] = P_f[m,n] + \Delta \cdot P_e[m,n] \tag{2.8}$$

Δ is a normalizing parameter chosen such that the average pressure equals the applied average pressure $\bar{P}$:

$$\Delta = \frac{\bar{P} - \text{mean}(P_f)}{\text{mean}(P_e)} \tag{2.9}$$

where the mean() operation is the arithmetic average of all the elements in the domain $\mathcal{D}$ for the feature-scale model. For the die-scale model, the averaging occurs over all elements within the die of interest $\mathcal{D}_{\text{die}}$, which is a subset of the entire domain $\mathcal{D}$ as described in Section 4.2.8.

2.6.5 Contact set determination

Similar to Xie [110], I take an iterative approach to determining the contact region $\mathcal{C}$. The contact set is given by

$$\mathcal{C} = \{(x,y) \in \mathcal{D} \,|\, w(x,y) \geq w_{\text{wafer}}(x,y)\} \tag{2.10}$$

Figure 2.6 shows examples of the rigid, rough topography $w_{\text{wafer}}(x, y)$ and of the deformed pad topography $w(x, y)$. The algorithm is shown in Figure 2.5 and, for the situations discussed here, typically requires three iterations through the loop for convergence.

The approach starts with an initial estimate for the contact set; while a good estimate may reduce the required number of iterations, it does not affect the final result. The algorithm then computes what the pressure distribution would have to be for this contact set, using the method discussed in Section 2.6.4. Now that the pressure distribution is known, the pad surface's deformation is calculated using the method from Section 2.6. Subsequently, the pad topography is used in conjunction with the wafer's topography to attain a refined estimate for the contact set. That includes checking if any new pad asperities are now contacting the wafer and subsequently removing any locations where the pressure is non-positive, since the contact is assumed as nonadhesive. This loop is repeated until the contact set estimate has converged.

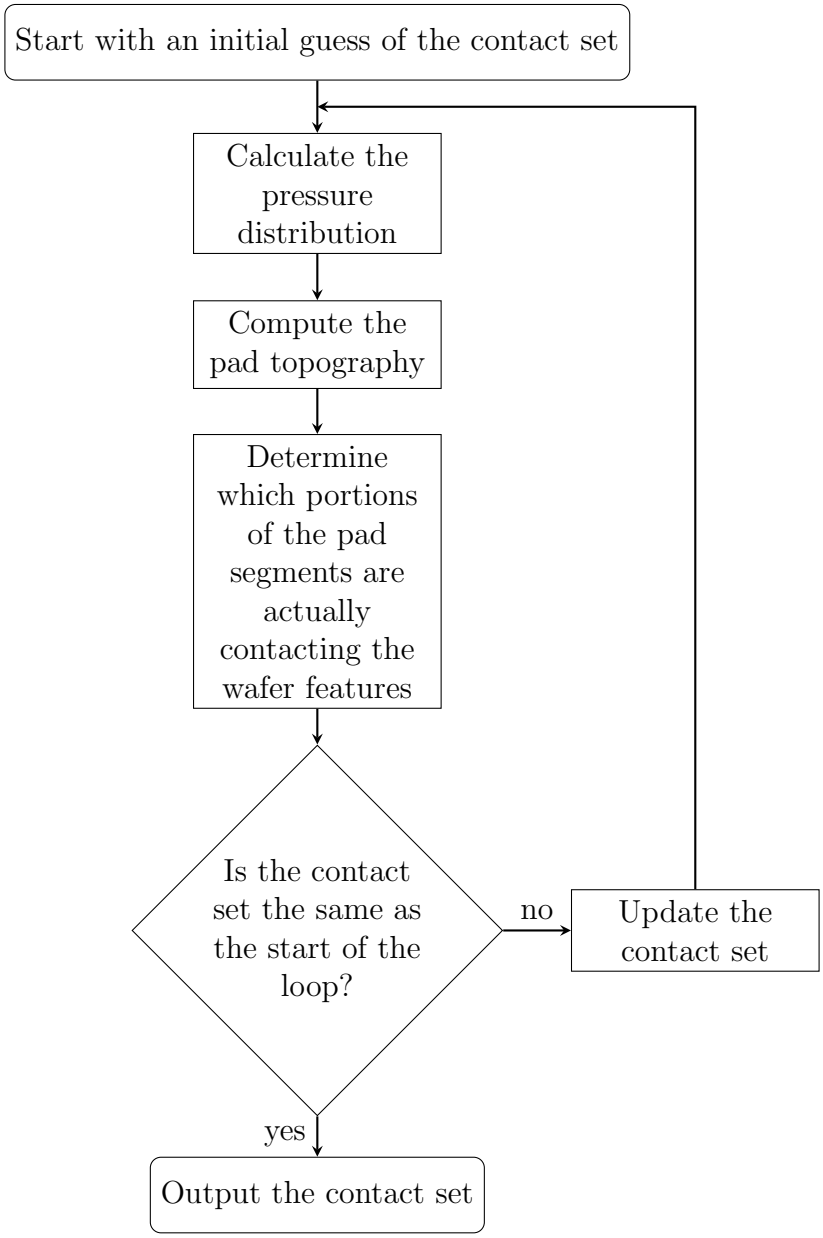


Figure 2.5: Algorithm for iteratively determining the contact set.

2.7 Incorporating surface roughness

The inclusion of surface roughness is crucial to calculating realistic contact pressures, especially inside of trenches. The wafer contains two types of roughness: macroroughness due to the presence of features ($\sim 0.2\,\mu m$ step heights) and microroughness ($\sim 0.02\,\mu m$). Because

the pad's microroughness ($\sim 40\,\mu\text{m}$) is much greater than the wafer's roughness, any microroughness within the active and trench regions is negligible, and those regions are represented as uniformly flat, with only the step height separating them. The starting (undeformed) pad topography is represented as

$$\hat{w}_{\text{pad}}(x,y) = \begin{cases} 0 & (x,y) \in \text{segment} \\ -\left(h_{\text{segment}} - t_{\text{subpad}} + b_{\text{subpad}}\right) \quad \text{(see Figure 2.1)} & (x,y) \in \text{groove} \end{cases} \tag{2.11}$$

Namely, the undeformed pad is represented as having flat segments, with grooves in between the segments that are deep enough that they will not contact the wafer.

The representation of pad asperities is critical to CMP modeling. Contact models are typically developed as having one smooth surface and one rough surface. As discussed in Nogi, Peng et al. [65], [66], the contact of two rough surfaces can be modeled as a smooth surface and an equivalent rough surface. This model proceeds by envisioning the pad segments as flat, and the wafer having roughness that accounts for both the wafer macroroughness $\hat{w}_{\text{wafer}}(x,y)$ and pad asperity profile $\hat{w}_{\text{asperities}}(x,y)$. Here, $\hat{w}_{\text{asperities}}(x,y)$ is defined to have only positive values, and $\min\left(\hat{w}_{\text{asperities}}(x,y)\right) := 0$. Namely, we have

$$w_{\text{wafer}}(x,y) = \hat{w}_{\text{wafer}}(x,y) - \hat{w}_{\text{asperities}}(x,y) + \max\left(\hat{w}_{\text{asperities}}(x,y)\right) \tag{2.12}$$

Figure 2.6 shows an example of this process (in cross-section), where the wafer topography is a superposition of the wafer pattern and the pad asperities, and the pad segments are represented as flat.

The pad texturing is acquired by scanning a used pad, as described in Section A.3. An example pad scan—containing two segments and two grooves—is shown in Figure 2.7. Caution needs to be taken when acquiring pad scans, as there are many types of possible measurement errors, including pad warpage. Maculotti discusses errors that occur when scanning surfaces (namely, the inclusion of artificial spikes) and a method for removing those errors [141]. Another common error is that the pad sample may not be flat; the scanner software was used to tilt-correct the examples shown here. As of now, a rigorous method for properly scanning pad surfaces for use in contact mechanics modeling has not been developed, and may be a source of possible errors.

2.7.1 Calculating the contact area ratio

I assume that when a pixel (grid element) is in contact, the entire area in that pixel is in contact, and the pressure is uniform throughout the pixel. I then use the pressure distribution to calculate the contact area ratio over the simulation domain, as follows:

$$\text{Contact area ratio} = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} \mathbf{1}_{\mathcal{C}}[m,n] \tag{2.13}$$

$\mathbf{1}_{\mathcal{C}}[m,n]$ is an indicator function that is 1 in the contact set and 0 outside of the contact set, as follows:

$$\mathbf{1}_{\mathcal{C}}[m,n] = \begin{cases} 1 & P[m,n] > 0 \\ 0 & \text{else} \end{cases} \tag{2.14}$$

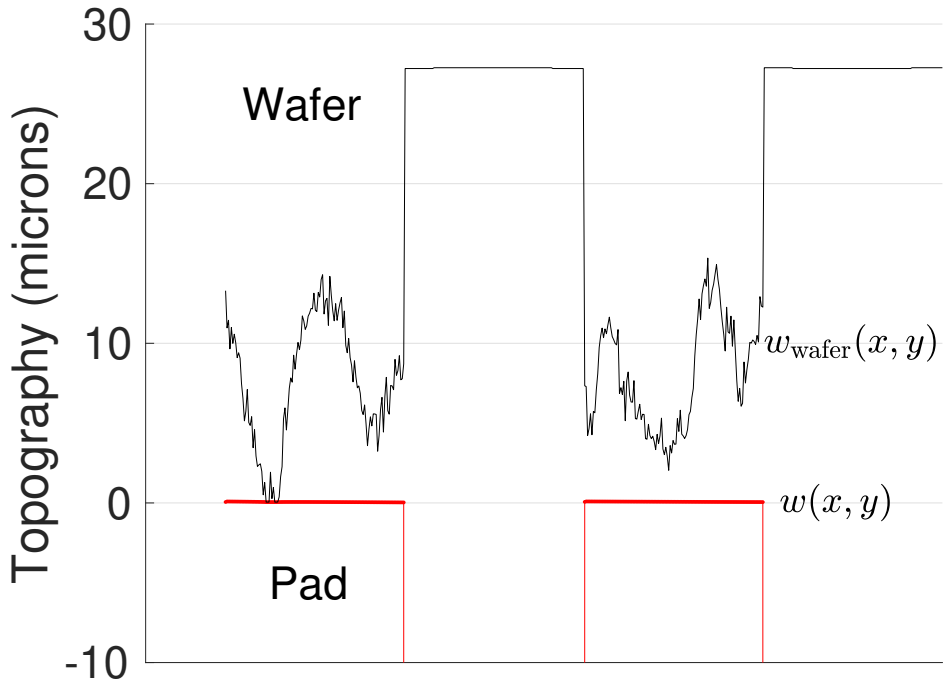


Figure 2.6: Plot of the contacting wafer and pad topographies. The wafer topography is a superposition of the wafer pattern and the pad roughness. The pad deforms in the contact zone.

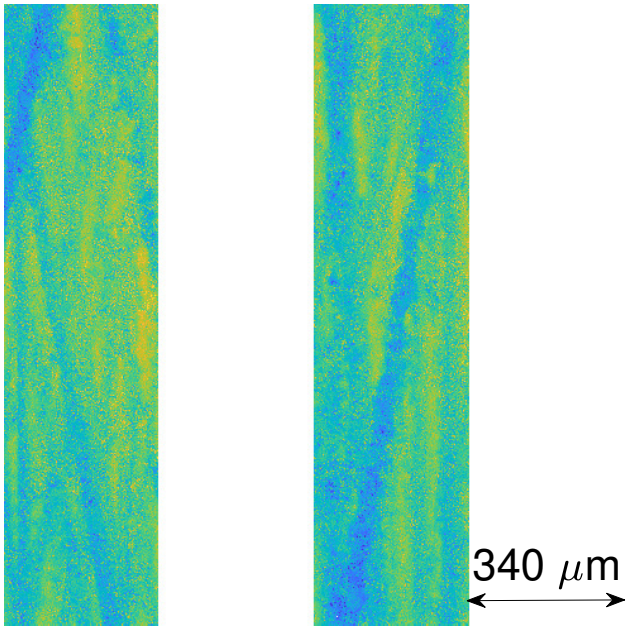


Figure 2.7: Surface scans of Pad A, after tilt-correction processing.

Note that from this definition of the contact area ratio, the maximum contact area ratio value that a segmented pad could have is not 1, but is limited by the segment width in relation to the sum of the segment size and pad groove width.

2.8 Computational requirements

The simulations presented here are run on a computational cluster. Each node contains an Intel Xeon E5 with four Nvidia K80 GPUs (2 CPU cores assigned to each GPU) and 64 GB of RAM. The feature-scale implementation (described in Chapter 3) has low memory requirements. Since it is highly parallelized, each task runs on one GPU and all available cores are used. Requesting additional nodes/GPUs from the cluster has a direct reduction in the required wall-clock time. A typical wall-clock time, when using a dozen of these nodes and matrix sizes of 256 $\times$ 256, is on the order of 30 minutes. The die-scale implementation (described in Chapter 4) uses only a single node at a time. When using matrix sizes of 1024 $\times$ 1024, it requires 32 GB of RAM. The most computationally intensive portion is creating the pad topography representations (described in Section 4.2.5). These pad topography representations may be computed once, saved, and subsequently used in simulations that require those same representations (i.e. the rotational speeds, wafer sweeping profile, and pad texturing did not change). Using this approach, a typical wall-clock time for simulating a 60 s polish is on the order of 3 days, instead of 7 days. Using computational resources with faster GPUs, such as the NVIDIA P100, leads to further decreases in the required wall-clock simulation time.

Chapter 3

Feature-scale model

3.1 Modeling goals

We require a model that can predict the planarization efficiency for features of a given size, shape, and step height. As this model will be run many times (to simulate various feature sizes, shapes, and heights), it must be computationally efficient. Further, it must capture the effects of the pad geometry. In particular, we expect there to be a transition in the behavior when the feature size goes from being smaller to being larger than the pad segment size. This is because if the feature size is larger than the pad segment size, then pad may more easily reach into trench areas, resulting in a drastic decrease in planarization efficiency. Finally, the contact area ratio between the pad and features is of particular importance. Strong agreement between the simulated contact area ratio and the contact area ratio as reported in the literature would provide greater confidence to this model.

3.2 Model description

The feature-scale model employs the contact mechanics framework described in Chapter 2, with inputs described in Section 3.4 and Section 3.4. In this implementation, the pad segments are envisioned to remain stationary, and the wafer features are represented as sweeping horizontally and rotating. To account for the various possible feature horizontal positioning and rotational orientations, multiple possible configurations are simulated.

The jth configuration's simulation output is the pressure distribution as a function of position $P(x, y)^{(j)}$. Section 3.5 describes how the contact mechanics framework is post-processed, to extract insights into feature-scale polishing behavior.

3.3 Feature topography representation

The features are represented as steps, with topography given by

$$\hat{w}_{\text{feature}}(x,y) = \begin{cases} 0 & (x,y) \in \text{active} \\ h_{\text{step}} & (x,y) \in \text{trench} \end{cases} \tag{3.1}$$

As discussed in Section 2.7, since any micro-roughness within the active and trench regions is negligible compared to the pad's roughness, those regions are represented as uniformly flat, with only the step height separating them.

The die layout described in Section A.2.1 contains many feature classes, which refers to features with a given size and shape. In particular, the 50, 100, 200, 500, 1000, and 2000 micron checkerboard and the 50, 100, and 200 micron 50%-line/space feature classes may be interesting to simulate since they demonstrate polish effects due to changing only the feature size. Each of these feature classes is simulated independently, following the methodology described below.

3.3.1 Wafer rotation and sweeping

The pad and wafer movement during polishing (both pad/wafer rotation and wafer sweeping) allows for various configurations under which material removal occurs. To capture this, the model simulates various relative angles between the features and pad segments. Namely, the model simulates the contact between the pad segments and features that have been rotated between 0° and 90°. Through a convergence study, I found that simulating 10 different angular configurations is sufficient. Furthermore, the model sweeps the features across the pad segments, with each simulation moving the features horizontally by one pixel. This means that there are 492 different horizontally swept configurations for Pad A (see Section A.1 for the pad details).

This results in 4920 different configurations to simulate the polishing of one feature class with Pad A. Figure 3.1 shows how each feature configuration is computed. First, for a feature with specific size and shape, a base configuration is determined. Then, the model uses MATLAB's `imrotate` to rotate the base configuration by a prescribed angle. Following that, using standard MATLAB array indexing, this rotated configuration is swept by a prescribed horizontal amount, resulting in the final configuration. Since each angle and horizontal position pairing is independent from all other pairings, the model employs the MATLAB Parallel Computing Toolbox to distribute simulations across computational nodes.

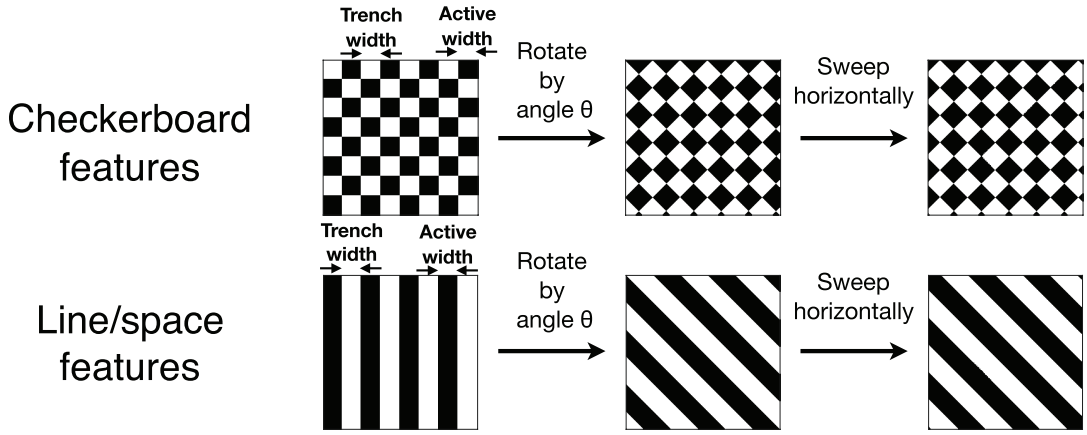


Figure 3.1: Diagram showing how the feature configurations are computed. For a specific feature class (with given size and shape), the base configuration is computed, then rotated and subsequently swept. Multiple feature configurations are computed, and simulation results averaged over all possible configurations.

3.4 Pad topography representation

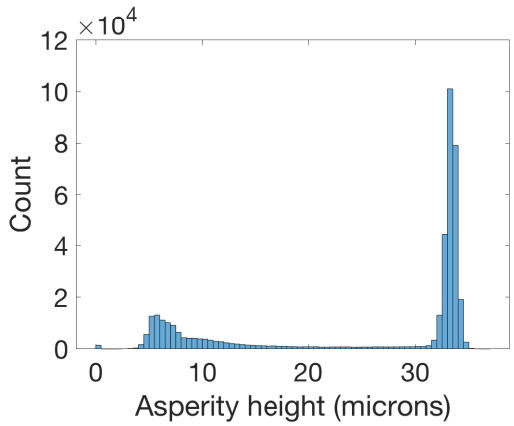
The number of pad segments contained in the simulation domain must be a power of two, so that the bulk segment deformation may be calculated using the framework described in Section 2.6. The number of pad segments needs to be chosen so that a sufficient number of features can be represented in the domain, and hence this decision depends on the segment size, groove size, and feature size. The results presented in this chapter contain domains with four pad segments, and further work is required to determine the predicted planarization efficiency's sensitivity to the number of features within the domain. The undeformed pad topography $\hat{w}_{\text{pad}}(x, y)$ is generated to have this macroroughness (as described in Equation 2.11) and the pad microroughness $\hat{w}_{\text{asperities}}(x, y)$ is experimentally scanned, for use in Equation 2.12. The pad and feature representations $w_{\text{feature}}(x, y)$, $\hat{w}_{\text{asperities}}(x, y)$, and $\hat{w}_{\text{pad}}(x, y)$ are used as input to the framework described in Chapter 2, with $\hat{w}_{\text{wafer}}(x, y) = \hat{w}_{\text{feature}}(x, y)$ in Equation 2.12. The pad topography $w(x, y)$ when the wafer applies a pressure distribution $P(x, y)$ is modeled using Equation 2.3.

3.4.1 Investigating the pad asperity distribution

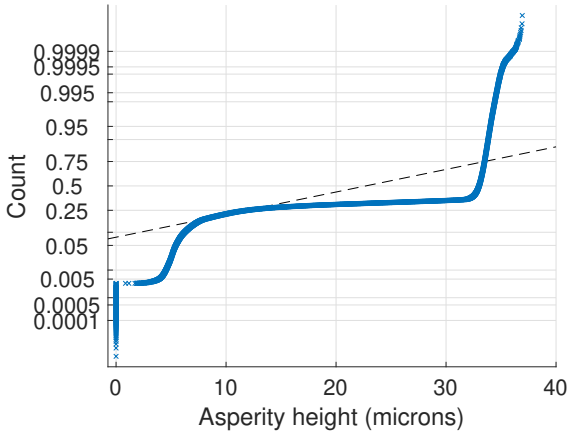
The pad surface scans contained in Figure 2.7 contain useful information about the pad asperities. Many previous contact mechanics analyses have employed theoretical distributions for the asperity heights, and it is worthwhile to check whether these particular measured surfaces can be approximated by such a theoretical distribution. Figure 3.2a contains the asperity height distribution for Pad A, and Figure 3.2b contains the associated normal prob-

ability plot. If these pad asperity heights were drawn from a normal distribution, then the points would form a straight line on this normal probability plot. Since the data deviate significantly from a linear relationship, it is clear that the measured asperity heights do not fall on a normal distribution.

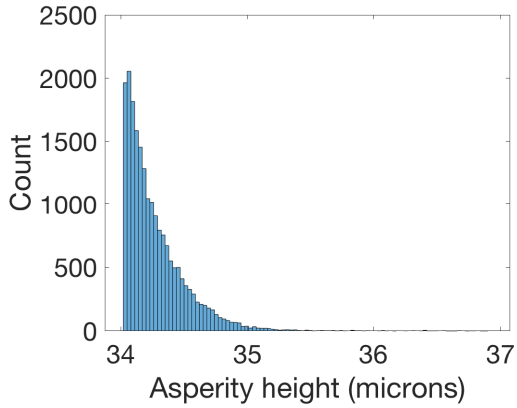
It is, however, only the tallest few asperity heights that contribute to polishing, and perhaps these tallest few asperities do possess a theoretical distribution. Figure 3.2c contains the asperity height distribution when analyzing only the tallest 5% of asperities. When investigating the probability plots for multiple well-known distributions including the normal, exponential, and logistic distributions (as shown in Figure 3.2), it is clear that these asperity heights are not well approximated by a well-known theoretical distribution. As such, direct use of the measured surfaces within contact mechanics simulations and not approximating the heights as having a particular distribution is likely to yield more accurate results.



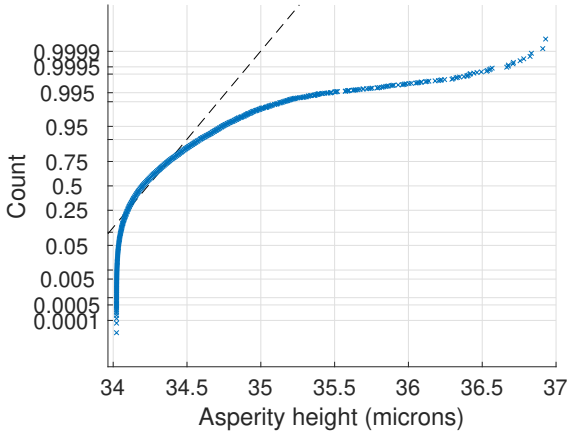
(a) Histogram (all asperities)



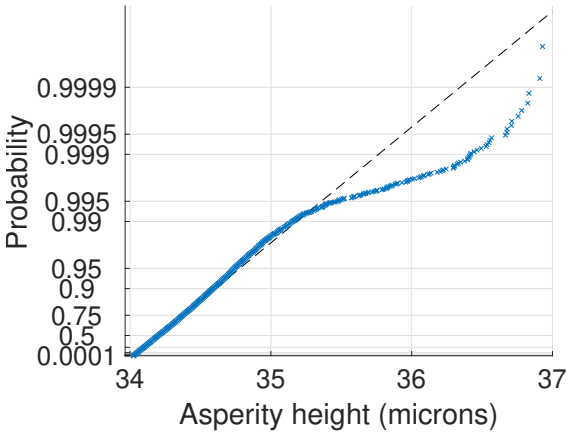
(b) Normal probplot (all asperities)



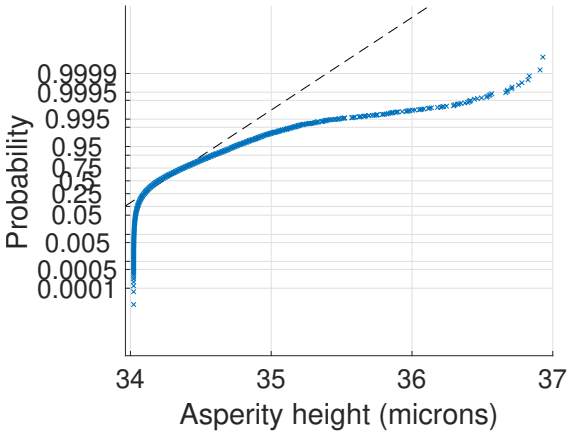
(c) Histogram (tallest asperities)



(d) Normal probplot (tallest asperities)



(e) Exponential probplot (tallest asperities)



(f) Logistic probplot (tallest asperities)

Figure 3.2: Histograms and probability plots for the measured asperity heights shown in Figure 2.7. (a) contains the histogram when all asperities are included and (b) shows the associated normal probability plot. (b) contains the histogram when only the tallest 5% of asperities are analyzed. (d) contains the normal (e) contains the exponential, and (f) contains the logistic probability plots when analyzing only the tallest 5% of asperities.

3.4.2 Pad material plasticity

We incorporate the nonlinear behavior of asperity deformation by setting a maximum pressure that an asperity point can withstand, similar to Terrell [94]. We first set the asperities into contact with a flat plate, and apply a down-pressure of p_0/ρ. ρ is the feature pattern density, and for the features simulated here, is equal to 0.5. If the pressure distribution reveals that an asperity point has a pressure exceeding the material's yield strength σ_{Y}, that point is clipped and the height is set equal to the height of the tallest of its eight neighboring points. The contact simulation is then iterated until all points in the contact set experience pressures less than or equal to the yield strength. Figure 3.3 contains a flowchart describing this algorithm. Equation 3.2 mathematically describes this flattening:

$$w_{\mathrm{asperities}}(x,y) = \begin{cases} \hat{w}_{\mathrm{asperities}}(x,y) & \text{if} \quad P(x,y) \leq \sigma_{\mathrm{Y}} \\ w^{\max}_{\mathrm{asperities}} & \text{if} \quad P(x,y) > \sigma_{\mathrm{Y}} \end{cases} \tag{3.2}$$

when the down-pressure is incrementally increased under displacement control until no additional points yield (i.e. convergence is reached). Here, $w^{\max}_{\mathrm{asperities}}$ represents the maximum asperity height. Then, in Equation 2.12, the flattened asperity profile $w_{\mathrm{asperities}}(x,y)$ is used in place of $\hat{w}_{\mathrm{asperities}}(x,y)$.

An example output from this glazing algorithm is shown in Figure 3.4. The flattening of asperities contains characteristics similar to the wafer-dominated pad surface profiles shown by Lawing [2], [3]. Our method is an approximation for both the glazing that a pad undergoes during polish and the flattening of asperities when loaded.

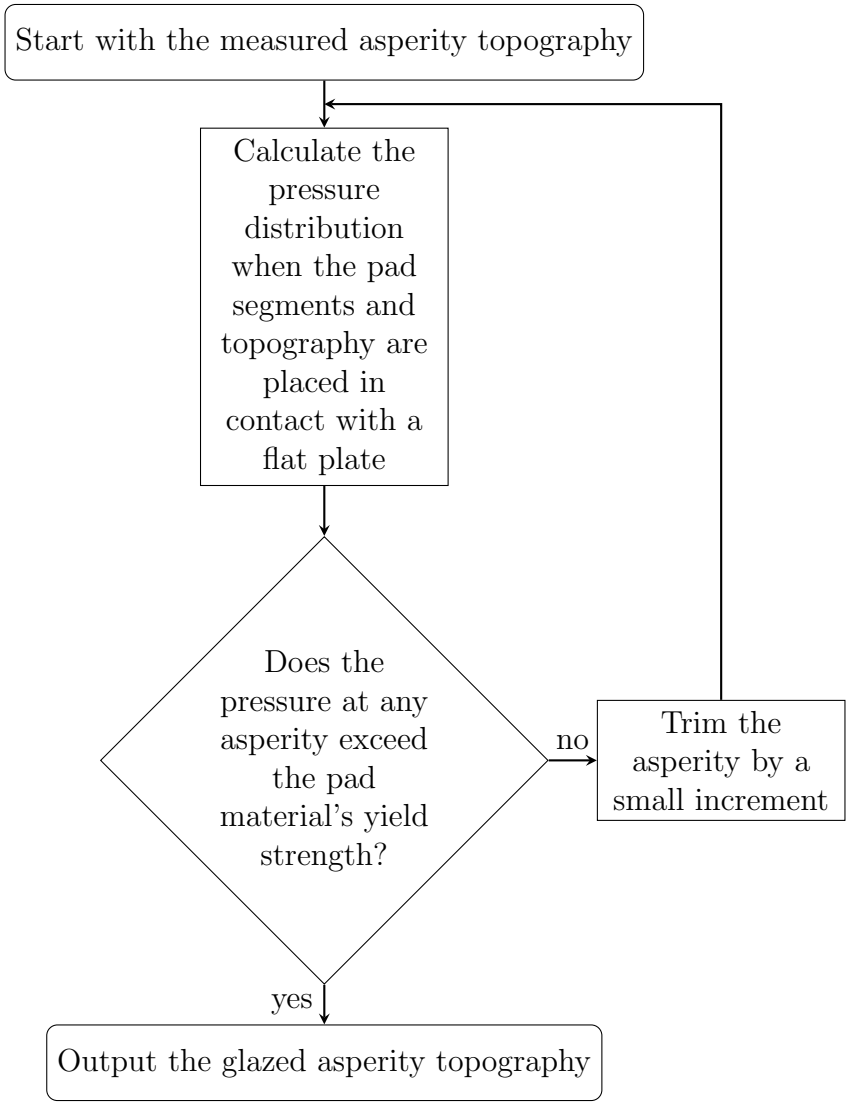


Figure 3.3: Algorithm for determining the flattened asperity profile.

