Advanced Modeling of Planarization Processes for Integrated Circuit Fabrication

by

Wei Fan

B.S., Peking University (2004)
M.S., Peking University (2007)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Electrical Engineering and Computer Science

at the
Massachusetts Institute of Technology

September 2012

Signature of Author: .....
Department of Electrical Engineering and Computer Science
August 6, 2012

Certified by: .....
Duane S. Boning
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by: .....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Advanced Modeling of Planarization Processes for Integrated Circuit Fabrication

by

Wei Fan

Submitted to the Department of Electrical Engineering and Computer Science
on August 6, 2012 in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

Planarization processes are a key enabling technology for continued performance and density improvements in integrated circuits (ICs). Dielectric material planarization is widely used in front-end-of-line (FEOL) processing for device isolation and in back-end-of-line (BEOL) processing for interconnection. This thesis studies the physical mechanisms and variations in the planarization using chemical mechanical polishing (CMP). The major achievement and contribution of this work is a systematic methodology to physically model and characterize the non-uniformities in the CMP process.

To characterize polishing mechanisms at different length scales, physical CMP models are developed in three levels: wafer-level, die-level and particle-level. The wafer-level model investigates the CMP tool effects on wafer-level pressure non-uniformity. The die-level model is developed to study chip-scale non-uniformity induced by layout pattern density dependence and CMP pad properties. The particle-level model focuses on the contact mechanism between pad asperities and the wafer. Two model integration approaches are proposed to connect wafer-level and particle-level models to the die-level model, so that CMP system impacts on die-level uniformity and feature size dependence are considered. The models are applied to characterize and simulate CMP processes by fitting polishing experiment data and extracting physical model parameters.

A series of physical measurement approaches are developed to characterize CMP pad properties and verify physical model assumptions. Pad asperity modulus and characteristic asperity height are measured by nanoindentation and microprofilometry, respectively. Pad aging effect is investigated by comparing physical measurement results at different pad usage stages. Results show that in-situ conditioning keeps pad surface properties consistent to perform polishing up to 16 hours, even in the face of substantial pad wear during extended polishing.

The CMP mechanisms identified from modeling and physical characterization are applied to explore an alternative polishing process, referred to as pad-in-a-bottle (PIB). A critical challenge related to applied pressure using pad-in-a-bottle polishing is predicted.

Thesis supervisor: Duane S. Boning
Title: Professor of Electrical Engineering and Computer Science

Acknowledgements

It has been five years since I first stepped on MIT campus. I cherish every day I experienced throughout my study at MIT. This work has been supported and guided by many people. I am truly grateful to all of them.

First, I must thank my advisor, Prof. Duane Boning, who has not only provided research support and guidance, but also advice and encouragement on my life and career. I have learned a lot of communication and collaboration skills from him. His passionate personality will inspire me throughout my life.

I would like to thank my thesis committee members, Prof. Jung-Hoon Chun and Prof. Judy Hoyt, for arranging meetings, and for their timely suggestions and feedback on my work.

This thesis consists of significant collaborations from both industry and academia. I am grateful to all my collaborators. Leslie Charns at JSR Micro in USA arranged the polishing experiments and provided the data for model fitting. Dr. Mansour Moinpour of Intel Corporation provided pad materials for my research. Dr. Yun Zhuang and Dr. Yasa Sampurno at the University of Arizona performed the pad aging experiment and prepared the pad samples for physical characterization. I would like to thank Prof. Ara Philipossian for many discussions on CMP, and it has been a great experience to collaborate with his group on CMP pad property study. I would also like to thank Dr. Len Borucki at Araca for sharing his experience and thoughts in CMP research.

My research has been constantly aided by the NanoLab and the Center for Materials Science and Engineering (CMSE) at MIT. I would like to thank Alan Schwartzman in the NanoLab for helping me perform nanoindentation on Hysitron Tribolndenter, and Tim McClure in the CMSE for helping me use the profilometer.

I am grateful to all the past and present members of the Boning group. I have built many strong friendships that will last forever. Specially, I would like to thank Joy Johnson and Hayden Taylor for their discussions and feedback during each process

subgroup meeting in the past five years. I would also like to thank Li Yu, John Lee and Albert Chang for sharing many interesting conversations about life.

In addition, I would like to thank Debb Hodges-Pabon, Mira Whiting and Grace Lindsay for their dedicated assistance.

A special thanks to my wife, Ying Wang, who has made my time in Boston so wonderful. Her love is always the best encouragement for my everyday life.

Finally I thank my parents for their support. I cherish every moment talking with them over the telephone and the internet. As their only child, I owe too much.

This work has been funded by SRC/SEMATECH Engineering Research Center for Environmentally Benign Semiconductor Manufacturing.

Contents

List of Figures

List of Tables

1 Introduction

This thesis describes a set of physical modeling and characterization approaches to understand the mechanisms and variations in the chemical mechanical polishing (CMP) process for integrated circuit (IC) fabrication. In this chapter, background of CMP application in IC fabrication is first discussed in Section 1.1. Section 1.2 then briefly reviews the CMP tool and material removal mechanism. The key challenges in CMP are summarized in Section 1.3. Section 1.4 briefly categorizes existing CMP models, and Section 1.5 points out the need for advanced modeling and characterization for CMP process. The contributions of this thesis and the thesis structure are introduced in Section 1.6 and Section 1.7 respectively.

1.1 CMP background

The chemical mechanical polishing (CMP) process is one of the key enabling technologies required to continue the shrinking of devices and interconnect structures in silicon integrated circuit fabrication [1]. It is widely used in the front end process for device isolation, in the back end process for interconnection, and in new process integration approaches for building advanced device structures [2].

CMP was originally introduced to semiconductor fabrication in silicon wafer manufacturing in the 1960s to achieve flatness of the silicon wafer surface [3]. It was first used within IC manufacturing lines to achieve improved transistor isolation [4]. An important and necessary application of the CMP process in IC fabrication came with building multilevel interconnections [5]. CMP was used to meet the planarization demands of both transistor and interconnect formation, and has been driven and developed by the industry to maintain silicon IC performance, density, and cost improvements through scaling down according to Moore's law [6]. Now CMP is widely used in both front end and back end processes.

1.1.1 CMP in inter-level dielectric planarization

For multilevel metal interconnects in very-large-scale integration (VLSI) technology, a key motivation for using CMP is to achieve planarity and to meet the stringent flatness requirements of the photolithography step. Although state-of-art photolithography tools are capable of refocusing after each exposure, extreme flatness with nanometer-scale height variations over a large chip area (about 20mm×20mm) are desired, depending on the depth of focus (DOF) of the optics system. As the feature size of silicon IC structures decreases, the inter-level dielectric (ILD) planarity requirement for CMP becomes even tighter [7].

Early CMP applications were largely driven by multilevel aluminum/oxide metallization schemes, together with tungsten vias and plugs. The SiO_2 inter-level dielectric planarization process flow is illustrated in Figure 1.1. The metal layer is first deposited; then the layer is patterned and etched to form desired structures; silicon oxide is deposited using conformal Chemical Vapor Deposition (CVD); lastly, the oxide layer is planarized using CMP. The global flatness allows an accurate photolithography step at each metal and via layer, and prevents topography from accumulating in the multilevel metal structures.

While copper has replaced aluminum in advanced IC fabrication, inter-level dielectric (ILD) CMP remains important in achieving planarity between the transistor formation and fabrication of multilevel copper interconnect. Furthermore, the basic mechanisms involved in single material CMP of SiO_2 are key to understanding planarization of other patterned IC structures.

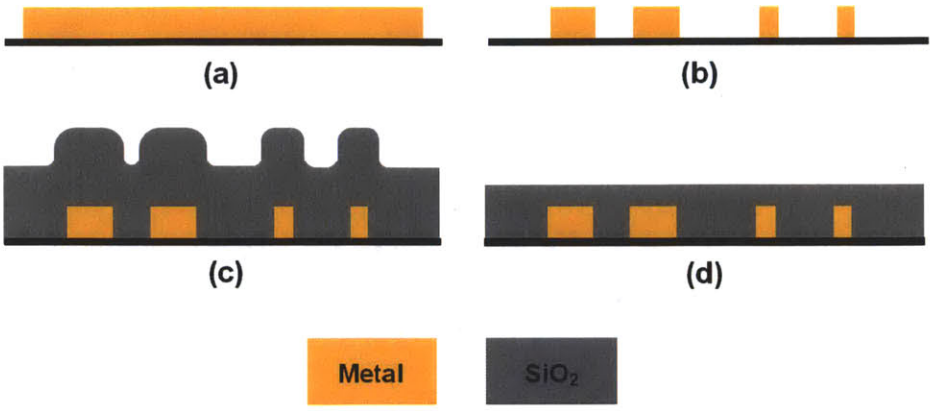


Figure 1.1: Inter-level dielectric planarization process: (a) metal layer deposition, (b) metal pattern and etch, (c) SiO2 dielectric layer deposition, (d) SiO2 dielectric layer CMP.

1.1.2 CMP in shallow trench isolation (STI)

Shallow trench isolation (STI) is the main isolation scheme for semiconductor manufacturing with active area pitches in the sub-0.25 μm regime [8]. STI is preferred because it has near zero field encroachment, good latch-up immunity, better planarity, and low junction capacitance. STI is also highly scalable, with the trench-fill capabilities being one major challenge to scaling. Figure 1.2 shows a typical STI process flow. First a thin pad SiO_2 layer and a blanket Si_3N_4 film are deposited on a flat silicon wafer. The isolation trenches are etched such that the desired trench depth (depth from silicon surface) is achieved (typical depth is 500 nm). Then a thick SiO_2 dielectric layer is deposited to fill the trenches. The CMP process is used to polish the overburden SiO_2 dielectric, down to the underlying nitride, where the nitride serves as a polishing stop layer. Whereas inter-level dielectric (ILD) planarization requires polishing of a single material, silicon dioxide, to achieve a desired flatness, STI CMP involves differential polishing of more than one material type. In addition to step-height reduction, selectivity of polishing to oxide and nitride layers is an important consideration in order to achieve desired planarity and the formation of in-laid isolation regions.

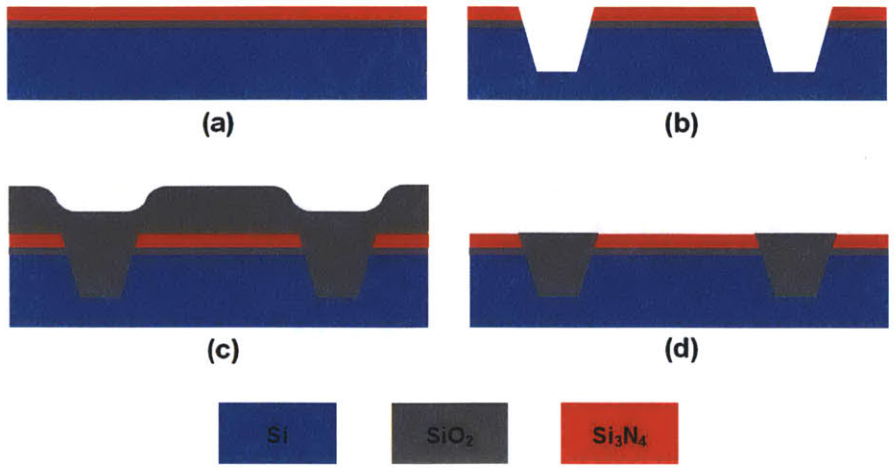


Figure 1.2: Shallow trench isolation process: (a) pad SiO2 and Si3N4 deposition, (b) shallow trench etch, (c) CVD SiO2 trench fill, (d) CMP planarization.

1.1.3 CMP in damascene process for multilevel copper interconnection

Multilevel copper interconnection is another critical element in advanced IC technologies [9]. CMP is the predominant fabrication technique because copper cannot be efficiently and cleanly dry-etched. There are several fabrication challenges in achieving high yield and economical copper wiring in key process steps including copper deposition, dielectric stack patterning, and planarization [10, 11]. Figure 1.3 illustrates a copper damascene process for interconnect. Trenches and vias are first patterned and etched within a low-k ILD stack. A thin layer of barrier material and copper seed layer are deposited. Then the copper wiring layer is electroplated from the seed. CMP is applied to remove excess copper and barrier material, leaving copper to form well-defined interconnect lines and vias.

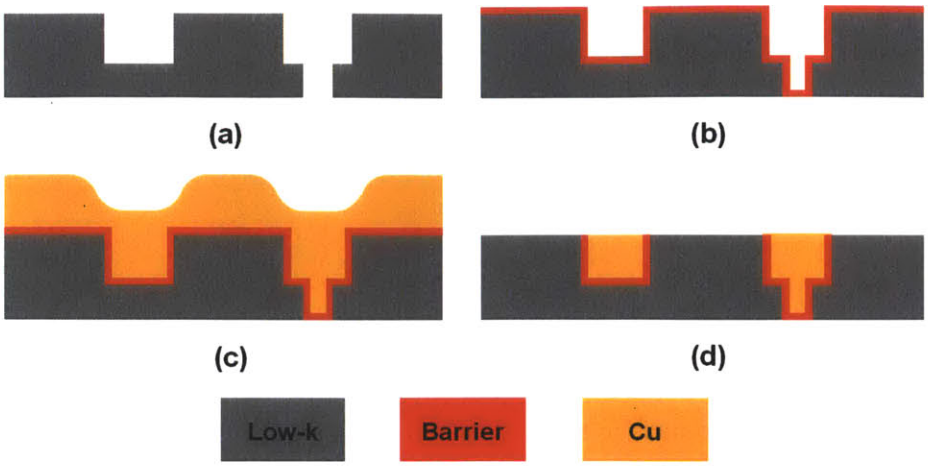


Figure 1.3: Damascene process of copper interconnection: (a) low-k dielectric trench and via etch; (b) barrier layer deposition; (c) copper interconnection layer plating; (d) CMP planarization.

1.1.4 Other applications of CMP

The three typical examples above demonstrate the ability of CMP to planarize the wafer surface and to build multilevel structures. The combination of planarization and damascene approaches enables CMP to be used in microelectronic manufacturing whenever a high degree of planarization is demanded, or inlaid material and structures are desired. As new materials and complex devices are introduced in IC fabrication, CMP finds many new applications, such as building advanced transistor structures, nonvolatile memories, silicon-on-insulator (SOI) processes [2], wafer bonding [12], and others. In addition to being used to planarize the IC devices, CMP has also been applied in fabrication of MEMS devices [13-16]. The ability to achieve planarity and form inlaid structures also makes CMP a critical step in making photonic crystals [17]. As the CMP process gains in popularity, stability, and reliability, more and more applications of this planarization technology will emerge. The challenges lie in the capability of CMP to handle new materials with a range of chemical and mechanical properties. An in-depth study of the physics in CMP will be helpful to guide further development and application of the technology.

1.2 CMP process

This section provides an overview of how CMP is implemented and how wafer surface material is removed.

1.2.1 CMP tool

A schematic view of a rotary CMP tool is shown in Figure 1.4. A wafer is held on a wafer carrier such that the surface to be polished faces a polishing pad, which is typically made from porous polyurethane, attached to a rotating platen. The wafer carrier is rotated in the same direction as the pad, while being pressed against the pad. Slurry composed of abrasive particles suspended in a chemical solution is delivered on the pad during polishing, and is transported to the pad-wafer interface by the pad.

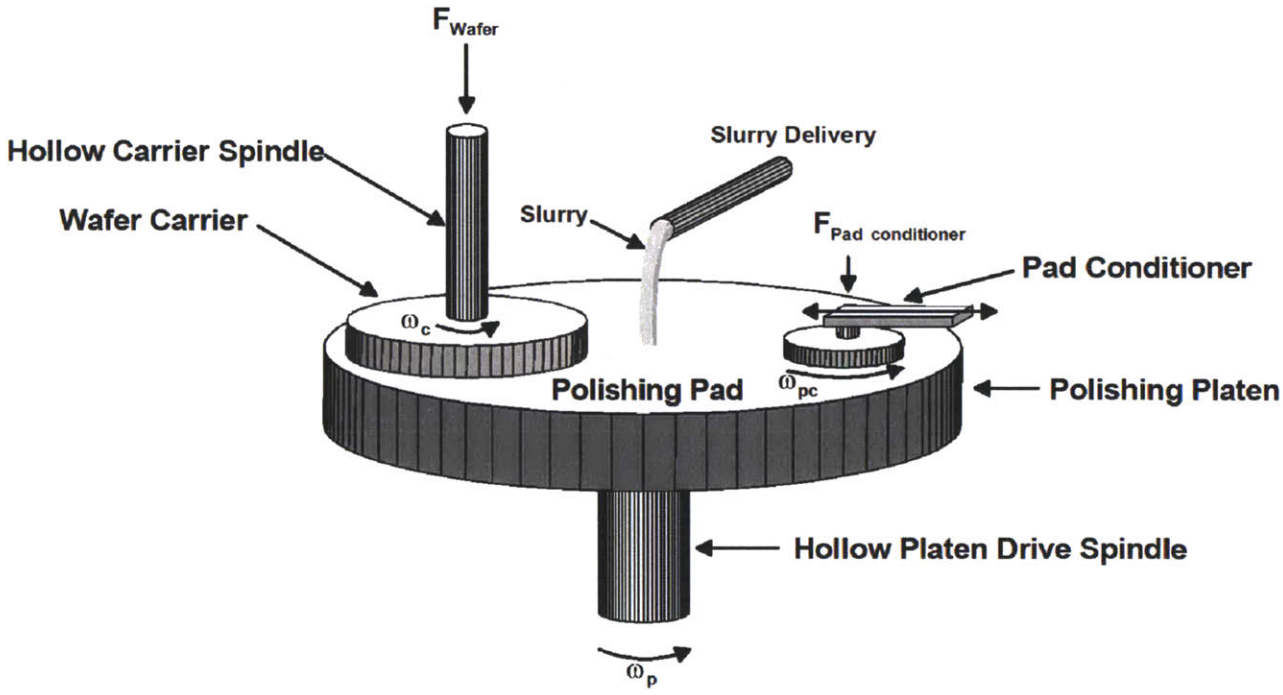


Figure 1.4: Schematic view of a rotary type CMP tool [18].

In addition to rotary CMP tools, there are also linear CMP tools that use a rotating wafer carrier contacting a CMP pad moving on a linear belt [19]. However, the

fundamental mechanism of removal (wafer surface contact with a CMP pad, aided by CMP slurry particles and chemistry) is similar.

One key criterion in designing a CMP polishing system is a uniform and consistent polishing spatially and in time, and this is achieved via a collection of subsystems [18]:

a) A **mechanical drive system** is able to control the relative surface velocity at the target speed. Sometimes the relative velocity, however, is intentionally set to vary across the wafer to compensate for other wafer level non-uniformity.

b) A **down force system** controls pressure distribution across the wafer. One approach is to divide the wafer area into a few co-center zones and apply different pressure on different zones.

c) A **thermal management system** is used to provide a stable and uniform temperature distribution during CMP. Temperature affects chemical reactions and has significant impact on oxide polishing as well as metal polishing [20, 21]. A spatial temperature variation causes non-uniformity removal rate, and an unstable temperature during CMP can result in over-polishing or under-polishing.

d) A **pad conditioning system** regenerates/dresses the polishing pad surface to a working condition via either in-situ (during polishing) or ex-situ (between polishing) approaches. Conditioning keeps the pad surface in a stable functioning state to ensure that the CMP process produces consistent performance.

e) A **slurry delivery system** tries to distribute slurry to the wafer-pad surface evenly and efficiently. Low slurry flow rate may cause lack of slurry in some regions, which can result in slow polishing rate, or even surface scratching due to the lack of lubrication. On the other hand, a high slurry flow rate will increase the cost significantly.

1.2.2 Material removal mechanism in CMP

CMP is a process that combines chemical reactions and mechanical forces in a synergistic way to remove surface materials and achieve desired planarity. It can be treated as chemically aided mechanical polishing, and material removal is believed to be primarily due to a three-body contact, as illustrated in Figure 1.5. First the wafer surface is modified and softened by the chemical solution, and the soft surface layer is removed by abrasive particles grabbed by the polishing pad. Without chemical modification, the wafer surface is too hard to be polished at appreciable rates; while without mechanical polishing, chemical modification and dissolution of the surface stops on its own (or does not activate planarization if it does proceed).

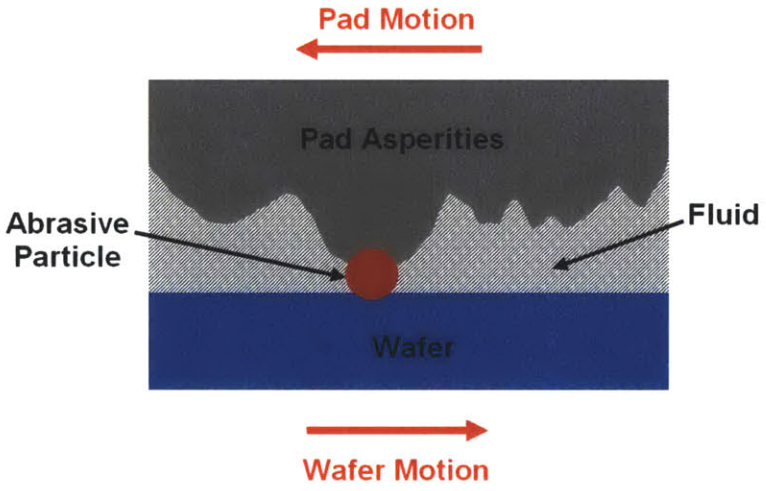


Figure 1.5: CMP material removal due to three-body contact.

In this simple picture, four components are involved: wafer surface, chemical solution, abrasive particles, and polishing pad.

a) **Wafer surface** is the object being polished, which can be a single material, such as silicon oxide or copper, or a mix of several materials, such as silicon oxide and silicon nitride in the later stages of the STI process.

b) **Chemical solution** is one of the main components of slurry. The chemical solution mixes and transports abrasive particles to the surface and carries wastes away

from the surface. Another function of the chemical solution is to modify and soften the wafer surface, thus it usually has a high pH value for polishing dielectrics and a low pH value for polishing metals.

c) **Abrasive particles**, the other component of slurry, remove the softened surface materials. For dielectric polishing, the abrasives are typically made of silica or ceria; while for metal polishing, they are typically made of silica or aluminum. The size of abrasives ranges from 50 nm to a few hundred nanometers [22].

d) **Polishing pad** transports fresh slurry to the wafer surface and carries removed debris away. The pad is crucial for the mechanical part of CMP. When the pad grabs abrasive particles, the pad addresses higher pressure on raised wafer surface areas and thus produces a higher removal rate on raised features, which results in the planarization effect.

The material removal rate is often described by Preston's equation [23],

$$RR = K \cdot P \cdot V \tag{1.1}$$

where $\boldsymbol{RR}$ is removal rate, $\boldsymbol{K}$ is a constant called Preston's coefficient, $\boldsymbol{P}$ is applied pressure on wafer surface, and $\boldsymbol{V}$ is relative velocity of the point on the surface of wafer versus the pad. Preston's equation is an empirical law first discovered in glass polishing. For most of the experiment results obtained in practice, especially in dielectric CMP, Preston's law provides a reasonably good fit. Preston's law suggests a linear dependence of removal rate on pressure and relative velocity; these contribute most of the mechanical dependencies in the CMP process. The rest of the polishing rate contributions, mainly chemical, are lumped in the constant $\boldsymbol{K}$. Preston's equation explains partly the planarization ability of CMP. The raised areas on the wafer surface compress the polishing pad more than the recessed areas, and the resulting higher localized pressure contributes to differential removal rates that flatten the topography.

1.2.3 Schematic scales in CMP

The CMP tool setup occurs over a macroscopic scale, for example, across an entire 300 mm wafer. However, material removal in CMP happens at a microscopic scale or even at a nanometer scale involving, for example, 30 nm abrasive particles. There is a 10^7 scale range between the two ends of these length scales. This subsection looks at the polishing behavior step by step across these orders of magnitude, as shown in Figure 1.6.

a) **Tool scale** (~ 100 mm). The CMP system is set up in this range according to wafer size, typically 100 mm, 150 mm, 200 mm and 300 mm. Reference pressure and relative velocity are controlled in this level.

b) **Pad-wafer contact scale** (~ 1 mm). Polishing pad bending and surface texture affects local pressure on the wafer surface in this range. Pad asperity height distribution is an additional key factor affect contact area and localized pressures.

c) **Abrasive trapping scale** (~ 10 μm) is typically the contact area between the wafer and a single pad asperity. Slurry abrasive particles spread into the contact area and are trapped between the pad and wafer.

d) **3-body contact scale** (~ 100 nm) is about a single abrasive particle size. The three interactive bodies are the wafer, abrasive particle and pad asperity. Wafer and particles are more rigid than a pad asperity, so that the main deformation is within the asperity.

e) **Material removal scale** (~ 1nm). Chemical reactions are of interest at this scale. Both wafer surface and abrasive particles are chemically modified. Deformation and material removal occur on both wafer surface and abrasive particles.