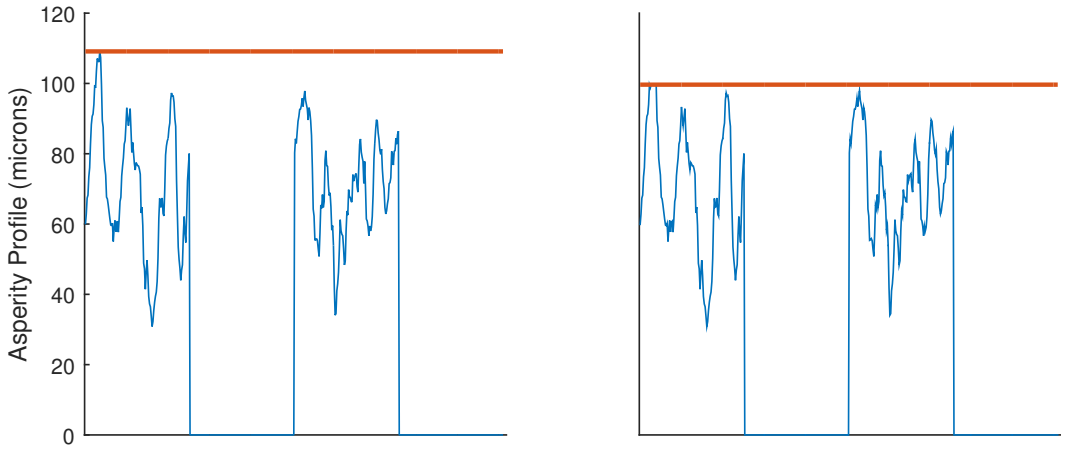


Figure 3.4: A cross section of as-measured asperities (from Pad A) contacting a flat plate, and the same asperities after they have been glazed.

3.5 Calculating the simulated planarization efficiency

The output of the contact mechanics simulation for the j-th pad-wafer configuration is the pressure distribution $P(x,y)^{(j)}$. To extract a planarization efficiency prediction for that feature class, the pressure distributions need to be averaged over the regions $\mathcal{D}^{(j)}_{\substack{\text{probe}\\\text{active}}}$ and $\mathcal{D}^{(j)}_{\substack{\text{probe}\\\text{trench}}}$. The locations of these probing regions must be evaluated for each configuration, since the location of the features changes. Section 4.2.11 discusses the importance of properly selecting this domain over which the pressure is probed.

For line/space features with sizes such that multiple features are captured in the simulation window, the pressures are averaged across the entirety of multiple active and trench regions. For checkerboard features, the averaging regions consists of all active and trench regions which lie in the central third of the simulation domain. The region-averaging is given by

$$\widehat{P}^{(j)}_{\text{active}} = \frac{1}{\text{Area}\left(\mathcal{D}_{\substack{\text{probe}\\\text{active}}}\right)} \int_{\mathcal{D}_{\substack{\text{probe}\\\text{active}}}} P(x,y)^{(j)}\, \mathrm{d}A \tag{3.3a}$$

$$\widehat{P}^{(j)}_{\text{trench}} = \frac{1}{\text{Area}\left(\mathcal{D}_{\substack{\text{probe}\\\text{trench}}}\right)} \int_{\mathcal{D}_{\substack{\text{probe}\\\text{trench}}}} P(x,y)^{(j)}\, \mathrm{d}A \tag{3.3b}$$

Then, the region-averaged pressures are averaged across all configurations, to attain

$$\widetilde{P}_{\text{active}} = \frac{1}{\#\text{ of configurations}} \sum_{j=1}^{\substack{\#\text{ of} \\ \text{configurations}}} \widehat{P}^{(j)}_{\text{active}} \tag{3.4a}$$

$$\widetilde{P}_{\text{trench}} = \frac{1}{\#\text{ of configurations}} \sum_{j=1}^{\substack{\#\text{ of} \\ \text{configurations}}} \widehat{P}^{(j)}_{\text{trench}} \tag{3.4b}$$

Subsequently, $\widetilde{P}_{\text{active}}$ and $\widetilde{P}_{\text{trench}}$ are used in Equation 2.2, to calculate the predicted planarization efficiency for that particular feature class. This process is repeated for additional feature classes.

3.6 Results: simulated pressure distribution and contact area ratio

For each of these configurations, the feature-scale model predicts the pressure distribution over the domain. An example of this pressure distribution, for one particular configuration, is shown in Figure 3.5. Since the pad is segmented, there is no contact where the pad grooves are located. The simulation shows that the contact is sparse (<0.4% of all pixels are contacting in this particular configuration). The predicted value for the contact area ratio aligns well with experimental observations from the literature, when recognizing that experiments from the literature use different materials and are performed at dry, room-temperature conditions. Section 1.9 further discusses the limitations associated with experimentally estimating the contact area ratio.

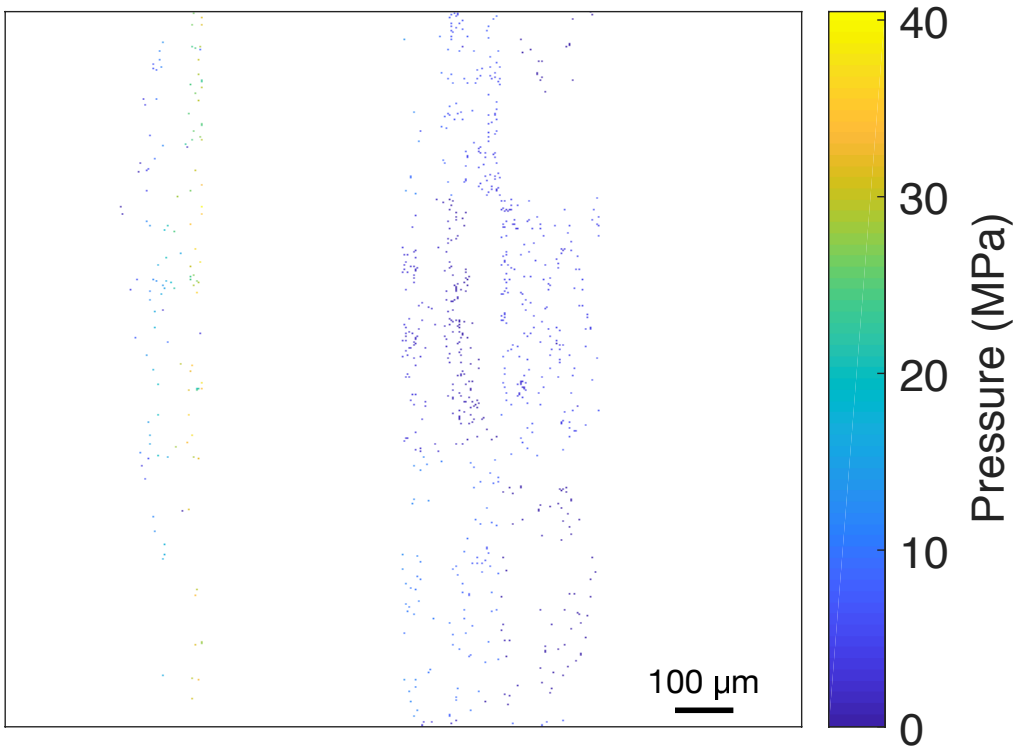


Figure 3.5: The simulated pressure distribution across the pad-wafer interface. White corresponds to regions outside of the contact set (i.e. zero pressure). The prediction shows sparse contact.

3.7 Results: planarization efficiency predictions

The feature-scale model has two parameters which can be tuned with experimental data until a best-fit parameter set for Pad A is attained. Section 3.7.1 discusses the influence of these two fitting parameters on the planarization efficiency vs step height curve, and Section 3.7.2 evaluates the predicted planarization efficiency's dependence on feature size (when using the best-fit parameters).

3.7.1 Effects of varying the simulation parameters

Figure 3.2 shows the effects of varying the simulation parameters. As shown in Figure 3.6a, increasing E_{surface} results in a higher predicted planarization efficiency, for a given feature step height. This matches both intuition and previous understanding of the CMP process, as harder pads are known to result in higher planarization efficiencies (because pad asperities are less likely to deform and reach into trenches). It is also helpful to inspect $h_{\substack{\text{step} \\ \text{critical}}}$, which

is the critical step height at which planarization efficiency dips to less than 1. Figure 3.6a shows that increasing E_{surface} results in a lower $h_{\substack{\text{step}\\\text{critical}}}$, for the same reason that harder pad asperities are less likely to deform and reach into trenches.

In contrast, increasing σ_{Y} results in a lower predicted planarization efficiency for a given step height and a higher $h_{\substack{\text{step}\\\text{critical}}}$, as shown in Figure 3.6b. This is because lower σ_{Y} values correspond to flatter pad surfaces and higher contact area ratios, which again restricts the pad asperities' ability to deform into trenches.

Increasing σ_{Y} by a unit amount results in a larger P.E. drop for smaller step heights than it does for larger step heights. Similarly, decreasing E_{surface} by a unit amount also results in a larger P.E. drop for smaller step heights than it does for larger step heights. Because of this, it is not yet possible to achieve strong agreement between simulations and experiments. Strong agreement is defined as having the predicted planarization efficiencies lie within the error bars, where the error bar sizes are determined as described in Section A.6. The simulation parameter set with the greatest experimental agreement (determined via manual tuning) is given in Table 3.1.

Table 3.1: Table of parameters used in feature-scale simulations with Pad A which yield the greatest experimental agreement, as shown in Figure 3.7.

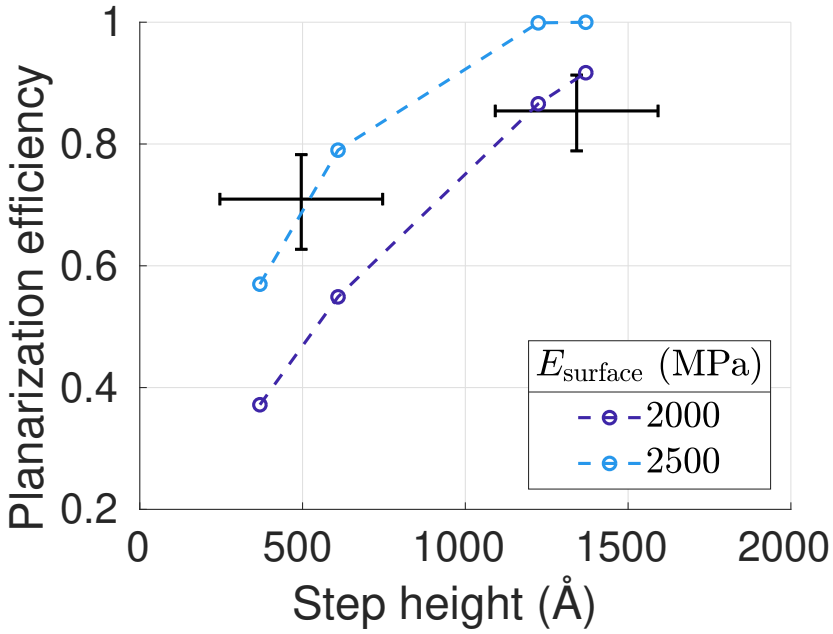
Parameter	**Value**	**Parameter description**
E_{surface}	2000 MPa	Pad segment material's surface modulus
σ_{Y}	100 MPa	Pad segment material's surface yield stress

3.7.2 Evaluating the predicted feature size dependence

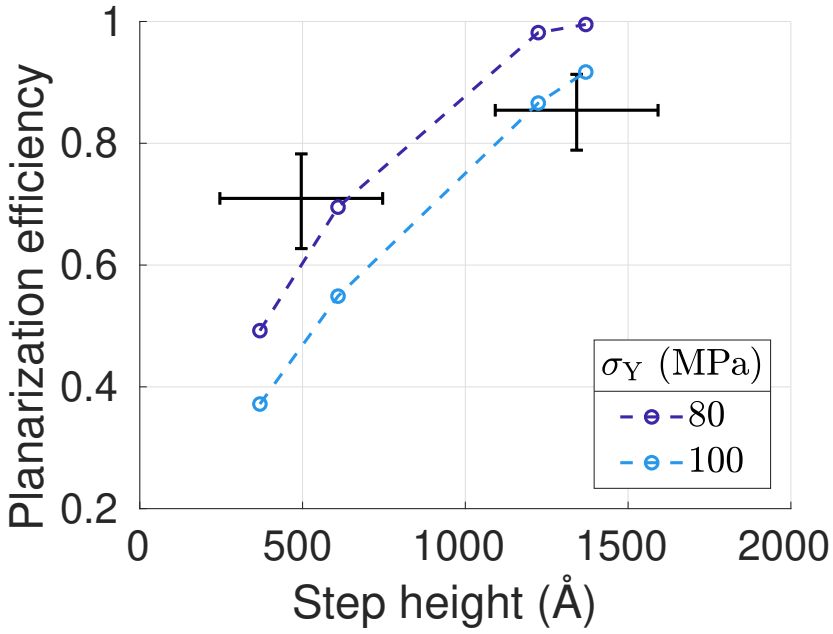
Figure 3.7 shows the planarization efficiency vs step height curves for two feature classes: 200 µm checkerboards and 500 µm checkerboards. The plot shows that smaller features are predicted to have a higher planarization efficiency, for a given step height. This matches intuition and previous understanding of the CMP process since the smaller trench openings make it more difficult for pad asperities to reach into the trenches.

3.8 Conclusion

This chapter has presented a feature-scale CMP model which predicts the planarization efficiency for a feature with given size, shape, and step height. The model employs a contact mechanics approach to simulate the pressure distribution between the pad and wafer and the behavior is averaged across all possible relative pad–wafer configurations. The pad deformation is decomposed into three modes—pad surface deformation due to local indentations, pad segment bulk downward displacement, and pad segment bulk rotation—and is



(a)



(b)

Figure 3.6: Plots of the planarization efficiency vs feature step height (for a 200 µm checkerboard feature class polished with Pad A). (a) shows the effects of changing E_{surface} (with $\sigma_{\text{Y}} = 80\,\text{MPa}$) and (b) shows the effects of changing σ_{Y} (with $E_{\text{surface}} = 2000\,\text{MPa}$). Circles and error bars correspond to the simulations and experiments, respectively.

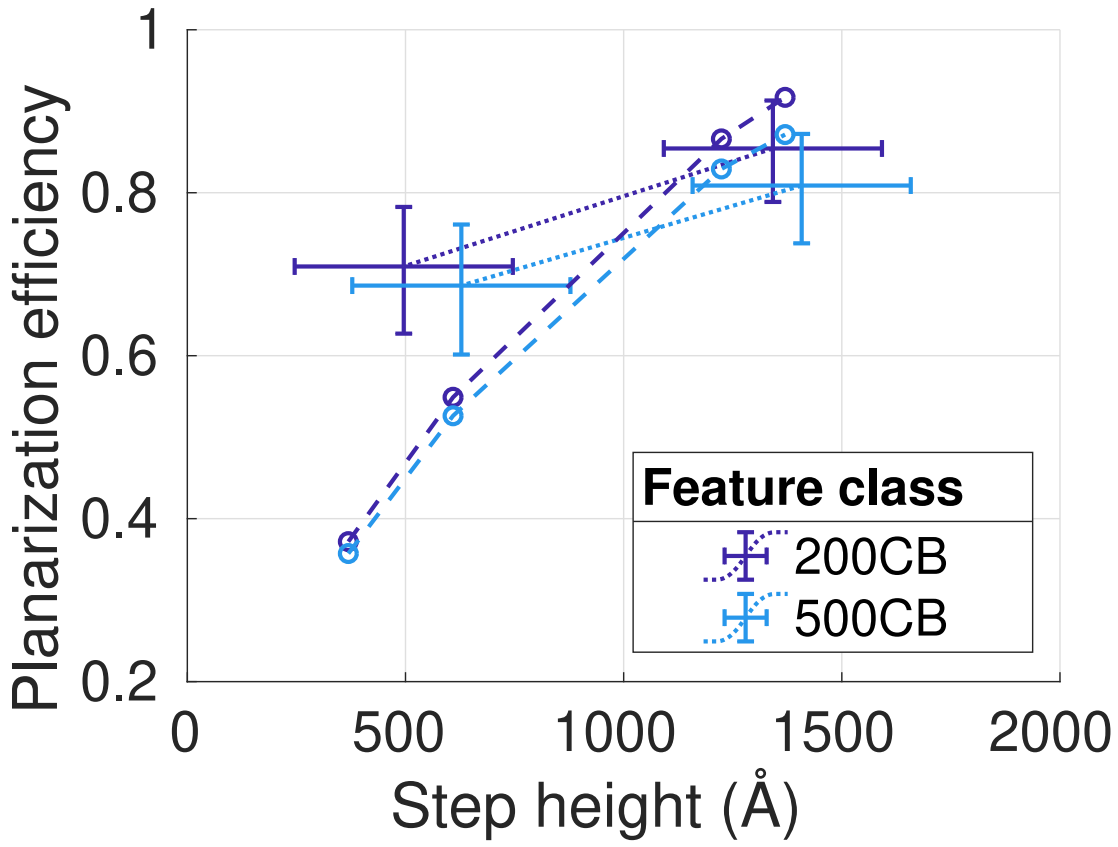


Figure 3.7: Plot of the planarization efficiency vs feature step height, for multiple feature classes polished with Pad A and when using the simulation parameters given in Table 3.1. The circles correspond to the simulations and the error bars correspond to the experiments.

computed using a computationally-efficient surface response approach. Critically, this model incorporates measured pad surface textures directly in contact mechanics simulations, which is essential to capturing the decrease in trench pressures compared to the active pressures. Pad asperity plasticity is modeled using an elastic-plastic material model to flatten the pad topography (preventing unrealistically low planarization efficiency predictions).

When comparing the model predictions against experimental data, there exists a significant difference between the predictions and experiments, suggesting that further model improvement is required. Regardless, the model at present captures significant trends, including the feature-size dependence, and it is a significant finding that this dependence can be captured using a solid-solid contact approach. Furthermore, the model is able to represent that pads with a higher surface modulus result in a higher planarization efficiency and that pads with a higher surface yield stress (and thus lower contact area) result in a lower planarization efficiency. The predicted contact area set is sparse, with a predicted contact area ratio less than 0.4%, matching experimental estimations from the literature.

These physically-based trends between the planarization efficiency, pad design, feature class, and feature step height can be used to create look-up tables. The look-up tables can then be incorporated into CMP models at other scales, including the die-scale model presented in Chapter 4.

Chapter 4

Die-scale model

4.1 Modeling goals

Given that the designs processed in CMP contain multiple features arranged in a die (and not solely features of a given size, shape, and height as considered in Chapter 3), we now need a model that takes that die design as input. The challenge is more difficult now, as the die may contain some features that are much smaller than the pad segment size, and some features that are much larger than the pad segment size.

The oxide topography depends on how long the die has been polished for. We now need to predict the oxide topography across the die as a function of the polish time. For strong confidence in the model, we need to compare the predicted oxide topography as measured on a particular feature within the die with an experimentally measured oxide topography. A strong model would have the predicted values lie within the 95% confidence intervals on the mean of the experimentally measured value. Further, we need to have at least as many experimental comparison points than we have model fitting parameters; one way to accomplish this is to perform these comparisons for multiple polish times and multiple feature types.

Whereas the feature-scale model discussed in Chapter 3 accounts for the various possible orientations between the pad segments and wafer features by representing the pad segments as stationary, and calculating the feature orientation by accounting for the wafer rotation and wafer horizontal translation, this die-scale does the opposite. Namely, in this die-scale model, the die on the wafer is represented as being stationary, and the pad segments' orientation is calculated, accounting for both the pad and wafer rotation and the wafer horizontal translation. This is because we now need to consistently track the die's topography (the oxide's thickness is the output of this model), making it simpler for the die to remain in an unrotated and untranslated state.

Once the model has been established as reasonably valid, we can use it to simulate the polishing of new die designs (for a given pad design) and then optimize the die design. Alternatively, we can use the model to simulate the polishing of new pad designs (for a given

die design) and then optimize the pad design. These optimization tasks are left as future work.

4.2 Modeling framework

The die-scale model uses the contact mechanics framework discussed in Section 2 to calculate the solid–solid pressure distribution between the pad and wafer. However, the die-scale model contains additional components, as it simulates polish material wear (i.e. the oxide topography will evolve over time). To accomplish this, I use a kinematics-based model to calculate the velocity distribution, a time-stepping approach to incrementally calculate the removal distribution, and a fitting process to determine the constant in the material removal law.

4.2.1 Time stepping: forward Euler method

We start with a representation of the die before polish (when $t = 0$ s and when the index $i = 0$). A forward Euler method tells us

$$h^{i+1}_{\text{active}} = h^{i}_{\text{active}} + \left.\frac{dh}{dt}\right|^{i}_{\text{active}} \Delta t \tag{4.1a}$$

$$h^{i+1}_{\text{trench}} = h^{i}_{\text{trench}} + \left.\frac{dh}{dt}\right|^{i}_{\text{trench}} \Delta t \tag{4.1b}$$

Assuming a Prestonian process, we have

$$\left.\frac{dh}{dt}\right|^{i}_{\text{active}} = -k \cdot v(x,y) \cdot P(x,y)^{i}_{\text{active}} \tag{4.2a}$$

$$\left.\frac{dh}{dt}\right|^{i}_{\text{trench}} = -k \cdot v(x,y) \cdot P(x,y)^{i}_{\text{trench}} \tag{4.2b}$$

4.2.2 Selecting an appropriate pixel size

The simulation domain is gridded into pixels, each of size $p \times p$. To select an appropriate value for p, there are two considerations: p must be small enough to resolve relevant feature groups in the die design, but p must not be so small that the overall number of elements becomes computationally taxing. Here, computationally taxing means that the simulation could require so much memory that it exceeds the available hardware or that the required time is too large. The allowable computation time depends on the usage context: simulations used within a computer-aided design tool may have time constraints of multiple hours while simulations used for verifying model accuracy may have time constraints of multiple days.

Reasonable matrix sizes for use with the computational resources described in Chapter 2.8 are $\mathcal{O}(1000) \times \mathcal{O}(1000)$.

For the die design described in Chapter A.2.1, a pixel size of $p = 50\,\mu\text{m}$ results in 310×240 elements.

4.2.3 Selecting an appropriate time step

When selecting an appropriate time step for simulations, there are three main considerations: the time step needs to allow all pixels in the simulation domain to potentially have contact, the simulations need to converge, and the simulations cannot require excessively large simulation times (no more than a few days). Too small of a time step, and the simulations can require several days, and too large of a time step and some pixels in the simulation domain might never see contact with pad asperities. A reasonable number of time increments for use with the computational resources described in Chapter 2.8 is $\mathcal{O}(10000)$.

There are two factors that affect how fast the wafer moves (and hence how small the time step must be): the wafer's horizontal translation and the wafer's rotation. I will start by analyzing the horizontal translation. The wafer translates horizontally a distance of $2 \cdot l_{\text{wafer}}$ with period T_{wafer}. That means the wafer translates with a speed of $\dfrac{2 \cdot l_{\text{wafer}}}{T_{\text{wafer}}}$. For the distance that a contacting asperity moves in a time step to not exceed the size of a pixel, that requires

$$\Delta t \leq \frac{p \cdot T_{\text{wafer}}}{2 \cdot l_{\text{wafer}}} \tag{4.3}$$

Next I analyze the wafer's rotation. A point on the wafer's edge moves at a speed of $\omega_{\text{wafer}} r_{\text{wafer}}$. For the distance that a contacting asperity moves in a time step to not exceed the size of a pixel, that requires

$$\Delta t \leq \frac{p}{\omega_{\text{wafer}} r_{\text{wafer}}} \tag{4.4}$$

When using the parameters from this experiment, the translation-based analysis suggests using $\Delta t \leq 0.003\,\text{s}$ and the rotation-based analysis suggests using $\Delta t \leq 2.45 \times 10^{-5}\,\text{s}$. The value provided by the rotation-based analysis would result in too many simulation increments, and too large a simulation time. The translation-based analysis, while resulting in a more reasonable number of required simulation increments, would still result in too large a simulation time. Instead, I selected a time step of $\Delta t = 0.0118\,\text{s}$, for a total polish time of 125 s, resulting in 10584 required increments. A convergence analysis (running the simulation with a time step equal to half this value) did not show meaningful differences in the predicted oxide thicknesses, suggesting that this time step value is sufficient.

4.2.4 Die topography representation

In this computational model, the oxide die's topography is represented as a matrix. It is necessary to define a zero reference point for the topography. While the choice of the zero

reference point is arbitrary, it is also convenient to set the zero at the bottom of the silicon features, as shown in Figure 4.1. The silicon features have a certain etch depth d_{etch}, which is assumed to be 1540 Å (based on the SEM measurements of a few features, as discussed in Section A.4). On the active portions of the features, a pad oxide and nitride film is present; the nitride is typically used as a stopper for the chemical mechanical polishing process. The thicknesses of these films were assumed to be $t_{\substack{\text{pad}\\\text{oxide}}} = 100$ Å and $t_{\text{nitride}} = 900$ Å.

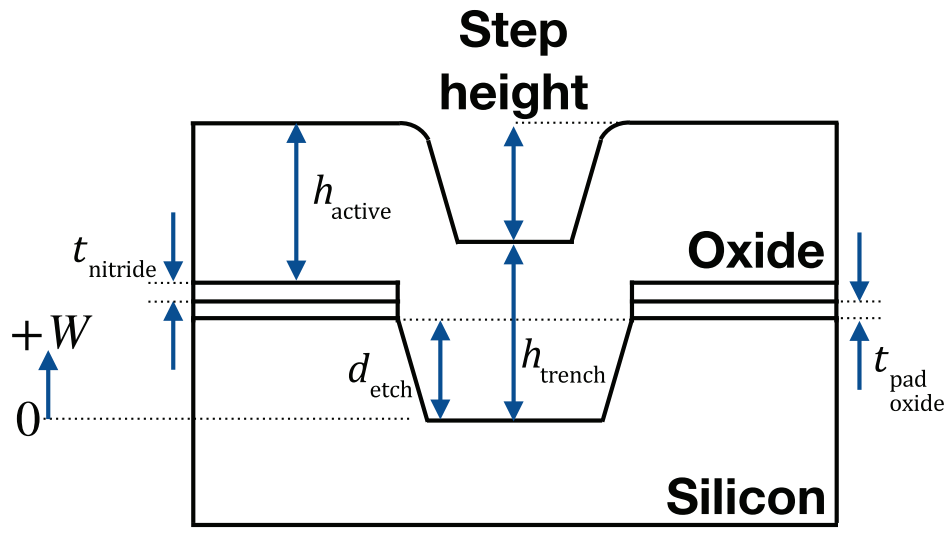


Figure 4.1: Schematic showing the terminology associated with an oxide feature. Shown in the schematic is an oxide thin film on a silicon feature before it has been polished. In the active regions, separating the oxide film and silicon feature, is a pad oxide film and a nitride film. The schematic shows the relevant dimension names.

The oxide film is conformally deposited on the nitride layer, and within the silicon trench. This conformal deposition process results in a step within the oxide film, whose height is known as the step height. The thickness of the oxide layer within the active region is labelled h_{active} and the thickness of the oxide layer within the trench region is labelled h_{trench}; these two quantities are the outputs that our simulation aims to predict. However, there are multiple features in a die. The test die discussed here has tens of thousands of features, and modern chips can have billions of features. We need to predict the oxide thicknesses for all features across the die—not only one—and so we represent the active and trench oxide thicknesses using matrices.

Note that there is variation in the silicon features' depths, and in the pad oxide and nitride films' thicknesses due to process variation (which may depend on the feature size). It is possible to experimentally measure d_{etch}, $t_{\substack{\text{pad}\\\text{oxide}}}$, and t_{nitride} across the die, to acquire values for use in simulation. That task is left as future work, and the assumption of those as constant values suffices for our purposes.

However, there is also variation in the pre-polish active and trench oxide thicknesses, and it is critical that we capture that spatial variation, since these oxide thicknesses are the simulations' output. To account for this variation in across-die oxide thicknesses, the oxide thickness was measured at multiple feature sites, as described in Section A.4. These experimental values were then used directly in simulations. Starting oxide thicknesses for feature sites that were not experimentally measured were assumed to have values equal to the mean thickness of features that were measured: 4517 Å for active regions and 4543 Å for trench regions.

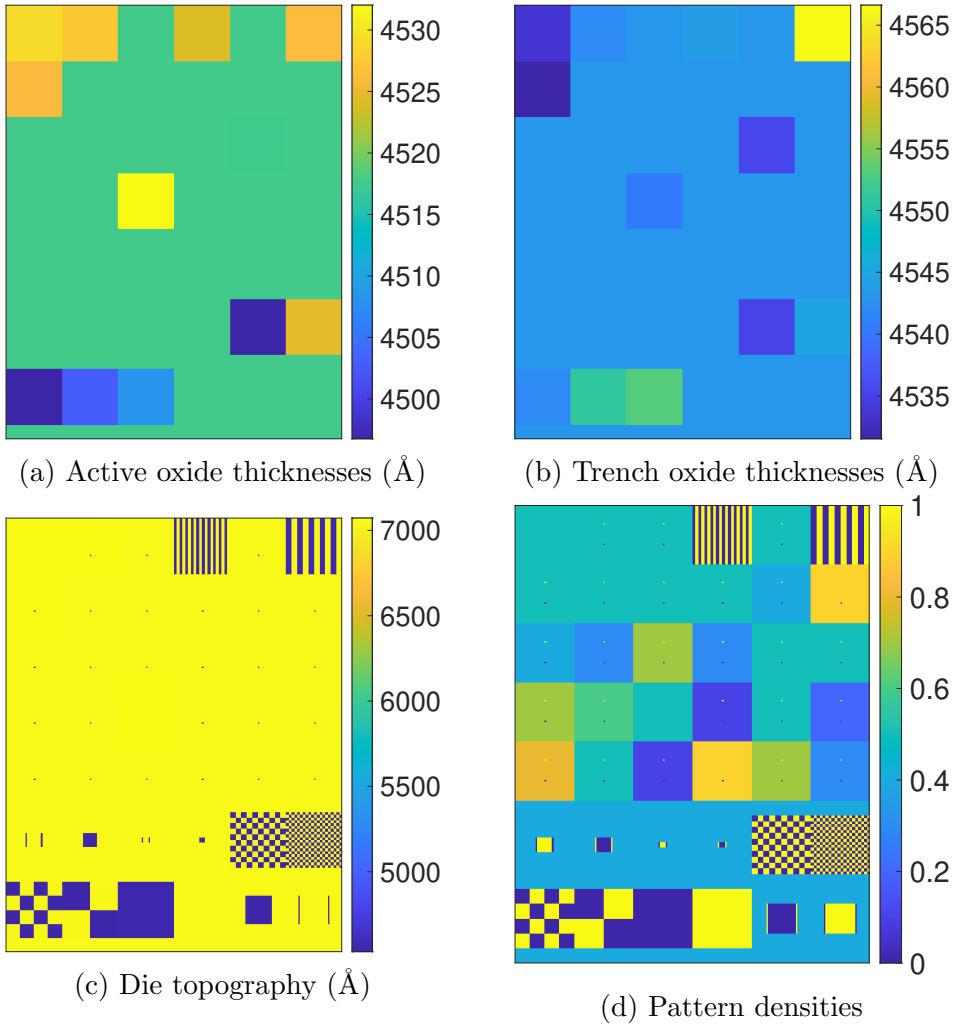
Section A.2.1 shows the test die design, including a description of the constituent features. From Figure 4.1, we see that

$$\begin{cases} h_{\text{step}}(x,y) = h_{\text{active}}(x,y) - h_{\text{trench}}(x,y) + t_{\text{nitride}} + t_{\substack{\text{pad}\\\text{oxide}}} + d_{\text{etch}} & \forall\,(x,y) \text{ s.t. } \rho(x,y) \neq 0 \\ 0 & \forall\,(x,y) \text{ s.t. } \rho(x,y) = 0 \end{cases} \tag{4.5}$$

and

$$\begin{cases} w(x,y) = h_{\text{active}}(x,y) + t_{\text{nitride}} + t_{\substack{\text{pad}\\\text{oxide}}} + d_{\text{etch}} & \forall\,(x,y) \text{ s.t. } \rho(x,y) \neq 0 \\ w(x,y) = h_{\text{trench}}(x,y) & \forall\,(x,y) \text{ s.t. } \rho(x,y) = 0 \end{cases} \tag{4.6}$$

where w is the oxide topography. Figure 4.2 shows the active oxide thicknesses, trench oxide thicknesses, die topography, and pattern densities which are used as input to this model (with a 50 micron grid size). We see that the across-die variation in both active and trench oxide thicknesses is less than 1% for this die, meaning that variation is unlikely to contribute significantly to the oxide evolution predictions for this die. However, CMP users may employ other oxide deposition processes, which may have higher oxide thickness variability, and the ability to incorporate that is already included in this model.



(a) Active oxide thicknesses (Å)

(b) Trench oxide thicknesses (Å)

(c) Die topography (Å)

(d) Pattern densities

Figure 4.2: Representation of the die's topography, including the pre-polish active oxide thickness, trench oxide thickness, oxide topography, and pattern density. Features with a critical dimension greater than or equal to 100 microns are fully resolved, and features smaller than that are lumped into pixels (with 50 micron size) with the corresponding pattern density.

4.2.5 Pad topography representation

Throughout the polish, the portion of the pad scan that is relevant to contact simulations changes. Here, our goal is to determine that portion of the pad scan that is relevant for use in contact simulations at a given polish time t.

I start by defining a Cartesian coordinate system, with the origin at the pad center and unit direction vectors $\hat{E}_x$ and $\hat{E}_y$. This coordinate system has an associated polar coordinate system with unit direction vectors $\hat{E}_r$ and $\hat{E}_\theta$. I then define a new polar coordinate system, with origin at the *wafer* center, to track the die's center position. This polar coordinate system has unit direction vectors $\hat{e}_r$ and $\hat{e}_\theta$.

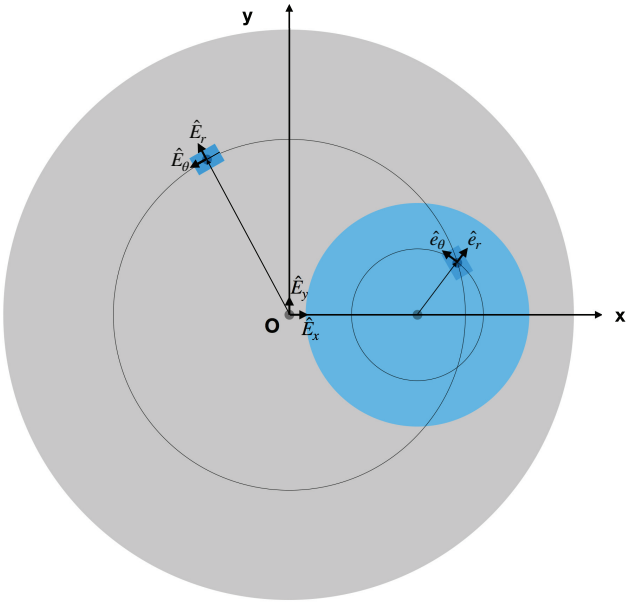


Figure 4.3: The coordinate systems associated with the pad and wafer. $\hat{E}_r$ is the unit direction vector in the radial direction, with the center of rotation at the pad center. $\hat{e}_r$ is the unit direction vector in the radial direction, with the center of rotation at the wafer center.

The wafer is constrained to move only along the x-direction. At time $t = 0\,\text{s}$, the spacing between the pad center and wafer center is $c_{\text{wafer},0}$. The pad platen is set to rotate with rotational speed N_{pad}, and the wafer carrier is set to rotate with rotational speed N_{wafer}. The pad's angular velocity is then $\omega_{\text{pad}} = 2\pi N_{\text{pad}}$ and the wafer's angular velocity is $\omega_{\text{wafer}} = 2\pi N_{\text{wafer}}$.

On this wafer, I select one die to analyze (shown in Figure 4.4a). That die's center is $z_{\text{die},0}$ away from the wafer center in the x-direction, and $y_{\text{die},0}$ away from the wafer center in the y-direction. At time $t = 0\,\text{s}$, the die has position $r_{\text{die},0}\hat{e_r} + \gamma_0\hat{e_\theta}$. I calculate that

$$\gamma_0 = \arctan\frac{y_{\text{die},0}}{z_{\text{die},0}} \tag{4.7a}$$

$$r_{\text{die},0} = \sqrt{z_{\text{die},0}^2 + y_{\text{die},0}^2} \tag{4.7b}$$

After t seconds of polish, the wafer has both rotated and moved along the x-axis (sweeping). I first analyze the rotation, and account for the sweeping in the subsequent step. After t seconds of rotation, the die's center has a new position: $z_{\text{die},1}$ away from the wafer center in the x-direction and $y_{\text{die},1}$ away from the wafer center in the y-direction (shown in Figure 4.4b).

$$z_{\text{die},1} = r_{\text{die},0}\cos(\gamma + \gamma_0) \tag{4.8a}$$

$$y_{\text{die},1} = r_{\text{die},0}\sin(\gamma + \gamma_0) \tag{4.8b}$$

I define γ as the angle that the die has rotated, with respect to the y-axis. I define ϕ as the angle which subtend the arc from the die's center position at 0 s to the die's center position at time t. Using Figure 4.4c, I show that $\phi = \gamma$. Figure 4.4c shows the die at its position at 0 s of polish, along with a line tangent to the die's rotational motion (shown in red) at this instant. The angle between this tangent line and the die's major axis (that is, the axis that goes through the center of the rectangle's smaller sides) at this instant is γ_0. Figure 4.4c also shows the die's position after t seconds of wafer rotation, along with a line tangent to the die's rotational motion (shown in green) at this instant. The angle between this tangent line and the die's major axis at this instant is $\phi+\gamma_0-\gamma$. Since the angle between the die's major axis and the radius of the die's motion does not change, the angle between the die's major axis and the tangent must also not change (since the radius and tangent line are perpendicular). This means that $\phi + \gamma_0 - \gamma = \gamma_0$ and hence $\phi = \gamma$.

I determine γ as

$$\gamma = \omega_{\text{wafer}} \cdot t \tag{4.9}$$

This means that, at time t, the die's center has coordinates $r_{\text{die},0}\hat{e_r} + (\gamma_0 + \gamma)\,\hat{e_\theta}$.

I now account for the wafer's translation, as shown in Figure 4.4d. The wafer's sweeping profile is specified on the polish tool. These simulations use a sweeping profile where the wafer sweeps a horizontal distance $2 \cdot l_{\text{wafer}}$ with period T_{wafer}. I assume the sweeping profile is a triangle wave. The sweep is centered so that the min for c_{wafer} is $c_{\text{wafer},0} - 0.5l_{\text{wafer}}$ and the max is $c_{\text{wafer},0} + 0.5l_{\text{wafer}}$. The distance between the wafer's center at time t and the wafer's center at time 0 is

$$s(t) = \frac{2 \cdot l_{\text{wafer}}}{T_{\text{wafer}}}\left(t - \frac{T_{\text{wafer}}}{2}\left\lfloor\frac{2t}{T_{\text{wafer}}} + 0.5\right\rfloor\right) * (-1)^{\left\lfloor\frac{2t}{T_{\text{wafer}}}+0.5\right\rfloor} \tag{4.10}$$

This means that the wafer center has an x-position of

$$c_{\text{wafer}} = c_{\text{wafer},0} - s(t) \tag{4.11}$$

Further, this means that after t seconds of wafer rotation and translation, the die's center has position: $z_{\text{die}} = z_{\text{die},1}$ away from the wafer center in the x-direction and $y_{\text{die}} = y_{\text{die},1}$ away from the wafer center in the y-direction. We now have coordinates for the die center on our Cartesian coordinate system:

$$x_{\text{die}} = c_{\text{wafer}} + z_{\text{die}} \tag{4.12a}$$

$$y_{\text{die}} = y_{\text{die},1} \tag{4.12b}$$

Our goal is to find the portion of the pad scan that will have position $x_{\text{die}}\hat{E}_x + y_{\text{die}}\hat{E}_y$ at time t. One option is to rotate the pad scan by angle δ, and determine which portion of the rotate pad scan lies underneath the die. However, since the pad scan is large (in terms of memory requirements), rotating this large matrix is computationally tasking (though possible). A quicker approach is to find the location of those asperities on the non-rotated pad scan, select a region from that pad scan, and rotate only that region a corresponding amount.

The die's center has position $\hat{r}\hat{E}_r + \sigma\hat{E}_\theta$ at time t. I find that

$$\hat{r} = \sqrt{x_{\text{die}}^2 + y_{\text{die}}^2} \tag{4.13a}$$

$$\sigma = \arctan\frac{y_{\text{die}}}{x_{\text{die}}} \tag{4.13b}$$

The portion of the pad scan that is relevant for contact simulations at time t has center-position $\hat{r}\hat{E}_r - (\sigma - \delta)\hat{E}_\theta$ (as shown in Figure 4.4e; this follows the parameterization for a clockwise circle). I find that

$$\delta = \omega_{\text{pad}} \cdot t \tag{4.14}$$

In the Cartesian coordinate system, the center-position is $a\hat{E}_x + b\hat{E}_y$, where

$$a = \hat{r}\cos(\sigma - \delta) \tag{4.15a}$$

$$b = -\hat{r}\sin(\sigma - \delta) \tag{4.15b}$$

Now that we have the center-position coordinates for the relevant pad region, we need to determine how that region is rotated. That region needs to be rotated an angle $270° + \sigma - \delta$ from the y-direction, counterclockwise (as shown in Figure 4.4f).

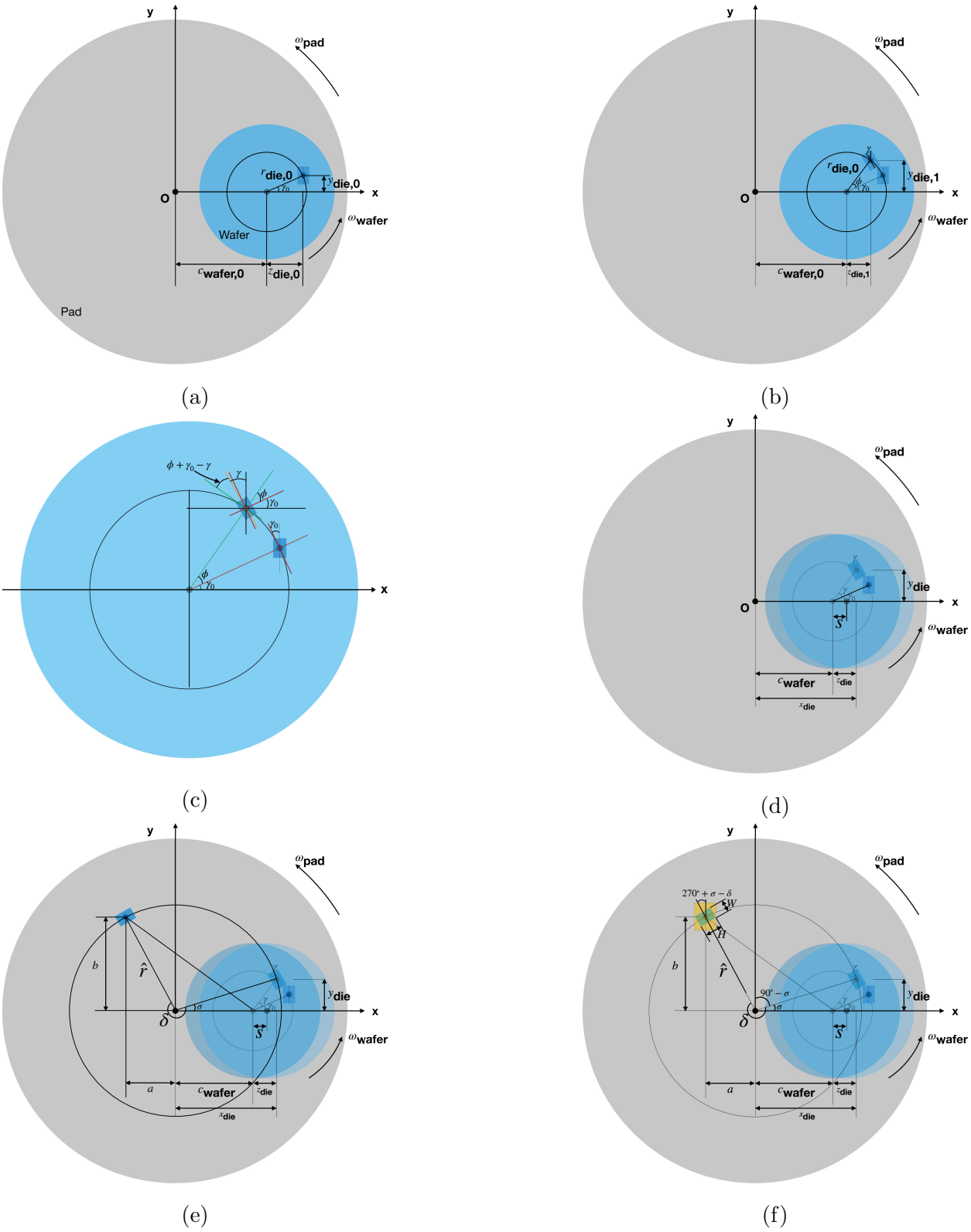


Figure 4.4: Diagrams showing the die's location on the wafer, and its motion throughout the polish, along with the location of the corresponding pad asperities

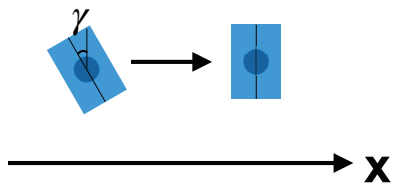


Figure 4.5: Schematic showing the representation of the pad scan that is relevant for simulations at a given instant. Since the simulation envisions the die as stationary throughout, the pad scan must be rotated a corresponding amount.

Note that these die-scale simulations envision the die as remaining stationary throughout the simulation, and the pad is what rotates. As shown in Figure 4.5, the pad scan must then be rotated by γ clockwise, so that the pad scan aligns with the stationary die.

Instead of rotating the pad scan $270^\circ + \sigma - \delta$ counterclockwise and then rotating it γ clockwise, it is simpler to rotate the pad scan $270^\circ + \sigma - \delta - \gamma$ counterclockwise.

Note that Equation 4.7 contains a singularity if the die is centered at the wafer center. This singularity is easily resolved by recognizing that, in this case, $\hat{r} = c_{\text{wafer}}$ and $\sigma = 0$.

Figure 4.6 shows some example portions of a pad scan that are relevant for polish simulations at the given simulation times.

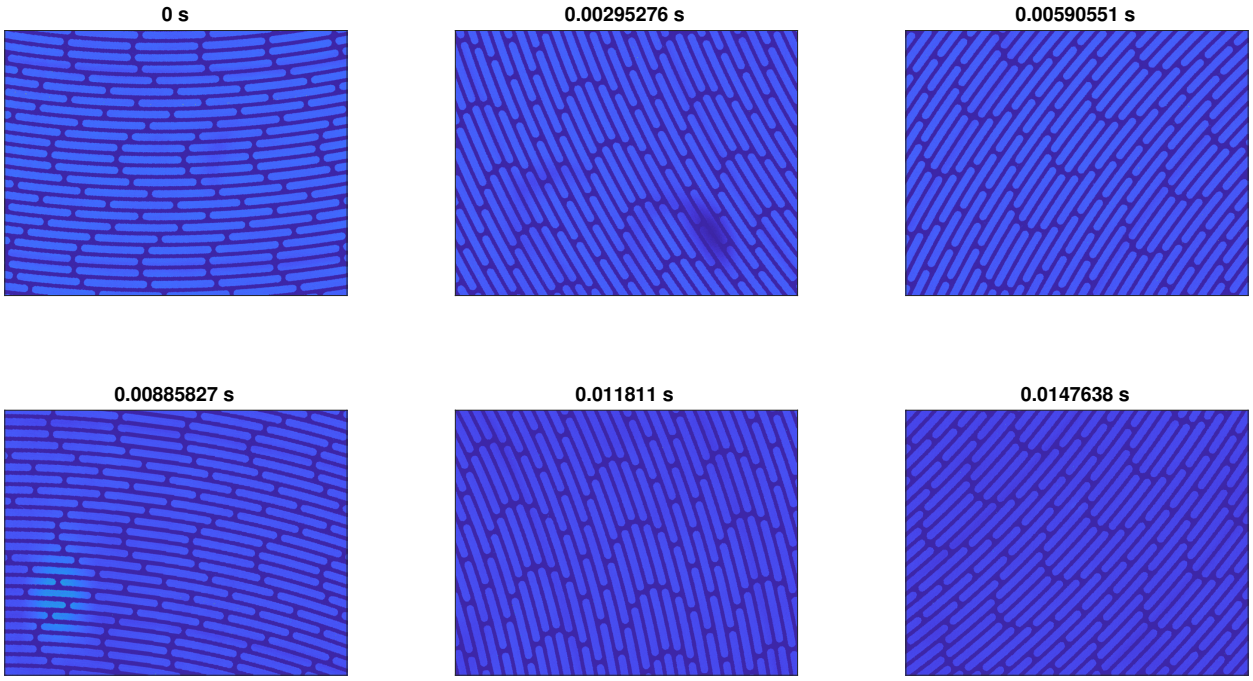


Figure 4.6: Six example portions of a pad scan that are relevant for polish simulations at the given simulation times.

4.2.5.1 Processing measured pad surfaces with a flattening algorithm

The white light scanning microscope used to measure entire pad surfaces (discussed in Section A.3) contains instances of unrealistically tall asperity spikes. These exceedingly tall asperities, if used directly in contact simulations, would dominate the polish and easily reach into trench regions. Instead, the pad scan must be processed to remove those unrealistically tall asperities. The pad surface material plasticity implemented in the feature-scale model (in Section 3.4.2) is designed for use with small pad portions and would be exceedingly computationally expensive for the large pad scan used in the die-scale model.

To flatten measured pad scans, asperities are envisioned to have a maximum allowable height $w^{\max}_{\text{asperities}}$. Any asperities with height greater than this value are trimmed to have a height equal to $w^{\max}_{\text{asperities}}$, and the pad surface is thereby flatter than before this processing step. $w^{\max}_{\text{asperities}}$ is treated as a fitting parameter, which controls both the maximum simulated pressures and the simulated contact area ratios. That is, a lower value of $w^{\max}_{\text{asperities}}$ corresponds to a flatter surface, and hence lower contact pressures and higher contact area ratios. It is important to note that $w^{\max}_{\text{asperities}}$ is defined in reference to a zero-height chosen by the scanning equipment, meaning that the value of $w^{\max}_{\text{asperities}}$ depends on the particular pad scan.

4.2.6 Calculating the relative velocity

The model requires the instantaneous relative velocity between the pad and wafer, for use in calculating the material removal rate. While the relative velocity tends not to vary much across the wafers (CMP tool users will select process parameters which keep the variation low), my results show that this variation can exceed 5%, and hence it is important to account for the relative velocity's spatial dependence. Further, as the polish time progresses (meaning the wafer rotates and sweeps), the distance between the die's center and the pad's center can go from being about 0 to about the pad's radius. That means that the relative velocity can have a large temporal variation. I calculate the relative velocity as a function of both the position in the simulation domain and the simulation time. The pad has angular frequency ω_{pad}, which is defined to be positive when the pad rotates counterclockwise (when viewing the pad from above). The wafer has angular frequency ω_{wafer}, which is also defined to be positive when the wafer rotates counterclockwise (when viewing the pad from above).

The instantaneous velocity of the pad consists of only rotational velocity, and is given by

$$v_{\text{pad}}(x,y) = -\omega_{\text{pad}} y \hat{E}_x + \omega_{\text{pad}} x \hat{E}_y \qquad \forall (x,y) \in \text{pad} \tag{4.16}$$

The instantaneous velocity of the wafer consists of both its rotational velocity and its translational velocity. The wafer's instantaneous velocity is given by

$$v_{\text{wafer}}(x,y,t) = \left(-\dot{s}(t) - \omega_{\text{wafer}} y\right) \hat{E}_x + \omega_{\text{wafer}} \left(x - c_{\text{wafer}}\right) \hat{E}_y \qquad \forall (x,y) \in \text{wafer} \tag{4.17}$$

Where $\dot{s}(t)$ is the component contributing to the translational velocity and is given by the time derivative of $s(t)$ (following Equation 4.10). Namely, we have

$$\dot{s}(t) = \frac{2 \cdot l_{\text{wafer}}}{T_{\text{wafer}}} \cdot (-1)^{\left\lfloor \frac{2t}{T_{\text{wafer}}} + 0.5 \right\rfloor} \tag{4.18}$$

However, for the experiments analyzed here, the wafer's translational speed is much smaller than its rotational speed (the translational speed is about 17 mm/s whereas the rotational speed is on the order of 2000 mm/s). Hence, the translational contribution is negligible and we will approximate the wafer's velocity as

$$v_{\text{wafer}}(x, y) \approx (-\omega_{\text{wafer}} y)\, \hat{E}_x + \omega_{\text{wafer}} \left(x - c_{\text{wafer}}\right) \hat{E}_y \qquad \forall (x, y) \in \text{wafer} \tag{4.19}$$

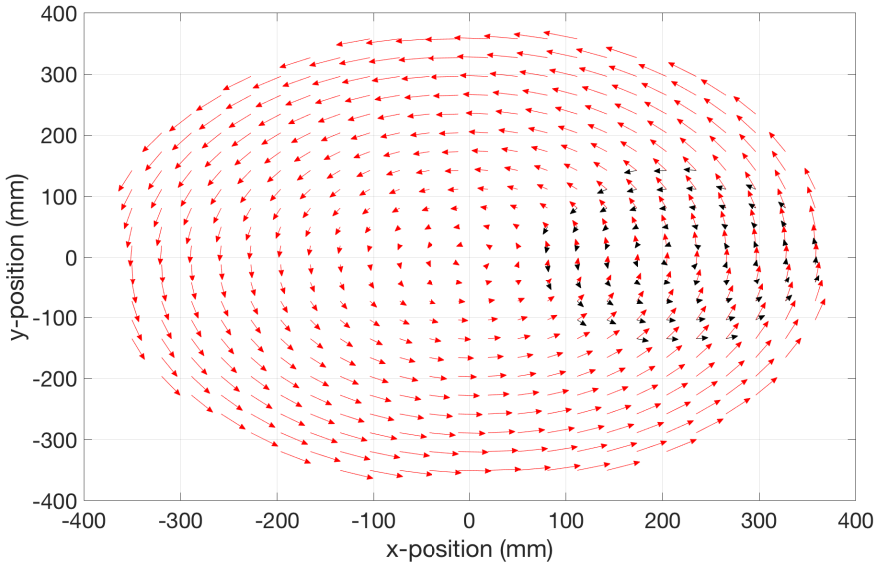
The instantaneous relative velocity between the pad and wafer is

$$\begin{aligned} v_{\text{relative}}(x, y) &= v_{\text{pad}}(x, y) - v_{\text{wafer}}(x, y) \\ &= y \left(\omega_{\text{wafer}} - \omega_{\text{pad}}\right) \hat{E}_x + \left(\omega_{\text{pad}} x - \omega_{\text{wafer}} x + \omega_{\text{wafer}} c_{\text{wafer}}\right) \hat{E}_y \qquad \forall (x, y) \in \text{wafer} \end{aligned} \tag{4.20}$$

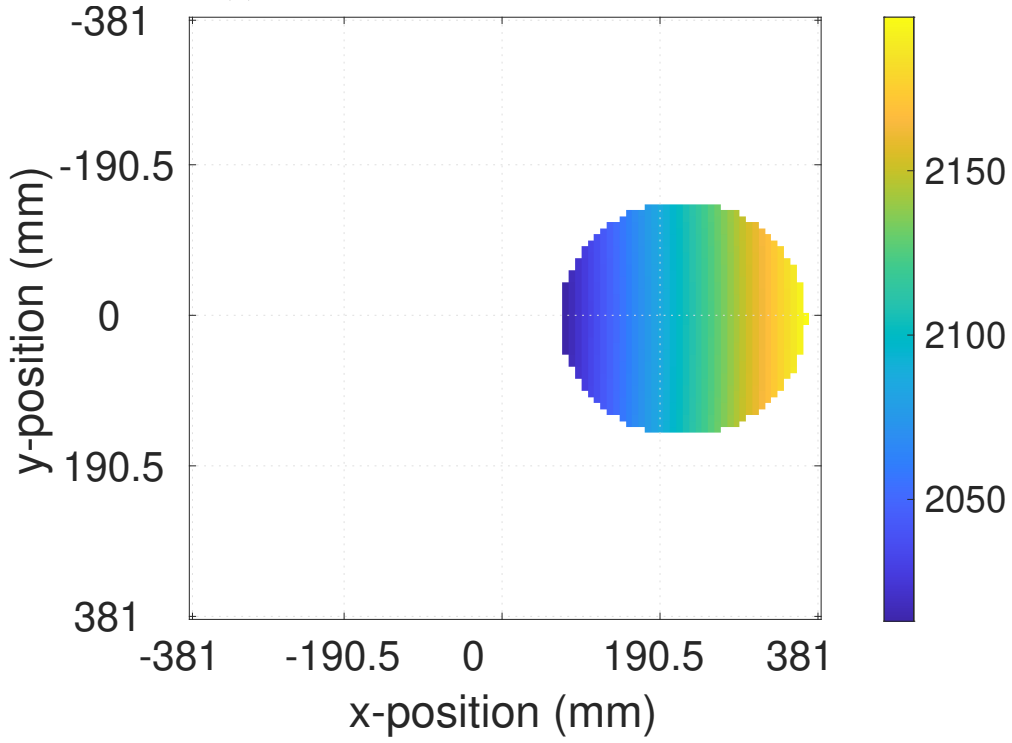
The magnitude of the relative velocity is then

$$\left| v_{\text{relative}}(x, y) \right| = \sqrt{\left(y \left(\omega_{\text{wafer}} - \omega_{\text{pad}}\right)\right)^2 + \left(\omega_{\text{pad}} x - \omega_{\text{wafer}} x + \omega_{\text{wafer}} c_{\text{wafer}}\right)^2} \qquad \forall (x, y) \in \text{wafer} \tag{4.21}$$

Figure 4.7 shows a visual representation of this relative velocity calculation, by showing the pad and wafer rotational speeds as vector fields, subtracting the two to attain the relative velocity within the wafer domain, and plotting the magnitude of the relative velocity (i.e. the relative speed). Figure 4.8 shows a cross-section of the relative speed through $y = 0$, for various possible platen rotation speeds and illustrates that the instantaneous relative speed may vary significantly across the wafer.