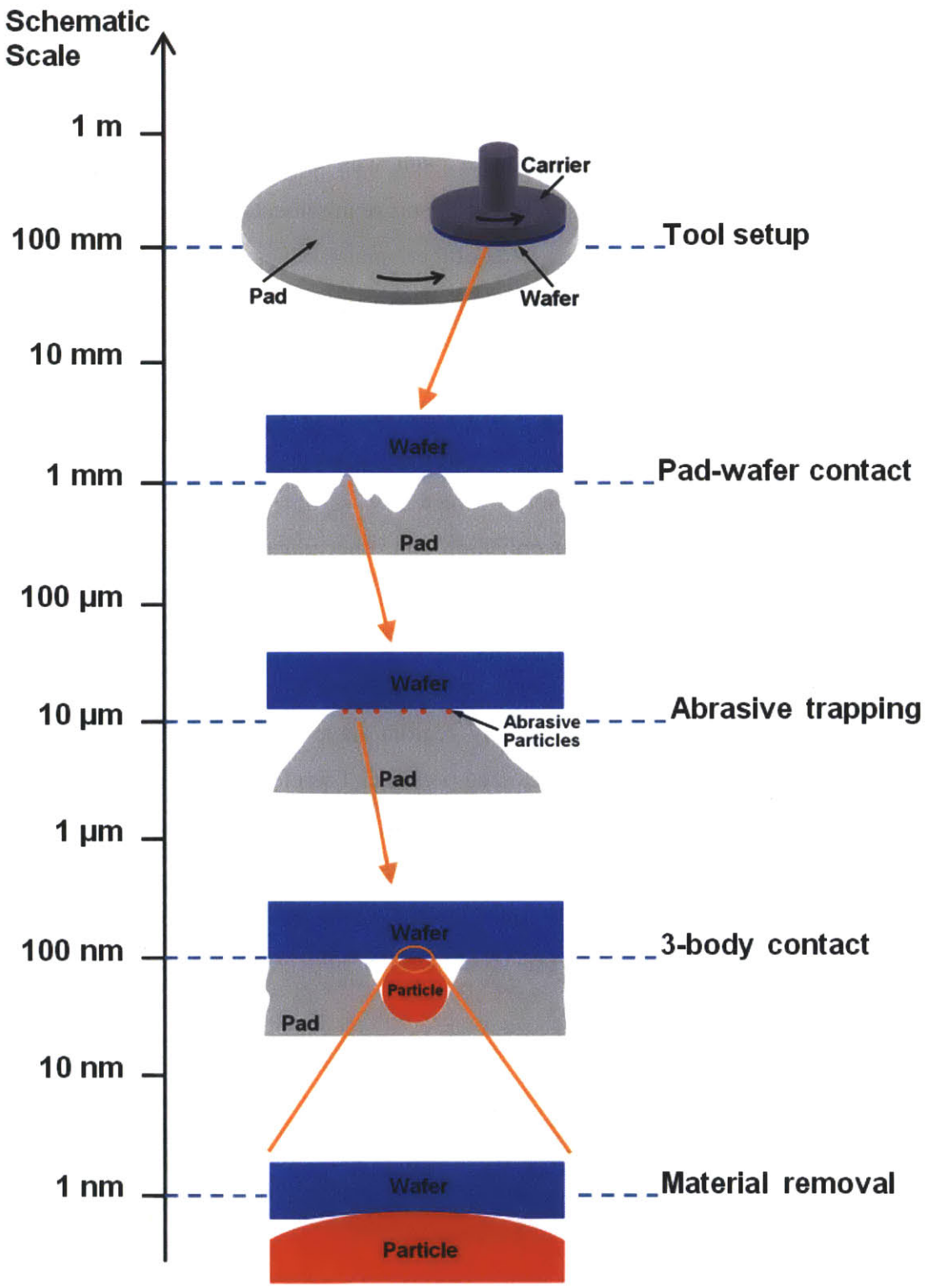


Figure 1.6: Schematic scales of interest in the CMP process.

1.3 The challenges of CMP

CMP faces tremendous challenges in current and future IC technology nodes. Many of the challenges are not new; they have been with CMP since the beginning, but continued scaling drives increasingly stringent requirements with respect to these challenges. This section summarizes some of these major issues in the CMP process.

1.3.1 Challenge of within-die non-uniformity

Pattern dependence is one of the main long lasting issues in CMP. The key definitions related to pattern dependence are the within-die non-uniformity (WIDNU), dishing, and erosion, which are heavily affected by layout pattern density and feature size. In inter-level dielectric (ILD) polishing and other planarization applications, film thickness and step height variation control is very important for the following photolithography steps. Figure 1.7 illustrates the non-uniform polishing result induced by pattern density difference. After the conformal deposition process, step structures are formatted within ILD layer. High density regions have larger up area to contact the polishing pad, which results lower localized pressure. Therefore, high density regions are planarized slower than low density area. Thickness difference occurs between high density regions and low density regions, which is considered as the chip-scale global non-planarity. CMP engineers have to understand and characterize pattern dependence properly so that they can modify the layout design rules and optimize the process to improve WIDNU.

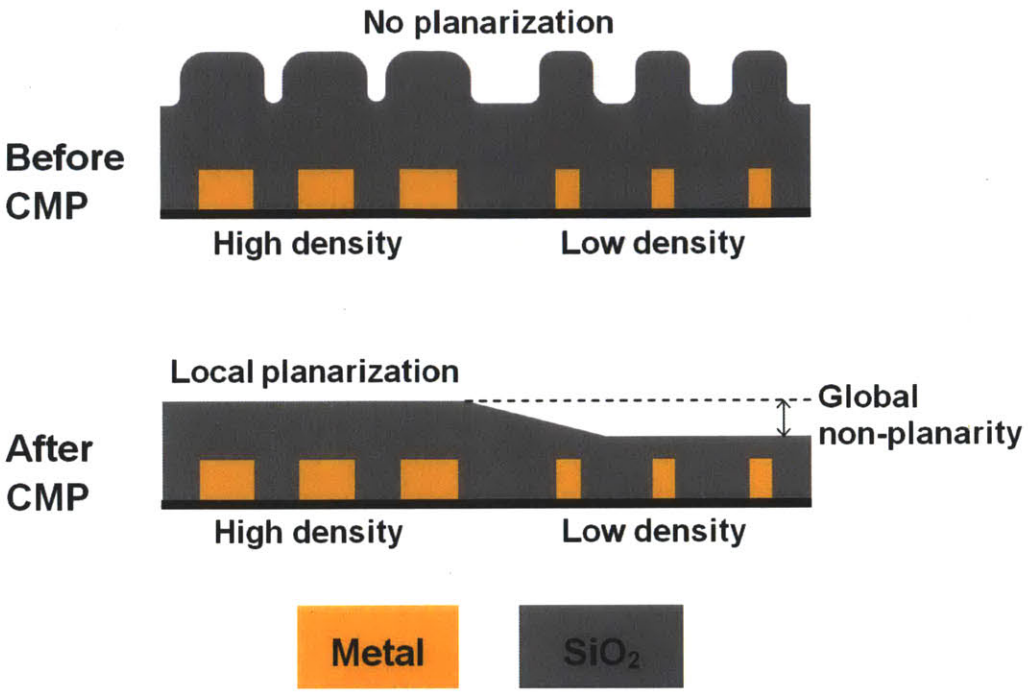


Figure 1.7: Pattern dependence in ILD CMP.

Dishing and erosion are two pattern dependent concerns in both STI CMP [24] and copper CMP [25, 26]. In STI CMP, dishing refers to the oxide loss relative to the level of the neighboring nitride space, and erosion refers to the nitride loss relative to the nitride level of the neighboring area. In copper CMP, dishing is defined as the copper loss relative to the level of the neighboring dielectric space, and erosion is the dielectric loss relative to the dielectric level of the neighboring area. Figure 1.8 shows dishing and erosion in copper CMP. Wide trenches or open structures usually have significant dishing, while fine trenches cause more erosion. Dishing of wide trenches or open structures is often considered to be a critical and insidious problem due to depth of focus issues in lithography from significant non-planarity, as well as electrical performance and yield impact of non-uniform thickness and topography.

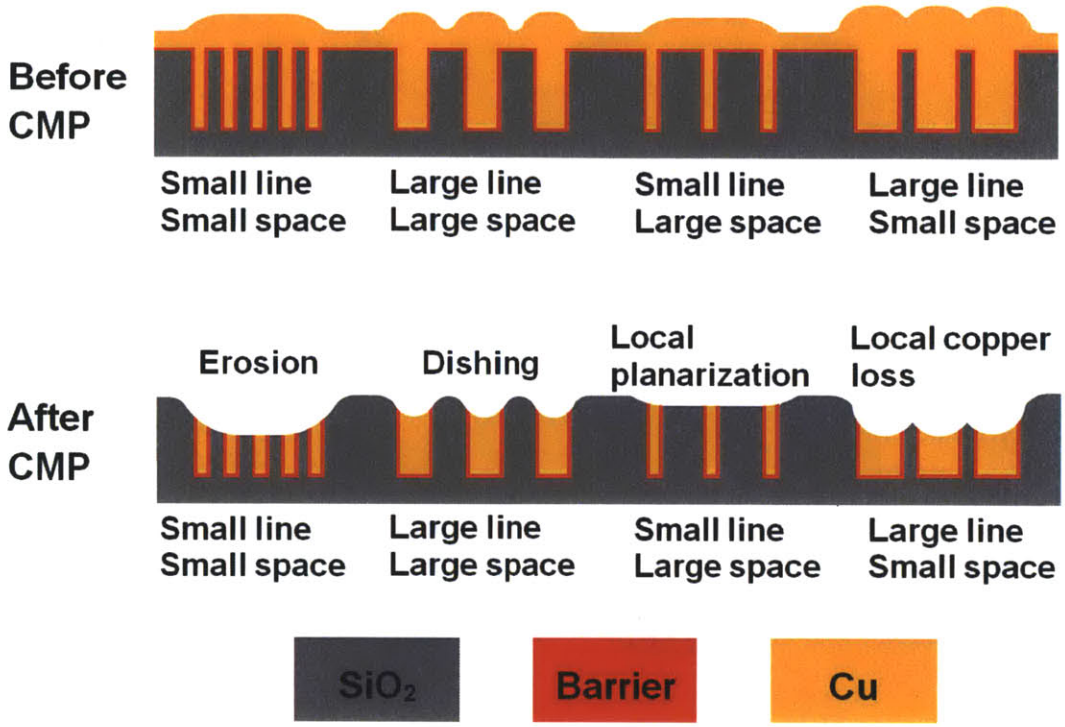


Figure 1.8: Pattern dependence in copper CMP.

1.3.2 Challenge of within-wafer non-uniformity

Another critical challenge in CMP is within-wafer non-uniformity (WIWNU) of material removal, which reduces the yield of the process [27]. The WIWNU also brings a systematic variation of the chip performance across the whole wafer. A proper understanding of the drives of WIWNU is important for the CMP process, in order to increase yield and improve chip quality. The WIWNU can be caused by non-uniform distributions of velocity and pressure from the polishing tool; consumable effects, such as polishing pad material, pad topography and slurry abrasive size distribution, and slurry flow non-uniformity, can also become sources of wafer level variation.

1.3.3 Other challenges

Within-die non-uniformity (WIDNU) and within-wafer non-uniformity (WIWNU) as discussed above are only two examples of CMP problems. There are other continuing

and growing challenges in CMP, such as defect rate, consumable cost, waste disposal, tool maintenance and environmental issues. CMP research is a wide area, and understanding of critical CMP issues is important for continued improvement of the process. Today's CMP is a jigsaw puzzle shown in Figure 1.9. Each new material process or new process integration approach usually requires a new CMP process, or at least solving a CMP jigsaw puzzle using known recipes. Every single piece in the puzzle may be a challenge under specific conditions in future development of IC processes. In Section 1.6, we identify the parts of this puzzle that this thesis seeks to address.

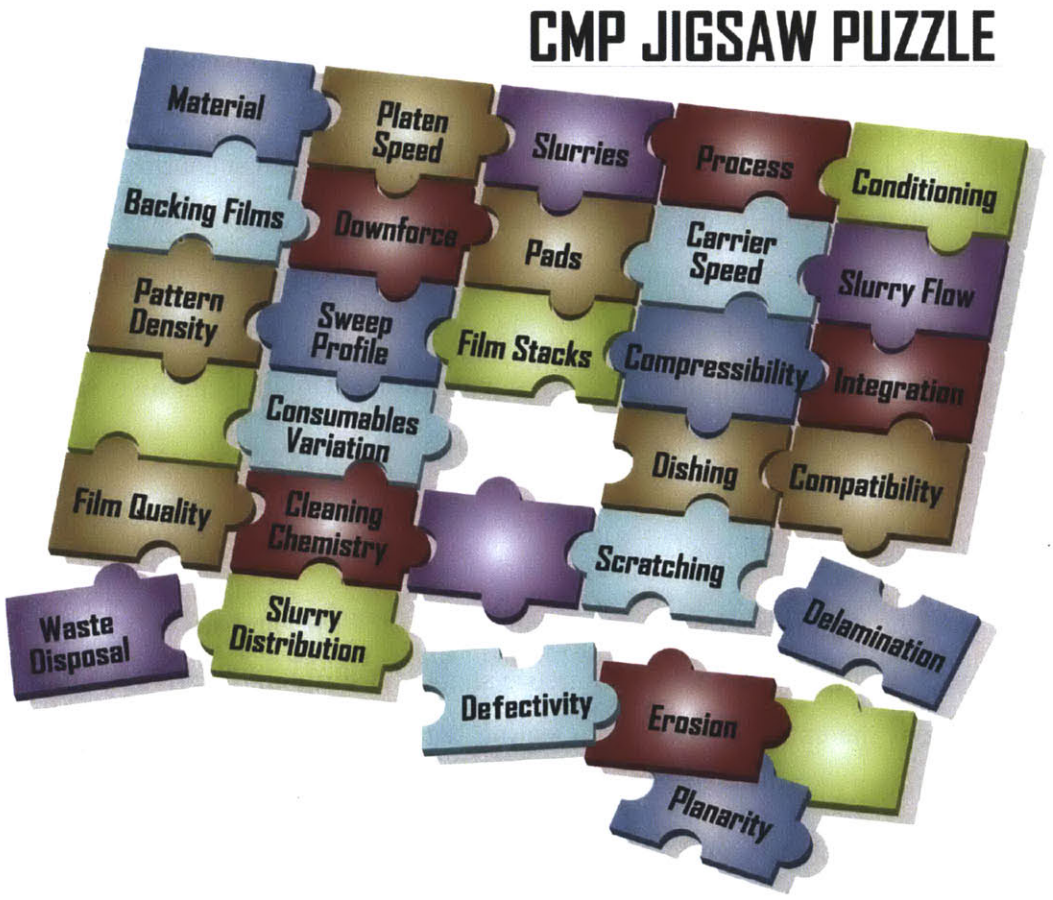


Figure 1.9: CMP jigsaw puzzle [28]: challenge is everywhere.

1.4 CMP models

Various CMP models are proposed in the literature to help understanding and improving the CMP process. According to the scale and usage in practice, they can be categorized into one of the three levels: particle-level, die-level and wafer-level.

1.4.1 Particle-level CMP models

Particle-level models seek to understand the material polishing mechanism of CMP and find the dependence of output variables, such as removal rate and surface quality [29, 30], on various input variables, such as applied pressure, chemical pH, abrasive size, and other consumable or process parameters. Physical understanding of the basic CMP mechanism enables better design and control of the process [31].

In particle-level models, CMP is usually studied in an ideal scenario: blanket wafer with single material, uniform chemical concentration, uniform abrasive size, etc. [32, 33]. Physical understanding can be approached empirically by isolating a few input and output variables and analyzing their dependence [34, 35], or theoretically by deriving models from basic physical assumptions [36, 37]. A particle-level model can also serve as the foundation to build die-level and wafer-level models.

1.4.2 Die-level CMP models

Die-level models focus on the planarization part of CMP, and study the polishing of one or more structures with known geometric shapes, at different locations within a chip or die, or across the entire die [38-40]. Dealing with a simple case, die-level models are able to focus in detail on how the existing structure features are planarizaed by modeling the transportation of CMP slurry chemicals and abrasives, pressure distribution, and/or other factors [41, 42]. In a real product die, however, there are millions to billions of individual structures; therefore modeling each of them separately is not feasible. Die-level models usually resort to statistical terms to describe and analyze the problem [18, 26, 43]. Die-level models often make assumptions such as uniform slurry flow across the

wafer, and benefit from the boundary condition that dies are arranged periodically on the wafer.

Die-level models help process engineers estimate process windows, identify potential weak spots of the polished chip, and modify CMP setup to improve the process. The models are particularly useful at the layout design stage. Chip designers can make their designs more fab-friendly with the feedback provided by these models, so that cost can be saved. This "Design for Manufacturing" (DFM) concept has been well adopted in the semiconductor manufacturing industry.

1.4.3 Wafer-level CMP models

Wafer-level models try to address the cases when the assumptions in die-level models fail due to tool limitations, such as non-uniform distribution of pressure [44, 45], slurry concentration [46, 47], and temperature [48]. Pressure distribution is highly non-uniform near the wafer edge, which results in a typical roll-off profile. Another cause of non-uniformity is that the dies near the wafer edge do not have some of their neighboring dies, resulting in different environments on the edge die. Slurry is a critical component of the CMP process; however, an even delivery of slurry across the wafer is difficult to achieve, which causes non-uniform slurry concentration [49]. Slurry transportation also has the effect of temperature cooling, and its variation can cause non-uniform temperatures across the wafer [50]. Wafer-level models help the tool manufacturers to design better polishing tools as well as help process engineers to better control the CMP process [51, 52].

1.5 The need for advanced modeling and characterization

As we have seen, many issues in the CMP process need to be studied and modeled quantitatively. One of the main motivations for modeling and characterization of CMP is to help selecting processing parameters. In the early days of CMP modeling, many empirical assumptions and parameters were made in the models. The empirical

parameters in these models are usually coupled with mixed process inputs, and it is not easy to separate and identify processing parameter effects based on empirical models. Another main motivation of modeling is to refine the designs on both the CMP user end (chip layout) and the CMP vendor end (CMP consumables). A successful CMP model should consider inputs from both ends and including parameters with physical meaning. Therefore, physically-based CMP models are eagerly desired.

CMP model development is generally focused on the particle-level, die-level or wafer-level. However, model simulation should not be limited to a single level. Multilevel model integration is needed to consider more effects, to understand the interaction between levels, and to make better polishing result predictions.

Physical characterization is required to verify model assumptions and test model reliability, especially for the models including consumable properties. A series of characterization approaches needs to be developed and applied.

1.6 Contributions of this thesis

This thesis contributes to both physical modeling of the CMP process, and to applications of these models. Physical measurement approaches are developed to characterize CMP pad properties and to verify model assumptions.

1.6.1 Developing physical CMP models

In this thesis, physical models are developed at three levels: wafer-level, die-level and particle-level. While the modeling approach is applicable to both dielectric and metal CMP, our work focuses on oxide polishing.

a) **Wafer-level model**. The within-wafer non-uniformity of the material removal rate has long been a concern in CMP, because wafer-level pressure distribution is non-uniform. We propose a physical wafer-level CMP model based on contact mechanics to address the pressure non-uniformity due to the polishing tool geometry, retaining ring

shape, and polishing pad properties. This model can be used to simulate blanket wafer polishing, or integrated with die-level models to simulate the implications on patterned wafer polishing.

b) **Die-level model**. In CMP process studies, an urgent need is to understand pattern density effects and to evaluate planarization performance at the chip scale. We adopt an explicit framework for die-level modeling of CMP, which abstracts the layout with of different pattern densities. Polishing performance is established by considering force responses from both the CMP pad bulk and from pad asperities. Pad properties are taken as model parameters. This model offers a potential characterization method for pad modulus and conditioning effect.

c) **Particle-level model**. A particle-level model is proposed to study the interaction between wafer surface and pad asperities using a Greenwood-Williamson approach [53]. Two main asperity properties are included in the model: asperity reduced modulus and asperity height distribution. Contact area percentage between the wafer and pad during CMP can be predicted. This model can be partly integrated into the physical die-level model to consider feature size effects.

The three single level models come together to help solve a simplified CMP jigsaw puzzle containing four key factors of the CMP process, as shown in Figure 1.10. The four factors of pad parameters, conditioning, applied pressure, and pattern dependency effects are attributed to the three levels based on the scale of interest and computational complexity, as listed in Table 1.1. Conditioning effects are addressed through impact on asperity height distribution in our models.

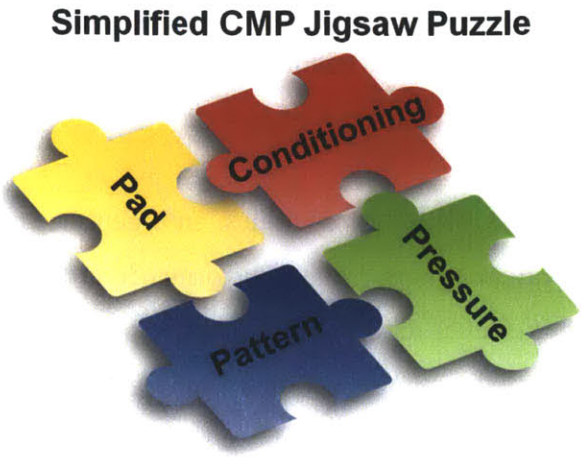


Figure 1.10: A simplified CMP jigsaw puzzle solved by the modeling work of this thesis.

Table 1.1: Polishing factors attributed to physical CMP models.

	Pattern density	Pressure	Pad modulus	Conditioning
Wafer-level	No	Yes	Yes	No
Die-level	Yes	Yes	Yes	Yes
Particle-level	No	Yes	Yes	Yes

Two extended die-level models are developed by integrating the three single level models in this thesis:

a) **Extended wafer-die-level model including wafer-level non-uniformity**. This model is derived by integrating the wafer-level model and the die-level model. Wafer-level pressure non-uniformity is implanted in the die-level model as the pressure boundary condition, so that CMP tool wafer-level impacts on die-level non-uniformity can be considered.

b) **Extended die-particle-level model including feature size effect**. This model is derived by integrating die-level model and particle-level model. Asperity shape is

considered when we calculate the contact pressure between asperity top and chip feature structure, so that asperity shape impact on feature size dependence of planarization is addressed.

1.6.2 Applying physical CMP models

Four engineering applications of the physical CMP models are demonstrated, together with our modeling methodology consisting of polishing experiments on standard testing wafers, model calibration (model parameter extraction), and model prediction (simulations with calibrated models).

a) **Model characterization of CMP pad properties**. Patterned oxide wafers are polished using CMP pads with intentionally modified bulk and surface properties. Physical die-level model is used to fit the polishing results and extract model parameters. CMP pad properties are related to model parameters. The effect of pad stiffness and conditioning disk diamond shape are investigated.

b) **CMP end-point strategy and within-die non-uniformity study**. Full chip model simulations are performed under two different end-point strategies: step height target strategy and up-area thickness target strategy. At each end-point, within-die non-uniformity is estimated.

c) **Evaluation of within-wafer non-uniformity impact on die-level planarization**. The extended wafer-die-level model is fit to patterned oxide wafer polishing results. Wafer-level pressure non-uniformity during polishing process is verified, and planarization results are compared across wafer center, middle and edge.

d) **Verification of feature size dependence in CMP**. The extended die-level model is fit to patterned wafer polishing results with a variety of both pattern densities and feature sizes. Feature size dependence is confirmed.

1.6.3 Physical characterization of CMP pad properties

In CMP, pad asperity modulus and asperity height distribution are two important properties that affect planarization. The coefficients of these two pad properties, asperity modulus and characteristic asperity height, are employed as CMP model parameters to understand polishing performance in this thesis. A physical CMP pad characterization approach based on nanoindentation and microprofilometry is demonstrated for pad surface property studies to verify underlying model assumptions.

Pad aging is an important factor in CMP, as typical processes suffer lot-to-lot, or even wafer-to-wafer, removal rate decay due to aging. A physical characterization approach is applied to evaluate pad aging effects. A pad aging experiment is run by polishing blanket oxide wafers up to 16 hours with in-situ conditioning. At different stages of this marathon test, physical characterization is performed at the same location on the pad and the measured results are compared. The measured asperity modulus and asperity height are applied in the particle-level CMP model to predict pad-wafer contact percentage.

1.6.4 Physical modeling of "pad-in-a-bottle"

The "pad-in-a-bottle" (PIB) approach to CMP is essentially a bottle of polymer beads which have similar chemical and mechanical properties as a polishing pad. The approach is hypothesized to be able to perform CMP by mixing in slurry to provide force response, so that no traditional pad is needed. Inspired by our particle-level model prediction and physical characterization, we know polishing is an accumulation of single material removal events, which we conjecture can be achieved by pad-in-a-bottle approach even in the absence of a traditional CMP pad. Therefore, we propose a simple physical model to study the behavior of the pad-in-a-bottle approach and estimate the relationship of applied pressure and material removal rate in this variant of CMP.