(a) Pad and wafer rotational velocities



(b) Magnitude of the relative velocity (mm/s)

Figure 4.7: Velocity fields. (a) shows the pad's rotational velocity (red) and the wafer's rotational velocity (black) as vector fields, where the length of the arrow corresponds to the rotational speed at that location. The difference between the pad and wafer's rotational velocities (within the wafer) is the instantaneous relative velocity, and the relative velocity's magnitude is plotted in (b).

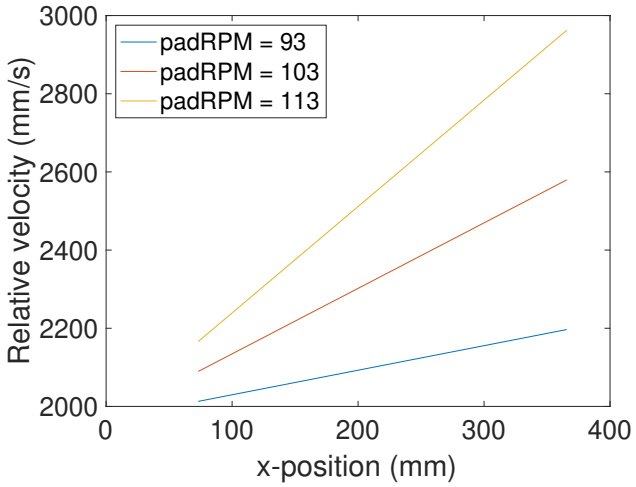


Figure 4.8: The instantaneous relative velocity (in magnitude) is plotted against the position on the pad (with $x = 0$ corresponding to the pad center. Various cases for the pad's rotational speed are shown, and the results indicate that increasing the pad's rotational speed not only increases the instantaneous relative speed, but it also increases the variation in speed across the wafer.

We need to extract the portion of the relative velocity field that is relevant to contact simulations (i.e. within the simulation domain $\mathcal{D}$ at a given time increment). The process is identical to that shown in Section 4.2.5, with the relative velocity field used in place of the measured pad surface scan, and with $\delta = 0$. An example of the relative velocity field at specified simulation time increments is shown in Figure 4.9.

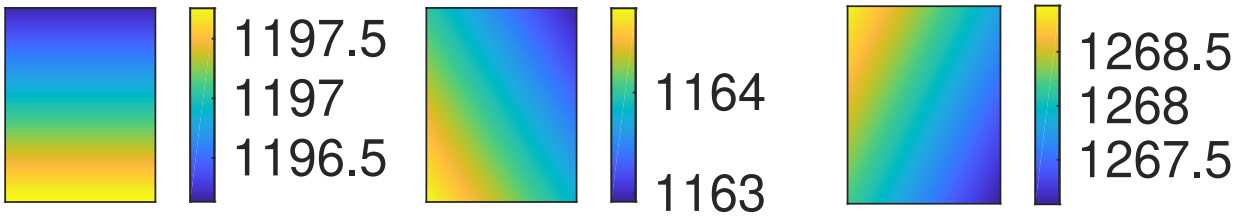


Figure 4.9: Relative velocity (mm/s) relevant for contact simulations at the following time increments: 0 s, 1.1811 s, and 2.3622 s of simulated polish.

4.2.7 Understanding the pad and wafer velocities' effect on the relative velocity

It is important to understand how the set pad and wafer rotational speeds impact the relative velocity distribution. There are two important quantities that describe the relative velocity

distribution: the range in relative velocity across the wafer and the average relative velocity across the wafer. It is typically desired to have a low range, so that polishing may be uniform across the wafer. The desired average velocity value, however, needs to be determined by the process engineer, as they need to balance high removal rates with stabilizing the process (i.e. not having the wafer hydroplane).

I studied these two summary statistics by predicting the relative velocity field for multiple pad rotational speed and wafer rotational speed combinations. Many insights can be drawn from scenarios where both the pad and wafer rotate in the same (counterclockwise) direction and this analysis also included scenarios where the pad and wafer rotate in different directions.

Figure 4.10 plots the variation in relative speed across the wafer. If the pad and wafer are both rotating counter clockwise, then high variation in relative speed occurs when the wafer's rotational speed is high but the pad's rotational speed remains low. If instead, the pad and wafer are set to rotate in different directions (say the pad is set to rotate counterclockwise and the wafer is set to rotate clockwise), then high variation in the relative velocity will occur with little sensitivity to the specified speeds. Figure 4.11 plots the average relative speed across the wafer. If the pad and wafer are both rotating counter clockwise, then increasing the pad's rotational speed by 10 rpm has a much larger effect on the average relative velocity than increasing the wafer's rotational speed by 10 rpm.

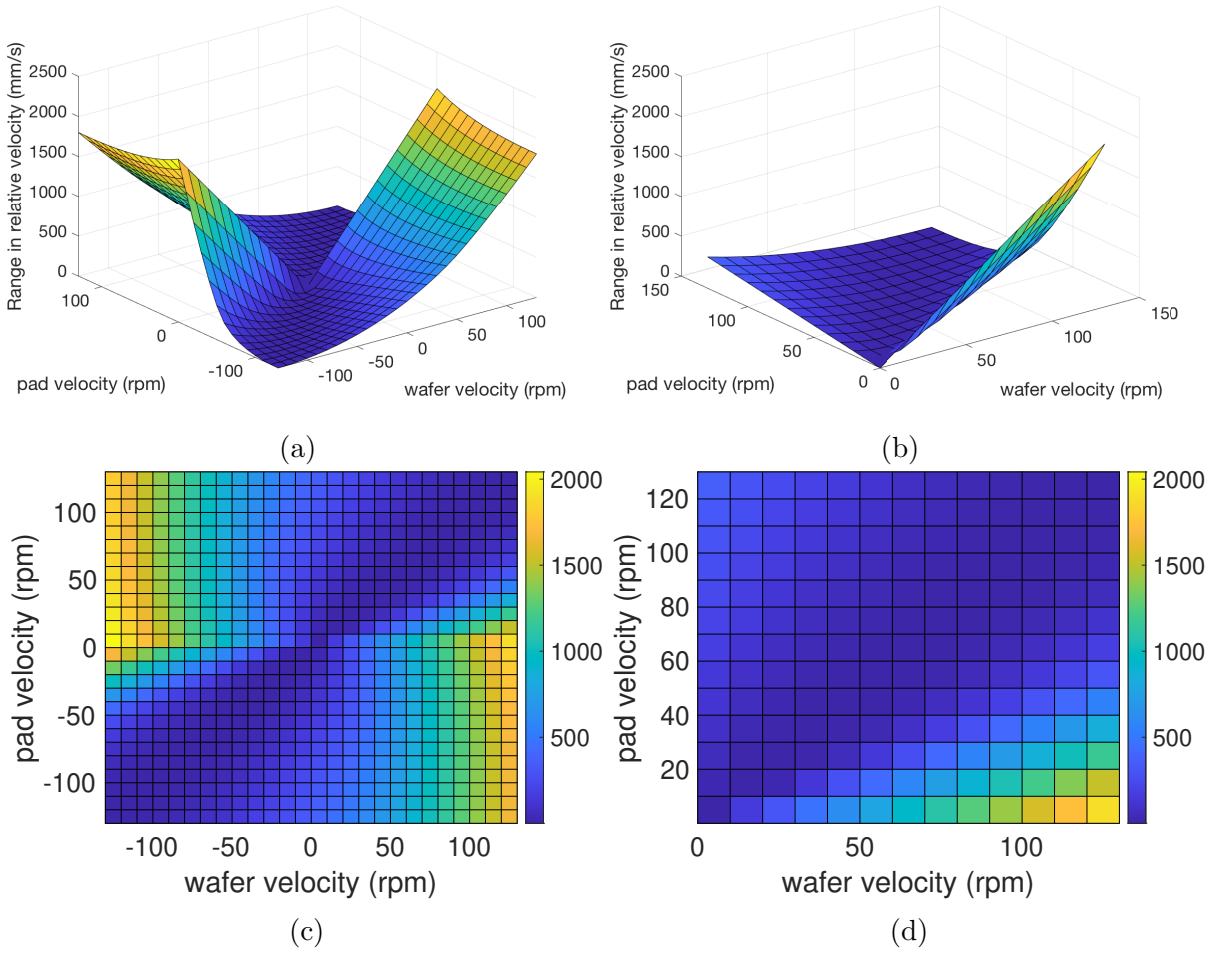


Figure 4.10: Visualizations of the variation in relative speed, for multiple pad velocity and wafer velocity combinations. In (a), situations when the pad rotates either clockwise or counterclockwise while the wafer rotates either clockwise or counterclockwise are included. In (b), only the situations when the pad and wafer both rotate counterclockwise are shown. (c) contains a 2D top view of (a) and (d) contains a 2D top view of (b), which aids in detecting combinations that lead to high variation in relative speed.

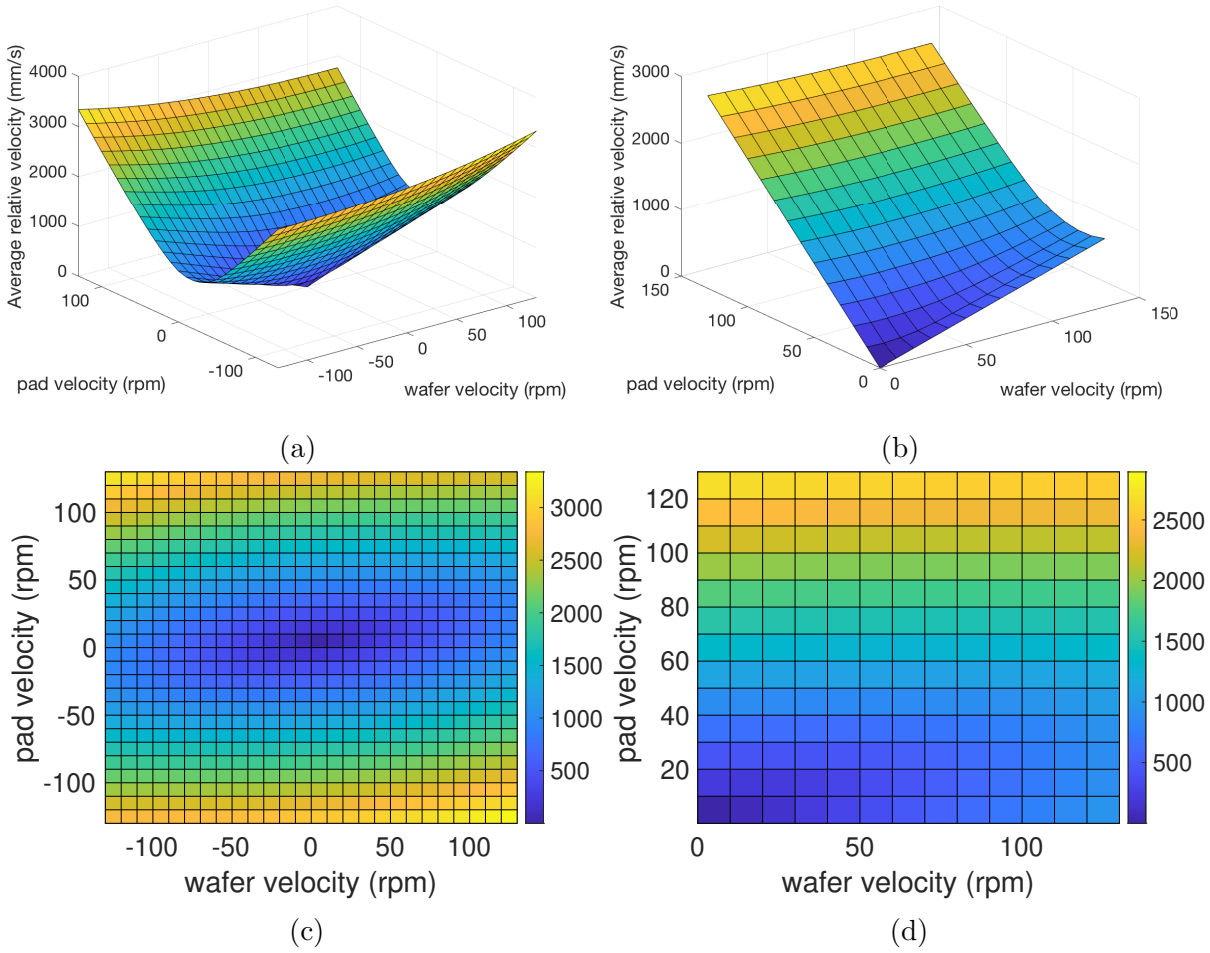


Figure 4.11: Visualizations of the average relative speed, for multiple pad velocity and wafer velocity combinations. In (a), situations when the pad rotates either clockwise or counterclockwise while the wafer rotates either clockwise or counterclockwise are included. In (b), only the situations when the pad and wafer both rotate counterclockwise are shown. (c) contains a 2D top view of (a) and (d) contains a 2D top view of (b), which aids in visualizing the average relative speed field.

Maury [133] provides equations for calculating the relative velocity across the wafer, when averaged across time. This time-averaged relative velocity is plotted against the radial position from the wafer center in Figure 4.12, for multiple values of the pad's rotational speed while the wafer's rotational speed is held constant. The plot shows that the time-averaged relative velocity fields have significantly less across-wafer variation than the instantaneous relative velocity fields. This is expected, since the purpose of rotating the wafer (and not only rotating the pad) is to provide this averaging. Rotating the pad and wafer about offset centers also helps reduce this variation. Figure 4.12 also shows that, for this particular

setup, setting the pad's rotational speed value close to the wafer's rotational speed results in the least amount of variation. Additionally, the time-averaged relative velocity remains relatively uniform near the wafer center, and rises towards the wafer edge.

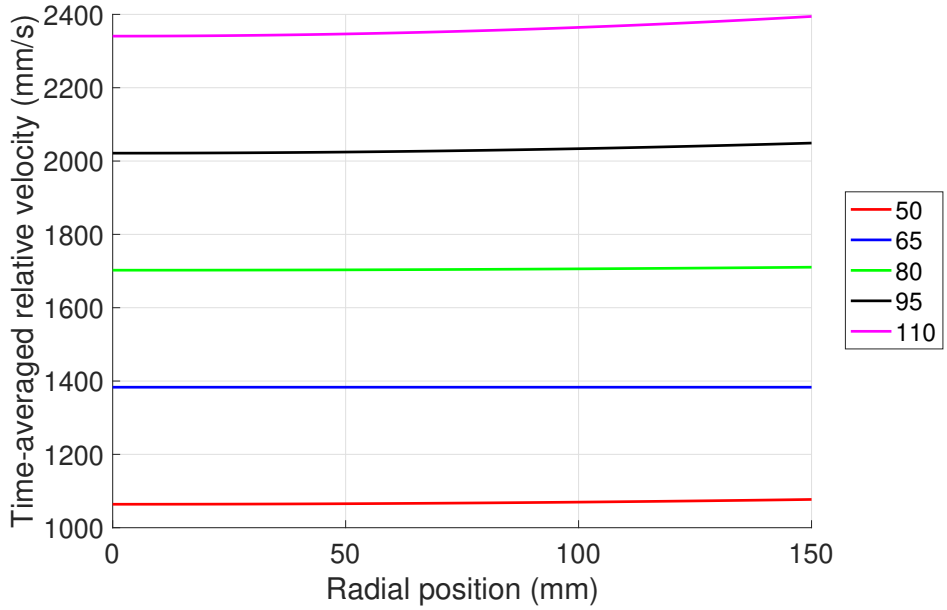


Figure 4.12: The time-averaged relative speed is plotted against the radial position from the wafer center, for various values of the pad's rotational speed. The wafer's rotational speed is held constant at 65 rpm.

4.2.8 Boundary conditions

The die-scale model needs to account for the periodic boundary conditions around the die's edge, since the die design are tessellated across the wafer and the die at the wafer center is completed surrounded by die's with the same pattern. Note that the wafer used in the experiments may contain spacer features between the dies, which are not accounted for here but may readily be modeled. Since the fast Fourier transform assumes periodic boundary conditions around the simulation domain, as described in Figure 4.12, the simulation domain is chosen to consist of a single die and the periodic boundary conditions are automatically imposed.

4.2.9 Incorporating multiple length scales

This die-scale model is useful for analyzing arrangements of features which span areas on the order of 10 mm x 10 mm. To simulate an area of this size, a rather coarse resolution is needed (here, $p = 50$ microns). This means that only features with size greater than $2p = 100$ microns are resolved within these simulations. However, there are features on the die which are smaller than 100 microns but which would still benefit from contact mechanics-based simulations using experimentally measured asperity profiles. To this end, this section presents a method for using the feature-scale model presented in Section 3 (and the finer scans associated with that model) in conjunction with this die-scale model (and the coarser scans associated with this model).

Now, for a given feature in the domain, the decision will be: which model needs to be used for calculating that feature's planarization efficiency? Figure 4.13 shows the decision tree. The decision of which model to use for calculating the feature's planarization efficiency depends on that feature's critical dimension. If that feature is sufficiently large that it can be adequately resolved within the die-scale model (here, greater than 100 microns), then the die-scale model is used directly for calculating that feature's planarization efficiency. If, instead, this feature is not large enough to be resolved within the die-scale model, then we have another branch: is this feature large enough to be resolved by the finer scanner? If so, we employ the feature-scale model for calculating the planarization efficiency (as a lookup table for this specific feature size and type, where the planarization efficiency vs step height curve is linearly interpolated to attain the planarization efficiency at any step height value). If, instead, the specific feature is so small that it cannot be resolved by either scanner, then we use a theoretical model for its planarization efficiency as presented in Section 4.2.10.

4.2.10 Determining the sub-pixel pressures

The die-scale model is implemented with a given grid size. In the present work, that grid size is set to 50 microns. However, there are many features in the die that are smaller than this grid size. It would be computationally intractable to use a grid that is fine enough to resolve even the smallest features. Instead, we select a grid size that is fine enough to resolve

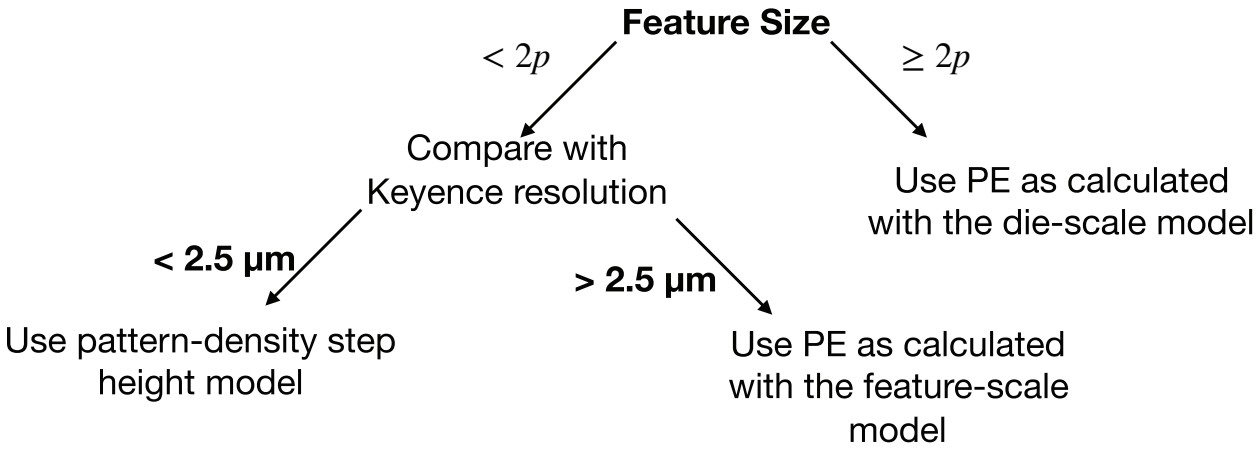


Figure 4.13: Decision tree for determining which model to use for calculating the planarization efficiency of a given feature set. The decision depends on the size of the particular feature, in relation to the die-scale simulation's pixel size and the finer scanner's resolution.

as many features on the die as possible but coarse enough so that the simulation runs in reasonable time.

For those features that are smaller than the grid size, we require a method for determining the pressure within those features' active and trench regions (herein called the sub-pixel pressures). This method is presented in [129]. We start with Equation 2.2 for the planarization efficiency of Prestonian materials

$$P.E. = 1 - \frac{P_{\text{trench}}}{P_{\text{active}}}$$

We then recognize that the sub-pixel active and trench pressures must sum to the total pixel pressure (when weighted by the pattern density), meaning that

$$P(x,y) = \rho P_{\text{active}}(x,y) + (1-\rho)\, P_{\text{trench}}(x,y) \tag{4.22}$$

By solving Equation 4.22 for P_{active}, substituting into Equation 2.2, and simplifying (and similarly solving Equation 4.22 for P_{trench}, substituting into Equation 2.2 and simplifying), we get

$$P_{\text{active}}(x,y) = \frac{P(x,y)}{1 + (\rho - 1)\, PE} \tag{4.23a}$$

$$P_{\text{trench}}(x,y) = \frac{(1 - PE)\, P(x,y)}{1 + (\rho - 1)\, PE} \tag{4.23b}$$

We now have equations for the pressures in active and trench regions, where those regions are smaller than the pixel size. These equations are in terms of the average pressure for that pixel, the pattern density in that pixel, and the planarization efficiency for the features

contained within that pixel. We now require a method for computing that planarization efficiency.

For very small features (in comparison to the pad segment size), the lateral feature size and feature shape are unlikely to affect the planarization efficiency. The feature step height, however, does affect the planarization efficiency, even for small features. We now incorporate a theoretical model for the planarization efficiency of small features.

Xie's pattern density step height (PDSH) model tells us that if the asperity height has an exponential probability density distribution, then

$$P_{\text{active}}(x,y) = \frac{1}{1+\rho\left(\exp\frac{h}{\lambda}-1\right)}P(x,y) \tag{4.24a}$$

$$P_{\text{trench}}(x,y) = \frac{\exp\frac{h}{\lambda}}{1+\rho\left(\exp\frac{h}{\lambda}-1\right)}P(x,y) \tag{4.24b}$$

Here, h is the step height of the features. It then follows from Equation 2.2 that

$$PE(x,y) = 1 - \frac{P_{\text{trench}}(x,y)}{P_{\text{active}}(x,y)} = 1 - \exp\left(-\frac{h}{\lambda}\right) \tag{4.25}$$

Figure 4.14 shows this relationship between the planarization efficiency and feature step height, for various values of λ, and compares it against experimental measurements of the planarization efficiency.

By substituting Equation 4.25 into Equation 4.23, we now have expressions for the pressure at active and trench regions for subpixel-sized features.

4.2.11 Selecting the areas on which to measure the oxide thicknesses

Selecting the regions over which to average the active and trench oxide thicknesses (for plotting) is an important task, as those thicknesses may vary across the feature cluster. For features with size $\geq 50\,\mu\text{m}$, the probing regions were chosen as the entirety of that feature's active and trench regions. This allows the plotted thicknesses to be averaged across the largest possible area, which minimizes any jaggedness in the predicted oxide thickness vs time curve that would result from the small percentage of time that each individual pixel experiences material removal (due to the low contact area ratio). While this averaging removes potential jaggedness in the predicted oxide thickness vs polish time curve, there is pixel-to-pixel jaggedness in each feature's simulated topography (as shown in Figure 4.17). This pixel-to-pixel jaggedness is an unphysical artifact that arises from the low contact area ratio, time step selection and simulation pixel size selection. Despite the pixel-to-pixel jaggedness, averaging across the entirety of the feature's active and trench regions produces smooth predicted oxide thickness vs time curves (as shown in Figure 4.15).

For features with size less than $50\,\mu\text{m}$, the experimental protocol specifies that measurements are performed within $50\,\mu\text{m} \times 50\,\mu\text{m}$ metrology patches, which are either entirely

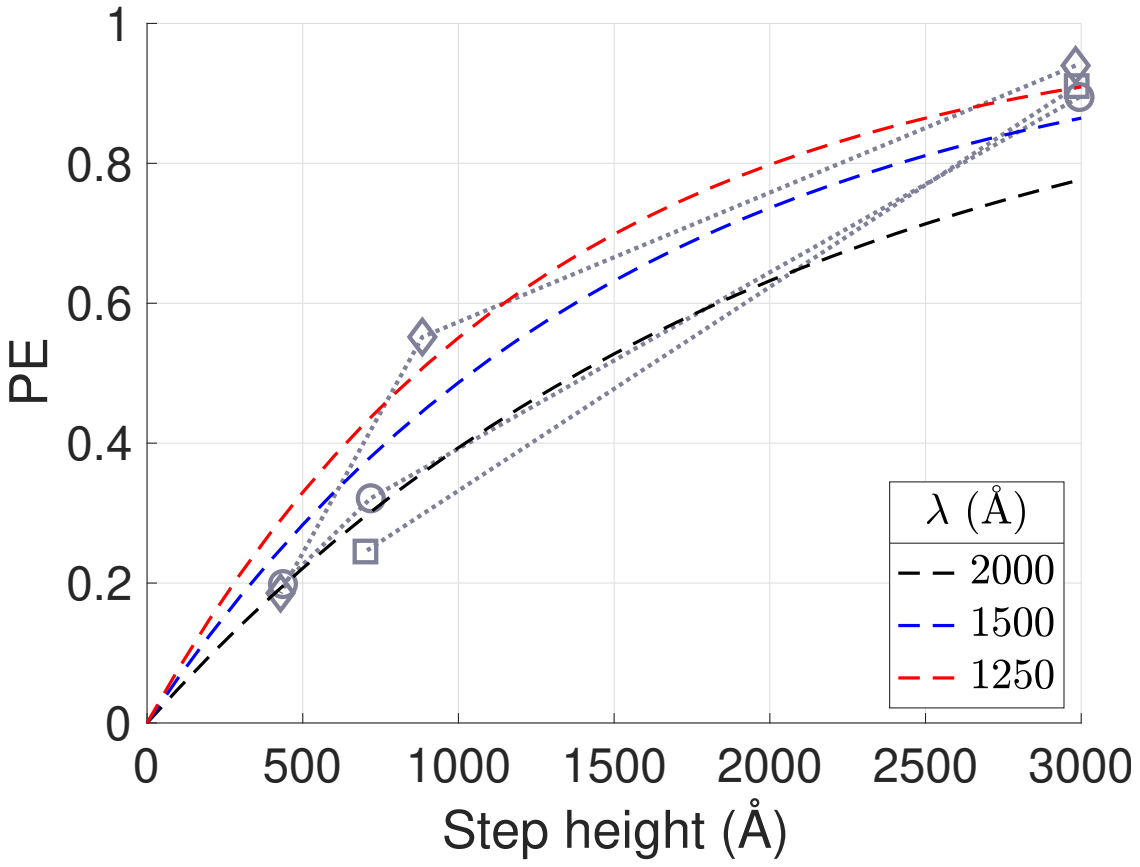


Figure 4.14: Plot of the planarization efficiency vs feature step heights (as given in Equation 4.25), for various values of λ. Also plotted are experimental measurements taken on small features (using a backward difference method with an assumption that the sum of the etched trench depth, nitride thickness, and pad oxide thickness equals 300 nm). Lines correspond to the simulations and symbols correspond to the experiments. Circle, square, and diamond symbols correspond to 10, 5, and 0.5 μm line/space feature classes, respectively.

active or trench regions and are surrounded by features of that class. These metrology patches are shown in Figure 4.2d, as yellow (active) or dark blue (trench) squares. The entirety of these metrology patches are chosen as the probing regions for the small features.

4.3 Results

Using this previously described method, several die-scale simulations were run, and those results are presented here. In the following subsections, I first present comparisons between simulations and experiments, and while reasonable agreement is shown, there is still opportunity for greater agreement. Following that, the model is used to simulate additional interesting scenarios for which experiments were not performed. Finally, a discussion of the simulation parameters and their effects on the polish predictions is presented.

4.3.1 Comparing simulated and experimentally measured oxide-thickness curves

The die-scale model contains four parameters: E_{surface}, $w^{\text{max}}_{\text{asperities}}$, λ, and k. Die-scale simulations using Pad Bwere run, and manually tuned until reaching the parameter set provided in Table 4.1. Predicted active and trench oxide thickness vs polish time curves, for multiple feature classes, are shown in Figure 4.15. Included on those plots are error bars representing 95% confidence intervals on the mean for the experimentally measured thicknesses. The plots show generally strong agreement between the simulations and experiments, for all four feature classes shown. The level of agreement varies as the polish time increases. In particular, the 1000 µm checkerboard simulations predict a significantly higher active oxide thickness at 75 s of polish than was measured in experiments. Additionally, the 2000 µm checkerboard simulations predict a significantly higher active oxide thickness at 50 s and 75 s of polish than was measured in experiments. At the time of writing, a set of values for simulation parameters had not been found that would improve the overall agreement, as improving the agreement at the 75 s polish times led to decreased agreement at the 25 s polish time.

It is important that the model captures the planarization efficiency's feature size dependence. Figure 4.16 shows a plot of the planarization efficiency vs feature class, for features of five different sizes. The plot shows that the predicted planarization efficiencies decrease as the feature size increases, which matches the experimental trend. The predictions are compared against experimentally measured planarization efficiencies (for four feature classes), and they agree with each other to within six percentage points—a rather strong agreement level.

It is important to note the possibility that the experimental metrology process introduces new process effects which are not modeled here. As described in Section A.4.2, since the wafer polishing is paused for measurement purposes, the pad is allowed to return to room temperature and the trench oxide thickness measurements differ from when less of these pauses are taken. This model does not attempt to capture pad temperature effects. Despite

these limitations, this die-scale model still achieves a high level of agreement with experiments and it is still useful for extracting CMP fundamental insights.

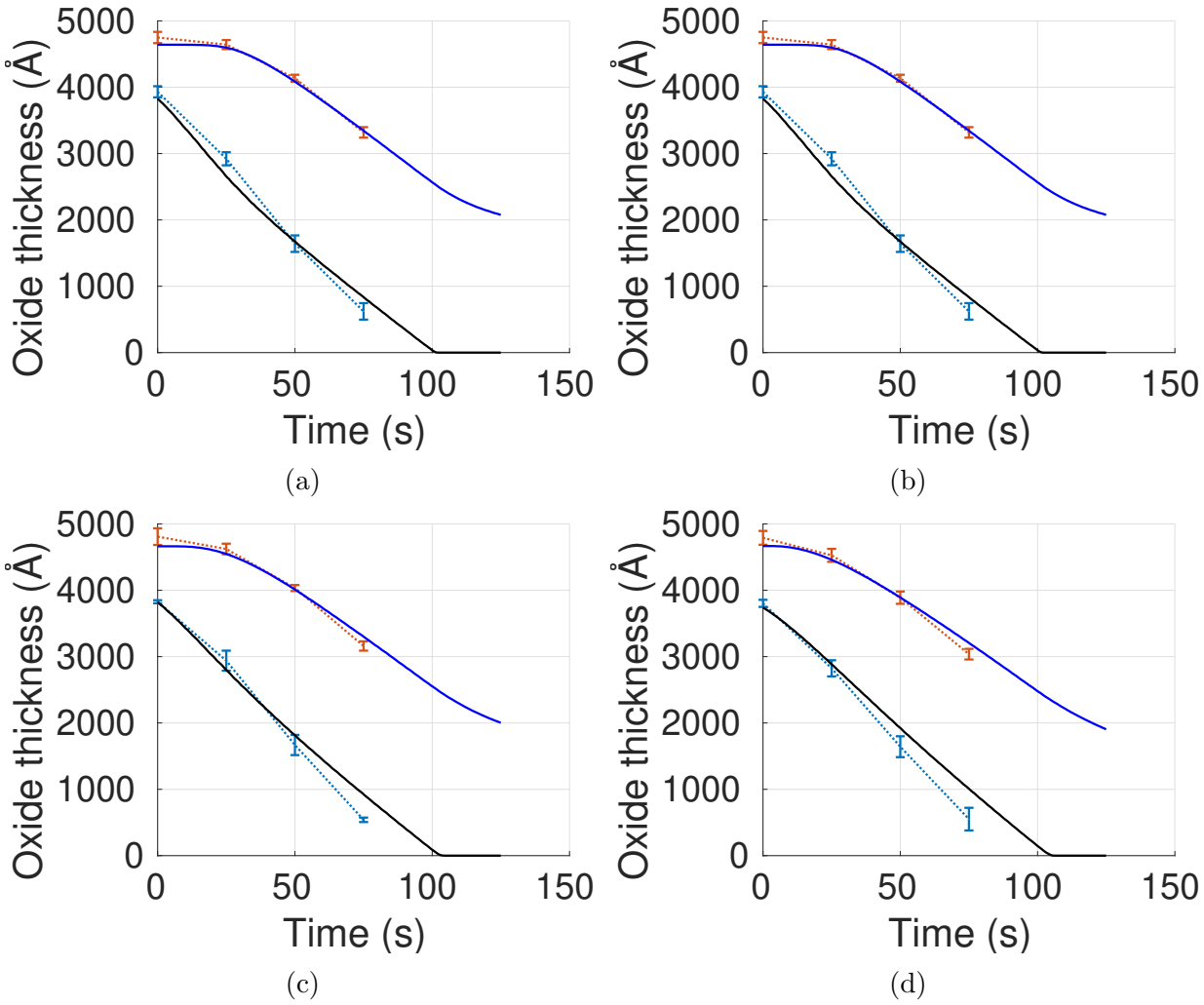


Figure 4.15: Plot of the oxide thickness vs polish time for the (a) 200 µm (b) 500 µm (c) 1000 µm and (d) 2000 µm checkerboard feature classes. Lines correspond to simulation predictions (black: active, blue: trench) and the error bars correspond to 95% confidence intervals on the mean (based on experimental measurements with Pad B for the active and trench regions).

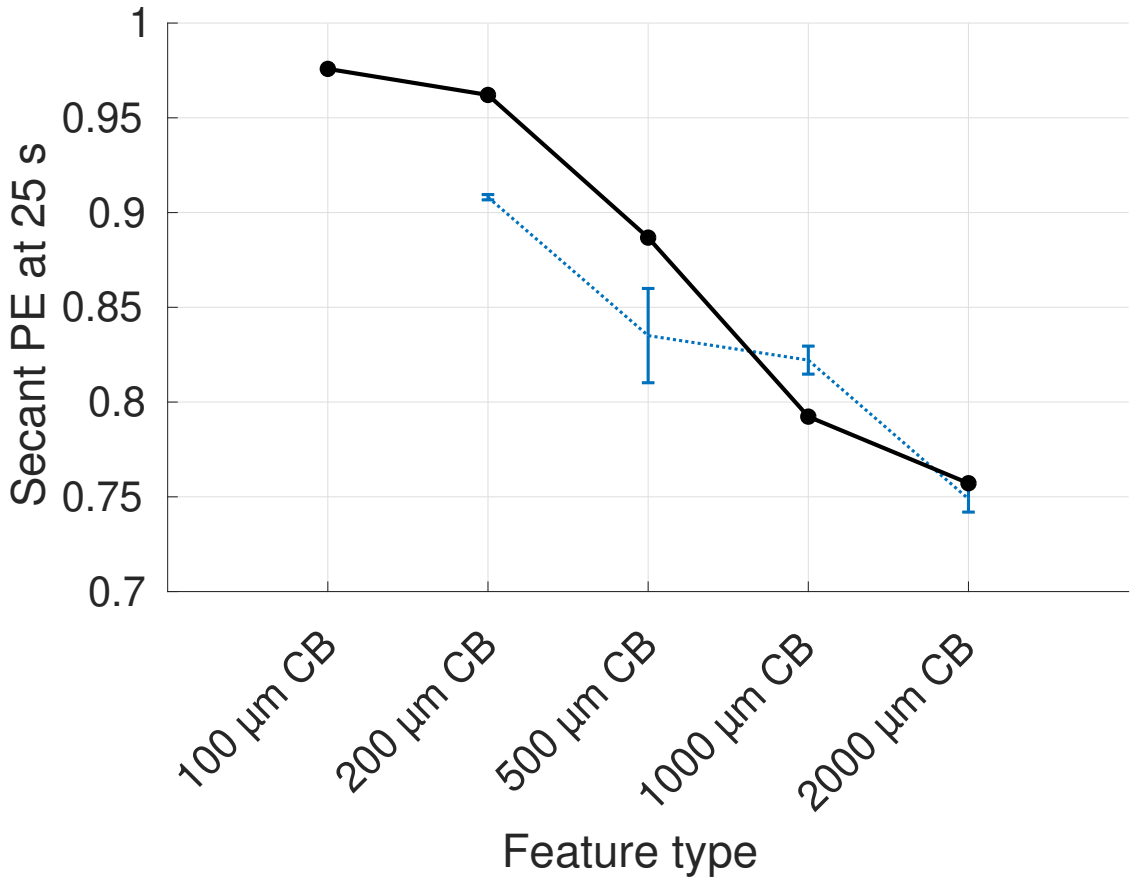


Figure 4.16: Plot of the planarization efficiency vs feature class. Error bars correspond to the 95% confidence intervals on the mean (from experiments with Pad B). The planarization efficiency is calculated using the secant slopes on the oxide thickness vs polish time curves from 0 to 20 s, for both the active and trench regions.

4.3.2 Visualizing the die topography's evolution over the polish time

Figure 4.17 shows the die topography at the start of the polish and after three different polish simulation times. The simulations show that the within-die range decreases as the polish continues, with the range being $\sim$ 2500 Å before the polish, and $\sim$ 700 Å after 75 seconds of polish (i.e. there is planarization). Note that there is one feature — a large active block — which is the slowest to polish (matching experimental observations).

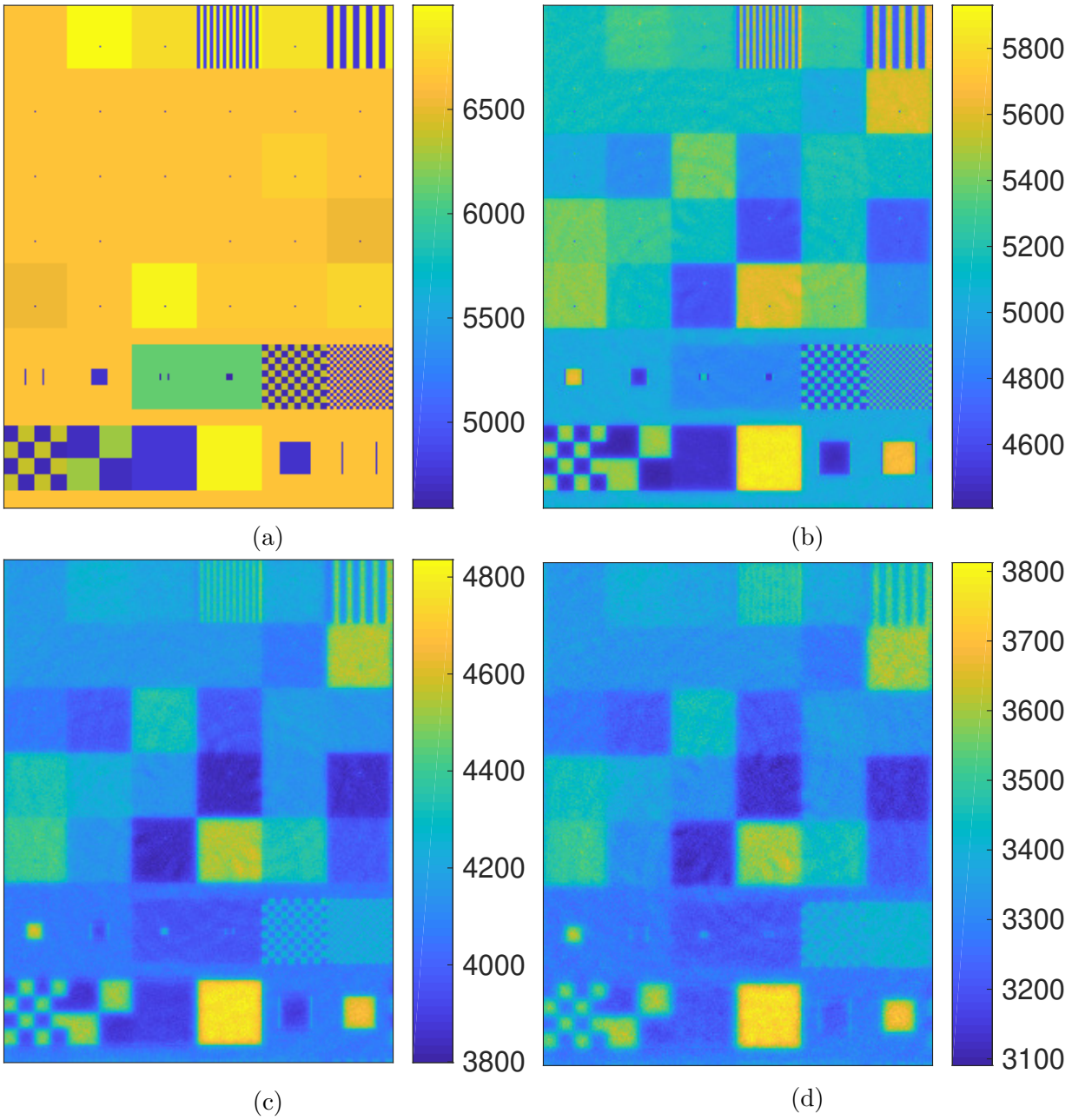


Figure 4.17: The simulated die topographies (in Å) after (a) 0 s, (b) 25 s, (c) 50 s, (d) 75 s of polishing with Pad B.

Table 4.1: Table of parameters used in die-scale simulations of Pad B which yield the greatest experimental agreement, as shown in Figure 4.15.

Parameter	Value	Parameter description
E_{surface}	400 MPa	Pad segment material's surface modulus
$w^{\max}_{\text{asperities}}$	50 µm	Height at which to trim pad asperities
λ	500 Å	Mean asperity height for use in the PDSH model
k	$2 \times 10^{-14}\,\text{Pa}^{-1}$	Coefficient in the material removal rate law

4.3.3 Simulating pad topographies with varying roughness

This CMP model may be particularly useful for extracting fundamental insights into the CMP process. One avenue for potential insights is understanding the pad surface roughness' effect on the oxide thickness vs polish time curve. To this end, simulations with three different pad surface textures were run. These pad surface textures were generated using an exponential distribution with parameter λ_1 for the asperity heights and the three values of λ_1 are 0 Å (i.e. perfectly flat), 500 Å (medium roughness), and 2000 Å (high roughness). The results, plotted in Figure 4.18, illustrate that flat pad surfaces result in high planarization efficiencies at low times, which does not match experimental observations. The flat pad segments are not able to contact the trench regions until more than 30 s of polish have elapsed. This suggests that incorporation of pad surface texturing is critical to accurately predicting the planarization efficiency, since it is the pad asperities that reach into the trenches. On the opposite end of the spectrum, a pad surface roughness that is too large results in planarization efficiencies that are too low compared to the experimental measurements. For this particular experiment, a value of $\lambda_1 = 500$ Å results in the strongest agreement with experiments.

4.3.4 Pad segment width effects

Next, I present the model's ability to capture pad segment width effects. To this end, I simulated polishing with four different pads, each with a different segment width w_{segment}, but having the same segment pitch $w_{\text{segment}} + w_{\text{groove}} = 2.05$ mm. The scenario where $w_{\text{segment}} = 1.35$ mm corresponds to the experiment with Pad B, and Pad B's surface texture was used for all four simulations (in the portions where the simulated pad contained segment material).

The model accurately predicts that rings with larger widths have higher planarization efficiencies, as shown in Figure 4.19. Moreover, pads with larger ring widths have smaller within-die active oxide thickness ranges, as shown in Figure 4.20 and Figure 4.21. Here, the within-die ranges are calculated at the polish time when the average active oxide thickness across the die is 100 Å. These finding indeed match both intuition and previous experimental observations, and it is important that this computational mechanics approach can capture these trends.

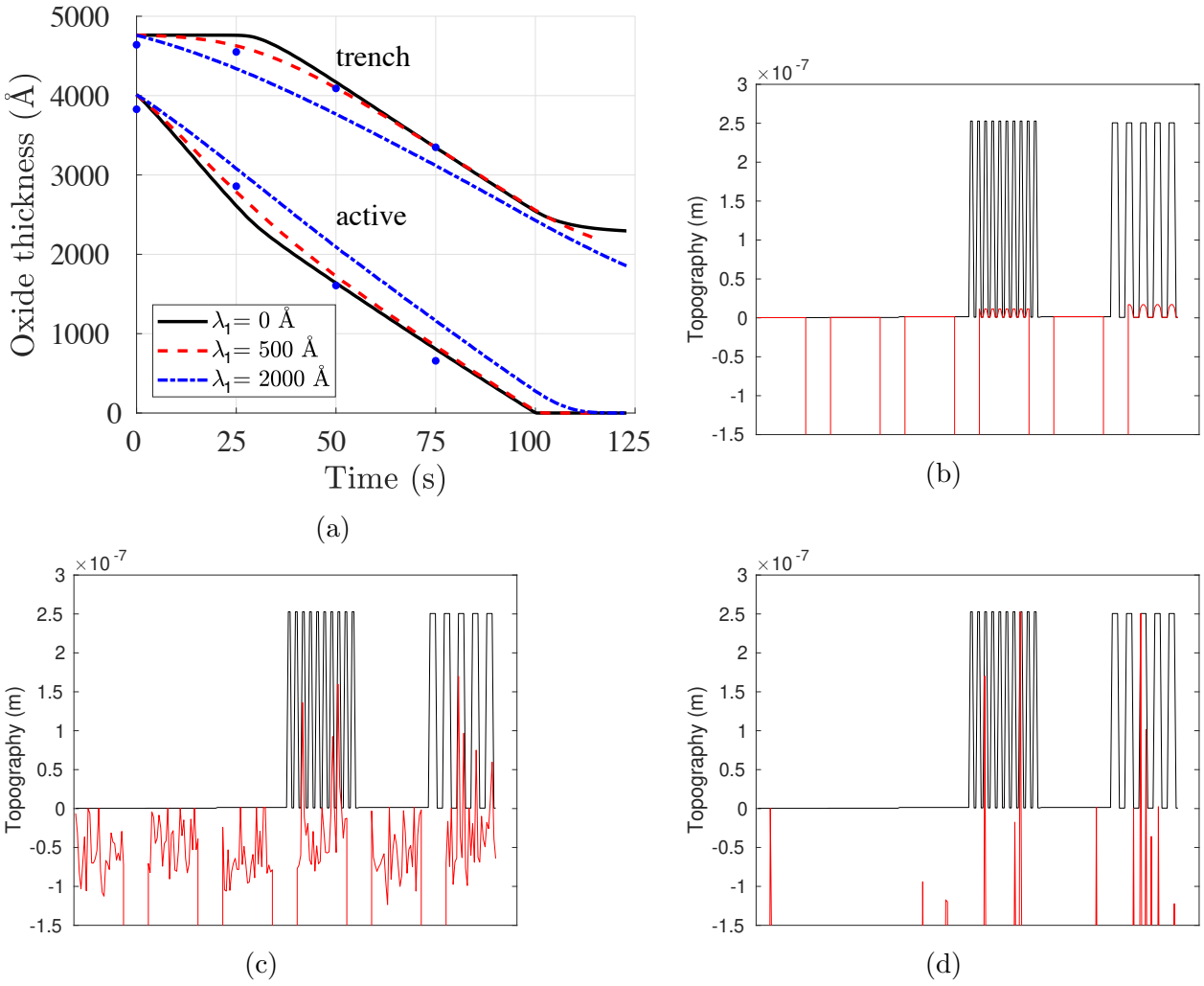


Figure 4.18: (a) Plot of the oxide thickness vs polish times, for the 200 µm checkerboard features. The pad surface texture was generated from an exponential distribution with parameter λ_1, and three different values of λ_1 are simulated. The circles represent the experimentally measured values. (b)–(d) show visualizations of the die (shown in black) contacting the pad (shown in red), in horizontal cross-section from the 50 µm checkerboards to the 200 µm line/spaces at $t = 0$ s. The corresponding values of λ_1 are (b) 0 Å, (c) 500 Å, and (d) 2000 Å.

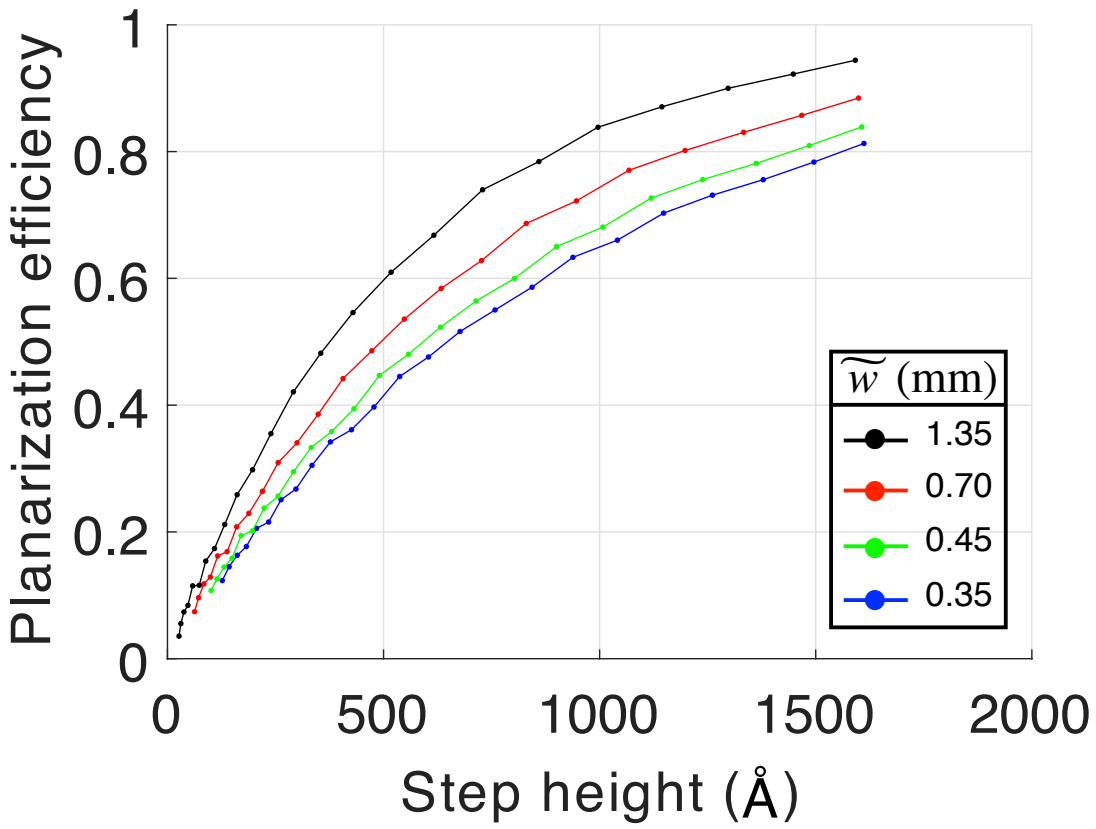


Figure 4.19: Plot of the simulated planarization efficiency against the feature step height (for the 200 µm checkerboard features). The 1.35 mm ring width scenario corresponds to Pad B.

4.3.5 Neighborhood effects: simulating various die designs

Next, I demonstrate the model's sensitivity to the input die design. I run simulations in which the die design changes slightly; the 2 mm x 2 mm area directly to the right of the 200 micron checkerboard features is varied from being (a) completely trench (b) line/space features with 25% pattern density (c) 50% line/space features (d) 75% line/space features (e) completely active. Figure 4.22 shows a cross section through the simulated 200 µm checkerboard topography after a 52 second polish. We see that the predicted oxide thickness—particularly for the region that was closest to the changing neighbor—depends on the neighbor's pattern density. The trend is consistent with intuition: having a neighbor with high pattern density results in slower material removal rates for the feature of interest (since high density regions tend to hold the pad away from adjacent regions), and hence a higher oxide topography. The model is sensitive to the pattern density of neighboring features; we see differences of ~ 300 Å between the two extreme cases.

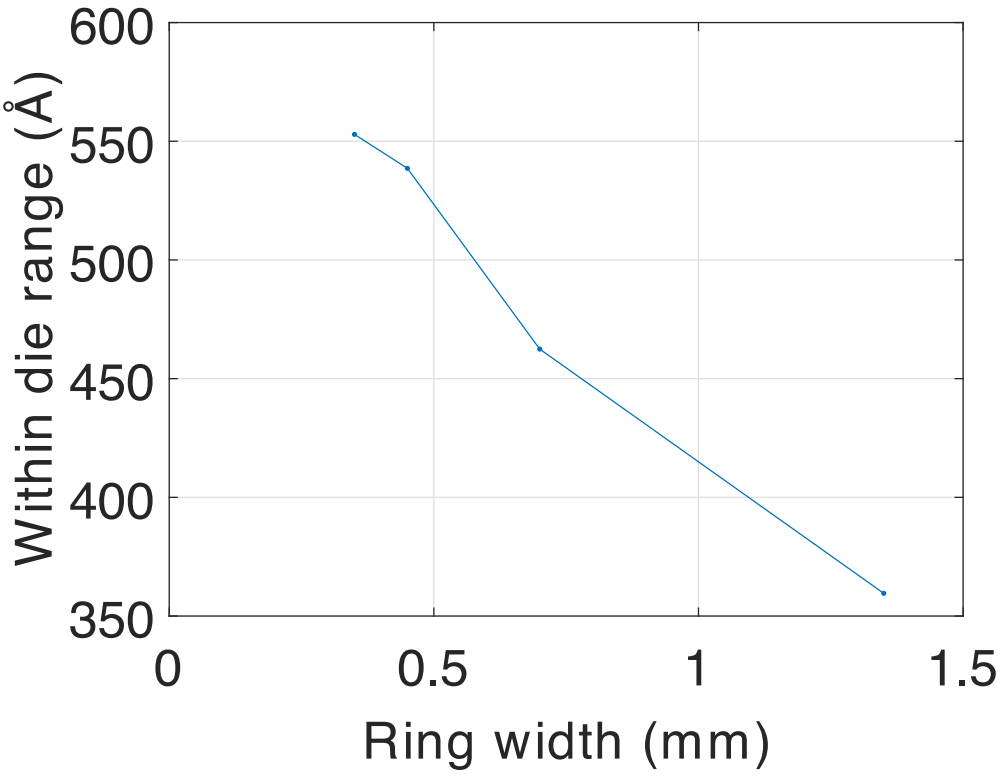


Figure 4.20: Plot of the within-die range against the simulated pad ring width w_{segment}. The within-die range is calculated at the instant when the average active oxide thickness across the die is 100 Å. The 1.35 mm ring-width scenario corresponds to Pad B.