1.7 Thesis structure

The remainder of this thesis has the following structure. Chapter 2 develops basic physical CMP models at three levels: wafer-level, die-level and particle-level, and presents two approaches to model integration to extend the die-level model. Chapter 3 presents physical model applications including model characterization of CMP pad properties, CMP end-point prediction, evaluation of wafer-level non-uniformity impact on die-level planarization, and verification of feature size dependence of patterned wafer CMP. Chapter 4 demonstrates two physical measurement approaches to characterize CMP pad properties. Pad aging effects are evaluated based on these measurements. Chapter 5 proposes a physical modeling approach for CMP with "pad-in-a-bottle". Finally, Chapter 6 concludes and suggests area for future research.

2 Physical modeling of CMP

In this thesis, we believe there is a high priority CMP modeling need at die-level for CMP users in the IC industry [39, 40, 54-57]. The within-die non-uniformity (WIDNU) is a major concern for both layout designers and process engineers. A die-level CMP model seeks to simulate the planarization of layout structures on a chip and predict the polished chip surface profile. Such a capability would provide layout designers with a way of making optimized pattern arrangements to realize a uniform polishing result. On the other hand, a die-level model would help CMP process engineers to choose an appropriate set of process parameters (such as pressure, conditioning force, CMP pad hardness, etc.) for a given layout design. Furthermore, a good die-level model should be compatible with wafer-level models and particle-level models for extended applications. Therefore, our core modeling is focused on the die-level, while our wafer-level and particle-level models provide inputs for better die-level predictions.

In this chapter, we first take a top-down view to introduce our physical CMP models at the wafer-level, die-level and particle-level. Then we extend the die-level model to include wafer-level or particle level effects. For convenience in mathematical representation, this chapter takes the positive Z-axis to correspond to a surface normal up through the face of the wafer; this gives the appearance of a "wafer face up" convention in all model derivations although we note that in most CMP processes, the wafer face is pressed downward into the polishing pad.

2.1 Physical wafer-level CMP model

The within-wafer non-uniformity (WIWNU) of the material removal rate has long been a concern in CMP. Pressure distribution is known to be highly non-uniform across the wafer surface. Non-uniform pressure distribution may result from the inherent discontinuities of the process tool geometry at the wafer edge. Figure 2.1(a) shows the wafer carrier configuration of a typical rotary CMP tool. The wafer carrier holds the wafer facing down, which is polished against the polishing pad. Figure 2.1(b)

schematically shows the geometry near the wafer edge. The wafer is surrounded by a retaining ring, which is usually a few millimeters away from the edge of the wafer. In a typical setting, different pressures are applied to the wafer and the ring, with the ring usually under higher pressure to prevent the wafer from slipping out. The pad bends around the wafer edge due to the existence of the gap and retaining ring, thus the wafer edge is polished non-uniformly due to a localized pressure affected by the retaining ring. The factors affecting the non-uniform pressure include pad modulus, pad thickness, retaining ring size, and reference pressures on the wafer and retaining ring. To understand the details, a physical wafer-level CMP model is developed based on contact mechanics. Retaining ring shape and CMP pad thickness effects are captured in the model.

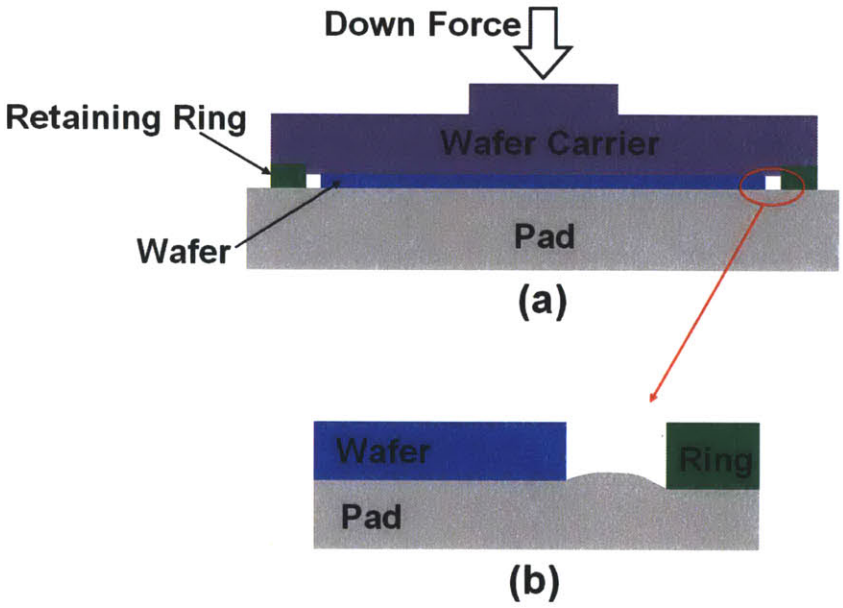


Figure 2.1: Wafer carrier configuration of CMP: (a) Wafer surrounded by retaining ring. (b) Pad deformation around wafer edge.

2.1.1 *Model derivation*

We first investigate the "thick pad" case: the applied pressure is low and the pad deflection is much smaller than the pad original thickness. The polishing pad thickness can be assumed to be infinite. We begin by adopting an analytical model based on the Boussinesq-Cerruti integral equations [58]. The roughness of contact surfaces is not considered. The CMP pad is assumed to be soft, while wafer and retaining ring are assumed to be rigid, so that only pad deflection needs to be calculated. The wafer and retaining ring sit are pressed into the pad with controllable applied reference pressures $P_{wafer,0}$ and $P_{ring,0}$, as shown in Figure 2.2. Here $z(x, y)$ is the profile of the rigid surfaces, both wafer and retaining ring. The deflected pad topography $w(x, y)$ and the contact pressure $p(x, y)$ are both defined as positive up into the pad material.

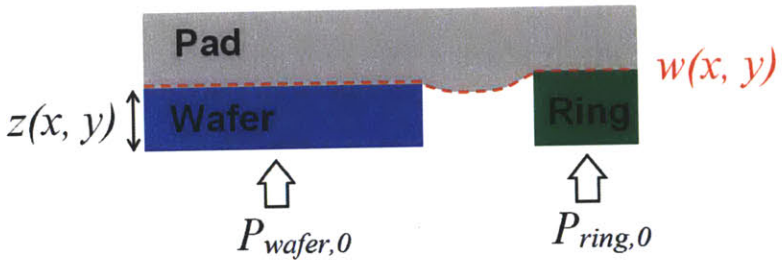


Figure 2.2: Pad surface deformation caused by applied pressure on wafer and retaining ring.

The pad is defined as a solid half-space from reference plane w_0, with Young's modulus E and Poisson's ratio v. Based on the Boussinesq integral equation, the pad deformation $u(x, y)=w(x, y)-w_0$ due to contact pressure can be expressed by,

$$u(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(x-x',y-y')\cdot p(x',y')dx'dy' \tag{2.1}$$

where G is the Green's function [59],

$$G(x,y)=\frac{1}{\pi E^{*}\sqrt{x^{2}+y^{2}}} \tag{2.2}$$

and E^* is the reduced modulus,

$$E^{*}=\frac{E}{1-\nu^{2}}\ . \tag{2.3}$$

This problem is subjected to boundary conditions,

$$\begin{cases} p(x,y)\geq 0 \\ w(x,y)\geq z(x,y) \\ \dfrac{1}{A_{wafer}}\iint\limits_{\substack{wafer\\surface}} p(x,y)dxdy = P_{wafer,0} \\ \dfrac{1}{A_{ring}}\iint\limits_{\substack{ring\\surface}} p(x,y)dxdy = P_{ring,0} \end{cases} \tag{2.4}$$

where A_{wafer} and A_{ring} are wafer area and retaining ring area, and $P_{wafer,0}$ and $P_{ring,0}$ are applied pressures on wafer and retaining ring.

Next we investigate "thin pad" case: the applied pressure is high and the pad deflection is affected by the underlying hard platen substrate. The pad is layered on a rigid platen, and under these conditions the pad thickness needs to be considered. The Green's function for finite thickness can only be written as a Taylor series expansion [60], and is difficult to use the Boussinesq integral method in this case. Therefore, we take a different approach to model the pad with finite thickness.

We solve this problem in the frequency domain using Fourier transforms of pressure and pad surface deflection,

$$\begin{cases} \widetilde{p}(\xi,\eta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-I(\xi x+\eta y)} p(x,y)dxdy \\ \widetilde{u}(\xi,\eta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-I(\xi x+\eta y)} u(x,y)dxdy \end{cases} \tag{2.5}$$

where $I = \sqrt{-1}$ is the imaginary unit, ξ and η denote the spatial angular frequencies corresponding to the x and y directions repectively. Using Papkovich-Neuber potentials [61], transformed influence coefficient $\widetilde{C_f}(\xi,\eta)$ can be found so that [62],

$$\widetilde{u}(\xi,\eta) = \widetilde{C}_f(\xi,\eta)\cdot\widetilde{p}(\xi,\eta) \tag{2.6}$$

Assuming the pad has finite thickness and the platen is a solid half-space, the transformed influence coefficient is expressed explicitly in the frequency domain as [63],

$$\widetilde{C}_f(\xi,\eta) = -\left(\frac{1-\nu_1}{\mu_1}\right)\left(1 + 4\alpha hke^{-2\alpha h} - \lambda ke^{-4\alpha h}\right)\alpha R \tag{2.7}$$

where

$$
\begin{cases}
u_1 = \dfrac{E_1}{2(1+v_1)} \\
u_2 = \dfrac{E_2}{2(1+v_2)} \\
\alpha = \sqrt{\xi^2 + \eta^2} \\
\lambda = 1 - \dfrac{4(1-v_1)}{1 + \dfrac{u_1}{u_2}(3-4v_2)} \\
k = \dfrac{\dfrac{u_1}{u_2} - 1}{\dfrac{u_1}{u_2} + (3-4v_1)} \\
R = -\dfrac{1}{[1-(\lambda + k + 4k\alpha^2 h^2)e^{-2\alpha h} + \lambda k e^{-4\alpha h}]\alpha^2}
\end{cases}
\tag{2.8}
$$

The subscripts ***1*** and ***2*** refer to the pad and the platen, respectively. The pad surface deflection is obtained by the inverse Fourier transform as follows,

$$
u(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{I(\xi x + \eta y)} \tilde{u}(\xi,\eta) d\xi d\eta \tag{2.9}
$$

The system of Eq. 2.5, 2.6 and 2.9 is also subject to boundary conditions of Eq. 2.4.

2.1.2 Computational approach

In order to perform a numerical analysis, the contact surfaces must be discretized. In the spatial domain, we mesh the wafer and pad surfaces on a $2N_x \times 2N_y$ grid of rectangular elements with pitch sizes Δx and Δy in x and y directions respectively, as shown in Figure 2.3(a). In the frequency domain, we use ξ and η to denote the frequencies in x and y directions, and a $2N_x \times 2N_y$ mesh grid of elements with pitch sizes $\frac{\pi}{N_x \Delta_x}$ and $\frac{\pi}{N_y \Delta_y}$ is used, as shown in Figure 2.3(b). We model contact pressures as being uniform within each element of this mesh grid, and take the displacement at the center of each element of the grid to represent that throughout the element. In this work,

the "element" means a region represented by a single number in a discretized map of pressure or topography. Let the pressure distribution and the surface displacements be denoted by $p(x_i, y_j)$ and $w(x_i, y_j)$ respectively, where $i = 0, 1, \cdots, (2N_x - 1)$ and $j = 0, 1, \cdots, (2N_y - 1)$. The discretized frequencies are denoted by ξ_m and η_n, where $m = 0, 1, \cdots, (2N_x - 1)$ and $n = 0, 1, \cdots, (2N_y - 1)$. The discretization can be expressed as

$$\begin{cases} x_i = i\Delta_x, & i = 0,1,\cdots,(2N_x - 1) \\ y_j = j\Delta_y, & j = 0,1,\cdots,(2N_y - 1) \end{cases} \tag{2.10}$$

in the spatial domain, and

$$\begin{cases} \xi_m = m\Delta_\xi, & m = 0,1,\cdots,(2N_x - 1) \\ \eta_n = n\Delta_\eta, & n = 0,1,\cdots,(2N_y - 1) \\ \Delta_\xi = \dfrac{\pi}{N_x \Delta_x} \\ \Delta_\eta = \dfrac{\pi}{N_y \Delta_y} \end{cases} \tag{2.11}$$

in the frequency domain.

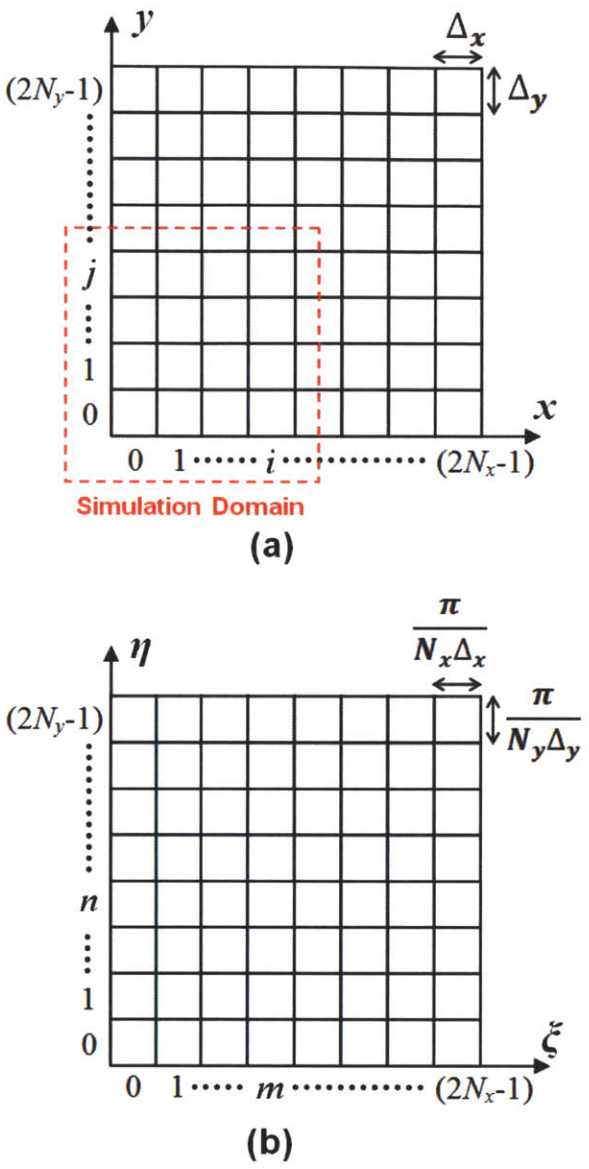


Figure 2.3: Discrete grid of the contact surface: (a) top view in space domain, (b) top view in frequency domain.

We enforce an assumption that the pressure distribution is periodic in space. Our assumption of periodicity makes it simple to use the discrete Fourier transform (DFT) as needed. Note that the pressure distribution and the surface displacements of the grid area $N_x \times N_y$ are considered to be our simulation domain, as indicated in Figure 2.3(a). However, the grid of $2N_x \times 2N_y$ is used for our calculation. This setup is used to protect

solutions from the false estimation by the DFT enforced periodicity. Outside the simulation domain, $p(x_i, y_j)$ is forced to zero in calculation [61].

We first calculate the "thick pad" case. The Green's function (Eq. 2.2) can be discretized by adopting the effect of a uniform unit pressure acting on a rectangular area analyzed by Love [64]. The discrete Green's function is expressed as

$$g(x_i, y_j) = \frac{1}{\pi E^*}\left[f(x_{i2}, y_{j2}) - f(x_{i1}, y_{j2}) - f(x_{i2}, y_{j1}) + f(x_{i1}, y_{j1})\right] \tag{2.12}$$

where

$$\begin{cases} f(x,y) = y\ln\left(x + \sqrt{x^2+y^2}\right) + x\ln\left(y + \sqrt{x^2+y^2}\right) \\ x_{i1} = i\Delta_x - \frac{1}{2}\Delta_x, \quad x_{i2} = i\Delta_x + \frac{1}{2}\Delta_x \\ y_{j1} = j\Delta_y - \frac{1}{2}\Delta_y, \quad y_{j2} = j\Delta_y + \frac{1}{2}\Delta_y \end{cases} \tag{2.13}$$

The discrete Boussinesq integral (Eq.2.1) is expressed as a discrete convolution,

$$u(x_i, y_j) = \sum_{i'=0}^{2N_x-1}\sum_{j'=0}^{2N_y-1} g(x_i - x_{i'}, y_j - y_{j'}) \cdot p(x_{i'}, y_{j'}) \tag{2.14}$$

The convolution can be computed efficiently using discrete Fourier transforms (DFT),

$$\begin{cases} u(x_i, y_j) = \Delta_\xi \Delta_\eta \sum_{m=0}^{2N_x-1} \sum_{n=0}^{2N_y-1} e^{I(\xi_m x_i + \eta_n y_j)} \tilde{u}(\xi_m, \eta_n) \\ \tilde{u}(\xi_m, \eta_n) = \tilde{g}(\xi_m, \eta_n) \cdot \tilde{p}(\xi_m, \eta_n) \\ \tilde{g}(\xi_m, \eta_n) = \Delta_x \Delta_y \sum_{i=0}^{2N_x-1} \sum_{j=0}^{2N_y-1} e^{-I(\xi_m x_i + \eta_n y_j)} g(x_i, y_j) \\ \tilde{p}(\xi_m, \eta_n) = \Delta_x \Delta_y \sum_{i=0}^{2N_x-1} \sum_{j=0}^{2N_y-1} e^{-I(\xi_m x_i + \eta_n y_j)} p(x_i, y_j) \end{cases} \tag{2.15}$$

We next calculate the "thin pad" case. The discrete transformed influence coefficient can be calculated directly using Eq. 2.7 with discrete frequencies as

$$\tilde{C}_f(\xi_m, \eta_n) = -\left(\frac{1-\nu_1}{\mu_1}\right)\left(1 + 4\alpha h k e^{-2\alpha h} - \lambda k e^{-4\alpha h}\right)\alpha R \tag{2.16}$$

The inverse Fourier transform Eq. 2.9 can be discretized as

$$\begin{cases} u(x_i, y_j) = \Delta_\xi \Delta_\eta \sum_{m=0}^{2N_x-1} \sum_{n=0}^{2N_y-1} e^{I(\xi_m x_i + \eta_n y_j)} \tilde{u}(\xi_m, \eta_n) \\ \tilde{u}(\xi_m, \eta_n) = \tilde{C}_f(\xi_m, \eta_n) \cdot \tilde{p}(\xi_m, \eta_n) \\ \tilde{p}(\xi_m, \eta_n) = \Delta_x \Delta_y \sum_{i=0}^{2N_x-1} \sum_{j=0}^{2N_y-1} e^{-I(\xi_m x_i + \eta_n y_j)} p(x_i, y_j) \end{cases} \tag{2.17}$$

In this calculation, the value of $\tilde{C}_f(0,0)$ is undefined. Gaussian quadrature integration over the elements near the origin can be used to compute the average influence coefficient [63], since $\tilde{C}_f(\xi_m, \eta_n)$ is singular but integrable around the origin. However, if we assume the pad material is incompressible (i.e. the Poisson ratio of pad material is not above 0.5), the value of $\tilde{C}_f(0,0)$ can be set to zero [65], with the result that the average value of the function in the spatial domain will be zero — consistent with conservation of the pad's volume.

How to choose "thick pad" or "thin pad" can be guided by comparing the kernel functions of these two cases, i.e., the displacement of a pad surface in response to unit pressure. As an example, the kernel function is evaluated and plotted in Figure 2.4 for four values of pad thickness. In the functions plotted here, the pad is assumed to have a Young's modulus of 100 MPa and a Poisson's ratio of 0.5, and each function is discretized at a 1 mm pitch in spatial domain and evaluated over a 400 mm by 400 mm region. These assumptions are relevant to the simulations reported later in this chapter. For h = 2 mm, h = 10 mm, and h = 50 mm, the functions are produced by Eq. 2.17; for elastic half-space, the function is evaluated by Eq. 2.15. As shown in Figure 2.4, the kernel function of h = 50 mm is close to the kernel function of half-space. We may assume when pad thickness is higher than 50 mm, the "thick pad" case could be applied. However, CMP pad thickness is usually a few millimeters. Thus, the "thin pad" assumption is more preferred in CMP.

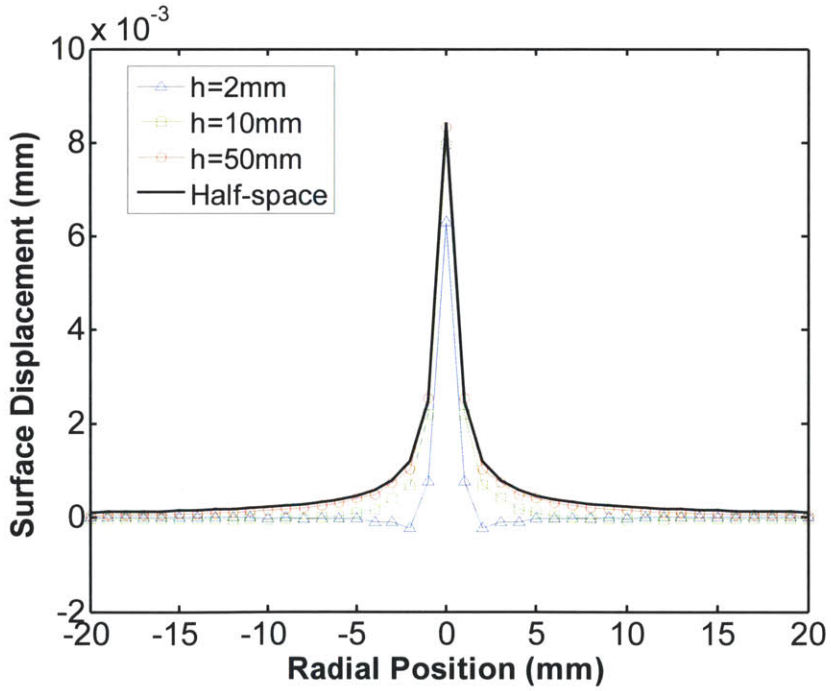


Figure 2.4: Comparison of kernel functions generated for different pad thicknesses.

Eq. 2.15 or Eq. 2.17 only gives us a relationship between pressure and displacement. To solve the problem subject to boundary conditions in Eq. 2.4, we need to determine the contact region Ω under applied pressures. The contact region Ω is defined such that

$p(x,y) > 0$ inside Ω, and $p(x,y) = 0$ outside Ω. The contact region Ω is not known in advance. However, it can be determined by an iterative method [62, 63, 65-67]. Nogi's [62] iterative method is adopted here and modified as illustrated in the flow chart of Figure 2.5. There are five important steps for the use of the present iterative computation method.

1) In the first iteration, we obtain an approximation of Ω from the geometric overlap region: we put $w(x,y)$ equal to $z(x,y)$ inside Ω and to w_0 outside Ω. Only in the first iteration, there is no need to solve Eq. 2.15 or Eq. 2.17 to obtain the pressure distribution since the displacements are known at all grid points.

2) After the first iteration, it will be found that the values of $p(x,y)$ near the periphery of the assumed contact region are negative, which is not permitted by the boundary condition of Eq. 2.4. These points are removed from the assumed contact region and the pressure distribution is solved from Eq. 2.15 or Eq. 2.17 by the biconjugate gradient stabilized method (BiCGSTAB) [68].

3) After the second iteration, the pressure distribution must be computed by solving Eq. 2.15 or Eq. 2.17 since the displacements are given only inside Ω. The displacements are then calculated everywhere to check that no contact could occur outside the assumed contact region. If so, these points are added to Ω, and $p(x,y)$ is solved again for the new assumed contact points. This process is repeated until boundary conditions $p(x,y) \geq 0$ and $w(x,y) \geq z(x,y)$ are satisfied.

4) If the wafer reference pressure boundary condition is not satisfied, the whole process from step 1 to step 3 must be repeated until the penetration reference plane w_0 is found such that the average wafer surface pressure is equal to the wafer reference pressure.

5) If the retaining ring reference pressure boundary condition is not satisfied, the whole sequence from step 1 to step 4 must be repeated until the proper rigid profile

$z(x, y)$ is found so that the average retaining ring pressure is equal to the retaining ring reference pressure.

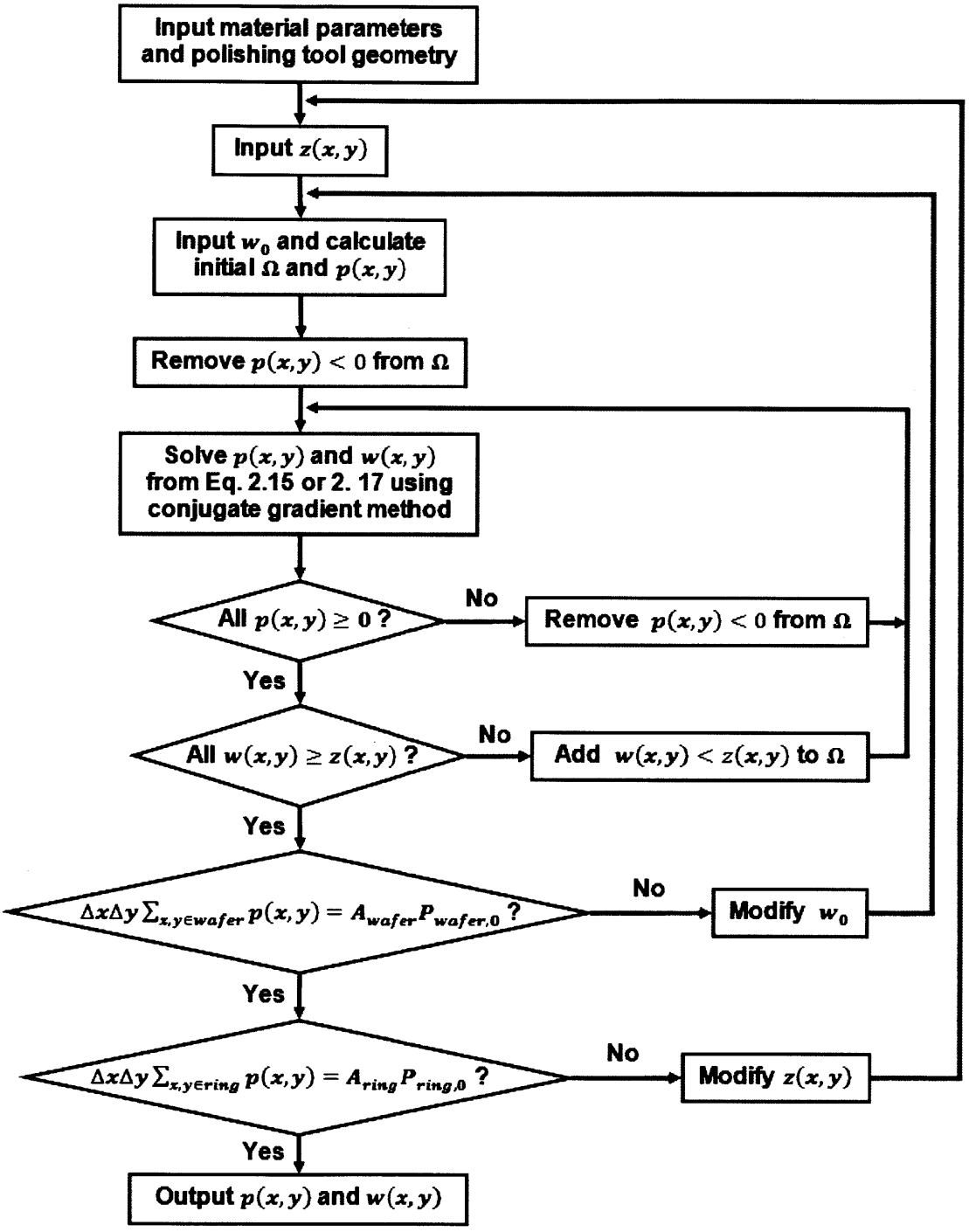


Figure 2.5: Flow chart of iterative program to calculate pressure distribution.

2.1.3 Simulation: pad modulus effect

This subsection investigates pad modulus effect on wafer-level non-uniformity. A 200 mm flat wafer contact is simulated for polishing pads with different Young's moduli: 50 MPa, 100 MPa, 200 MPa and 400 MPa. The retaining ring is set to be 2 mm away from the wafer edge. The applied wafer reference pressure and retaining ring reference pressure are 5 psi and 6 psi, respectively. As shown in Figure 2.6, the pressure distributions are the same for all moduli above, while the pad deformation profiles are different. Modulus is the key factor in pad bending, without affecting the pressure distribution. We can see that the pad surface is squeezed into the gap between wafer and retaining ring, shown as negative displacement in Figure 2.6(b). However, wafer-level pressure non-uniformity cannot be tuned by changing pad modulus. A high pressure concentration is observed at the wafer edge in Figure 2.6(a), which can explain the edge roll-off profile of polished wafers [69]. Although modulus does not have strong impact on wafer-level pressure uniformity, it still needs to be well designed or selected in order to control the retaining ring penetration for preventing wafer slip out, as shown in Figure 2.6(b).

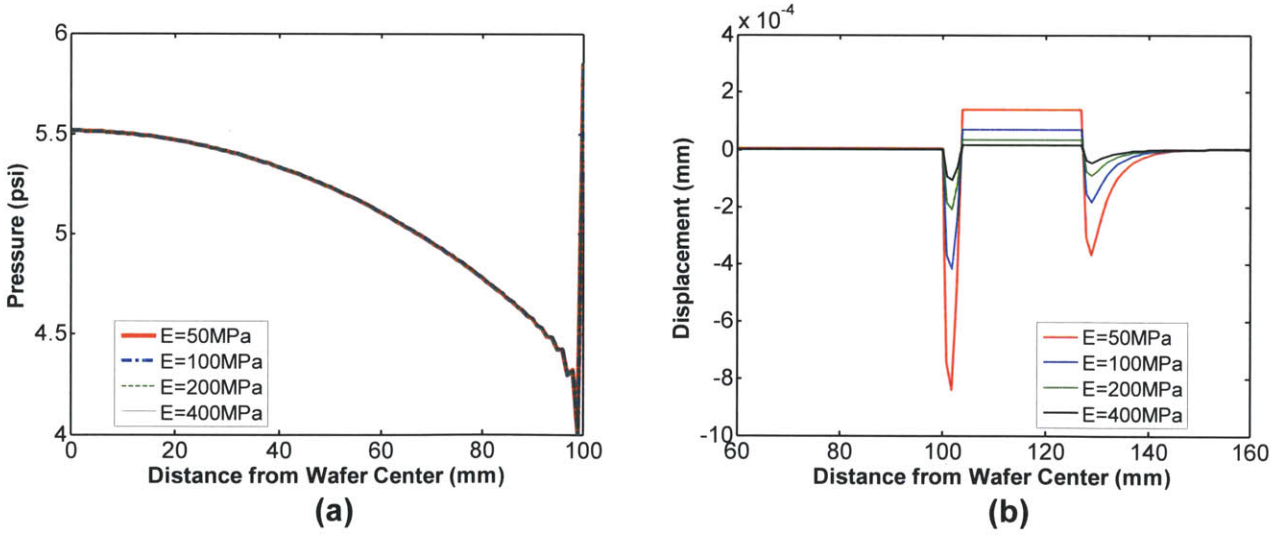


Figure 2.6: Pad modulus effect on wafer-level non-uniformity: (a) pressure distribution along wafer radius; (b) pad surface displacement near wafer edge.

The wafer-level pressure non-uniformity suggested by the simulations is quite large. In Section 3.4, we will extract wafer-level pressure non-uniformity by fitting wafer-level polishing data, and discuss the wafer-level pressure non-uniformity impacts on die-level planarization.

2.1.4 Simulation: pad thickness effect

Pad thickness is changing during or across many cycles of polishing and conditioning in a CMP process [70]. Understanding pad thickness impact on wafer-level non-uniformity is thus important. A 200 mm flat wafer contact is simulated for polishing pads with different thickness, 2 mm, 3 mm, 4 mm and 5 mm. The pad modulus is 100 MPa. The retaining ring is set to be 3 mm away from the wafer edge. The applied wafer reference pressure and retaining ring reference pressure are 5 psi and 6 psi, respectively. Figure 2.7 shows the pressure distributions on the wafer surface and the retaining ring surface. When the pad is thin, high pressure concentration occurs in wafer center region.

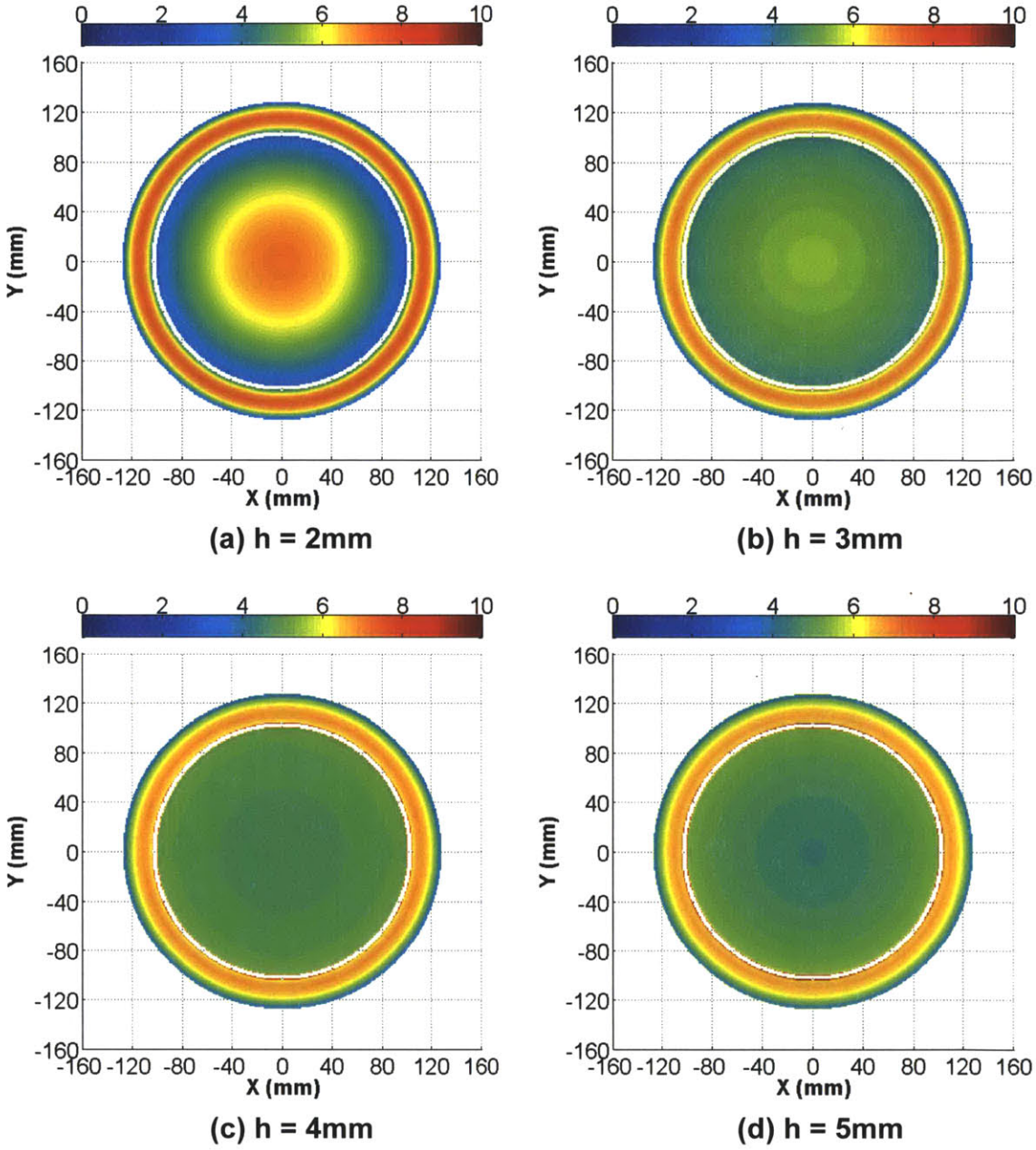


Figure 2.7: Wafer-level pressure (psi) distributions of different pad thickness.

Figure 2.8 shows how the pad thickness impacts wafer-level non-uniformity. Thin pads have high pressure at wafer center, while thick pads have low pressure at wafer center (Figure 2.8(a)). An optimized pad thickness is possible obtain a uniform wafer-

level pressure distribution; in this case, the optimized thickness could be between 3 mm and 4 mm. Thick pads allow more retaining ring penetration on pad surface, as shown in Figure 2.8(b). Therefore, thick pads provide better wafer slip protection.

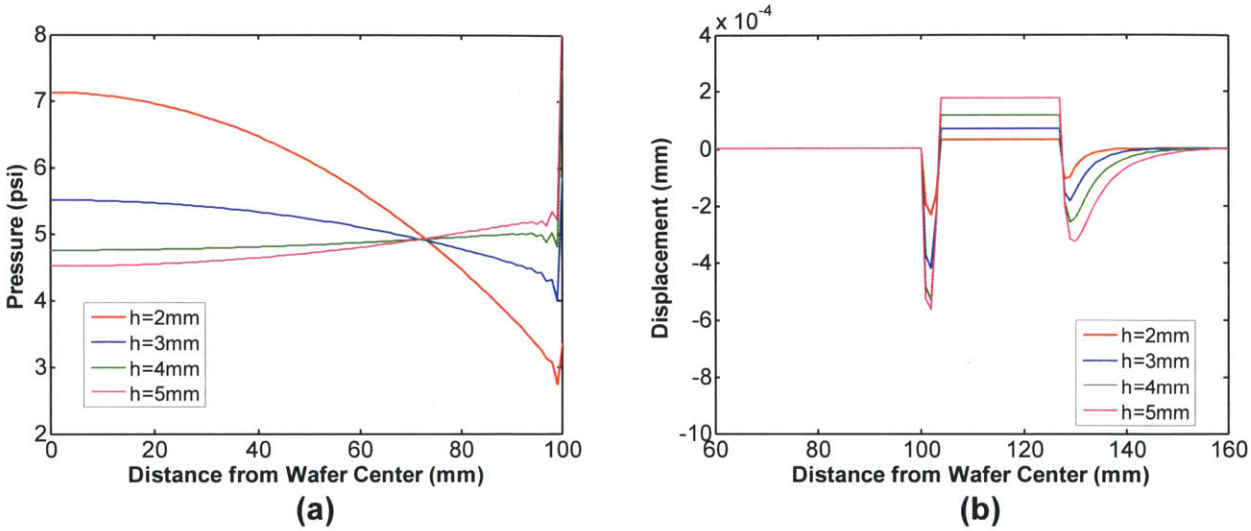


Figure 2.8: Pad thickness effect on wafer-level non-uniformity: (a) pressure distribution along wafer radius; (b) pad surface displacement near wafer edge.

2.1.5 Simulation: retaining ring gap effect

The subsection examines the retaining ring gap impact on wafer-level pressure distribution. A 200 mm flat wafer contact is simulated on a polishing pad with modulus of 100 MPa. The applied wafer reference pressure and retaining ring reference pressure are 5 psi and 6 psi, respectively. The retaining ring gap is set to be 1 mm, 2 mm, 3 mm and 4 mm separately. Figure 2.9 shows the relationship between retaining ring gap and pressure distribution. As the gap becomes bigger, higher pressure concentration is induced in the wafer center region. The wafer-level pressure uniformity can be tuned by varying the gap size, as shown in Figure 2.10(a). It is possible to optimize the gap size to get a uniform pressure distribution; in this case, the optimized gap size could be between 1 mm and 2 mm. When the retaining ring is closer to the wafer, the pad penetration becomes less, as shown in Figure 2.10(b).

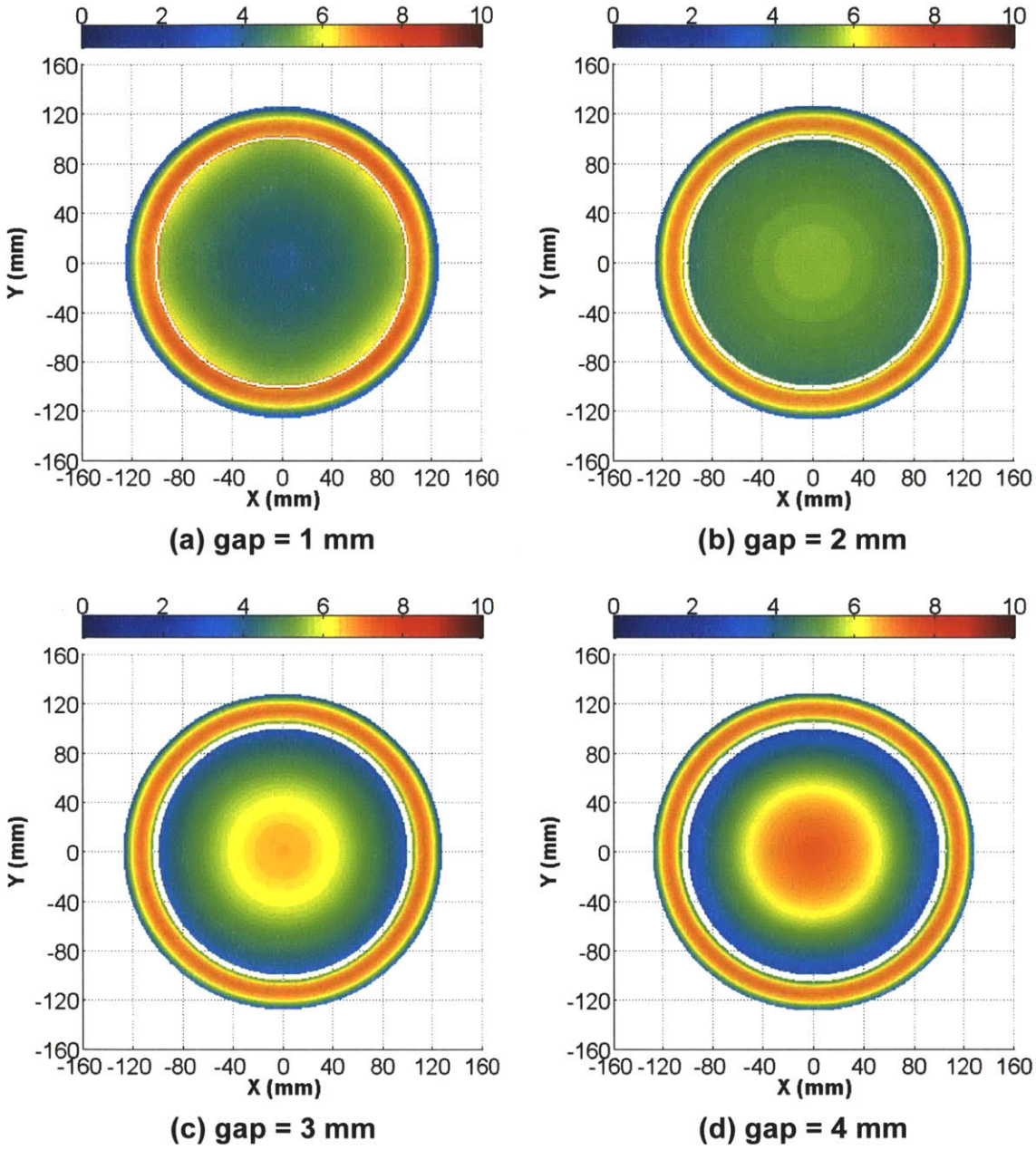


Figure 2.9: Wafer-level pressure (psi) distributions of different retaining ring gaps.

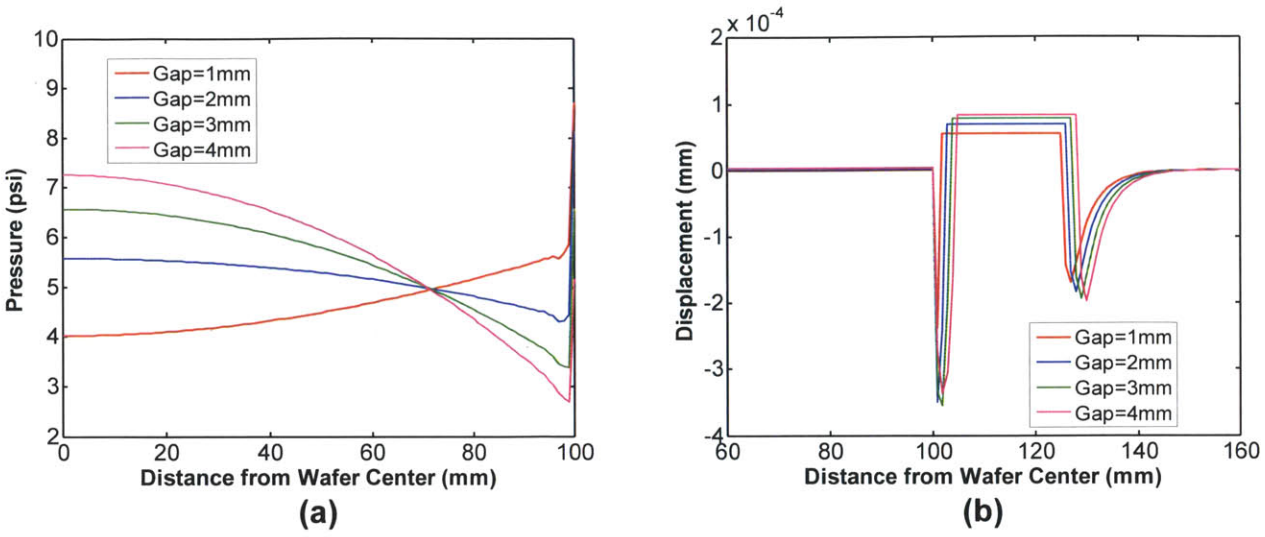


Figure 2.10: Retaining ring gap effect on wafer-level non-uniformity: (a) pressure distribution along wafer radius; (b) pad surface displacement near wafer edge.

2.1.6 Simulation: retaining ring reference pressure effect

We consider applied retaining ring pressure effects in this subsection. A 200 mm flat wafer simulation is run on a 3 mm thick polishing pad. The pad modulus is set to be 100 MPa. The retaining ring gap is 3 mm. We set the wafer reference pressure at 5 psi, while the retaining ring reference pressure is varied at 6 psi, 7 psi, 8 psi and 9 psi. As shown in Figure 2.11, when we increase the applied pressure on the retaining ring, the wafer center pressure is reduced. We can clearly see that the wafer-level pressure non-uniformity is affected by the retaining ring pressure in Figure 2.12(a). A fairly uniform pressure is achieved when we apply 7 psi pressure on the retaining ring. Therefore, changing retaining ring pressure is a possible approach to tune the wafer pressure non-uniformity.

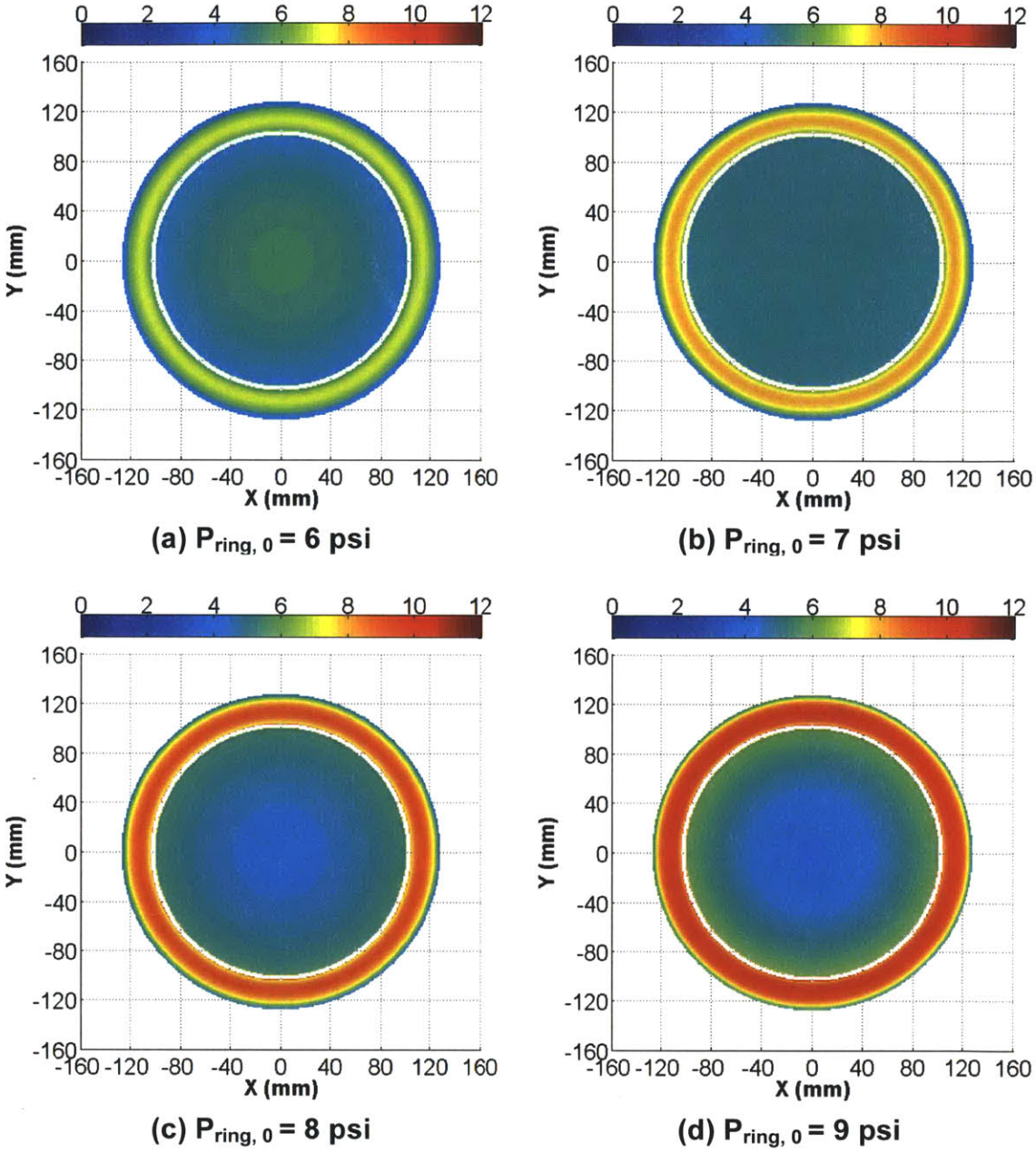


Figure 2.11: Wafer-level pressure (psi) distributions of different retaining ring reference pressures.