4.3.6 Simulating additional pad designs and opportunities for model improvement

The preceding subsections discussed simulations with Pad B, which yielded strong agreement with experimental data. While the framework is able to model Prestonian processes (like in the case with Pad B), at the time of writing, the model is not yet able to adequately predict polishing for non-Prestonian processes. I now present an example non-Prestonian process using Pad C. Section 2.2 contains discussion of alternate material removal laws; here, I will present the methodology using a power material removal rate equation:

$$R(x,y) = k_1 \cdot (v(x,y) \cdot P(x,y))^{k_2} \tag{4.26}$$

where k_1 and k_2 are constants. Alternate rate equations may still be considered. With the removal rate law from Equation 4.26, the die-scale model now contains five parameters: E_{surface}, $w^{\max}_{\text{asperities}}$, λ, k_1, and k_2. Equation 2.2 for calculating the planarization efficiency

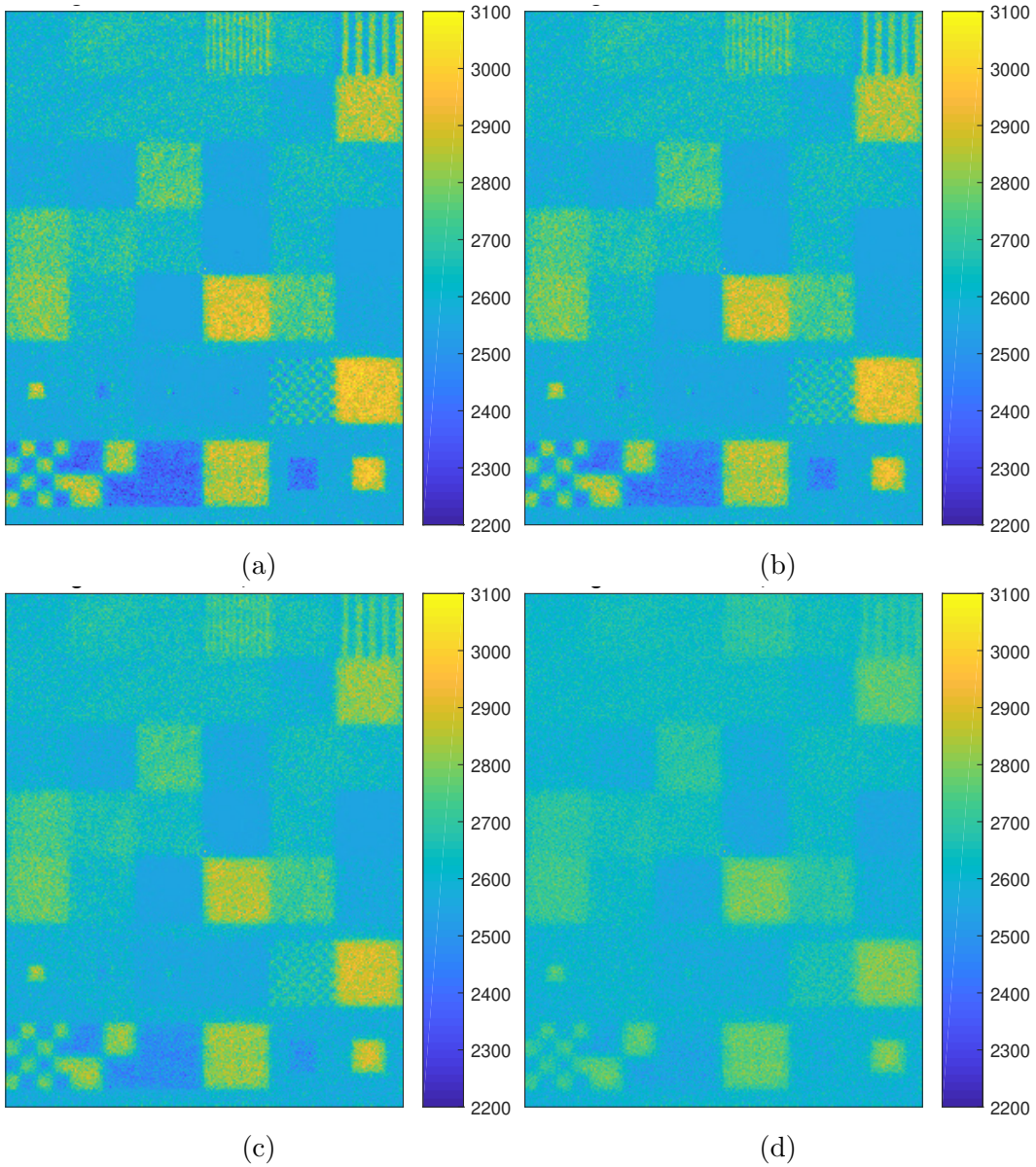


Figure 4.21: Visualization of the simulated die topography (in Å) when varying the pad ring width. Plots are shown at the instant when the average active oxide thickness across the die is 100 Å. The corresponding ring widths and simulated polish times are (a) 0.35 mm, 99.0 s (b) 0.45 mm, 98.7 s (c) 0.70 mm, 98.3 s (d) 1.35 mm, 98.0 s. Scenario (d) corresponds to Pad B.

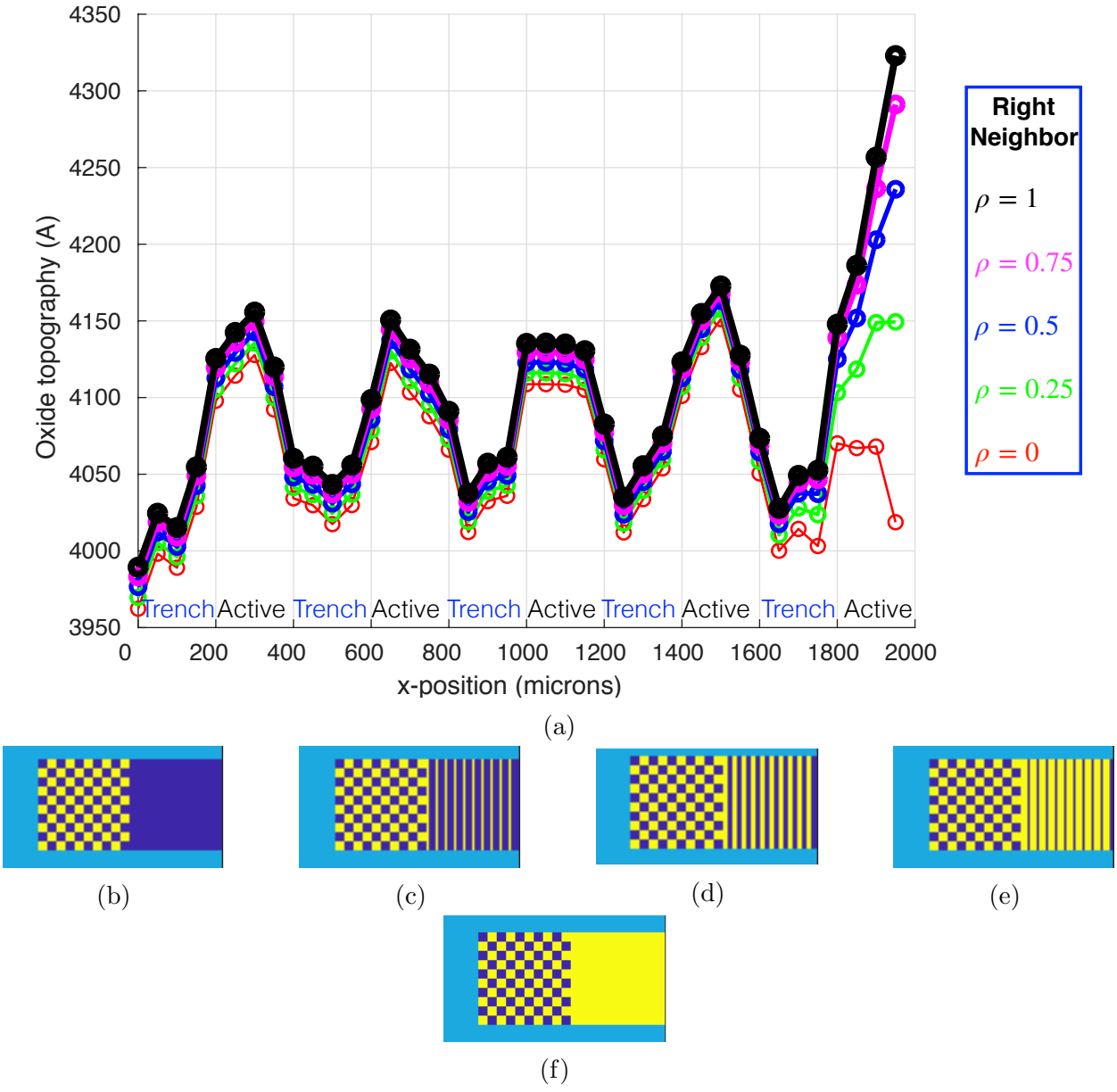


Figure 4.22: Simulations of the oxide topography when the average pattern density of the 200 µm checkerboard feature's right neighbor is varied. (a) Plot of the oxide topography against the x-position along the 200 µm checkerboard features, where x=2000 µm corresponds to the edge of the 200 µm checkerboard feature which is closest to the right neighbor. The five different scenarios are shown in (b)–(f) and the corresponding right neighbor average pattern densities are (b) 0 (c) 0.25 (d) 0.5 (e) 0.75 (f) 1.0. Simulations use the Pad B design and experimental parameters.

needs to be adjusted to account for this material removal law, as follows:

$$P.E. = 1 - \left(\frac{P_{\text{trench}}}{P_{\text{active}}}\right)^{k_2} \tag{4.27}$$

That means that Equation 4.25 is updated to:

$$PE(x,y) = 1 - \exp\left(-\frac{h \cdot k_2}{\lambda}\right) \tag{4.28}$$

Accordingly, Equation 4.23 is also updated to:

$$P_{\text{active}}(x,y) = \frac{P(x,y)}{\rho + (1-\rho)\,(1-PE)^{\frac{1}{k_2}}} \tag{4.29a}$$

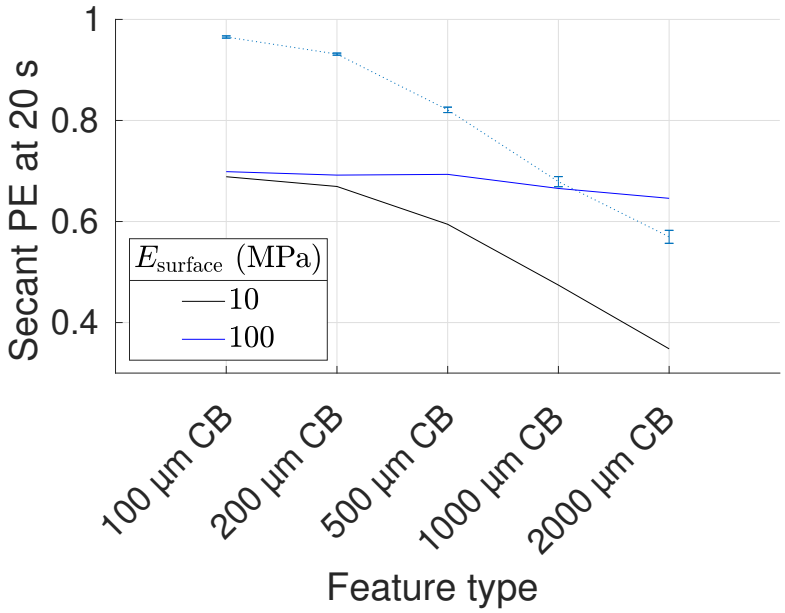
$$P_{\text{trench}}(x,y) = \frac{P(x,y)}{\rho + (1-\rho)\,(1-PE)^{\frac{1}{k_2}}}\,(1-PE)^{\frac{1}{k_2}} \tag{4.29b}$$

The simulation parameter set can be tuned to achieve high accuracy for one specific feature class, such as for the 500 µm feature class shown in Figure 4.23b. However, the simulations for Pad C have not been able to adequately predict the planarization efficiency's dependence on feature size, leading to insufficient accuracy for the remaining feature classes. To increase the amount of feature size dependence, the parameter set may be adjusted. Figure 4.23 shows that lowering E_{surface} (and thereby increasing the contact area ratio) leads to a larger predicted feature size dependence. In particular, Figure 4.23a shows that lowering E_{surface} from 100 MPa to 10 MPa results in the change in planarization efficiency (calculated from the 100 µm checkerboard features to the 2000 µm checkerboard features) going from ∼5% to ∼35%. This suggests that lowering E_{surface} can lead to greater predicted feature size dependence, though the overall values for the predicted planarization efficiencies are also affected. Figure 4.24 shows two additional parameters that may be adjusted to increase the overall model agreement: decreasing k_2 and decreasing $w^{\text{max}}_{\text{asperities}}$ both lead to an increase in the predicted planarization efficiency.

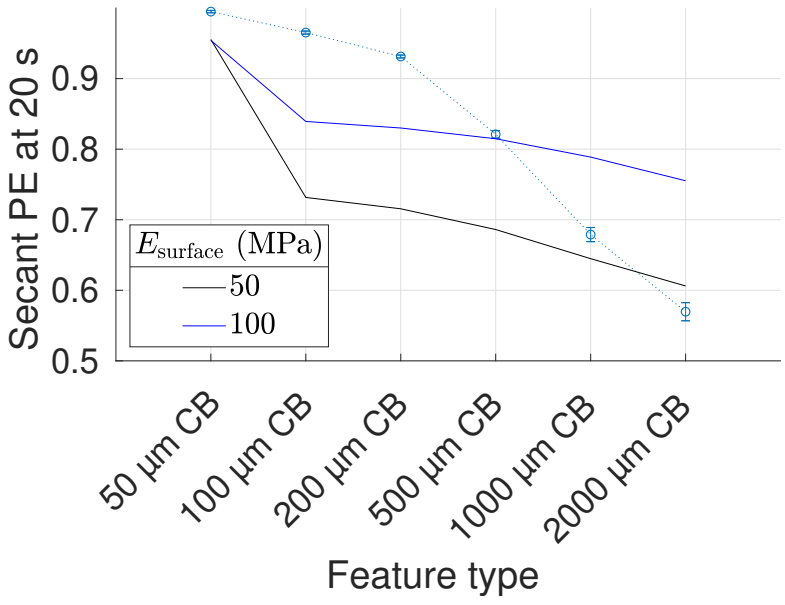
Figure 4.23 and Figure 4.24 also show a large deviation in behavior between the 50 µm and 100 µm checkerboard features. This sharp change in predicted planarization efficiency is a result of the different models used for these features (see Figure 4.13). These large deviation are not physical, and greater agreement between the different models is required.

4.4 Conclusion

This chapter has presented a die-scale CMP model which predicts the die topography's evolution over the course of a polish. This die-scale model expands upon the feature-scale model presented in Chapter 3, to allow simulation of arrangements of features. The particular die design is provided as input to the model.

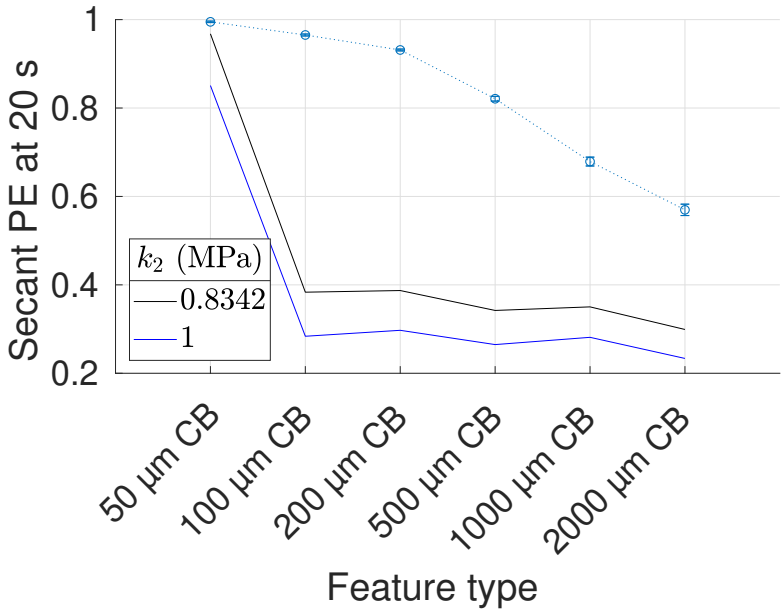


(a)

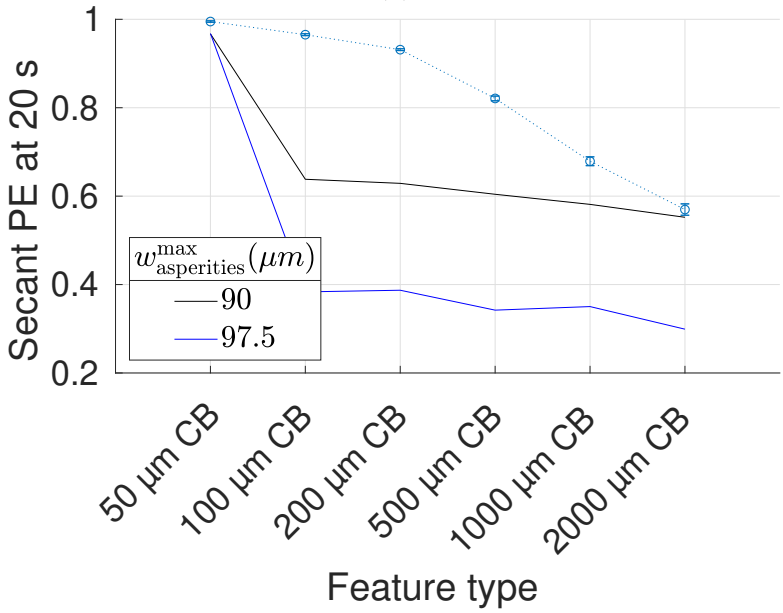


(b)

Figure 4.23: Plot of the planarization efficiency vs feature class, for various values of E_{surface}. In (a), the pad texture was generated from an exponential distribution with parameter $\lambda_1 = 2000\,\text{Å}$ while (b) uses a measured pad scan. Error bars correspond to the 95% confidence intervals on the mean (from experiments using Pad C). The planarization efficiency is calculated using the secant slopes on the oxide thickness vs polish time curves from 0 to 20 s, for both the active and trench regions.



(a)



(b)

Figure 4.24: Plot of the planarization efficiency vs feature class. (a) shows the effect of increasing k_2 and (b) shows the effect of increasing $w^{\max}_{\text{asperities}}$. Error bars correspond to the 95% confidence intervals on the mean (from experiments using Pad C). The planarization efficiency is calculated using the secant slopes on the oxide thickness vs polish time curves from 0 to 20 s, for both the active and trench regions.

The model employs a contact mechanics approach to simulate the pressure distribution between the pad and die. For the two pads discussed here, only one pad deformation mode is significant: pad surface deformation due to local indentations. The pad deformation is computed using a surface response approach and fast Fourier transforms. Notably, this model incorporates entire pad surface texture scans directly in contact mechanics simulations, and may be the first to do so. The model can also be used in tandem with finer pad scans, to create a multi-scale model.

When comparing the model predictions against experimental data, strong agreement was attained with Pad B's experiments. However, further work is required to reach strong agreement with Pad C's experiments. The model was shown to capture pad segment width dependence, feature-size dependence, and neighborhood effects. It is significant that these three dependencies can be accounted for by using a solid–solid contact approach. Additionally, the model predicts that the pad surface texture has a dominant effect on the polish, which reaffirms that proper pad conditioning and pad surface texture maintenance is vital for CMP processes. The model is also demonstrated using a non-Prestonian material removal law, and may readily be adapted with other removal rate equations.

With this model for predicting the active and trench oxide thicknesses as a function of the polish time, die design, pad design, and experimental tool parameters, new pad designs may be simulated. In the future, simulation results may be used to inform new pad designs, to select appropriate pads for a particular die, and to optimize die layout.

Chapter 5

Conclusions and future work

In Chapter 3, I have developed a feature-scale model of the planarization efficiency, as a function of the feature size, shape, and height. The model analyzes the solid-solid contact pressures at various possible pad-feature configurations, and then averages across these configurations to calculate an average planarization efficiency. In this manner, the effects of wafer horizontal sweeping and rotation are accounted for. I decompose the pad behavior into three deformation modes: surface elasticity, bulk segment downward displacement, and bulk segment rotation and generate surface response functions for each of these deformation modes. By using the surface response function approach, the pad deformation computations are much more computationally efficient than a purely finite element method approach would be. Results show that asperities are the dominant source of planarization, and for sufficiently large features, asperities are able to reach within the trench regions, causing the planarization efficiency to drop. Asperity tips cannot be treated as perfectly sharp, as this would result in only a few contact points and exceedingly high contact stresses. Instead, I model the asperities' nonlinear mechanical behavior as elastic, perfectly-plastic, which leads to flattened asperities and contact scenarios that match experimental results in the literature. A solid–solid contact model of CMP is sufficient to capture the planarization efficiency's dependence on feature size.

The die-scale model expands the methods used in the feature-scale model, and simulates the polishing of feature arrangements, not solely features of the same size and shape. It integrates various length-scales, by employing the feature-scale model to predict the planarization efficiency for features that are too small to be resolved by the die-scale model's resolution. For features that are even smaller than the feature-scale model's resolution (dicated by the pad scanner's resolution), a theoretical model known as the pattern density step height model is used. Significant insight into how the die layout affects polishing is provided.

The modeling work presented here is particularly important for the following reasons:

- The interactions between neighboring pad asperities are accounted for. This is done by computing asperity deformations using a boundary element approach. Pad asperity mechanics are not simply represented as independent springs.

- Pad texturing is represented as realistically as possible, by inputting measured pad scans directly into the contact mechanics models. The pressure distributions predicted by simulations using these measured pad asperities differ greatly from those predicted by models that assume flat pad surfaces, since the pressure is concentrated within a small area determined by the tallest asperities. Analysis showed that these pad asperities had distributions that were not well modeled by commonly used theoretical asperity distributions such as Gaussian or exponential distributions.

- The pad mechanical behavior is intricately represented as a composite material behavior. The material properties and sizes of each constituent material are used within finite element simulations to generate pad surface response functions. Those pad surface response functions are then used in convolution operations to determine the total pad deformation in a computationally efficient approach. The pad is not simply lumped to behave as a layer with an effective modulus.

- This analysis is performed for patterned wafers whose layout contains various feature sizes and types. The starting topography for this patterned wafer is measured and used in simulations. This analysis does not solely focus on blanket wafers, which, while easier to manufacture, test, and measure, possess polish behavior that might differ from that of patterned wafers.

- The die-scale model can be deployed as a multi-scale model, by creating planarization efficiency lookup tables with the feature-scale model. In this way, pad scans at two length-scales—a finer, lower-area and a coarser, larger-area scan—are used in conjunction with each other.

Looking forward, there remain significant areas for additional modeling work. Further work may focus on building this modeling framework, towards achieving greater agreement between simulations and experiments. It remains interesting to investigate dies with new layouts, which may provide more insight than the test die used here. This model can be expanded to investigate retaining ring effects and to account for the various applied pressure zones. A significant research thrust is in the optimization of process design variables; while this work focused on prediction, the die design and pad design can be optimized to achieve a user's process goals. The problem of the conditioning disk's sweep path remains an open problem; the disk's design may be improved and the sweeping profile may be optimized, leading to more uniform pad textures and higher planarization efficiencies.

The framework presented here opens new avenues to envisage the CMP process and its modeling. Above all else, proper representation of pad asperities is crucial to representing the pad–wafer interface and the pressures that develop at the contact zones.

Bibliography

[1] Z. Liu, G. Muldowney, and R. Palaparthi, "Measurement and Analysis of Single-Pass Conditioning Furrows in CMP Pad Surfaces," *ECS Journal of Solid State Science and Technology*, vol. 4, no. 11, P5118–P5126, 2015. DOI: 10.1149/2.0181511jss.

[2] A. S. Lawing, "Pad conditioning and pad surface characterization in oxide chemical mechanical polishing," in *Materials Research Society Symposium Proceedings*, vol. 732, 2002. DOI: 10.1557/PROC-732-I5.3.

[3] A. S. Lawing, "Polish Rate, Pad Surface Morphology and Pad Conditioning in Oxide Chemical Mechanical Polishing," in *Chemical Mechanical Planarization : proceedings of the International Symposium*, Electrochemical Society Proceedings Series, 2002, pp. 46–60.

[4] H. Kim, S. Hong, C. Shin, Y. Jin, D. H. Lim, J. yong Kim, H. Hwang, and T. Kim, "Investigation of the pad-conditioning performance deterioration in the chemical mechanical polishing process," *Wear*, vol. 392-393, pp. 93–98, 2017. DOI: 10.1016/j.wear.2017.07.019.

[5] R. Srinivasan, P. V. Dandu, and S. V. Babu, "Shallow Trench Isolation Chemical Mechanical Planarization: A Review," *ECS Journal of Solid State Science and Technology*, vol. 4, no. 11, P5029–P5039, 2015. DOI: 10.1149/2.0071511jss.

[6] T. Gan, "Modeling of Chemical Mechanical Polishing for Shallow Trench Isolation," Master's thesis, Massachusetts Institute of Technology, 2000.

[7] J. Choi, S. Tripathi, D. A. Hansen, and D. A. Dornfeld, "Chip Scale Prediction of Nitride Erosion in High Selectivity STI CMP," in *CMP-MIC*, Fremont, CA, 2006.

[8] M. R. Oliver, Ed., *Chemical-Mechanical Planarization of Semiconductor Materials*. Springer-Verlag Berlin Heidelberg GmbH, 2004, ISBN: 9783540431817. DOI: 10.1007/978-3-662-06234-0.

[9] J.-G. Park, T.-Y. Kwon, R. P. Venkatesh, and B.-J. Cho, "CMP Defects; Their Detection and Analysis on Root Causes," *ECS Transactions*, vol. 44, no. 1, pp. 559–564, 2012. DOI: 10.1149/1.3694369.

[10] Y. Y. Lin, F. S. Tsai, L. C. Hsu, H. K. Hsu, C. Y. Li, Y. Y. Ke, C. W. Huang, J. M. Chen, S. J. Chang, T. Y. Lee, E. Chen, and C. Y. Cheng, "Fast and accurate defect classification for CMP process monitoring," *ASMC (Advanced Semiconductor Manufacturing Conference) Proceedings*, vol. 2019-May, pp. 224–228, 2019. DOI: 10.1109/ASMC.2019.8791750.

[11] C. Liu, B. Dai, and C. Yeh, "Characterization of the Chemical-Mechanical Polishing Process Based on Nanoindentation Measurement of Dielectric Films," *Journal of The Electrochemical Society*, vol. 142, no. 9, pp. 3098–3104, 1995. DOI: 10.1149/1.2048695.

[12] H. Lee, D. Lee, and H. Jeong, "Mechanical aspects of the chemical mechanical polishing process: A review," *International Journal of Precision Engineering and Manufacturing*, vol. 17, no. 4, pp. 525–536, 2016. DOI: 10.1007/s12541-016-0066-0.

[13] D. Zhao and X. Lu, "Chemical mechanical polishing: Theory and experiment," *Friction*, vol. 1, no. 4, pp. 306–326, 2013. DOI: 10.1007/s40544-013-0035-x.

[14] D. Zhao, T. Wang, Y. He, and X. Lu, "Kinematic optimization for chemical mechanical polishing based on statistical analysis of particle trajectories," *IEEE Transactions on Semiconductor Manufacturing*, vol. 26, no. 4, pp. 556–563, 2013. DOI: 10.1109/TSM.2013.2281218.

[15] Y. Mu, Y. Zhuang, Y. Sampurno, X. Wei, T. Ashizawa, H. Morishima, and A. Philipossian, "Effect of pad groove width on slurry mean residence time and slurry utilization efficiency in CMP," *Microelectronic Engineering*, vol. 157, pp. 60–63, 2016. DOI: 10.1016/j.mee.2016.02.035.

[16] S. Lee, S. Jeong, K. Park, H. Kim, and H. Jeong, "Kinematical modeling of pad profile variation during conditioning in chemical mechanical polishing," *Japanese Journal of Applied Physics*, vol. 48, no. 12, 2009. DOI: 10.1143/JJAP.48.126502.

[17] S. Choi, F. M. Doyle, and D. A. Dornfeld, "Material Removal Mechanism during Copper Chemical Mechanical Planarization Based on Nano-Scale Material Behavior," *ECS Journal of Solid State Science and Technology*, vol. 6, no. 5, P235–P242, 2017. DOI: 10.1149/2.0071705jss.

[18] A. J. Khanna, P. Jawali, D. Redfield, R. Kakireddy, A. Chockalingam, D. Benvegnu, M. Yang, S. Rozo, J. Fung, M. Cornejo, I. Abramson, M. Yamamura, Z. Yuan, and R. Bajaj, "Methodology for pad conditioning sweep optimization for advanced nodes," *Microelectronic Engineering*, vol. 216, no. March, 2019. DOI: 10.1016/j.mee.2019.111101.

[19] H. Lee and S. Lee, "Investigation of pad wear in CMP with swing-arm conditioning and uniformity of material removal," *Precision Engineering*, vol. 49, pp. 85–91, 2017. DOI: 10.1016/j.precisioneng.2017.01.015.

[20] X. X. Ban, H. Y. Zhao, S. J. Zhao, R. Q. Xie, Y. W. Gu, X. L. Zhu, D. F. Liao, L. Li, and Z. D. Jiang, "Effect of geometry error on accuracy of large-diameter pads used for CMP dressing," *International Journal of Advanced Manufacturing Technology*, vol. 100, no. 5-8, pp. 1505–1520, 2019. DOI: 10.1007/s00170-018-2797-9.

[21] A. J. Khanna, V. R. Kakireddy, J. Fung, P. Jawali, M. Yamamura, N. Baradanahalli Kenchappa, V. Hariharan, D. Redfield, and R. Bajaj, "High-Performance Pad Conditioning (HPPC) Arm for Augmenting CMP Performance," *ECS Journal of Solid State Science and Technology*, vol. 9, no. 6, p. 064 012, 2020. DOI: 10.1149/2162-8777/aba9fd.