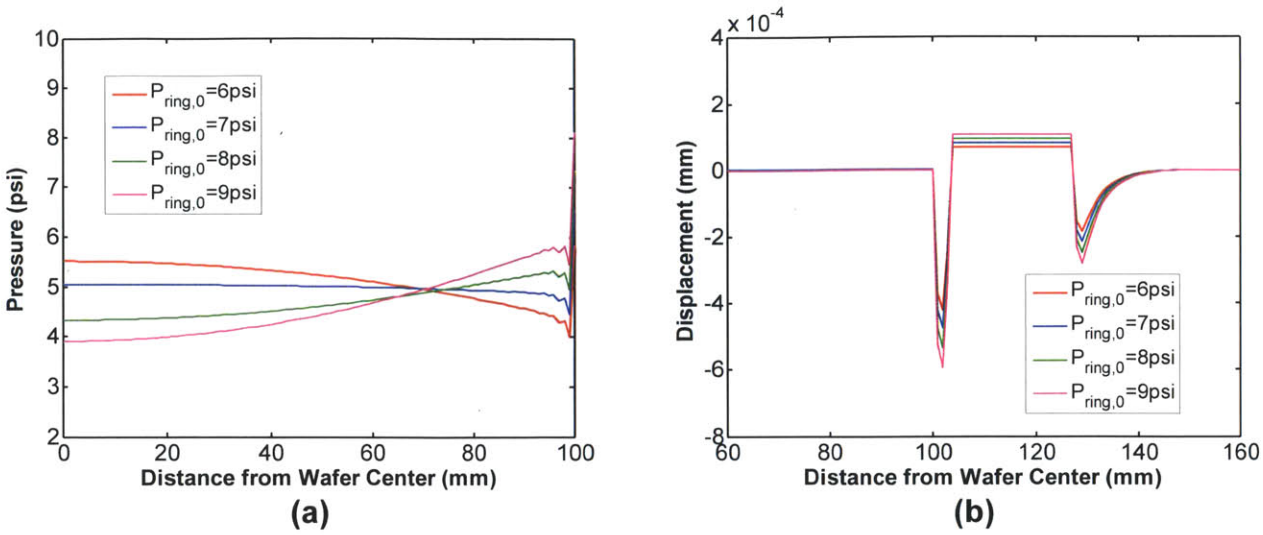


Figure 2.12: Retaining ring pressure effect on wafer-level non-uniformity: (a) pressure distribution along wafer radius; (b) pad surface displacement near wafer edge.

2.1.7 Simulation: non-centered wafer position effect

During the CMP process, the wafer may not sit in the center of the retaining ring. As shown in Figure 2.13, the rotating platen pushes the wafer towards the direction of rotation. The wafer is dynamically non-centered in the retaining ring, and the gap between wafer edge and the retaining ring is asymmetric. The instantaneous pressure distribution on the wafer surface is affected by the wafer position.

A 200 mm flat wafer contact is simulated on a 3 mm thick polishing pad with modulus of 100 MPa. The applied wafer reference pressure and retaining ring reference pressure are 5 psi and 6 psi, respectively. The same retaining ring size is used for centered and non-centered wafer positions. The retaining ring gap is set to be 2 mm for the centered wafer position. In the non-centered situation, the wafer is assumed to touch the retaining ring on the right side, while the gap on the left side is 4 mm.

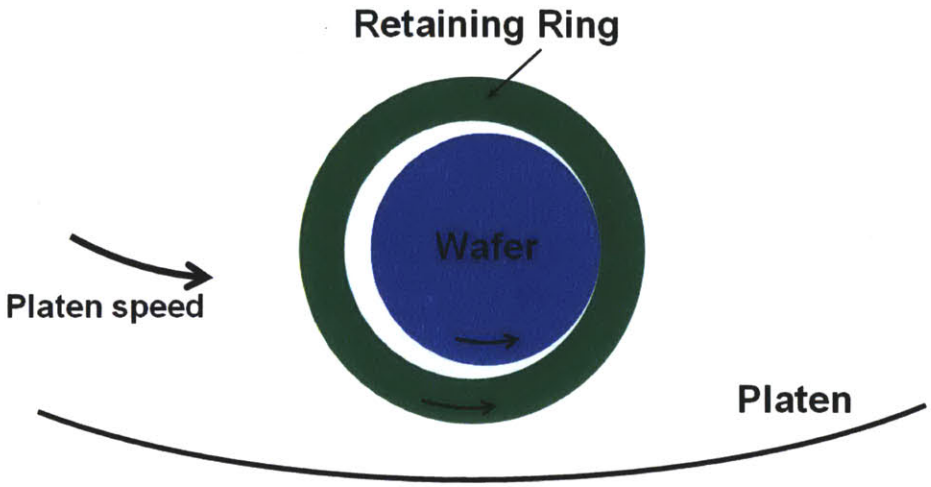


Figure 2.13: Top view of the non-centered wafer position during the CMP process.

Figure 2.14 compares the instantaneous wafer-level pressure distribution of centered and non-centered wafer positions. We can clearly see that the instantaneous pressure distribution is asymmetric when the wafer is non-centered in Figure 2.14(b). On the right side, where the wafer edge is in contact with the retaining ring, the pressure is highly concentrated. A radial time-averaged pressure distribution based on rotation of the wafer around its center is calculated, and compared in Figure 2.15 to the centered wafer-position pressure distribution. We can see that the wafer-level non-uniformity will be affected by the wafer position during polishing, in the case where the wafer is not centered within the retaining ring. Future work should consider dynamic fluid pressure [71-73] and other dynamic effects [27].

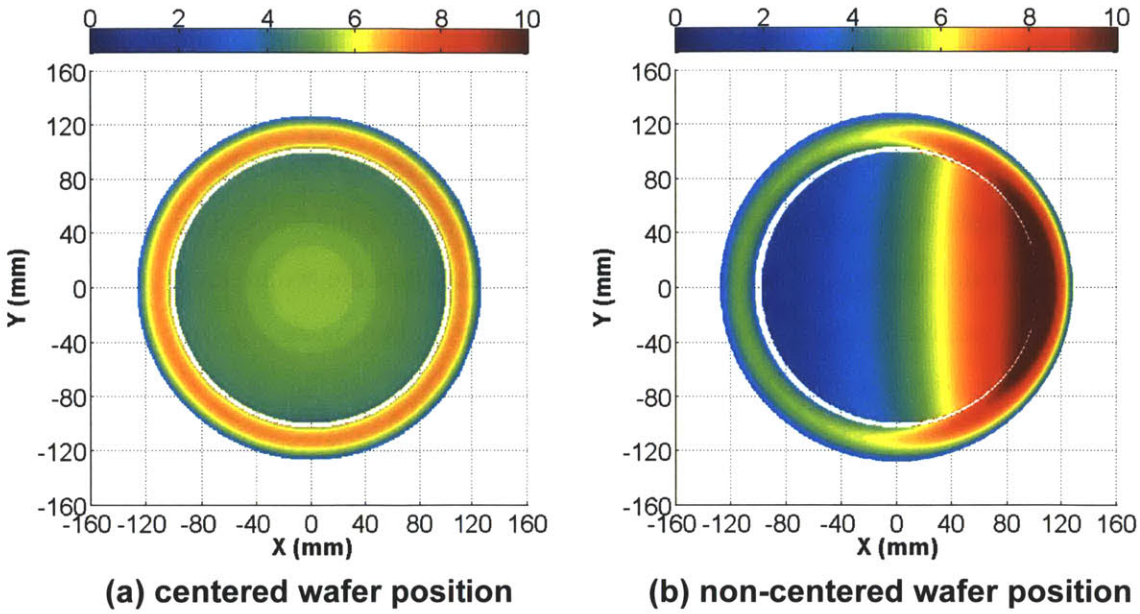


Figure 2.14: Instantaneous wafer-level pressure (psi) distributions of centered and non-centered wafer position in the retaining ring.

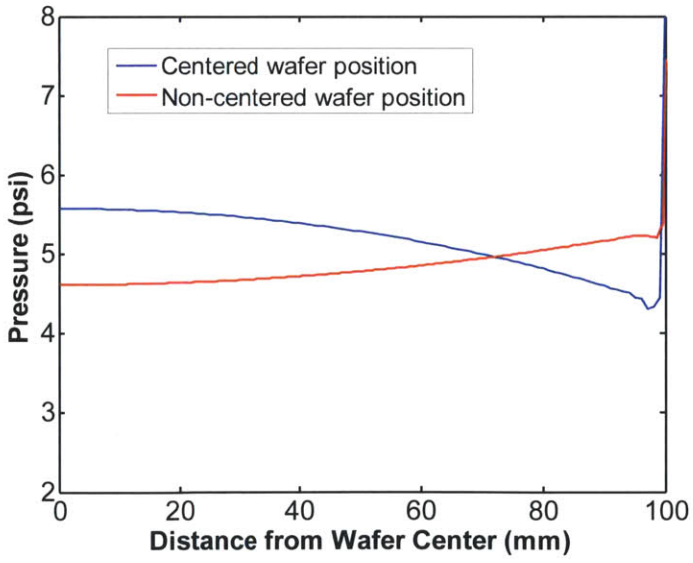


Figure 2.15: Wafer position effect: radial time-averaged pressure distribution.

2.1.8 *Summary of physical wafer-level CMP model*

A wafer-level CMP model is developed to understand wafer-level pressure non-uniformity related to retaining ring geometry and process conditions. The computational approach is demonstrated. Model simulation suggests that the wafer-level pressure non-uniformity can be tuned by optimizing pad thickness, retaining ring gap or retaining ring pressure, but not by pad modulus.

2.2 Physical die-level CMP model

Dielectric CMP is utilized in both front-end (shallow trench isolation) and back-end (pre-metal and inter-metal dielectric) processes in IC manufacturing. Planarization of patterned topography is very important to enable following photolithography and process integration. Planarization results rely on many polishing parameters such as pressure and pad modulus and the pattern layout itself. This section presents a physical die-level CMP model to understand the relationships of die-level pressure distribution, pad modulus, pad surface topography and layout pattern density. This model takes the same assumptions as Xie's model [18]. However, Xie's model involves the biconjugate gradient stabilized method (BiCGSTAB), which is computationally intensive at the die-level. In this work, the model is refined in derivation and simplified in computation.

2.2.1 *Model derivation*

We assume that the polishing pad is elastic and can be decomposed into pad bulk and pad asperities from a certain reference plane, as shown in Figure 2.16. The bulk material can be treated as an elastic body, deforming in response to long range wafer height differences. The surface asperities come in contact with the wafer surface, and the compression of the asperities depends on both the wafer surface profile and pad bulk bending. In this die-level model, the "pad bulk" means the pad top region connecting to asperities. This is different from the deep or entire pad bulk in the wafer-level model; here the "pad bulk" still belongs to pad surface, which may have different properties than

the deep pad bulk due to the porous structure of pad body, sub-pad laminations, or other pad stack effects.

Figure 2.16: Pad structure assumption in physical die-level CMP model. The whole pad is comprised of bulk and asperities.

Figure 2.17 illustrates the model framework. The wafer is set to sit face down in the real process, and the wafer surface is pressed down onto the polishing pad. For convenience, the surface normal of the wafer is taken as the positive Z direction, corresponding to the conventional "wafer face up" mathematical representation. During CMP, the die surface and pad surface are constantly in contact. The wafer topography is assumed to have step or height varying structures arising from the chip layout. The wafer surface profile can be described by up area thickness of oxide film, $z_u(x, y)$, as shown in Figure 2.17(a). Pad bulk long range bending $w(x, y)$ is the main response to the wafer surface profile. Pad asperities are compressed between pad bulk and wafer, as shown in Figure 2.17(b); these asperities can be treated as springs. Both up area and down area of the step structure on the wafer surface are in contact with asperities.

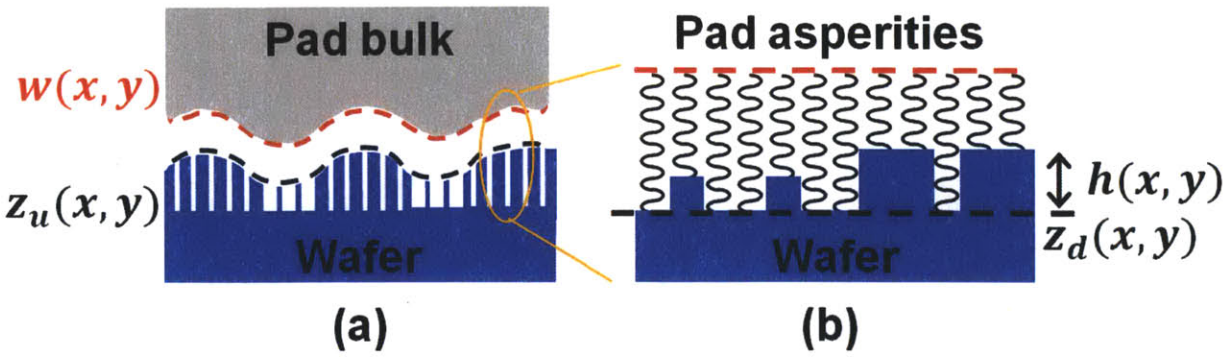


Figure 2.17: Framework of physical die-level model: (a) wafer surface profile and pad long-range deformation; (b) local step structures and pad asperity compression.

The model derivation takes the following approach:

1) **Modeling of pad bulk.** The pad bulk is an elastic body, which can be modeled using the same contact wear model as in the wafer-level model described in Section 2.1. The pad is defined as a solid half-space, with Young's modulus E_0 and Poisson's ratio v. Here we adopt Eq. 2.1 to calculate pad bulk deformation and rewrite it as

$$w(x,y) - w_0 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(x-x', y-y') \cdot p(x', y')dx'dy' \tag{2.18}$$

where G is the Green's function,

$$G(x,y) = \frac{1}{\pi E_0^{\,*}\sqrt{x^2+y^2}} \tag{2.19}$$

E_0* is the reduced modulus,

$$E_0^{\,*} = \frac{E_0}{1-v^2} \tag{2.20}$$

and w_0 is the reference plane of starting deformation. The boundary conditions applied to Eq. 2.18 is

$$\begin{cases} p(x,y) \geq 0 \\ w(x,y) \geq z_u(x,y) \\ \dfrac{1}{A_{chip}} \iint\limits_{\substack{chip \\ surface}} p(x,y)dxdy = P_0 \end{cases} \tag{2.21}$$

where A_{chip} is the whole chip area and P_0 is the applied reference pressure.

2) **Modeling of pad asperities.** The asperities can be assumed to have negligible width [18, 74] and an exponential height distribution [75]. Eq. 2.22 defines the probability that the asperity height lies between l and $l + \Delta l$,

$$\phi(l) = \frac{1}{\lambda} e^{-\frac{l}{\lambda}} \tag{2.22}$$

where λ is called the characteristic asperity height. At location (x,y), the distance between wafer profile $z_u(x,y)$ and pad bulk profile $w(x,y)$ is $d(x,y) = w(x,y) - z_u(x,y)$. So asperities of height l larger than $d(x,y)$ will be compressed and the compression amount is $l - d(x,y)$. All of the asperities are assumed to be ideal springs and follow Hooke's law [76], i.e., the exerting force is proportional to the compressed amount. The expected value of up area pressure $p_u(x,y)$ can be estimated by averaging across all of the asperities as follows:

$$\begin{aligned} p_u(x,y) &= \int_{d(x,y)}^{\infty} k \cdot \{l - d(x,y)\} \cdot \phi(l) dl \\ &= k \cdot \Phi\big(w(x,y) - z_u(x,y)\big) \end{aligned} \tag{2.23}$$

where k is an equivalent spring constant [18] and $\Phi(z)$ is a derived accumulative height distribution function of asperity compression, defined as

$$\Phi(z) = \int_z^{\infty} (l - z) \cdot \phi(l) dl \tag{2.24}$$

$\Phi(z)$ can be calculated once the probability distribution of asperity height is known, and it is a strictly decreasing function and approaches zero at infinity. Since we assume the asperity height distribution as Eq. 2.22, $\Phi(z)$ is given by

$$\Phi(z) = \lambda \cdot e^{-\frac{z}{\lambda}} \tag{2.25}$$

Therefore, the up area pressure can be calculated by Eq. 2.23 and Eq. 2.25.

When a feature of step height $h(x, y)$ is pressed against the polishing pad, as shown in Figure 2.17(b), the down area pressure can be derived in a similar fashion as for Eq. 2.23, obtaining

$$\begin{aligned} p_d(x, y) &= \int_{d(x,y)+h(x,y)}^{\infty} k \cdot \{l - d(x, y) - h(x, y)\} \cdot \phi(l) dl \\ &= k \cdot \Phi(w(x, y) - z_d(x, y)) \end{aligned} \tag{2.26}$$

where $z_d(x, y) = z_u(x, y) - h(x, y)$ is down area thickness.

The total local pressure is the sum of the two pressures weighted by pattern density $\rho(x, y)$ which is the area fraction of the up area. So we get

$$\begin{aligned} p(x, y) &= \rho(x, y) \cdot p_u(x, y) + (1 - \rho(x, y)) \cdot p_d(x, y) \\ &= \rho(x, y) \cdot k \cdot \Phi(w(x, y) - z_u(x, y)) \\ &\quad + (1 - \rho(x, y)) \cdot k \cdot \Phi(w(x, y) - z_u(x, y) + h(x, y)) \end{aligned} \tag{2.27}$$

Combining Eq. 2.23, 2.26 and 2.27, we relate pressures to step height $h(x,y)$ and characteristic asperity height λ as

$$\begin{cases} p(x,y) = k\left[\rho(x,y) + (1-\rho(x,y))e^{-\frac{h(x,y)}{\lambda}}\right]\cdot \lambda e^{-\frac{w(x,y)-z_u(x,y)}{\lambda}} \\ p_u(x,y) = \dfrac{e^{\frac{h(x,y)}{\lambda}}}{1+\rho(x,y)\cdot\left(e^{\frac{h(x,y)}{\lambda}}-1\right)}p(x,y) \\ p_d(x,y) = \dfrac{1}{1+\rho(x,y)\cdot\left(e^{\frac{h(x,y)}{\lambda}}-1\right)}p(x,y) \end{cases} \tag{2.28}$$

3) **Force balance from pad bulk and pad asperities.** The pressure from pad bulk and the pressure from asperities need to be equal to satisfy force balance. So the overall pressure distribution can be obtained by equating the efforts of the two parts above together: the elastic pad bulk, which is described by Eq. 2.18, and the asperities with exponential height distribution, which is described by Eq. 2.28. The pressure and deflection interactions between wafer surface topography and CMP pad are therefore described by

$$\begin{cases} p(x,y) = k\left[\rho(x,y) + (1-\rho(x,y))e^{-\frac{h(x,y)}{\lambda}}\right]\cdot \lambda e^{-\frac{w(x,y)-z_u(x,y)}{\lambda}} \\ w(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(x-x',y-y')\cdot p(x',y')dx'dy' + w_0 \end{cases} \tag{2.29}$$

This problem is subject to boundary conditions of Eq. 2.21.

4) **Modeling of CMP process.** To calculate the wafer topography evolution during CMP process, pattern density $\rho(x,y)$ needs to be extracted from the chip layout. With

initial values of up area thickness and step height, the die-level pressure distribution $p(x,y)$ can be obtained by solving Eq. 2.29. Once $p(x,y)$ is solved, $p_u(x,y)$ and $p_d(x,y)$ can be calculated by Eq. 2.28. Then we utilize Preston's equation [23] with local pressures $p_u(x,y)$ and $p_d(x,y)$ to calculate the instantaneous material removal rates of up area and down area as

$$\begin{cases} \dfrac{dz_u(x,y)}{dt} = -K_0 \dfrac{p_u(x,y)}{P_0} \\ \dfrac{dz_d(x,y)}{dt} = -K_0 \dfrac{p_d(x,y)}{P_0} \end{cases} \tag{2.30}$$

where $K_0 = K_p P_0 V_0$ is the blanket removal rate under reference pressure P_0, K_p is Preston's coefficient and V_0 is the assumed constant relative velocity between pad surface and wafer surface. Up area thickness, down area thickness and step height can be dynamically updated in time steps using Eq. 2.30.

2.2.2 Remarks on model parameters

Unlike traditional and current semi-empirical die-level CMP models [77-80], planarization length (PL) is no longer used as a parameter in this physical die-level model. Conceptually, the planarization length (PL) in die-level CMP is the distance at which the CMP polishing pad no longer interacts with a localized step height, and thus we do not preferentially remove material from raised regions; instead, the entire die surface continues to polish with the same removal rate [81]. Traditional die-level models use planarization length to define an effective density window size [39, 40]; in the window, weighted average pattern density is calculated for the chip layout and used in model simulation instead of the real local layout pattern density to consider interactions between neighboring structures and the polishing pad. Most oxide CMP processes have a planarization length on the order of 3 to 5 mm from model fitting [81]. As a multi-functional model parameter, planarization length is affected by many polishing factors including pad modulus, pattern density and feature size. The physical meaning or impact of these multiple factors cannot be separated from the planarization length.

This model presented in this thesis makes the meaning of the physical parameters clear by avoiding the use of planarization length. There are three key parameters in this physical die-level CMP model: blanket removal rate K_0, pad effective modulus E_0^*, and characteristic asperity height λ. The blanket removal rate is affected by many CMP tool, consumable, and process parameters, such as the CMP system reference pressure. The effective modulus is related to properties of the pad bulk, and is hypothesized to most strongly impact within-die uniformity and layout pattern density effects, resulting from long range pad bending due to differential removal rates in different die pattern density regions. The characteristic asperity height reflects the distribution of pad asperity heights, and is hypothesized to most strongly impact the feature scale step height reduction.

2.2.3 Computational approach

The numerical analysis approach is the same as in the wafer-level model presented in subsection 2.1.2. However, for the die-level model, a periodic boundary condition is desired to represent the periodic arrangement of dies on a wafer, so we do not need to put a void mesh region as wafer-level model (Figure 2.3). In the spatial domain, we mesh the wafer and pad surfaces on an $N_x \times N_y$ grid of rectangular elements with pitch sizes Δx and Δy in x and y directions respectively, as shown in Figure 2.18(a). In the frequency domain, we use ξ and η to denote the frequencies in x and y directions, and an $N_x \times N_y$ mesh grid of elements with pitch sizes $\frac{2\pi}{N_x \Delta_x}$ and $\frac{2\pi}{N_y \Delta_y}$ is used, as shown in Figure 2.18(b).

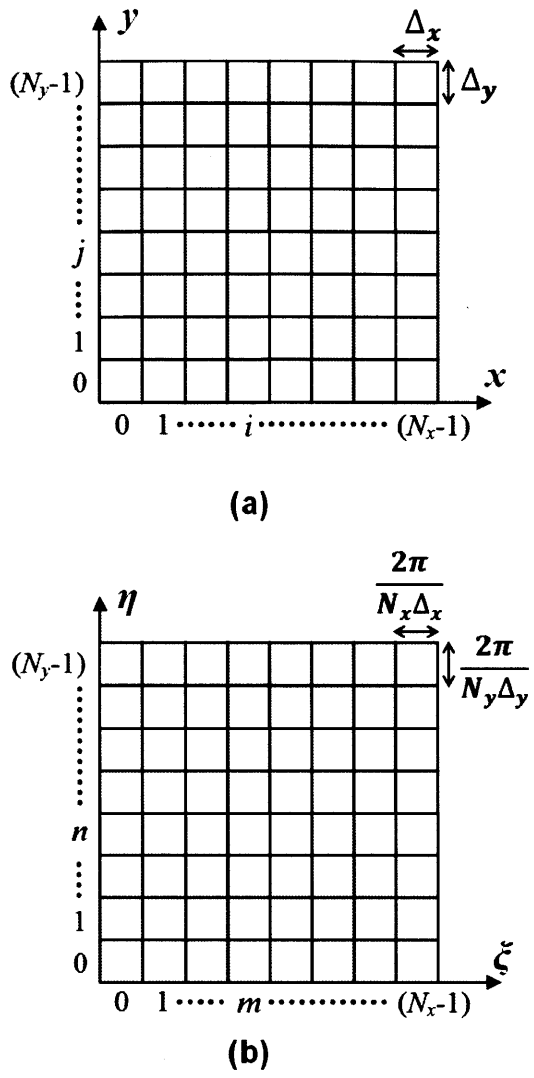


Figure 2.18: Discretization grid of the die-level model: (a) top view in spatial domain, (b) top view in frequency domain.

Let the pressure distribution and the surface displacements be denoted by $p(x_i, y_j)$ and $w(x_i, y_j)$ respectively, where $i = 0, 1, \cdots, (N_x - 1)$ and $j = 0, 1, \cdots, (N_y - 1)$. The discretized frequencies are denoted by ξ_m and η_n, where $m = 0, 1, \cdots, (N_x - 1)$ and $n = 0, 1, \cdots, (N_y - 1)$. The discretization can be expressed as

$$\begin{cases} x_i = i\Delta_x, & i = 0,1,\cdots,(N_x - 1) \\ y_j = j\Delta_y, & j = 0,1,\cdots,(N_y - 1) \end{cases} \tag{2.31}$$

in the spatial domain, and

$$\begin{cases} \xi_m = m\Delta_\xi, \quad m = 0,1,\cdots,(N_x - 1) \\ \eta_n = n\Delta_\eta, \quad n = 0,1,\cdots,(N_y - 1) \\ \Delta_\xi = \dfrac{2\pi}{N_x \Delta_x} \\ \Delta_\eta = \dfrac{2\pi}{N_y \Delta_y} \end{cases} \tag{2.32}$$

in the frequency domain. The discrete Green's function is expressed as

$$g(x_i, y_j) = \frac{1}{\pi E_0^*}\left[f(x_{i2}, y_{j2}) - f(x_{i1}, y_{j2}) - f(x_{i2}, y_{j1}) + f(x_{i1}, y_{j1})\right] \tag{2.33}$$

where $f(x, y)$ has the same definition as in Eq. 2.13. The discrete Boussinesq integral (Eq. 2.18) is expressed as a discrete convolution,

$$w(x_i, y_j) - w_0 = \sum_{i'=0}^{N_x-1}\sum_{j'=0}^{N_y-1} g(x_i - x_{i'}, y_j - y_{j'}) \cdot p(x_{i'}, y_{j'}) \tag{2.34}$$

Then Eq. 2.29 can be written in discrete form as

$$\begin{cases} p(x_i, y_j) = k\left[\rho(x_i, y_j) + (1 - \rho(x_i, y_j))e^{-\frac{h(x_i, y_j)}{\lambda}}\right] \cdot \lambda e^{-\frac{w(x_i, y_j) - z_u(x_i, y_j)}{\lambda}} \\ w(x_i, y_j) = \sum_{i'=0}^{N_x-1}\sum_{j'=0}^{N_y-1} g(x_i - x_{i'}, y_j - y_{j'}) \cdot p(x_{i'}, y_{j'}) + w_0 \end{cases} \tag{2.35}$$

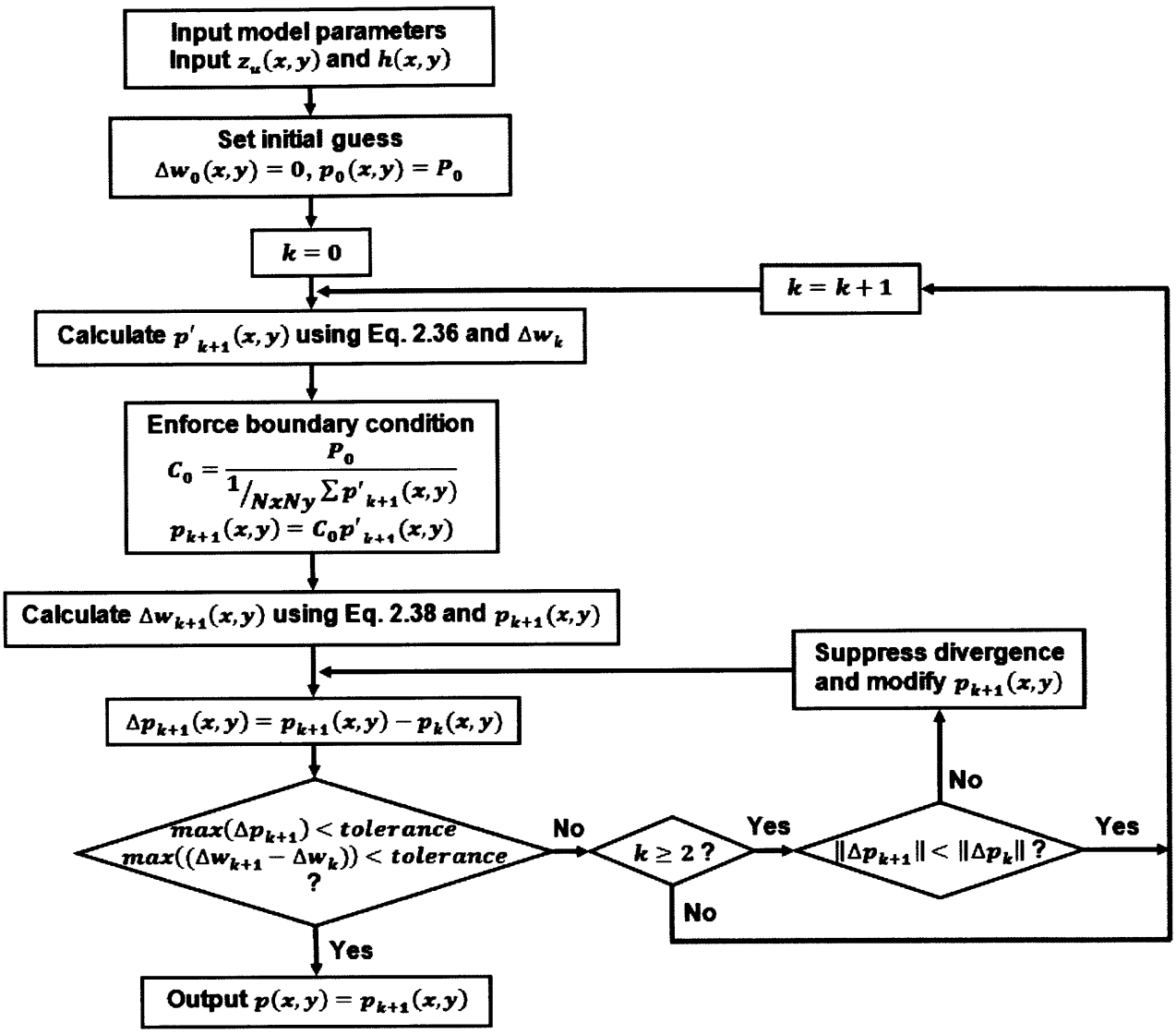


Figure 2.19: Flow chart of iterative program to calculate die-level pressure distribution.

Note that when the reference boundary condition is applied, spring constant k and pad bulk displacement $w(x_i, y_j)$ cannot be solved independently. as they are coupled unknowns. The average displacement $\overline{w} = \frac{\sum w(x_i,y_j)}{N_x N_y}$ relies on the spring constant of asperities, since k is a linear scale factor. However, Eq. 2.35 is still solvable if we

combine $\bar{w}$ and k into a single unknown $C_0 = ke^{-\frac{\bar{w}}{\lambda}}$. For mathematical convenience, w_0 is set to zero. Eq. 2.35 is expressed in a system of equations as follows:

$$p'(x_i, y_j) = \left[\rho(x_i, y_j) + (1 - \rho(x_i, y_j))e^{-\frac{h(x_i, y_j)}{\lambda}} \right] \cdot \lambda e^{-\frac{\Delta w(x_i, y_j) - z_u(x_i, y_j)}{\lambda}} \tag{2.36}$$

$$p(x_i, y_j) = C_0 p'(x_i, y_j) \tag{2.37}$$

$$\Delta w(x_i, y_j) = \sum_{i'=0}^{N_x-1} \sum_{j'=0}^{N_y-1} g(x_i - x_{i'}, y_j - y_{j'}) \cdot [p(x_{i'}, y_{j'}) - P_0] \tag{2.38}$$

where $\Delta w(x_i, y_j) = w(x_i, y_j) - \bar{w}$ is called the adjusted displacement. Under our pressure boundary condition, this system of equations can be solved by non-linear Richardson iteration [82]. The iterative program is explained by the flow chart of Figure 2.19. Discrete Fourier transforms (DFT) in Eq. 2.15 are also suitable for computing Eq. 2.38, which is rearranged as

$$\begin{cases} \Delta w(x_i, y_j) = \Delta_\xi \Delta_\eta \sum_{m=0}^{N_x-1} \sum_{n=0}^{N_y-1} e^{I(\xi_m x_i + \eta_n y_j)} \Delta \tilde{w}(\xi_m, \eta_n) \\ \Delta \tilde{w}(\xi_m, \eta_n) = \tilde{g}(\xi_m, \eta_n) \cdot \tilde{p}(\xi_m, \eta_n) \\ \tilde{g}(\xi_m, \eta_n) = \Delta_x \Delta_y \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} e^{-I(\xi_m x_i + \eta_n y_j)} g(x_i, y_j) \\ \tilde{p}(\xi_m, \eta_n) = \Delta_x \Delta_y \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} e^{-I(\xi_m x_i + \eta_n y_j)} [p(x_i, y_j) - P_0] \end{cases} \tag{2.39}$$

Once the pressure distribution $p(x_i, y_j)$ is solved from the iterative method, time-stepped chip topography evolution is enabled by Eq. 2.28 and 2.30.

2.2.4 Simulation: pattern density dependence

This subsection examines the pattern density dependence of CMP, using the simulation framework of the previous section. The MIT standard oxide CMP characterization layout [77] is used to run the die-level simulation, as shown in Figure 2.20(a). The initial die topography is assumed to have 2000 nm up area thickness and 800 nm step height through the whole die. The layout and initial topography assumption matches the physical pattern on the SKW7-2 oxide CMP test wafer [83]. We fix the reference pressure at 5 psi and assume that the blanket removal rate at the reference pressure is 200 nm/min. The pad bulk reduced modulus is set to be 300 MPa and the characteristic asperity height is 100 nm. A polishing process of 150 seconds is simulated.

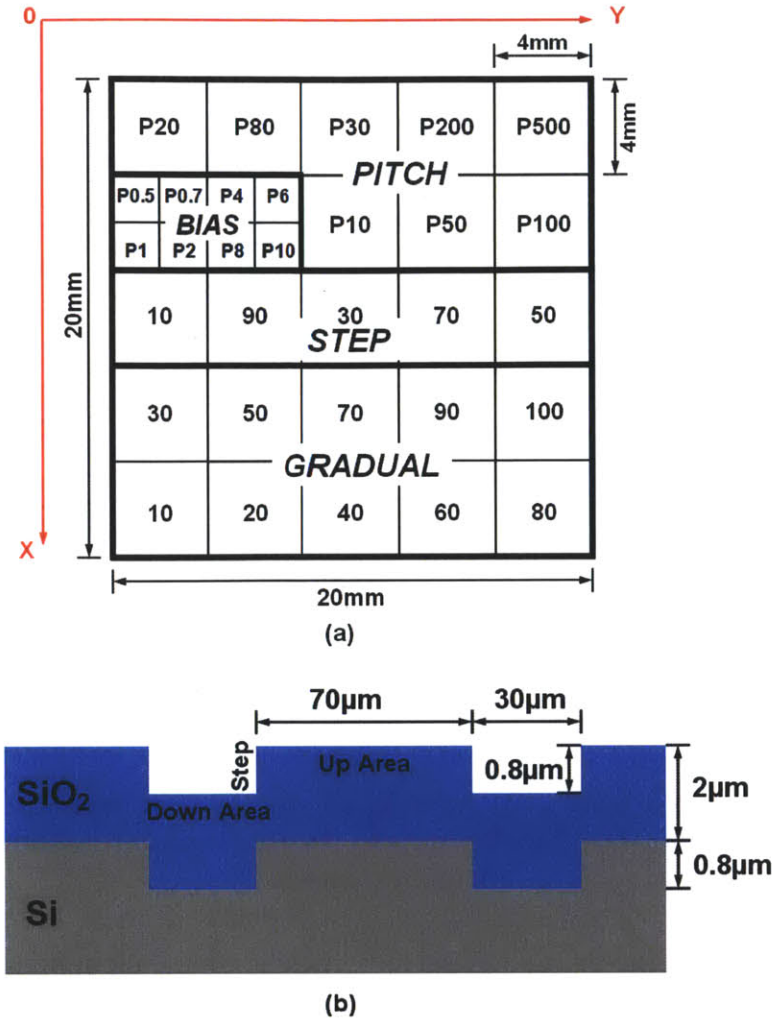


Figure 2.20: Pattern type and pattern density within a die of SKW7-2 wafer: (a) Layout of a die on SKW7-2 wafer (MIT standard oxide CMP characterization layout). A "P" preceding a number indicates a pitch structure with 50% density, with the number following in microns. All other numbers are localized densities, with the number indicating the density. Density structures have a fixed 100 micron pitch. (b) Topography of the 70% STEP array in a die.

A strong pattern density (PD) dependence is observed from Figure 2.21. Center points of STEP arrays are selected as monitor sites. We can see that both material removal (up area thickness reduction) and local planarization (step height reduction) are faster in lower density areas. This is caused by higher local pressure in these low density up areas.

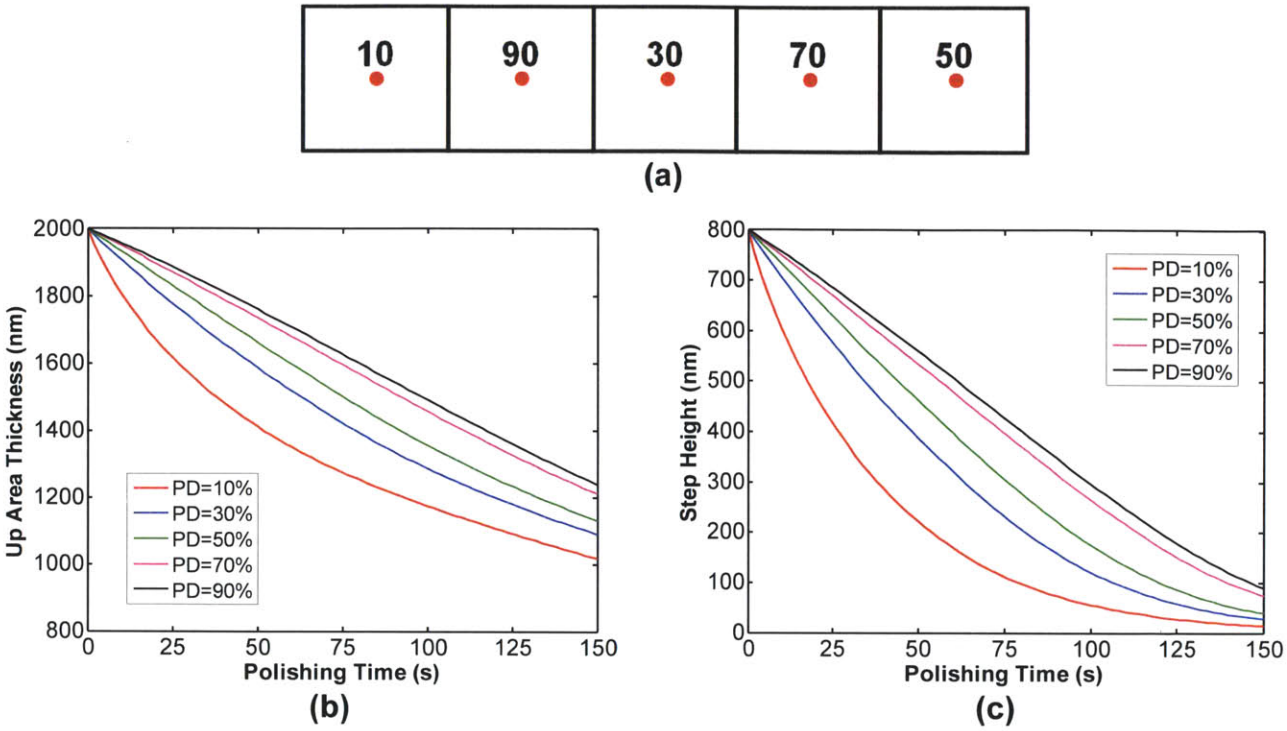


Figure 2.21: Pattern density (PD) dependence of CMP process: (a) Monitor sites in STEP arrays of MIT standard layout; (b) Up area thickness evolution; (c) Step height evolution.

Polishing performance is also affected by neighboring pattern density. Three sites are monitored respectively at left edge, center point and right edge of the 50% array, as shown in Figure 2.22(a). Comparing to the center point, up area material removal and step height reduction at the left edge are slower, while those at the right edge are faster. The difference is due to the influence from neighboring arrays. The left edge site is next to the 70% array, so the "effective" pattern density is higher. The right edge site is next to the 10% array in another die on the wafer, so the "effective" pattern density is lower. Physically, this is the result of the long-range pad bending, which couples the force response of neighboring patterns. As in subsection 2.2.2, we do not use planarization length in this model, and effective pattern density is not calculated. However, the model still includes the spatial averaging effect, since the pad bulk pressure response is computed by the Boussinesq integral (Eq. 2.18), which is a spatial convolution.

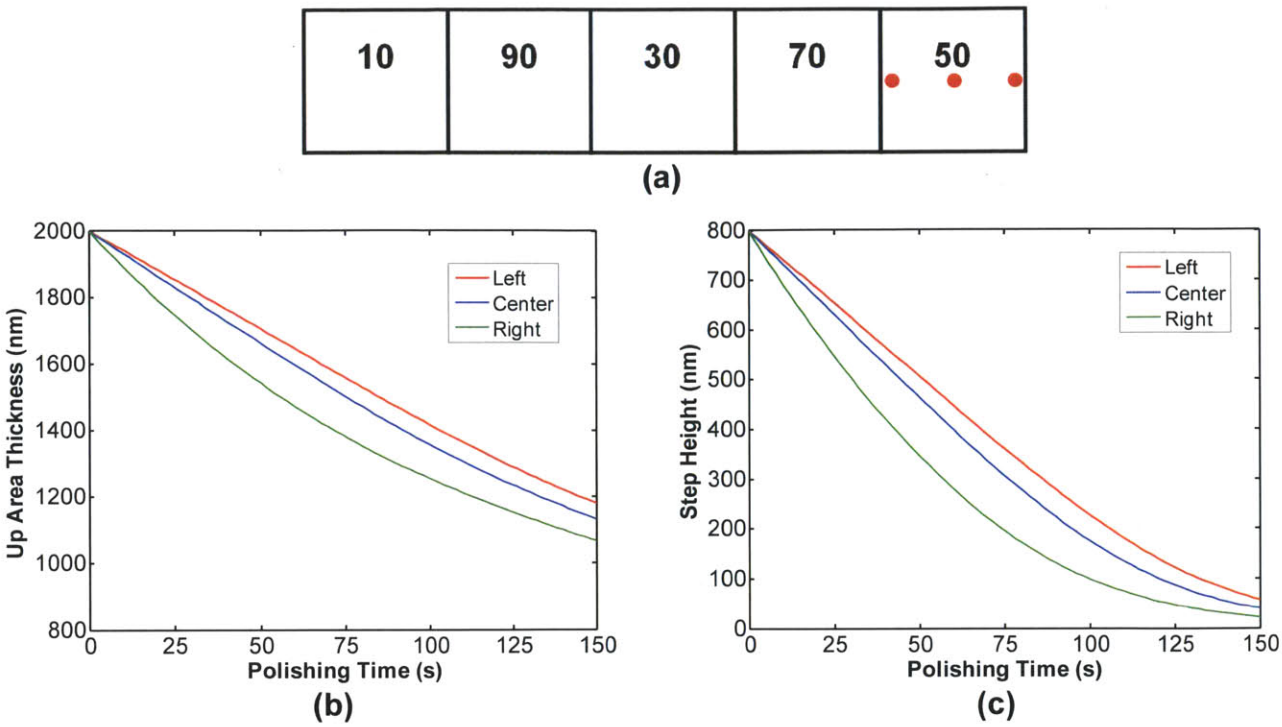


Figure 2.22: Neighboring pattern density (pad long-range bending) effects of CMP process: (a) Monitor sites in STEP arrays of MIT standard layout; (b) Up area thickness evolution; (c) Step height evolution.

2.2.5 Simulation: pad bulk modulus effect

This subsection investigates pad bulk modulus impacts on planarization. The same die layout, initial topography, reference pressure, blanket removal rate and characteristic asperity height are used as in Subsection 2.2.4. We vary the pad bulk reduced modulus at 100 MPa, 200 MPa, 300 MPa and 400 MPa. The polishing process is simulated up to 150 seconds.

Planarization efficiency (PE) is a useful parameter to describe local planarization capability, which is defined as

$$PE = 1 - \frac{AR_d}{AR_u} \tag{2.40}$$

where AR_u and AR_d are up area removal amount and down area removal amount, respectively. If there is no down area removal, the planarization efficiency equals to 1, which indicates that pure step reduction is realized. On the opposite extreme, when the planarization efficiency is 0, removal amounts in up area and down area are the same; no step reduction is achieved. Figure 2.23 shows that step height evolution and planarization efficiency are similar for different pad bulk modulus at the center site of 50% array. Thus pad modulus does not have strong impact on local planarization.

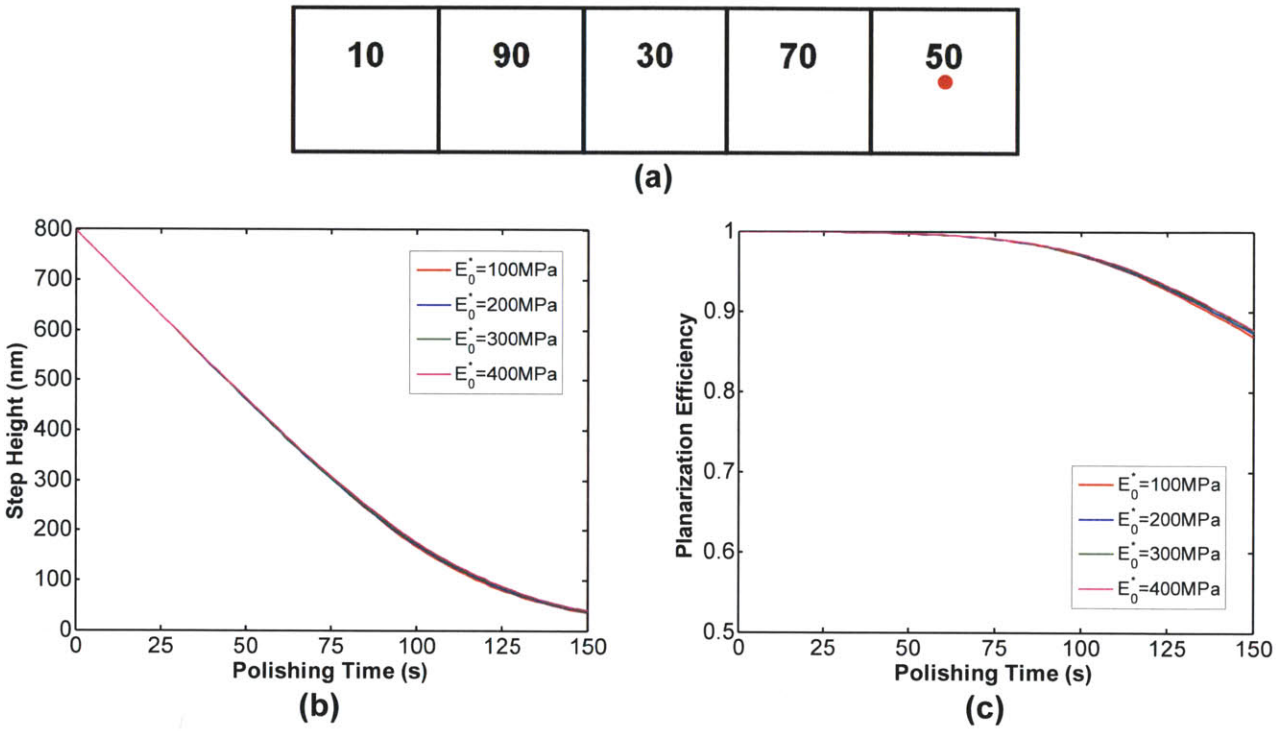


Figure 2.23: Local planarization results of different pad bulk reduced moduli: (a) Monitor site in 50% STEP array of MIT standard layout; (b) Step height evolution; (c) Planarization efficiency evolution.

To describe global planarization, we define a parameter called nominal range (NR) as the difference between the up area oxide thickness of the 90% pattern-density array center and that of the 10% pattern-density array center, shown in Figure 2.24(a). The initial nominal range is 0 since we assume the same up area thickness across the whole chip. As the polishing process starts, nominal range goes up because of higher removal

rate in the 10% array and lower removal rate in 90% array. As polishing proceeds further in time, nominal range will drop, because the 10% array becomes a "recessed" region and takes less pressure than before. The nominal range is a useful output to reflect pad bulk modulus impact on within-die non-uniformity. As shown in Figure 2.24, low pad modulus induces a high nominal range, e.g., a large within-die thickness non-uniformity. To combat this effect, pads with higher modulus are of interest, as they are subject to less long-range bending to achieve better global planarization.