[22] A. J. Khanna, M. Yamamura, V. R. Kakireddy, A. Chockalingam, P. Jawali, N. Baradanahalli Kenchappa, V. Hariharan, D. Redfield, and R. Bajaj, "Investigation of the Impact of Pad Surface Texture from Different Pad Conditioners on the CMP Performance," *ECS Journal of Solid State Science and Technology*, vol. 9, no. 6, p. 064 011, 2020. DOI: 10.1149/2162-8777/aba726.

[23] A. J. Khanna, V. R. Kakireddy, J. Fung, M. Yamamura, P. Jawali, A. Chockalingam, N. Baradanahalli Kenchappa, D. Redfield, and R. Bajaj, "Engineering Surface Texture of Pads for Improving CMP Performance of Sub-10 nm Nodes," *ECS Journal of Solid State Science and Technology*, vol. 9, no. 10, p. 104 003, 2020. DOI: 10.1149/2162-8777/abbcb5.

[24] T. Matsunaga, M. Uneda, Y. Takahashi, K. Shibuya, Y. Nakamura, D. Ichikawa, and K. I. Ishikawa, "Influence into platen and polishing pad surface temperature on removal rate in sapphire-chemical mechanical polishing," in *2015 International Conference on Planarization/CMP Technology (ICPT)*, 2015, pp. 1–4, ISBN: 9781619565104.

[25] A. J. Khanna, R. Kakireddy, P. Jawali, A. Chockalingam, D. Redfield, R. Bajaj, J. Fung, M. Cornejo, M. Yamamura, Z. Yuan, C. Orilall, B. Fu, G. Ganapathi, F. C. Redeker, and N. B. Patibandla, "Impact of Pad Material Properties on CMP Performance for Sub-10nm Technologies," *ECS Journal of Solid State Science and Technology*, vol. 8, no. 5, P3063–P3068, 2019. DOI: 10.1149/2.0121905jss.

[26] N. Suzuki, A. Kato, M. Asaba, and E. Shamoto, "Primary study on nonlinear viscoelasticity measurement of chemical mechanical polishing pads by applying on-machine compression test," *Japanese Journal of Applied Physics*, vol. 53, no. 5 SPEC. ISSUE 2, 2014. DOI: 10.7567/JJAP.53.05GB01.

[27] M. Uneda, T. Omote, K. Shibuya, Y. Nakamura, D. Ichikawa, and K. I. Ishikawa, "Evaluation of in-plane microdeformation distribution characteristics of polishing pad surface," *Japanese Journal of Applied Physics*, vol. 52, no. 05FC01, 2013. DOI: 10.7567/JJAP.52.05FC01.

[28] Z. Li, E. Baisie, X. Zhang, and Q. Zhang, "Diamond disc pad conditioning in chemical mechanical polishing," *Advances in Chemical Mechanical Planarization (CMP)*, pp. 383–412, 2022. DOI: 10.1016/b978-0-12-821791-7.00014-9.

[29] K. Lee, J. Seo, and U. Paik, *Preparation and characterization of slurry for CMP*. 2022, pp. 323–354, ISBN: 9780128217917. DOI: 10.1016/b978-0-12-821791-7.00005-8.

[30] L. M. Cook, "CMP Consumables 11 : Pad," *Semiconductors*, vol. 63, 2000.

[31] L. Cook, *CMP pads and their performance*, Second Edi, d. Elsevier Ltd., 2022, pp. 567–590, ISBN: 9780128217917. DOI: 10.1016/b978-0-12-821791-7.00011-3.

[32] S. Li, G. Gaudet, F. Sun, and A. Naman, "ILD CMP with Silica Abrasive Particles: Interfacial Removal Kinetics and Effect of Pad Surface Textures," *Journal of The Electrochemical Society*, vol. 157, no. 11, H1061, 2010. DOI: 10.1149/1.3486806.

[33] R. Bajaj, D. Redfield, M. C. Orilall, B. Fu, A. J. Khanna, J. G. Fung, A. Chockalingam, M. Yamamura, V. R. R. Kakireddy, G. E. Menk, and N. B. Patibandla, "Polishing Pads Produced By An Additive Manufacturing Process," United States Patent 20210107116A1, 2021.

[34] A. J. Khanna, J. G. Fung, P. Jawali, R. Bajaj, A. W. Manzonie, N. B. Kenchappa, V. R. R. Kakireddy, J. An, J. Kim, and M. Yamamura, "Polishing Pads Having Selectively Arranged Porosity," United States Patent 20210187693A1, 2021.

[35] N. B. Kenchappa, R. Popuri, A. Chockkalingam, P. Jawali, S. Jayanath, D. Redfield, and R. Bajaj, "Soft Chemical Mechanical Polishing Pad for Oxide CMP Applications," *ECS Journal of Solid State Science and Technology*, vol. 10, no. 1, p. 014 008, 2021. DOI: 10.1149/2162-8777/abdc40.

[36] B. Bhushan, *Introduction to Tribology*, 2nd Editio. John Wiley & Sons, 2013, ISBN: 9781119944539.

[37] J. Seo, "A review on chemical and mechanical phenomena at the wafer interface during chemical mechanical planarization," *Journal of Materials Research*, vol. 36, no. 1, pp. 235–257, 2021. DOI: 10.1557/s43578-020-00060-x.

[38] A. Philipossian and S. Olsen, "Fundamental Tribological and Removal Rate Studies of Inter-Layer Dielectric Chemical Mechanical Planarization," *Japanese Journal of Applied Physics, Part 1: Regular Papers and Short Notes and Review Papers*, vol. 42, no. 10, pp. 6371–6379, 2003. DOI: 10.1143/JJAP.42.6371.

[39] R. Han, Y. Sampurno, S. Theng, F. Sudargho, Y. Zhuang, and A. Philipossian, "Application of the Stribeck+ Curve in Silicon Dioxide Chemical Mechanical Planarization," *ECS Journal of Solid State Science and Technology*, vol. 6, no. 4, P161–P164, 2017. DOI: 10.1149/2.0241704jss.

[40] M. Bahr, Y. Sampurno, R. Han, and A. Philipossian, "Improvements in Stribeck Curves for Copper and Tungsten Chemical Mechanical Planarization on Soft Pads," *ECS Journal of Solid State Science and Technology*, vol. 6, no. 5, P290–P295, 2017. DOI: 10.1149/2.0241705jss.

[41] G. Diaz, L. Peckler, Y. Sampurno, and A. Philipossian, "Communication—Tribology of Retaining Rings in Chemical Mechanical Planarization," *ECS Journal of Solid State Science and Technology*, vol. 7, no. 5, P266–P268, 2018. DOI: 10.1149/2.0211805jss.

[42] H. Liang and G. Helen Xu, "Lubricating behavior in chemical-mechanical polishing of copper," *Scripta Materialia*, vol. 46, no. 5, pp. 343–347, 2002. DOI: 10.1016/S1359-6462(01)01249-0.

[43] T. Kasai and B. Bhushan, "Physics and tribology of chemical mechanical planarization," *Journal of Physics Condensed Matter*, vol. 20, no. 22, 2008. DOI: 10.1088/0953-8984/20/22/225011.

[44] T. K. Doy, K. Seshimo, K. Suzuki, A. Philipossian, and M. Kinoshita, "Impact of Novel Pad Groove Designs on Removal Rate and Uniformity of Dielectric and Copper CMP," *Journal of The Electrochemical Society*, vol. 151, no. 3, G196, 2004. DOI: 10.1149/1.1646144.

[45] L. Borucki, Y. Sampurno, and A. Philipossian, "Pad-Wafer-Slurry Interface Information from Force Data," *ECS Journal of Solid State Science and Technology*, vol. 8, no. 5, P3133–P3144, 2019. DOI: 10.1149/2.0211905jss.

[46] D. Bozkaya and S. Müftü, "A Material Removal Model for CMP Based on the Contact Mechanics of Pad, Abrasives, and Wafer," *Journal of The Electrochemical Society*, vol. 156, no. 12, H890, 2009. DOI: 10.1149/1.3231691.

[47] S. Kim, N. Saka, and J.-H. Chun, "Pad Scratching in Chemical-Mechanical Polishing: The Effects of Mechanical and Tribological Properties," *ECS Journal of Solid State Science and Technology*, vol. 3, no. 5, P169–P178, 2014. DOI: 10.1149/2.027405jss.

[48] Y. Sampurno, F. Sudargho, Y. Zhuang, T. Ashizawa, H. Morishima, and A. Philipossian, "Effect of cerium oxide particle sizes in oxide chemical mechanical planarization," *Electrochemical and Solid-State Letters*, vol. 12, no. 6, pp. 191–194, 2009. DOI: 10.1149/1.3098401.

[49] S. Jeon, J. Lee, S. Hong, H. Seo, Y. Cho, P. Liu, K. Park, and T. Kim, "Investigation of abrasive behavior between pad asperity and oxide thin film in chemical mechanical planarization," *Materials Science in Semiconductor Processing*, vol. 138, p. 106 280, 2022. DOI: 10.1016/j.mssp.2021.106280.

[50] W. Fan and D. Boning, "Multiscale modeling of chemical mechanical planarization (CMP)," in *Advances in Chemical Mechanical Planarization (CMP)*, Second Edition, Elsevier Ltd., 2022, pp. 143–174, ISBN: 9780128217917. DOI: 10.1016/b978-0-12-821791-7.00019-8.

[51] S. Kim, N. Saka, and J. H. Chun, "The effect of pad-asperity curvature on material removal rate in chemical-mechanical polishing," *Procedia CIRP*, vol. 14, no. Ic, pp. 42–47, 2014. DOI: 10.1016/j.procir.2014.03.014.

[52] S. Kim, N. Saka, and J. H. Chun, "The role of pad topography in chemical-mechanical polishing," *IEEE Transactions on Semiconductor Manufacturing*, vol. 27, no. 3, 2014. DOI: 10.1109/TSM.2014.2335156.

[53] S. W. Du, Y. T. Li, J. J. Song, and H. P. Qi, "Wafer Non-Uniformity Analysis during Chemical Mechanical Polishing Based on Finite Element Method," *Advanced Materials Research*, vol. 160-162, pp. 1518–1523, 2010. DOI: 10.4028/www.scientific.net/AMR.160-162.1518.

[54] D. Zhao, T. Wang, Y. He, and X. Lu, "Effect of zone pressure on wafer bending and fluid lubrication behavior during multi-zone CMP process," *Microelectronic Engineering*, vol. 108, pp. 33–38, 2013. DOI: 10.1016/j.mee.2013.03.042.

[55] N. Suzuki, Y. Hashimoto, H. Yasuda, S. Yamaki, and Y. Mochizuki, "Prediction of polishing pressure distribution in CMP process with airbag type wafer carrier," *CIRP Annals - Manufacturing Technology*, vol. 66, no. 1, pp. 329–332, 2017. DOI: 10.1016/j.cirp.2017.04.088.

[56] K. L. Johnson, *Contact Mechanics*. Cambridge University Press, 1985. DOI: 10.1017/CBO9781139171731.

[57] K. Komvopoulos, "Elastic-plastic finite element analysis of indented layered media," *Journal of Tribology*, vol. 111, no. 3, pp. 430–439, 1989. DOI: 10.1115/1.3261943.

[58] B. Bhushan, "Contact mechanics of rough surfaces in tribology: Multiple asperity contact," *Tribology Letters*, vol. 4, no. 1, pp. 1–35, 1998. DOI: 10.1023/A:1019186601445.

[59] G. G. Adams, S. Müftü, and N. M. Azhar, "A Nano-Scale Multi-Asperity Contact and Friction Model," in *Proceedings of IMECE2002*, ASME, 2002, pp. 455–462. DOI: 10.1115/IMECE2002-39305.

[60] H. Ghaednia, X. Wang, S. Saha, Y. Xu, A. Sharma, and R. L. Jackson, "A Review of Elastic-Plastic Contact Mechanics," *Applied Mechanics Reviews*, vol. 69, no. 6, 2017. DOI: 10.1115/1.4038187.

[61] J. A. Greenwood and J. B. P. Williamson, "Contact of Nominally Flat Surfaces," *Proceedings of the Royal Society London A*, vol. 295, no. 1442, pp. 300–319, 1966. DOI: 10.1098/rspa.1966.0242.

[62] K. Komvopoulos and D. H. Choi, "Elastic finite element analysis of multi-asperity contacts," *Journal of Tribology*, vol. 114, no. 4, pp. 819–831, 1992. DOI: 10.1115/1.2920955.

[63] M. N. Webster and R. S. Sayles, "A numerical model for the contact of layered elastic bodies with real rough surfaces," *Journal of Tribology*, vol. 108, no. 3, pp. 314–320, 1986. DOI: 10.1115/1.3261185.

[64] S. J. Cole and R. S. Sayles, "A Numerical Model for the Contact of Layered Elastic Bodies With Real Rough Surfaces," *Journal of Tribology*, vol. 114, no. 2, pp. 334–340, 1992. DOI: 10.1115/1.2920892.

[65] T. Nogi and T. Kato, "Influence of a Hard Surface Layer on the Limit of Elastic Contact—Part I: Analysis Using a Real Surface Model," *Journal of Tribology*, vol. 119, no. 3, p. 493, 1997. DOI: 10.1115/1.2833525.

[66] W. Peng and B. Bhushan, "A Numerical Three-Dimensional Model for the Contact of Layered Elastic/Plastic Solids With Rough Surfaces by a Variational Principle," *Journal of Tribology*, vol. 123, no. 2, pp. 330–342, 2001. DOI: 10.1115/1.1308004.

[67] S. Cai and B. Bhushan, "A numerical three-dimensional contact model for rough, multilayered elastic/plastic solid surfaces," *Wear*, vol. 259, no. 7-12, pp. 1408–1423, 2005. DOI: 10.1016/j.wear.2005.02.014.

[68] W. W. Chen, S. Liu, and Q. J. Wang, "Fast Fourier Transform Based Numerical Methods for Elasto-Plastic Contacts of Nominally Flat Surfaces," *Journal of Applied Mechanics*, vol. 75, no. 1, p. 011 022, 2008. DOI: 10.1115/1.2755158.

[69] Z. Wang, W. Wang, Y. Hu, and H. Wang, "A Simplified Numerical Elastic-Plastic Contact Model for Rough Surfaces," in *Advanced Tribology*, Springer, Berlin, Heidelberg, 2009, pp. 159–166. DOI: 10.1007/978-3-642-03653-8_54.

[70] K. Komvopoulos and N. Ye, "Three-dimensional contact analysis of elastic-plastic layered media with fractal surface topographies," *Journal of Tribology*, vol. 123, no. 3, pp. 632–640, 2001. DOI: 10.1115/1.1327583.

[71] K. Komvopoulos and N. Ye, "Elastic-Plastic Finite Element Analysis for the Head-Disk Interface With Fractal Topography Description," *Journal of Tribology*, vol. 124, no. 4, p. 775, 2002. DOI: 10.1115/1.1467088.

[72] W. Yan and K. Komvopoulos, "Contact analysis of elastic-plastic fractal surfaces," *Journal of Applied Physics*, vol. 84, no. 7, pp. 3617–3624, 1998. DOI: 10.1063/1.368536.

[73] P. Prokopovich and V. Starov, "Adhesion models: From single to multiple asperity contacts," *Advances in Colloid and Interface Science*, vol. 168, no. 1-2, pp. 210–222, 2011. DOI: 10.1016/j.cis.2011.03.004.

[74] V. L. Popov, Q. Li, I. A. Lyashenko, and R. Pohrt, "Adhesion and friction in hard and soft contacts: theory and experiment," *Friction*, vol. 9, no. 6, pp. 1688–1706, 2021. DOI: 10.1007/s40544-020-0482-0.

[75] D. Huang, X. Yan, R. Larsson, and A. Almqvist, "Boundary element method for the elastic contact problem with hydrostatic load at the contact interface," *Applied Surface Science Advances*, vol. 6, no. October, p. 100 176, 2021. DOI: 10.1016/j.apsadv.2021.100176.

[76] Y. Tan, L. Zhang, L. Ma, and X. Lv, "Modeling and simulation of dynamical contact of asperities between flat rough surfaces," *Advances in Mechanical Engineering*, vol. 8, no. 10, pp. 1–12, 2016. DOI: 10.1177/1687814016673498.

[77] K. Zhang, G. Li, J. Z. Gong, and M. Zhang, "Normal contact stiffness of rough surfaces considering oblique asperity contact," *Advances in Mechanical Engineering*, vol. 11, no. 1, pp. 1–14, 2019. DOI: 10.1177/1687814018824471.

[78] K. Zhang, Z. Wei, and Y. Chen, "Contact analysis of measured surfaces based on modified interacting and coalescing Hertzian asperities (ICHA) model with oblique contact asperities," *Advances in Mechanical Engineering*, vol. 13, no. 7, pp. 1–16, 2021. DOI: 10.1177/16878140211033090.

[79] S. Li, Q. Yao, Q. Li, X. Q. Feng, and H. Gao, "Contact stiffness of regularly patterned multi-asperity interfaces," *Journal of the Mechanics and Physics of Solids*, vol. 111, pp. 277–289, 2018. DOI: 10.1016/j.jmps.2017.10.019.

[80] H. Taylor, Y. C. Lam, and D. Boning, "A computationally simple method for simulating the micro-embossing of thermoplastic layers," *Journal of Micromechanics and Microengineering*, vol. 19, no. 7, p. 075 007, 2009. DOI: 10.1088/0960-1317/19/7/075007.

[81] H. Taylor and R. O'Rorke, "A computational design framework for two-layered elastic stamps in nanoimprint lithography and microcontact printing," *Journal of Applied Physics*, vol. 125, no. 9, 2019. DOI: 10.1063/1.5081495.

[82] J. Cen and K. Komvopoulos, "On the mechanics of metal imprinting by nominally flat and patterned rigid surfaces," *International Journal of Solids and Structures*, vol. 206, pp. 426–435, 2020. DOI: 10.1016/j.ijsolstr.2020.10.003.

[83] B. Stine, D. Ouma, R. Divecha, D. Boning, J. Chung, D. Hetherington, I. Ali, and O. Nakagawa, "A closed-form analytic model for ILD thickness variation in CMP processes," in *Proc. CMP-MIC*, 1997, pp. 266–273.

[84] B. Stine, D. Ouma, R. Divecha, D. Boning, J. Chung, D. Hetherington, C. Harwoo, O. Nakagawa, and Soo-Young Oh, "Rapid characterization and modeling of pattern-dependent variation in chemical-mechanical polishing," *IEEE Transactions on Semiconductor Manufacturing*, vol. 11, no. 1, pp. 129–140, 1998. DOI: 10.1109/66.661292.

[85] D. O. Ouma, D. S. Boning, J. E. Chung, W. G. Easter, V. Saxena, S. Misra, and A. Crevasse, "Characterization and modeling of oxide chemical-mechanical polishing using planarization length and pattern density concepts," *IEEE Transactions on Semiconductor Manufacturing*, vol. 15, no. 2, pp. 232–244, 2002. DOI: 10.1109/66.999598.

[86] Y. Zhao and L. Chang, "A micro-contact and wear model for chemical–mechanical polishing of silicon wafers," *Wear*, vol. 252, no. 3-4, pp. 220–226, 2002. DOI: 10.1016/S0043-1648(01)00871-7.

[87] R. Chen, S. Li, Z. Wang, and X. Lu, "Mechanical model of single abrasive during chemical mechanical polishing: Molecular dynamics simulation," *Tribology International*, vol. 133, no. November 2018, pp. 40–46, 2019. DOI: 10.1016/j.triboint.2018.12.018.

[88] B. Vasilev, S. Bott, R. Rzehak, R. Liske, and J. W. Bartha, "A method for characterizing the pad surface texture and modeling its impact on the planarization in CMP," *Microelectronic Engineering*, vol. 104, pp. 48–57, 2013. DOI: 10.1016/j.mee.2012.10.007.

[89] Q. Xu and L. Chen, "A Feature-Scale Greenwood–Williamson Model for Metal Chemical Mechanical Planarization," *Journal of Electronic Materials*, vol. 42, no. 8, pp. 2630–2640, 2013. DOI: 10.1007/s11664-013-2601-4.

[90] J. J. Vlassak, "A model for chemical-mechanical polishing of a material surface based on contact mechanics," *Journal of the Mechanics and Physics of Solids*, vol. 52, no. 4, pp. 847–873, 2004. DOI: 10.1016/j.jmps.2003.07.007.

[91] J. Seok, C. P. Sukam, A. T. Kim, J. A. Tichy, and T. S. Cale, "Multiscale material removal modeling of chemical mechanical polishing," *Wear*, vol. 254, no. 3-4, pp. 307–320, 2003. DOI: 10.1016/S0043-1648(03)00022-X.

[92] D. Boning and W. Fan, "Characterization and Modeling of CMP Pad Asperity Response in CMP," *Mat. Res. Soc. Symp. Proc.*, vol. 1249, 2010. DOI: 10.1577/PROC-1249-E05-04.

[93] T. Laursen, I. Kim, J. Schlueter, and S. Runnels, "Modeling of Feature-Scale Planarization in STI CMP Using MESA," in *Materials Research*, vol. 671, 2001. DOI: 10.1557/PROC-671-M4.2.

[94] E. J. Terrell and C. F. Higgs, "A Particle-Augmented Mixed Lubrication Modeling Approach to Predicting Chemical Mechanical Polishing," *Journal of Tribology*, vol. 131, no. 1, p. 012 201, 2009. DOI: 10.1115/1.2991173.

[95] Y. C. Wang and T. S. Yang, "Modeling and calculation of slurry-chemistry effects on chemical - Mechanical planarization with a grooved pad," *Journal of Engineering Mathematics*, vol. 63, no. 1, pp. 33–50, 2009. DOI: 10.1007/s10665-008-9240-4.

[96] G. Srivastava and C. F. Higgs, "A fast, experimentally validated, particle augmented mixed lubrication framework to predict CMP," *Materials Research Society Symposium Proceedings*, vol. 1560, pp. 67–72, 2013. DOI: 10.1557/opl.2013.970.

[97] G. Srivastava, "A Macro-scale, Tribological Modeling Framework for Simulating Multiple Lubrication Regimes and Engineering Applications," PhD thesis, Carnegie Mellon University, 2015.

[98] D. Wang, S. Shao, C. Yan, W. Cai, and X. Zeng, "Feature-scale simulations of particulate slurry flows in chemical mechanical polishing by smoothed particle hydrodynamics," *Communications in Computational Physics*, vol. 16, no. 5, pp. 1389–1418, 2014. DOI: 10.4208/cicp.261213.030614a.

[99] A. T. Kim, J. Seok, J. A. Tichy, and T. S. Cale, "A Multiscale Elastohydrodynamic Contact Model for CMP," *Journal of The Electrochemical Society*, vol. 150, no. 9, G570–G576, 2003. DOI: 10.1149/1.1598215.

[100] K. S. Chen, H. M. Yeh, J. L. Yan, and Y. T. Chen, "Finite-element analysis on wafer-level CMP contact stress: Reinvestigated issues and the effects of selected process parameters," *International Journal of Advanced Manufacturing Technology*, vol. 42, no. 11-12, pp. 1118–1130, 2009. DOI: 10.1007/s00170-008-1672-5.

[101] T. Wang and X. Lu, "Numerical and experimental investigation on multi-zone chemical mechanical planarization," *Microelectronic Engineering*, vol. 88, no. 11, pp. 3327–3332, 2011. DOI: 10.1016/j.mee.2011.08.011.

[102] T. Wang, X. Lu, D. Zhao, and Y. He, "Contact stress non-uniformity of wafer surface for multi-zone chemical mechanical polishing process," *Science China Technological Sciences*, vol. 56, no. 8, pp. 1974–1979, 2013. DOI: 10.1007/s11431-013-5245-y.

[103] E. J. Terrell, R. B. Comes, and C. F. Higgs, "Analysis of feature-scale wear in chemical mechanical polishing: Modeling and experiments," *Tribology Letters*, vol. 37, no. 2, pp. 327–336, 2010. DOI: 10.1007/s11249-009-9524-5.

[104] C. Gray, "Detecting Pad-Wafer Contact in CMP Using Dual Emission Laser Induced Fluorescence," PhD thesis, Tufts University, 2008.

[105] C. Yu, Z. Wang, F. Sun, S. Lu, L. M. Keer, and Q. J. Wang, "A Deterministic Semi-Analytical Model for the Contact of a Wafer and a Rough Bi-Layer Pad in CMP," *ECS Journal of Solid State Science and Technology*, vol. 2, no. 9, P368–P374, 2013. DOI: 10.1149/2.017309jss.

[106] L. Borucki, Y. Sampurno, S. Theng, and A. Philipossian, "Transient copper removal rate phenomena with implications for polishing mechanisms," in *Advances in Chemical Mechanical Planarization (CMP)*, Second Edition, Elsevier Ltd., 2022, pp. 237–257, ISBN: 9780128217917. DOI: 10.1016/b978-0-12-821791-7.00007-1.

[107] E. Baisie, B. Lin, X. Zhang, and Z. Li, "Finite Element Analysis (FEA) of Pad Deformation Due to Diamond Disc Conditioning in Chemical Mechanical Polishing (CMP)," in *ECS Transactions*, vol. 34, 2011, pp. 633–638, ISBN: 9781607682356. DOI: 10.1149/1.3567650.

[108] E. Baisie, Z. C. Li, and X. H. Zhang, "Finite Element Analysis (FEA) of Pad Deformation Due to Diamond Disc Conditioning in Chemical Mechanical Polishing (CMP)," in *Proceedings of the ASME 2012 International Manufacturing Science and Engineering Conference collocated with the 40th North American Manufacturing Research Conference and in participation with the International Conference on Tribology Materials and Process*, Notre Dame, Indiana, USA, 2012, pp. 209–215. DOI: 10.1115/MSEC2012-7364.

[109] C. Feng, C. Yan, J. Tao, X. Zeng, and W. Cai, "A Contact-Mechanics-Based Model for General Rough Pads in Chemical Mechanical Polishing Processes," *Journal of The Electrochemical Society*, vol. 156, no. 7, H601, 2009. DOI: 10.1149/1.3133238.

[110] X. Xie, "Physical Understanding and Modeling of Chemical Mechanical Planarization in Dielectric Materials," PhD thesis, Massachusetts Institute of Technology, 2007, p. 268.

[111] W. Fan, "Advanced Modeling of Planarization Processes for Integrated Circuit Fabrication," PhD thesis, Massachusetts Institute of Technology, 2012, p. 225.

[112] W. Fan, D. Boning, L. Charns, H. Miyauchi, H. Tano, and S. Tsuji, "Study on Stiffness and Conditioning Effects of CMP Pad Based on Physical Die-Level CMP Model," *Journal of The Electrochemical Society*, vol. 157, no. 5, H526, 2010. DOI: 10.1149/1.3369963.

[113] J. Luo and D. A. Dornfeld, *Integrated Modeling of Chemical Mechanical Planarization for Sub-Micron IC Fabrication*. New York: Springer-Verlag Berlin Heidelberg, 2004, ISBN: 9783642061158. DOI: 10.1007/978-3-662-07928-7.

[114] L. J. Borucki, T. Sun, Y. Zhuang, D. Slutz, and A. Philipossian, "Pad Topography, Contact Area and Hydrodynamic Lubrication in Chemical-Mechanical Polishing," in *Materials Research Society Symposium Proceedings*, vol. 1157, 2009, E01–02. DOI: 10.1557/PROC-1144-LL16-04.

[115] T. Sun, Y. Zhuang, L. Borucki, and A. Philipossian, "Characterization of pad-wafer contact and surface topography in chemical mechanical planarization using laser confocal microscopy," *Japanese Journal of Applied Physics*, vol. 49, no. 6 PART 1, pp. 0 665 011–0 665 014, 2010. DOI: 10.1143/JJAP.49.06650.

[116] T. Sun, Y. Zhuang, L. Borucki, and A. Philipossian, "Optical and mechanical characterization of chemical mechanical planarization pad surfaces," *Japanese Journal of Applied Physics*, vol. 49, no. 4 PART 1, pp. 0 465 011–0 465 015, 2010. DOI: 10.1143/JJAP.49.046501.

[117] T. Sun, L. Borucki, Y. Zhuang, Y. Sampurno, F. Sudargho, X. Wei, S. Anjur, and A. Philipossian, "Investigating effect of conditioner aggressiveness on removal rate during interlayer dielectric chemical mechanical planarization through confocal microscopy and dual emission ultraviolet-enhanced fluorescence imaging," *Japanese Journal of Applied Physics*, vol. 49, no. 2 Part 1, 2010. DOI: 10.1143/JJAP.49.026501.

[118] Y. Jiao, Y. Zhuang, X. Liao, L. J. Borucki, A. Naman, and A. Philipossian, "Effect of Temperature on Pad Surface Contact Area in Chemical Mechanical Planarization," *ECS Solid State Letters*, vol. 1, no. 2, N13–N15, 2012. DOI: 10.1149/2.016202ssl.

[119] A. Isobe, M. Akaji, and S. Kurokawa, "Proposal of new polishing mechanism based on feret's diameter of contact area between polishing pad and wafer," *Japanese Journal of Applied Physics*, vol. 52, no. 12, 2013. DOI: 10.7567/JJAP.52.126503.