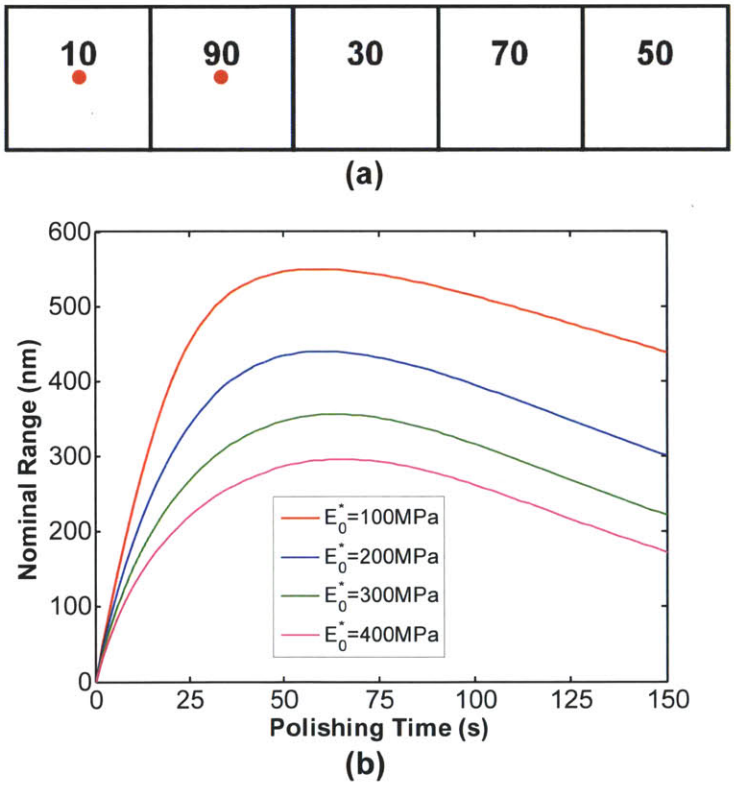


Figure 2.24: global planarization results of different pad bulk reduced moduli: (a) Nominal range monitor sites; (b) Nominal range evolution.

2.2.6 *Simulation: asperity height effect*

We consider pad asperity height effects in this subsection. The same die layout, initial topography, reference pressure, blanket removal rate and pad bulk reduced modulus are used as subsection 2.2.4. We vary the pad characteristic asperity height at 100nm, 150nm and 200nm. Polishing process is simulated up to 150 seconds.

When asperities are taller (the characteristic asperity height parameter λ is larger), slower step height reduction and lower planarization efficiency are observed from Figure 2.25. This relationship indicates that asperity height has a strong impact on local planarization. Taller asperities can touch down areas earlier during polishing so that early down area removal occurs, making the local planarization slow.

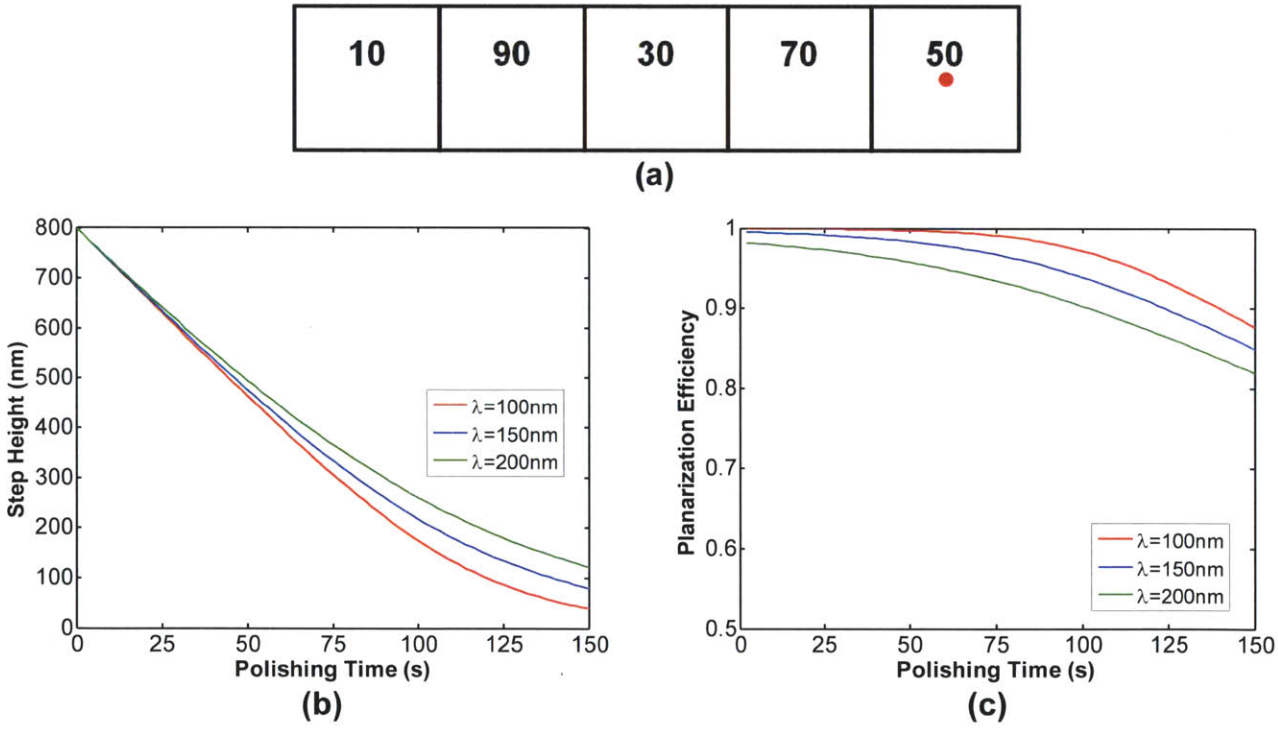


Figure 2.25: Local planarization results of different characteristic asperity heights: (a) Monitor site in 50% STEP array of MIT standard layout; (b) Step height evolution; (c) Planarization efficiency evolution.

For global planarization, asperity height has a minor effect. At the beginning of polishing, nominal range differences occur in Figure 2.26. However, when the process lasts a longer time, the nominal range goes to a similar final value, which means similar within-die non-uniformity.

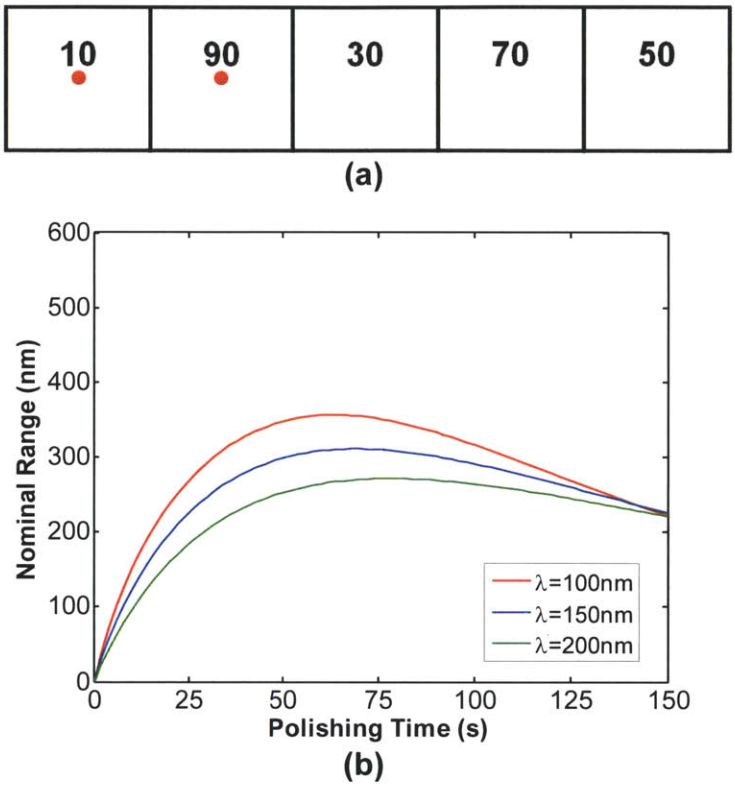


Figure 2.26: Global planarization results of different characteristic asperity heights: (a) Nominal range monitor sites; (b) Nominal range evolution.

2.2.7 Summary of physical die-level CMP model

A die-level CMP model is developed to understand die-level non-uniformity of CMP. A computational approach using discrete Fourier transform and iterative program is demonstrated, and model simulation cases are studied to understand pattern density dependencies. Physical parameters are considered by simulation and related to polishing results: pad bulk modulus has major impact on global planarization, while asperity height has major impact on local planarization.

2.3 Physical particle-level CMP model

Particle-level models usually focus on two important outputs of CMP: material removal rate and surface qualities (surface defects and scratching) [18]. However, in this thesis, a microscopic or particle-level model is developed to understand the interaction between wafer surface and pad asperities and the contact mechanism. Our model focuses on the $1\sim10\mu m$ scale, in order to better understand the contact between wafer and pad surfaces, including the fraction of the pad involved in that contact.

2.3.1 Model assumptions

The Greenwood-Williamson approach can be used to analyze rough surface contact. Before we follow this approach, assumptions need to be made to simplify pad surface profile. Recent works have reported measurements of the asperity height distributions on pad surfaces [84-86]. The asperity height distribution depends on both pad material and pad conditioning. Although the distribution varies, the active parts of asperities that create pad-wafer contact are only the few highest peaks (usually less than 5% of total pad area), as shown in Figure 2.27. Therefore, to understand the pad-wafer interaction and predict the contact between pad and wafer, we need to focus on those highest asperity peaks.

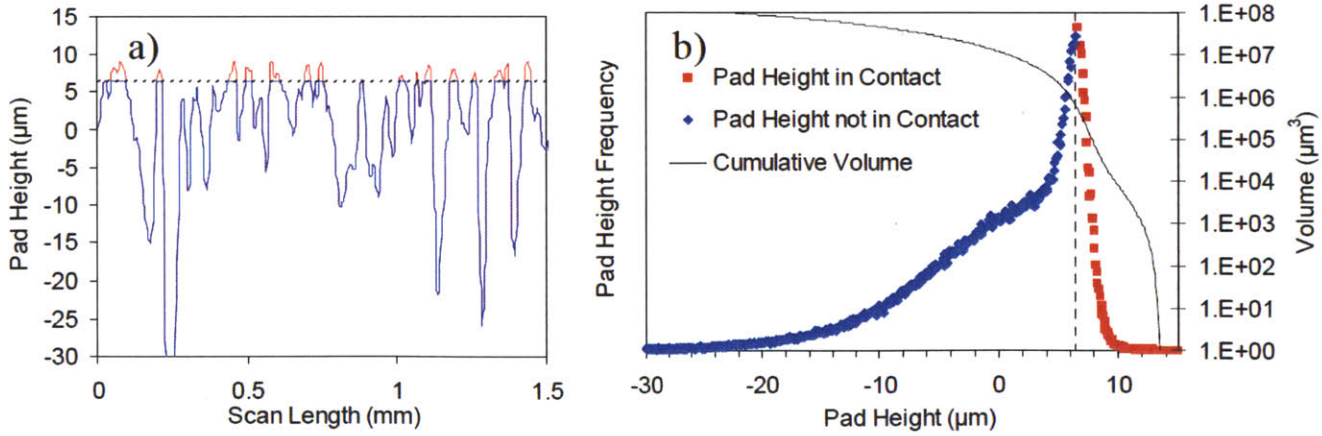


Figure 2.27: Active contact peaks of the surface profile on a glazed CMP pad [87]: (a) Line scan; (b) Pad height probability distribution.

Figure 2.28 illustrates the pad surface profile simplification we will use. We start with a measured pad surface and define the mean height as a reference plane where asperities are counted (Figure 2.28(a)), i.e., peaks above this reference plane are counted as asperities (Figure 2.28(b)). Then we denote the active plane, above which asperity elastic deformation is considered to make pad-wafer contact, i.e., through contact with active peaks. Assuming each active peak is independent of its neighbors according to the Greenwood-Williamson model, the "active peaks" and the pad underneath them are considered to be force responsive as individual posts (Figure 2.28(c)). Finally, the post heads are assumed to be spherical with a constant average radius of curvature (Figure 2.28(d)). This simplified pad profile containing active posts is ready to use in a Greenwood-Williamson approach.

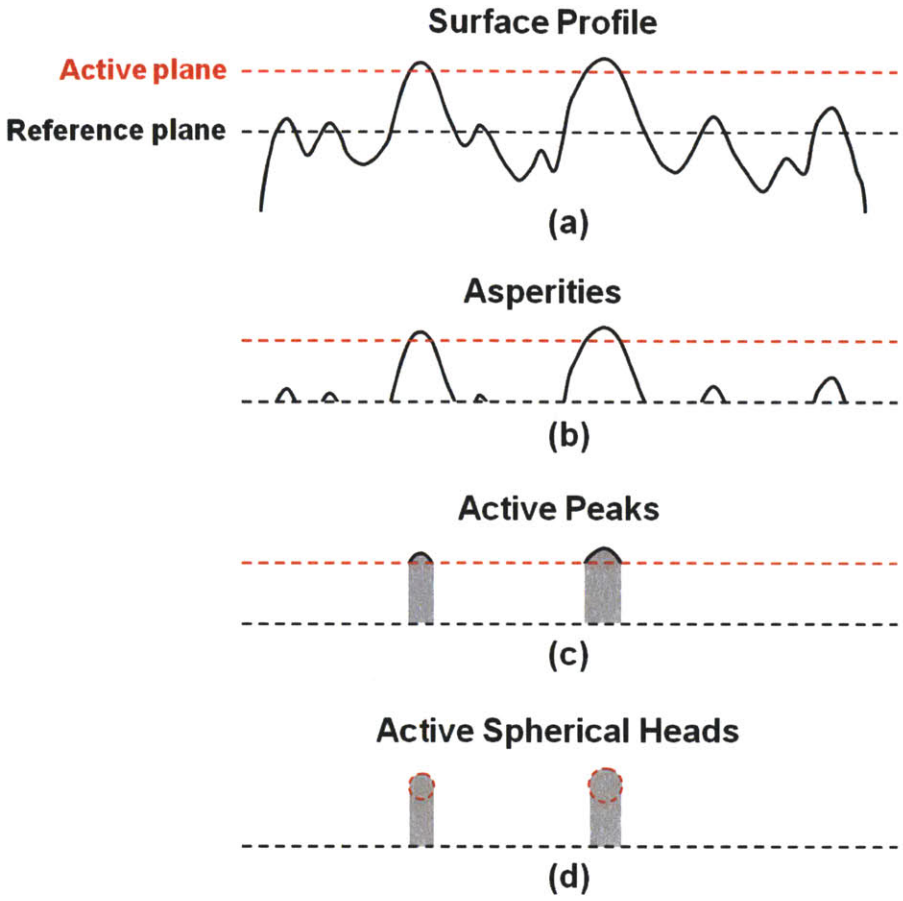


Figure 2.28: CMP pad surface profile simplification: (a) define reference plane and active parts of asperities; (b) find active asperities; (c) define active asperities as elastic posts; (d) define the shape of active post heads.

2.3.2 Model derivation

The model derivation can be broken down to two steps. First, we solve the elastic deformation problem of a single asperity (active peak) with height h when pressed upon the wafer surface. Wafer face up mathematical convention is used, and the wafer surface is assumed to be flat. If the asperity deformation is δ, we can express the following terms as functions of δ, as illustrated in Figure 2.29: the contact area $a(\delta)$, the single asperity force load $L(\delta)$, and the pressure distribution in the contact area $P(x, y; \delta)$. Second, we assume an asperity height distribution or probability density function $\phi(h)$, i.e., the number of asperities per unit area with height between h and $h + dh$ is $\phi(h)dh$. If the

distance between the wafer and the nominal surface of the pad is d, the asperities with height larger than d will be in contact with the wafer surface. The number of asperities in contact is

$$n = N\int_{d}^{\infty} \phi(h)\, dh \tag{2.41}$$

where N is the total number of active asperities. For the asperity with height $h > d$, the deformation is $\delta = h - d$. The total contact area is

$$A = N\int_{d}^{\infty} a(h-d)\phi(h)\, dh \tag{2.42}$$

For an applied force of F_0, the distance d can be obtained as

$$F_0 = N\int_{d}^{\infty} L(h-d)\phi(h)\, dh \tag{2.43}$$

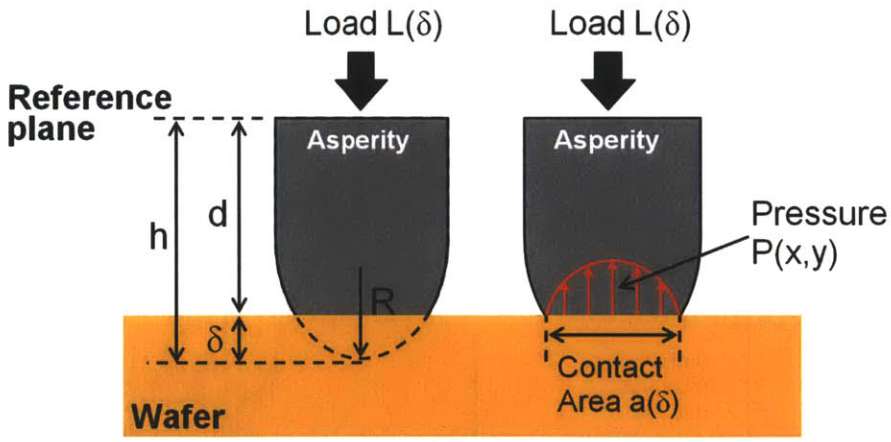


Figure 2.29: A single asperity (active peak) being compressed.

Greenwood [88] assumes that the asperities have spherical tops with the same radius R, and the contact is Hertzian [58]. Based on the same assumptions as Greenwood and using the Hertzian results, the contact area $a(\delta)$, the single asperity force load $L(\delta)$ and the pressure distribution in the contact area $P(x, y; \delta)$ can be expressed as [58]:

$$\begin{cases} a(\delta) = \pi R \delta \\ L(\delta) = \dfrac{4}{3} E_a R^{\frac{1}{2}} \delta^{\frac{3}{2}} \\ P(x, y; \delta) = P_c \left(1 - \dfrac{\pi(x^2 + y^2)}{a(\delta)} \right), \quad \pi(x^2 + y^2) \le a(\delta) \end{cases} \tag{2.44}$$

where E_a is the reduced modulus of the asperity and $P_c = \frac{3L(\delta)}{2a(\delta)}$ is the pressure at the center of the contact peak. Here it is assumed that the wafer material is much more rigid than the pad asperity.

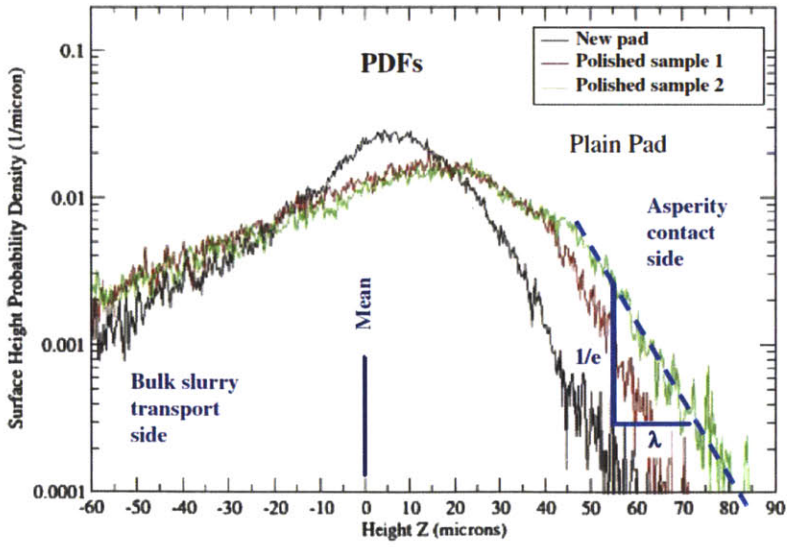


Figure 2.30: Example of characteristic asperity height (λ) extraction from interferometry data of conditioned CMP pad [84].

Measurements have shown (in Figure 2.30) that the asperity height distribution of a conditioned CMP pad follows an exponential decay for large asperity heights [84],

$$\phi(h)=\frac{1}{\lambda}e^{-\frac{h}{\lambda}} \tag{2.45}$$

where λ is the characteristic asperity height. Then the number of asperities in contact n, total contact area A, and the applied force F_0 can be determined by plugging Eq. 2.45 into Eq. 2.41 to 2.43 as

$$\begin{cases} n = Ne^{-\frac{d}{\lambda}} \\ A = N\pi R\lambda e^{-\frac{d}{\lambda}} \\ F_0 = E_a N\sqrt{\pi R\lambda^3}e^{-\frac{d}{\lambda}} \end{cases} \tag{2.46}$$

Assume the total pad area of interest is A_0. The nominal area of each active asperity is $A_s = \frac{A_0}{N}$. The area of a single active peak is $A_p = 4R^2$. Here we define a fraction parameter, β, called asperity occupation rate, that indicates the degree or intensity of packing of asperities together..

$$\beta = \frac{A_p}{A_s} = \frac{4NR^2}{A_0} \tag{2.47}$$

This parameter depends on both active plane selection and asperity radius of curvature. It can be used as a monitor to check the asperity independence assumption of the Greenwood-Williamson model: if β is fairly large, this model may not give a good estimation, as the asperities may then be close enough that they interact.

Considering the total area, we get a reference pressure as

$$P_0 = \frac{F_0}{A_0} = \frac{\beta E_a}{4}\sqrt{\frac{\pi \lambda^3}{R^3}} e^{-\frac{d}{\lambda}} \tag{2.48}$$

Since material removal only happens in the contact area during polishing [18], an important pad surface property is the contact area percentage under the applied reference pressure. Combining Eq. 2.46 and 2.48, we have

$$f(P_0) = \frac{P_0 \beta}{E_a}\sqrt{\frac{\pi R}{\lambda}} \tag{2.49}$$

Thus, we can now relate the contact area percentage to key pad surface geometry and mechanical properties.

2.3.3 Model trend

Using this model, we can calculate an example of pad-wafer contact. We choose asperity reduced modulus E_a = 300 MPa, asperity radius of curvature R = 30 μm, characteristic asperity height λ = 5 μm and asperity occupation rate β = 0.1. The contact percentage increases linearly according to applied reference pressure, as shown in Figure 2.31 and given by Eq. 2.49. If we fix the reference pressure at 5 psi and vary asperity modulus, the result is shown in Figure 2.32. Asperities with high modulus create less contact area, because the asperities are hard to deform.

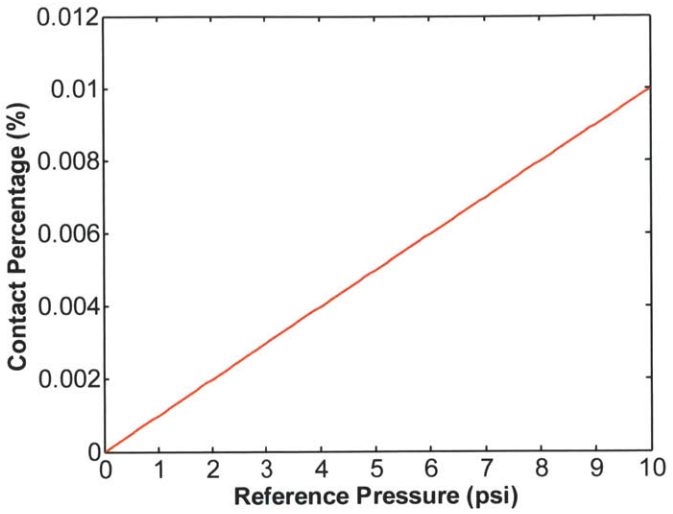


Figure 2.31: Pad-wafer contact percentage vs. applied reference pressure.

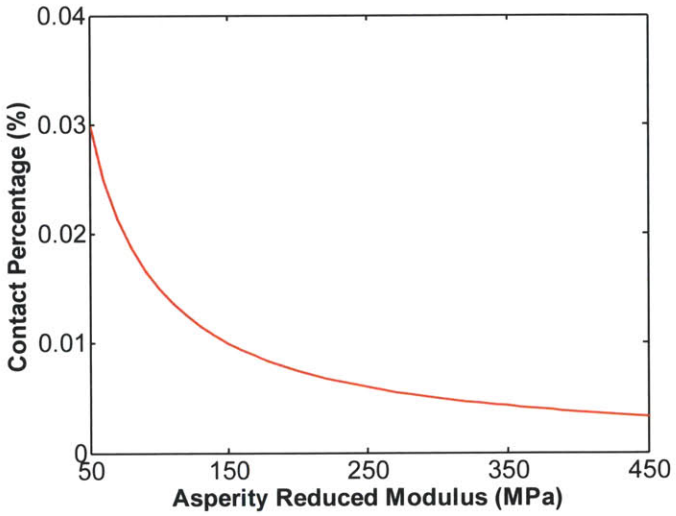


Figure 2.32: Pad-wafer contact percentage vs. asperity reduced modulus.