[120] C. L. Elmufdi and G. P. Muldowney, "A Novel Optical Technique to Measure Pad-Wafer Contact Area in Chemical Mechanical Planarization," in *Materials Research Society Symposium Proceedings*, vol. 914, 2006. DOI: 10.1557/PROC-0914-F12-06.

[121] C. L. Elmufdi and G. P. Muldowney, "The impact of diamond conditioning on surface contact in CMP pads," *Materials Research Society Symposium Proceedings*, vol. 991, 2007. DOI: 10.1557/proc-0991-c01-02.

[122] L. Wang, P. Zhou, Y. Yan, R. Kang, and D. Guo, "Physically-based modeling of pad-asperity scale chemical-mechanical synergy in chemical mechanical polishing," *Tribology International*, vol. 138, no. March, pp. 307–315, 2019. DOI: 10.1016/j.triboint.2019.05.046.

[123] M. Uneda, Y. Maeda, K.-i. Ishikawa, K. Ichikawa, T. Doi, T. Yamazaki, and H. Aida, "Relationships between Contact Image Analysis Results for Pad Surface Texture and Removal Rate in CMP," *Journal of The Electrochemical Society*, vol. 159, no. 2, H90–H95, 2011. DOI: 10.1149/2.008202jes.

[124] M. Uneda, Y. Maeda, K. Shibuya, Y. Nakamura, D. Ichikawa, K. Fujii, and K. I. Ishikawa, "Influence of pad surface asperity on removal rate in chemical mechanical polishing of large-diameter silicon wafer applied to substrate of GaN-based LEDs," *Sensors and Materials*, vol. 26, no. 6, pp. 435–445, 2014. DOI: 10.18494/sam.2014.975.

[125] B. Jiang and G. P. Muldowney, "Computational solid mechanics modeling of asperity deformation and pad-wafer contact in CMP," *Materials Research Society Symposium Proceedings*, vol. 991, pp. C01–05, 2007. DOI: 10.1557/proc-0991-c01-05.

[126] T. Wang, X. Lu, D. Zhao, Y. He, and J. Luo, "Optimization of design of experiment for chemical mechanical polishing of a 12-inch wafer," *Microelectronic Engineering*, vol. 112, pp. 5–9, 2013. DOI: 10.1016/j.mee.2013.05.010.

[127] R. B. Comes, E. J. Terrell, and C. F. Higgs, "Pad deflection-based model of chemical-Mechanical polishing for Use in CAD IC layout," *IEEE Transactions on Semiconductor Manufacturing*, vol. 23, no. 1, pp. 121–131, 2010. DOI: 10.1109/TSM.2009.2039182.

[128] R. Tian, D. F. Wong, and R. Boone, "Model-based dummy feature placement for oxide chemical-mechanical polishing manufacturability," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 20, no. 7, pp. 902–910, 2001. DOI: 10.1109/43.931037.

[129] B. Salazar, M. Yamamura, R. Kakireddy, S. Jayanth, A. Chockalingam, R. Bajaj, and H. Taylor, "Die-scale modeling of planarization efficiency using segmented CMP pads: analyzing the effects of asperity topography," in *International Conference on Planarization/CMP Technology*, Hsinchu, Taiwan, 2019.

[130] B. Salazar, R. Kakireddy, S. Jayanth, A. Chockalingam, R. Bajaj, and H. Taylor, "Computational Modeling of CMP Pads: Analyzing the Pad Geometry's Impact on Planarization Efficiency," *ECS Meeting Abstracts*, vol. MA2021-01, no. 20, p. 821, 2021. DOI: 10.1149/ma2021-0120821mtgabs.

[131] F. W. Preston, "The Theory and Design of Plate Glass Polishing Machines," *Journal of Glass Technology*, vol. 11, no. 44, pp. 214–256, 1927.

[132] W. Tseng and Y. Wang, "Re-examination of Pressure and Speed Dependences of Removal Rate during Chemical-Mechanical Polishing Processes," *Journal of The Electrochemical Society*, vol. 144, no. 2, pp. L15–L17, 1997. DOI: 10.1149/1.1837417.

[133] A. Maury, D. Ouma, D. Boning, and J. Chung, "A Modification To Preston's Equation and Impact on Pattern Density Effect Modeling," in *Conference for Advanced Metallization and Interconnect Systems for ULSl Applications*, San Diego, CA, 1997.

[134] H. Hocheng, H. Y. Tsai, and Y. T. Su, "Modeling and Experimental Analysis of the Material Removal Rate in the Chemical Mechanical Planarization of Dielectric Films and Bare Silicon Wafers," *Journal of The Electrochemical Society*, vol. 148, no. 10, G581, 2001. DOI: 10.1149/1.1401087.

[135] S. R. Runnels and L. M. Eyman, "Tribology Analysis of Chemical-Mechanical Polishing," *Journal of The Electrochemical Society*, vol. 141, no. 6, pp. 1698–1701, 1994. DOI: 10.1149/1.2054985.

[136] S. R. Runnels, "Feature-Scale Fluid-Based Erosion Modeling for Chemical-Mechanical Polishing," *Journal of The Electrochemical Society*, vol. 141, no. 7, p. 1900, 1994. DOI: 10.1149/1.2055024.

[137] D. Castillo-Mejia and S. Beaudoin, "A Locally Relevant Prestonian Model for Wafer Polishing," *Journal of The Electrochemical Society*, vol. 150, no. 2, G96, 2003. DOI: 10.1149/1.1532330.

[138] H. Lee and H. Jeong, "A wafer-scale material removal rate profile model for copper chemical mechanical planarization," *International Journal of Machine Tools and Manufacture*, vol. 51, no. 5, pp. 395–403, 2011. DOI: 10.1016/j.ijmachtools.2011.01.007.

[139] H. S. Lee, H. D. Jeong, and D. A. Dornfeld, "Semi-empirical material removal rate distribution model for SiO2 chemical mechanical polishing (CMP) processes," *Precision Engineering*, vol. 37, no. 2, pp. 483–490, 2013. DOI: 10.1016/j.precisioneng.2012.12.006.

[140] W. Wu, J. Owino, A. Al-Ostaz, and L. Cai, "Applying Periodic Boundary Conditions in Finite Element Analysis," in *SIMULIA Community Conference*, Providence, 2014, pp. 707–719.

[141] G. Maculotti, G. Genta, D. Quagliotti, M. Galetto, and H. N. Hansen, "Gaussian process regression-based detection and correction of disturbances in surface topography measurements," *Quality and Reliability Engineering International*, no. January, pp. 1–18, 2021. DOI: 10.1002/qre.2980.

[142] N. Y. Nguyen, Y. Tian, and Z. W. Zhong, "Modeling and simulation for the distribution of slurry particles in chemical mechanical polishing," *International Journal of Advanced Manufacturing Technology*, vol. 75, no. 1-4, pp. 97–106, 2014. DOI: 10.1007/s00170-014-6132-9.

Appendix A

Experimental methods

A.1 Pad designs

Three different pads were used in this work. Table A.1 provides descriptions about these three pads and the parameters used in the experiments.

Table A.1: Table of pads used in this work and the associated pad design and experimental parameters. The pad design variable names correspond with those shown in Table 2.1. The experimental data are courtesy of Applied Materials, Inc.

Parameter	**Pad A**	**Pad B**	**Pad C**
Segment design	Concentric rings	Concentric rings	Segmented spirals
E_1	2500 MPa	790 MPa	69.3 MPa
E_2	1 MPa	55 MPa	105 MPa
w_{segment}	0.34 mm	1.35 mm	1.35 mm
w_{groove}	0.34 mm	0.68 mm	0.68 mm
h_{segment}	1 mm	0.5 mm	0.5 mm
b_{subpad}	2 mm	2 mm	1 mm
t_{subpad}	2.5 mm	2.5 mm	1.5 mm
$\bar{P}$	13 789.5 Pa	13 789.5 Pa	6894.8 Pa
N_{pad}	93 rpm	93 rpm	93 rpm
N_{wafer}	87 rpm	87 rpm	87 rpm
Paused for metrology?	No	Yes	Yes

A.2 Wafer processing

Wafers were processed using standard lithographic techniques and a conformal deposition of oxide. Experiments were run on an Applied Materials Reflexion LK Prime using 300 mm

wafers, 760.2 mm diameter pads, and a ceria-based slurry. For the experiments shown in Chapter 3, a 3M^{TM} A82 conditioning disk with a 74 µm nominal diamond size was used. The wafer was swept across the pad such that the wafer center to pad center distance varied from 165.1 mm to 215.9 mm over a 3 second period. The applied pressures, platen rotational speeds, and wafer carrier rotational speeds are listed in Table A.1.

A.2.1 Mask Design

The 15.5 mm × 12 mm die contains 2 mm by 2 mm blocks of features of a given size and shape. The die design is tessellated across the wafer. Our analysis focuses on the features with line/space or checkerboard shapes, and with critical dimensions ranging from 10 µm to 2000 µm.

	A	B	C	D	E	F
7	50x50 µm checkerboard	50% 10 µm pitch L/S	50% 0.4 µm pitch L/S	50% 200 µm pitch L/S	50% 4 µm pitch L/S	50% 400 µm pitch L/S
6	50% 20 µm pitch L/S	50% 1.0 µm pitch L/S	50% 2 µm pitch L/S	50% 0.2 µm pitch L/S	40% staggered blocks (DRAM)	90% 1.0 µm pitch L/S
5	40% staggered blocks (DRAM)	30% staggered blocks (DRAM)	70% 1.0 µm pitch L/S	30% 1.0 µm pitch L/S	50% 1.0 µm pitch L/S	50% staggered blocks (DRAM)
4	70% staggered blocks (DRAM)	60% staggered blocks (DRAM)	50% 100 µm pitch L/S	10% 1.0 µm pitch L/S	50% 0.2 µm pitch L/S	20% staggered blocks (DRAM)
3	80% staggered blocks (DRAM)	50% 0.9 µm Line / vary Space	10% 0.9 µm Line / vary Space	90% 0.9 µm Line / vary Space	70% 0.9 µm Line / vary Space	30% 0.9 µm Line / vary Space
	40% staggered blocks (DRAM) - 0.5mm spacer					
2	500x500 µm active block 40% DRAM	500x500 µm trench block 40% DRAM	200x200 µm active block 40% DRAM	200x200 µm trench block 40% DRAM	200x200 µm checkerboard	100x100 µm checkerboard
	40% staggered blocks (DRAM) - 0.5mm spacer					
1	500x500 µm checkerboard	1000x1000 µm checkerboard	2000x2000 µm trench only	2000x2000 µm active only	1000x1000 µm trench block 40% DRAM	1000x1000 µm active block 40% DRAM
	40% staggered blocks (DRAM) - 0.5mm spacer					

Figure A.1: The test die pattern, which is tessellated across the wafer. The die contains features of various shapes, including line/space features, checkerboard features, and features similar to DRAM. For checkerboard features, the size of one square—and not the pitch—is provided. For line/space features, the feature pattern density and pitch are provided. Figure provided courtesy of Applied Materials, Inc.

A.3 Measuring pad surface topographies

Two different scanning systems were used to measure pad textures in this work: one system to scan small pad portions at high resolution and another system to scan entire pads at coarse resolution. Both systems were used on pad specimens that had been conditioned and used in polishing experiments (courtesy of Applied Materials, Inc.).

A Keyence VK4 was used to perform 3D laser scanning microscopy on used pad specimens. The specimens were punched out of a region of the pad located 190.5 mm away from the pad center. The profile scans had sizing sufficient to encapsulate between two and four pad segments across. The scanner's lateral resolution adjusts with its overall size, to keep memory requirements manageable. With a scan size of 1.36 mm × 1.36 mm, the scanner provided asperity profile data with a lateral resolution of 2.782 µm. The Keyence Analyzer software was used to perform tilt correction, which corrects for possible sloping of the pad sample.

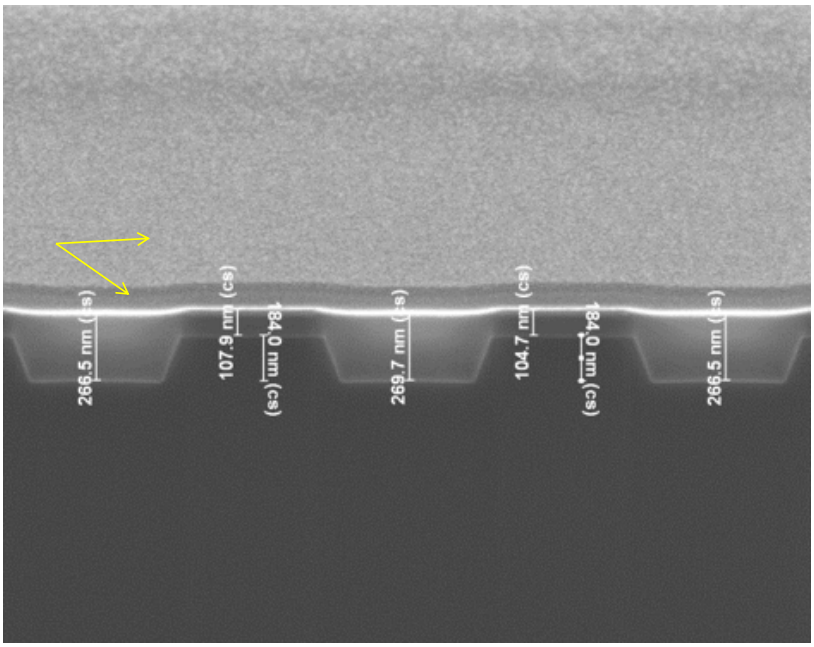
A Solarius white light interferometer was used to scan entire pads after removal from the polishing platen. The scanner was able to scan the entire 762 mm diameter pad with a 25 µm lateral resolution. The scanner's output was processed with an averaging filter to produce a pad surface texture with 50 µm lateral resolution for use in Chapter 4.

It is important to note that both the pad punch-outs and the entire pads may exhibit warpage, which may affect the scans and simulations. A more rigorous method for scanning pad topographies while keeping the pad flat is left as future work.

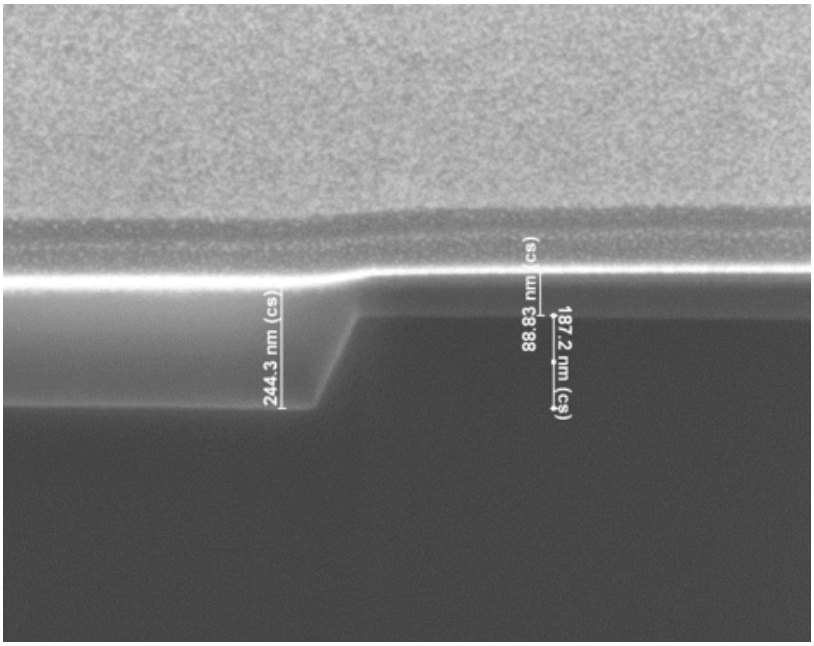
A.4 Measuring die oxide topographies

A Nova i550 ellipsometry tool was used to measure the oxide thicknesses for active and trench regions, at multiple features and die locations. For the experimental results discussed in Chapter 3, a KLA Tencor HRP-340 profilometer was used to measure the wafer topographies prior to polishing.

For the experimental results discussed in Chapter 4, SEM micrographs were used to extract the etched feature depths d_{etch}, the nitride film thickness t_{nitride} and the pad oxide thickness $t_{\substack{\text{pad}\\\text{oxide}}}$. It is well known that these dimensions may depend on the feature size and shape. Indeed, Figure A.2 shows SEM micrographs of 1 µm and 100 µm line-space features, and $t_{\text{nitride}} + t_{\substack{\text{pad}\\\text{oxide}}}$ is almost 20 nm (almost 20%) smaller for the 100 µm feature compared to the 1 µm feature. Despite this feature-scale non-uniformity, this work proceeds by assuming all features to have $d_{\text{etch}} + t_{\text{nitride}} + t_{\substack{\text{pad}\\\text{oxide}}} = 254\,\text{nm}$ (obtained by averaging across multiple feature measurements). This assumption is necessary because it was not possible to obtain SEM micrographs for all features, and it would be unfeasible to measure all wafers used in this study (to account for potential wafer-to-wafer variation). Future studies may investigate the simulation's sensitivity to this starting die topography assumption.



(a) 1 µm line-space



(b) 100 µm line-space

Figure A.2: SEM micrographs of STI features, with dimensions of the etched trench and nitride + pad oxide films shown. (a) shows a set of 1 µm line-space features and (b) shows a 100 µm line-space feature. The arrows point to coatings used for metrology purposes. Figure provided courtesy of Applied Materials, Inc.

The experimental results shown in Chapter 4 were obtained by sequentially polishing a wafer: a wafer was polished for a set amount of time, measured, polished for an additional amount of time, and the cycle repeated until the total number of measurement times was obtained. In contrast, the experimental results shown in Chapter 3 were obtained by polishing multiple wafers, each for a predetermined and different amount of time. The benefit to sequentially polishing a single wafer is that it requires fewer materials and there are no wafer-to-wafer variation effects. The drawback is that there may be startup effects that get compounded and affect the polish behavior (as discussed in Section A.4.2). All wafers were measured prior to polishing as well.

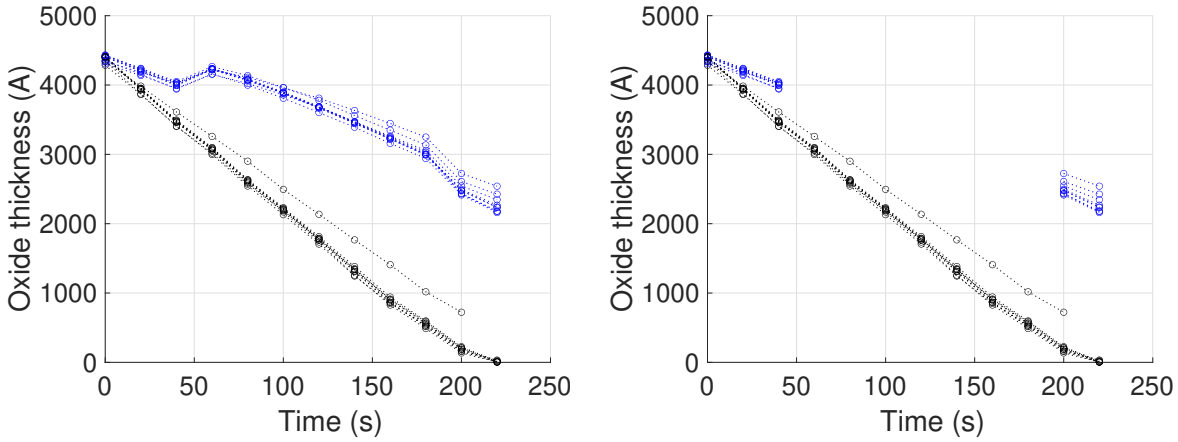
A.4.1 Messy data: detecting erroneous measurements

An important step when analyzing a dataset is to confirm that all data points are physically realistic. Visualization of the oxide thickness vs time curves revealed instances of unphysical measurement values, namely where the oxide thickness increased in comparison to a previous measurement time. These unphysical measurements were typically limited to only a few features and did occur at multiple die locations. The cause of these erroneous measurements is unknown, and those data points are removed from the dataset.

Figure A.3a shows an example where unphysical measurements were collected on a 2000 μ feature. In particular, this example plot shows a trench oxide profile with a vertical shift from 60 s to 180 s. A careful decision needs to be made: is it the measurements between 60 s and 180 s that are erroneous, or is it the remaining data points that are erroneous? To answer that question, the data points need to be compared with measurements from other features, and it is expected that (for a constant polish time) the trench oxide thickness for a larger feature should be smaller than for a smaller feature. Using this analysis, Figure A.3b shows the processed results without those erroneous measurements.

A.4.2 Variation in experimental measurements: thickness measurements depend on the number of time splits

Experiments suggest that pausing a polish for measurement purposes may affect the polish behavior. Figure A.4 shows a pair of experiments, where one experiment contained six measurement times (4 pauses) and the other experiment contained 3 measurement times (1 pause). The plot shows a significant difference in the trench oxide thickness measurements, despite the two experiments being performed under identical conditions. This suggests that there may be start-up effects that compound in the scenario with more measurement times. A possible mechanism for this is related to the pad temperature: when pausing the polish, the pad temperature is returned to room temperature and hence the pad does not soften as much. This means that the pad is not able to reach into trenches to the extent that it would if the pad temperature was allowed to build. A second possible explanation is related to slurry: it is well known that it may take a few seconds at the start of the polish for the slurry to equilibrate in the pad-wafer interface [142].



(a) Unprocessed (b) Processed

Figure A.3: Oxide thickness vs polish time plots for a 2000 µm feature. Black corresponds to active measurements and blue corresponds to trench measurements. The various lines correspond to replicate measurements performed on neighboring dies. (a) shows the raw measurements, including nonphysical positive slopes in the trench thickness curves. (b) shows the same measurements after processing to remove erroneous measurements.

It is important to recognize that experiments may have such start-up effects which the simulation does not currently capture. Additionally, when comparing simulation predictions to experimental results, it is important to recognize that the experimental results themselves may have large variations due to the measurement procedure.

A.5 Material characterization

Dogbone samples of the pad segment and subpad materials were fabricated and tested in uniaxial tension on an Instron machine. For the pad designs discussed in Chapter 3, the test results indicated the segment material and subpad material had an average tensile elastic modulus of 2500 MPa and 1 MPa, respectively. In finite-element simulations, I assume the high-modulus material to have Poisson's ratio of 0.3, and the low-modulus material to have Poisson's ratio of 0.45 (based on the Poisson's ratios of similar polymers). Nanoindentation tests on a Hysitron TriboIndenter TI 900 with a conical tip indicated the high modulus material to have a surface modulus of 3800 MPa. Tensile tests indicated the high modulus material to have a yield strength of $\sigma_Y = 58\,\text{MPa}$. Note that the uniaxial tests were performed under tensile conditions, whereas the pad experiences compressive loading, which is significant since the mechanical properties of polymers may differ when loaded under pure compression versus when loaded under pure tension. Furthermore, these tests were performed without the presence of slurry and at room temperature, which could further affect

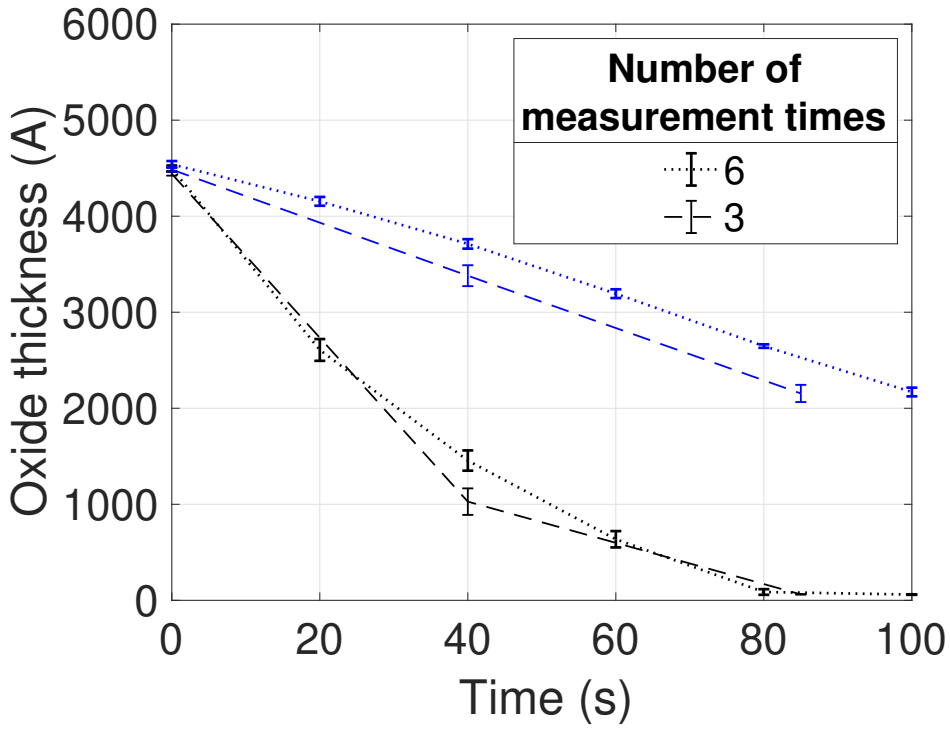


Figure A.4: Plot of the oxide thickness vs polish time, as measured for two experiments. Black symbols correspond to the active region and blue symbols correspond to the trench region. In the experiment that used six measurement times, the trench oxide thicknesses measurements are higher than in the experiment that used only three measurement times. Both experiments held all other parameters constant. Error bars correspond to 95% confidence intervals on the mean.

the mechanical properties. Despite those limitations, these experimental material characterizations are still helpful for estimating the simulation parameters, and those parameters may be further refined in simulation with parameter tuning.

A.6 Calculating planarization efficiency from experiments

Calculating the planarization efficiency using experimental measurement data collected at discrete times is a delicate task, since the planarization efficiency decreases over the polish time. Here, I provide a method for estimating the planarization efficiency using those discrete time measurements. This method assumes that the active and trench oxide thicknesses were measured at all measurement times, and that the starting feature step height was measured before polishing. This method is used in Chapter 3.

The goal here is to determine the planarization efficiency and step height for individual feature classes. The fact that measurements occur at multiple times means that the feature step heights will vary, and it is important to capture the planarization efficiency's dependence on the step height. For a particular feature, the active and trench oxide thickness measurements at a given measurement time t^i are h^i_{active} and h^i_{trench} and the initial step height is $h^{i=0}_{\text{step}}$. Here, $i = 0$ corresponds to the start of the polish, and thus $t^{i=0} = 0\,\text{s}$. The step height at a given measurement time is given by

$$h^i_{\text{step}} = h^{i=0}_{\text{step}} + \left(h^i_{\text{active}} - h^{i=0}_{\text{active}}\right) - \left(h^i_{\text{trench}} - h^{i=0}_{\text{trench}}\right) \tag{A.1}$$

Calculating the planarization efficiency requires knowledge of the material removal rate. Using a central difference method, the removal rates are calculated as

$$R^i_{\text{active}} = \frac{h^{i+1}_{\text{active}} - h^{i-1}_{\text{active}}}{t^{i+1} - t^{i-1}} \tag{A.2a}$$

$$R^i_{\text{trench}} = \frac{h^{i+1}_{\text{trench}} - h^{i-1}_{\text{trench}}}{t^{i+1} - t^{i-1}} \tag{A.2b}$$

Then, using Equation 1.1, the planarization efficiency at a given measurement time is given by

$$P.E.^i = 1 - \frac{R^i_{\text{trench}}}{R^i_{\text{active}}} = 1 - \frac{h^{i+1}_{\text{trench}} - h^{i-1}_{\text{trench}}}{h^{i+1}_{\text{active}} - h^{i-1}_{\text{active}}} \tag{A.3}$$

In both Equation A.2 and Equation A.3, due to the use of the central difference method, the number of measurement times at which these equations may be used is 2 fewer than the total number of measurement times, and hence we have $i \in \{1, 2, ..., q\}$, where q is the total number of measurement times.

It is important to place error bars on these experimentally determined values. I present a simple method for determining upper and lower bounds on the step height and planarization efficiency, and a more rigorous, statistical method could be used with more replicate measurements. This method proceeds by assuming that the metrology tools have an error of $\pm 50\,\text{Å}$ (based on previous experience with the tools). That error then propagates to the planarization efficiency as follows:

$$P.E.^i\Big]_{\substack{\text{Lower}\\\text{bound}}} = 1 - \frac{h^{i+1}_{\text{trench}} - h^{i-1}_{\text{trench}} - 100\,\text{Å}}{h^{i+1}_{\text{active}} - h^{i-1}_{\text{active}} + 100\,\text{Å}} \tag{A.4a}$$

$$P.E.^i\Big]_{\substack{\text{Upper}\\\text{bound}}} = 1 - \frac{h^{i+1}_{\text{trench}} - h^{i-1}_{\text{trench}} + 100\,\text{Å}}{h^{i+1}_{\text{active}} - h^{i-1}_{\text{active}} - 100\,\text{Å}} \tag{A.4b}$$

For the step height, each of the 5 measured quantities in Equation A.1 has a $\pm 50\,\text{Å}$ error, and hence the step height has a $\pm 250\,\text{Å}$ error. Note that the planarization efficiency must be between 0 and 1 (inclusive) and the step height must be nonnegative.