Asperity height distribution also affects the contact percentage. Figure 2.33 shows that the contact percentage decreases when the characteristic asperity height increases. This is because when there is a wide asperity height distribution (large λ), only a smaller number of the tallest asperities have the chance to contact the wafer, and bear the load with aggregate smaller contact area.

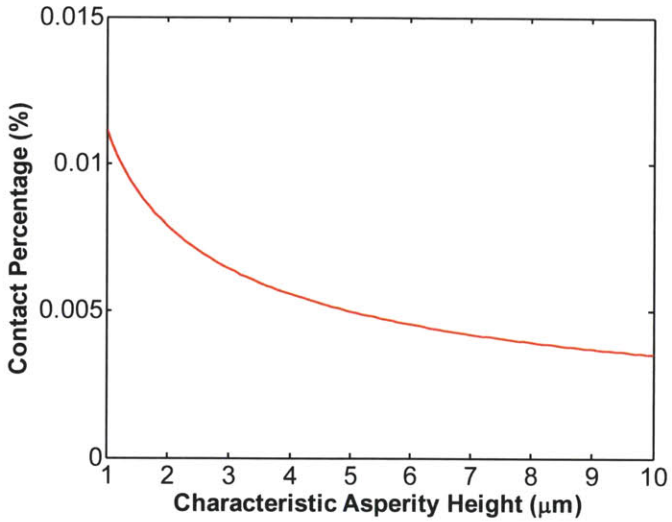


Figure 2.33: Pad-wafer contact percentage vs. characteristic asperity height.

2.3.4 Summary of physical particle-level CMP model

A physical model is proposed to understand the contact mechanism between CMP pad asperities and the wafer. Two main asperity properties are included in the model, asperity reduced modulus and asperity height distribution. Contact percentage between wafer and pad can be predicted by the model once we know the asperity properties. The model assumptions about asperity profile can be integrated into the die-level CMP model to include feature size effect; this integration is discussed in Section 2.5..

2.4 Model integration: extended wafer-die-level model

In a CMP process with wafer-scale pressure non-uniformity, the severity of the pattern density effect is a function of the die location on the wafer [89]. Thus, an integrated wafer-die-level polishing model is required to fully understand the effectiveness of the process for a given planarization requirement. Ouma [89] previously

proposed an empirical wafer and die joint model to relate wafer-level non-uniformity to die-level planarization. No physical model we are aware of has yet attempted to account for pattern dependent effects across the whole wafer. This section proposes an approach for model integration of wafer-level and die-level physical models. An extended wafer-die-level model is developed to include wafer-level pressure non-uniformity impact on die-level planarization non-uniformity.

2.4.1 Integration approach

The basic physical wafer-level and die-level models are developed in sections 2.1 and 2.2. The computational problems are also solved for each level. The proposed model integration approach focuses on the pressure boundary conditions as the mechanism to connect the two levels. Figure 2.34 illustrates the integration approach. Each single die has a specific position on the wafer. When the wafer-level pressure distribution is calculated, the average single die pressure is known at that single die position. The local average pressure for that die can be used as the die level model boundary conditions (Eq. 2.21) as

$$\frac{1}{A_{chip}} \iint_{\substack{chip \\ surface}} p(x,y)dxdy = P_{local}(X,Y) \tag{2.50}$$

where (x,y) indicates grid position within a die, (X,Y) indicates the die position on the wafer and $P_{local}(X,Y)$ is the local pressure calculated from the wafer-level model. Since the wafer-level pressure distribution is non-uniform, the wafer-level pressure impact will pass to the die-level polishing result through the boundary condition.

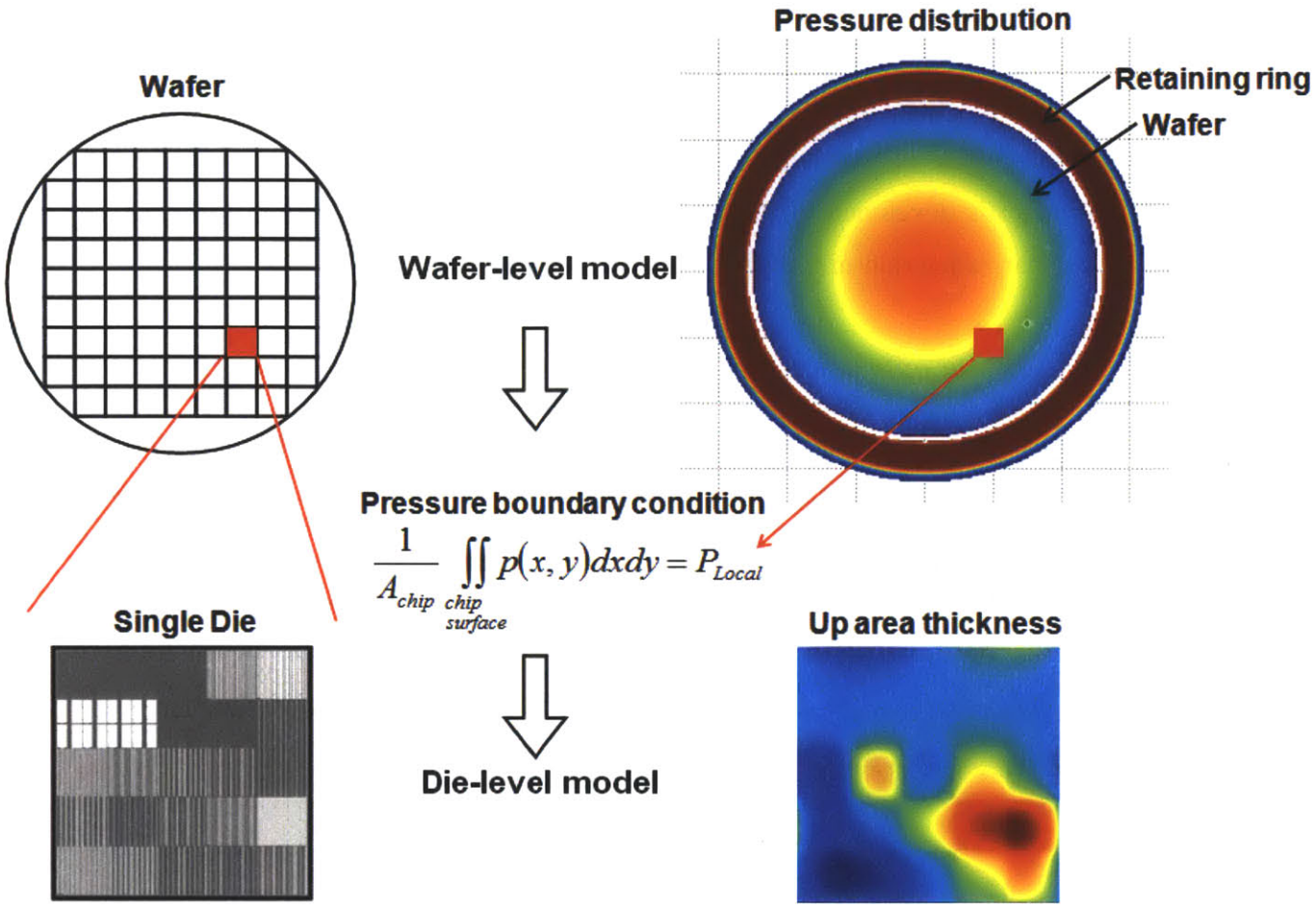


Figure 2.34: Model integration of wafer-level and die-level models through pressure boundary condition.

2.4.2 Remarks on model parameters

The integrated model combines all of the model parameters from the wafer-level and die-level cases, including CMP pad thickness, pad modulus, blanket removal rate and characteristic asperity height. We assume pad body modulus E at the wafer-level and pad bulk modulus E_0 at the die-level, respectively. Physical measurements have shown that pad material has an indent depth dependent modulus [70, 90]; that is, when pad surface deformation is deeper towards the body, the equivalent pad modulus is lower. So we allow these two modulus parameters to be different in the integrated model to capture the

indent depth dependence. Even though this indent dependence only occurs in the surface of the pad polymer matrix, a different pad body modulus is still helpful to account for the pad porous body and lamination effects in the macro scale force response. Therefore, when we calculate the pressure distribution at the wafer-level, we use measured pad body modulus E, because this level involves macro scale deformation. When we calculate the pressure distribution at the die-level, we use pad bulk modulus E_0, since step features on a die only induce micro scale pad deformations.

However, the different moduli assumption is only useful for simulation when the pad body modulus is measurable. As discussed in subsection 2.1.3, the pad body modulus does not change the wafer-level pressure distribution; rather, it works as a scaling factor of the pad deformation. The pressure distribution is, however, affected by pad thickness and retaining ring setup. Thus model fitting of polishing data having wafer-level pressure non-uniformity will not enable us to extract the pad body modulus. In this case, the pad body modulus is approximated to be equal to the pad bulk modulus.

2.4.3 Simulation: wafer-level pressure non-uniformity impact on die-level planarization

To illustrate the integrated model, a test simulation is run for a 200 mm SKW7-2 wafer (Figure 2.20). The polishing pad is assumed to be 3 mm thick, and the retaining ring sits 2 mm away from the wafer edge. Applied pressures on wafer and retaining ring are 5 psi and 6 psi, respectively. The wafer-level blanket removal rate under reference pressure is 250 nm/min. The characteristic asperity height is 100 nm. The pad body modulus (wafer-level) and pad bulk modulus (die-level) are assumed to be 100 MPa and 225 MPa separately.

Three dies are monitored on the wafer. Figure 2.35 shows the wafer-level pressure non-uniformity and locations of monitor dies. The center die has the highest wafer-level local pressure, while the edge die has the lowest. The wafer-level non-uniformity impacted within-die planarization results are compared in Figure 2.36. We see that the within-die step reduction is slower in the edge die than in the middle or center die.

Therefore, in this case the wafer-level pressure non-uniformity will require longer polishing times for edge dies to achieve the planarization target (a final step height). Simulations in section 2.1 show that higher wafer edge pressure is possible under some wafer-level tool setups; thus, by changing the tool setup, it should be possible to compensate and increase the edge planarization rate to improve uniformity.

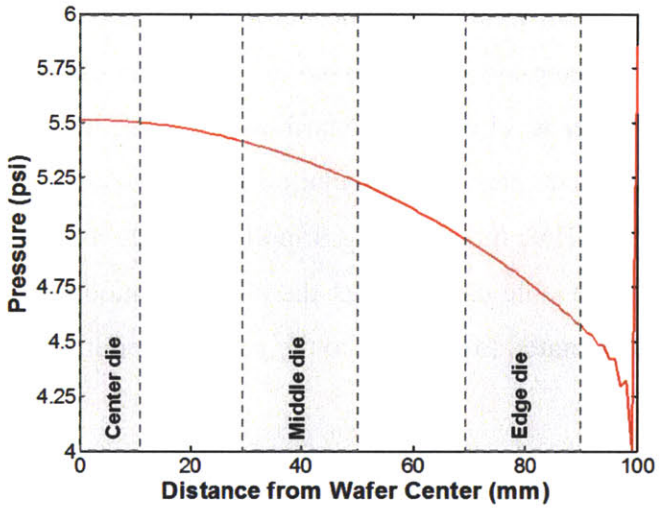


Figure 2.35: Wafer-level pressure distribution and covering range of monitor dies along wafer radius.

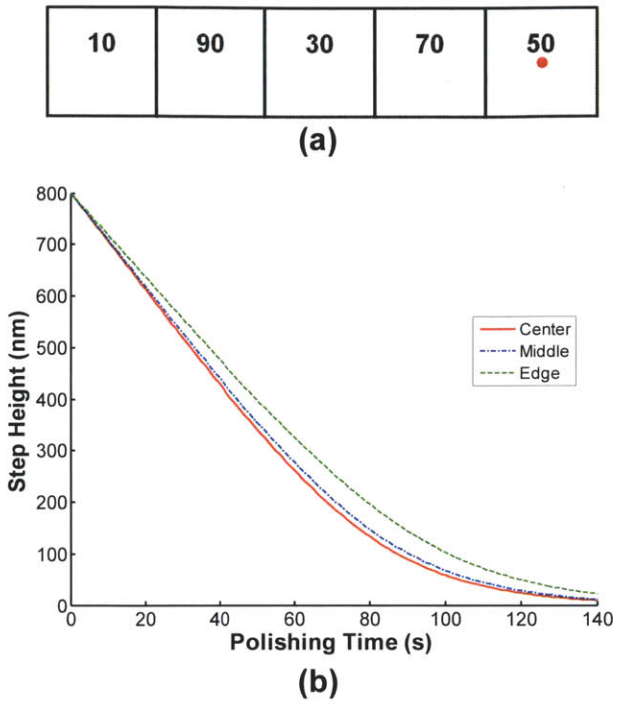


Figure 2.36: Within-die local planarization results from different dies on the wafer: (a) Monitor site in 50% STEP array of a single die; (b) Step height evolution of different dies.

2.4.4 Summary of wafer-die-level model integration

An integration approach for physical wafer-level model and die-level model is accomplished by connecting the pressure boundary condition between the wafer-level and the die-level. The effect of wafer-level pressure spatial non-uniformity on die-level planarization is estimated.

2.5 Model integration: extended die-particle-level model

The original physical die-level model (section 2.2) and the extended wafer-die-level model (section 2.4) have focused on the pattern density dependence of planarization, since it is known to be the dominant source of die-level variation [78]. However, a significant non-uniformity arising from different layout feature size is also observed in

oxide CMP [91]. Recent die-level model improvements make efforts to address this feature size effect, based on empirical die-level models [80, 92]. A physically based die-level model considering both pattern density effect and feature size effect is desired. This section develops an extended physical die-particle-level model by integrating the original physical die-level model and the particle-level model. Both pattern density and pitch size are included in the extended model.

2.5.1 Integration approach

This model integration is based on relating the feature size dependence to asperity size and asperity shape. Figure 2.37 illustrates the contact between asperities and different feature sizes. When the feature size is large, asperities can touch both up area and down area of the step structure, as seen in Figure 2.37(a). If the feature size is small, asperities will only touch the up area as Figure 2.37(b). Although both features have the same pattern density, small feature planarization is faster than large feature, because no down area removal occurs until a later polishing stage.

The shape of a step feature cross section on a chip is not an ideal rectangular shape during CMP, but rather has a rounding or roll-off of the sharp corners. Vasilev [80] proposed a parabolic shape approximation as shown in Figure 2.38, which can be merged with a Greenwood-Williamson approach by including the curvatures of up and down areas of each feature. Here we adopt Vasilev's approximation and use it to connect our die-level model and particle-level model.

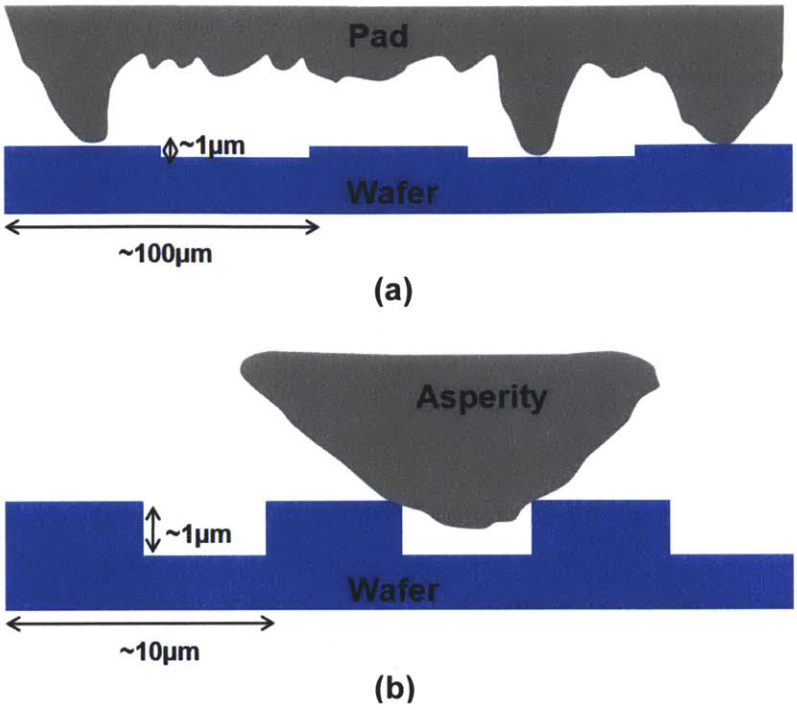


Figure 2.37: Contact between asperities and features with 50% density on a chip: (a) large feature size; (b) small feature size.

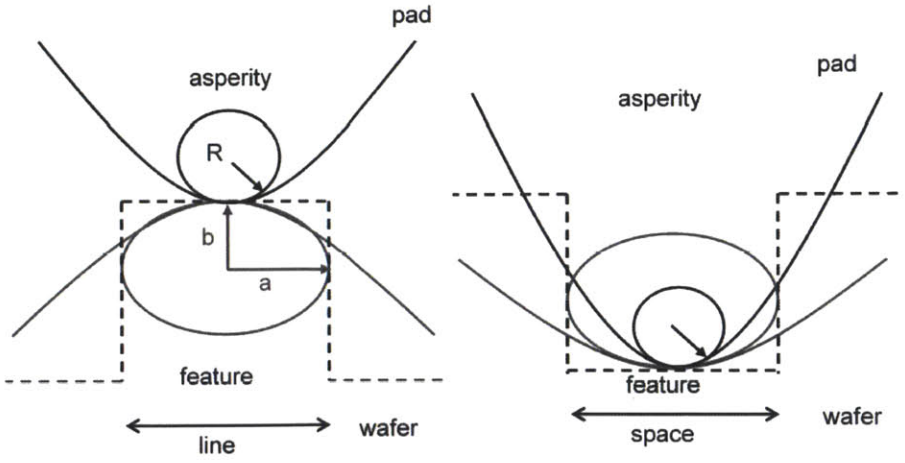


Figure 2.38: Geometry of the contact between a pad asperity and a feature on the die [80]. Both shapes are assumed to be described by parabolic curves in the vicinity of the point of first contact.

By using the assumption of Figure 2.38, the effective curvatures of the contact surfaces in up area and down area are calculated as

$$\begin{cases} \kappa^U = \kappa_{asp} + \dfrac{4\alpha h_s}{line^2} = \dfrac{1}{R_{asp}} + \dfrac{4\alpha h}{line^2} \\ \kappa^D = \kappa_{asp} - \dfrac{4\alpha h_s}{space^2} = \dfrac{1}{R_{asp}} - \dfrac{4\alpha h}{space^2} \end{cases} \tag{2.51}$$

where κ_{asp} is asperity top curvature, R_{asp} is asperity radius of curvature, h is step height and α is a geometric fit parameter to account for deviation of the real structure shape from the parabolic approximation. When $\alpha = 1$, the feature structure is ideal parabolic; when $\alpha < 1$, the feature structure is close to a rectangular step structure; when $\alpha > 1$, the feature structure is sharper than a parabolic curve and becomes similar to a triangular shape. The effective curvatures are dynamically changed during polishing due to the step height reduction, which represents the feature shape change.

Assuming Hertzian contact [58] and using the particle-level model derivation in Eq. 2.41-2.44, the asperity response force F_U and F_D, asperity contact area A_U and A_D, and asperity contact number n_U and n_D in both up area and down area of the step feature can be expressed as

$$\begin{cases} F_U = \dfrac{4}{3}\dfrac{E_a}{\sqrt{\kappa_U}} N\rho \int_d^\infty (z-d)^{\frac{3}{2}} \phi(z) dz \\ F_D = \dfrac{4}{3}\dfrac{E_a}{\sqrt{\kappa_D}} N(1-\rho) \int_{d+h}^\infty (z-d-h)^{\frac{3}{2}} \phi(z) dz \end{cases} \tag{2.52}$$

$$\begin{cases} A_U = \dfrac{\pi}{\kappa_U} N\rho \int_d^\infty (z-d)^{\frac{3}{2}} \phi(z) dz \\ A_D = \dfrac{\pi}{\kappa_D} N(1-\rho) \int_{d+h}^\infty (z-d-h)^{\frac{3}{2}} \phi(z) dz \end{cases} \tag{2.53}$$

$$\begin{cases} n_U = N\rho\int_d^{\infty}(z-d)^{\frac{3}{2}}\phi(z)dz \\ n_D = N(1-\rho)\int_{d+h}^{\infty}(z-d-h)^{\frac{3}{2}}\phi(z)dz \end{cases} \tag{2.54}$$

where N is the total number of asperities, E_a is the asperity reduced modulus, ρ is pattern density, d is the reference distance between pad bulk and wafer surface, and $\phi(z)$ is the asperity height distribution. Using the same exponential asperity height distribution as in the die-level model given by $\phi(z) = \frac{1}{\lambda}e^{-\frac{z}{\lambda}}$, Eq. 2.52-2.54 can be derived as

$$\begin{cases} F_U = \rho\dfrac{e^{\frac{h}{\lambda}}F_T\sqrt{\kappa_D}}{\sqrt{\kappa_U}(1-\rho)+e^{\frac{h}{\lambda}}\sqrt{\kappa_D}\rho} \\ F_D = (1-\rho)\dfrac{F_T\sqrt{\kappa_D}}{\sqrt{\kappa_U}(1-\rho)+e^{\frac{h}{\lambda}}\sqrt{\kappa_D}\rho} \end{cases} \tag{2.55}$$

$$\begin{cases} A_U = \rho\dfrac{e^{\frac{h}{\lambda}}F_T\sqrt{\kappa_D}\sqrt{\pi}}{E_a\sqrt{\lambda}\sqrt{\kappa_U}\left(\sqrt{\kappa_U}(1-\rho)+e^{\frac{h}{\lambda}}\sqrt{\kappa_D}\rho\right)} \\ A_D = (1-\rho)\dfrac{F_T\sqrt{\kappa_U}\sqrt{\pi}}{E_a\sqrt{\lambda}\sqrt{\kappa_D}\left(\sqrt{\kappa_U}(1-\rho)+e^{\frac{h}{\lambda}}\sqrt{\kappa_D}\rho\right)} \end{cases} \tag{2.56}$$

$$\begin{cases} n_U = \rho \dfrac{e^{\frac{h}{\lambda}} F_T \sqrt{\kappa_D}\sqrt{\kappa_U}}{E_a \sqrt{\pi}\lambda^{\frac{3}{2}}\left(\sqrt{\kappa_U}(1-\rho)+e^{\frac{h}{\lambda}}\sqrt{\kappa_D}\rho\right)} \\ n_D = (1-\rho)\dfrac{F_T \sqrt{\kappa_D}\sqrt{\kappa_U}}{E_a \sqrt{\pi}\lambda^{\frac{3}{2}}\left(\sqrt{\kappa_U}(1-\rho)+e^{\frac{h}{\lambda}}\sqrt{\kappa_D}\rho\right)} \end{cases} \tag{2.57}$$

where $F_T = F_U + F_D$ is the total local force on the die from the asperity response. Then we can find the average force acting on an asperity $\overline{F_{asp}^U}$ and $\overline{F_{asp}^D}$, the average contact area for an asperity $\overline{A_{asp}^U}$ and $\overline{A_{asp}^D}$, and the average pressure under an asperity $\overline{P_{asp}^U}$ and $\overline{P_{asp}^D}$ as

$$\begin{cases} \overline{F_{asp}^U} = \dfrac{F^U}{n^U} = \dfrac{E_a \sqrt{\pi}\lambda^{\frac{3}{2}}}{\sqrt{\kappa^U}} \\ \overline{F_{asp}^D} = \dfrac{F^D}{n^D} = \dfrac{E_a \sqrt{\pi}\lambda^{\frac{3}{2}}}{\sqrt{\kappa^D}} \end{cases} \tag{2.58}$$

$$\begin{cases} \overline{A_{asp}^U} = \dfrac{A^U}{n^U} = \dfrac{\lambda}{\kappa^U}\pi \\ \overline{A_{asp}^D} = \dfrac{A^D}{n^D} = \dfrac{\lambda}{\kappa^D}\pi \end{cases} \tag{2.59}$$

$$\begin{cases} \overline{P_{asp}^U} = \dfrac{\overline{F_{asp}^U}}{\overline{A_{asp}^U}} = \dfrac{E_a\sqrt{\kappa^U}\sqrt{\lambda}}{\sqrt{\pi}} \\ \overline{P_{asp}^D} = \dfrac{\overline{F_{asp}^D}}{\overline{A_{asp}^D}} = \dfrac{E_a\sqrt{\kappa^D}\sqrt{\lambda}}{\sqrt{\pi}} \end{cases} \tag{2.60}$$

We notice that the average force acting on an asperity, the average contact area for an asperity and the average pressure under an asperity are independent of total local force F_T. In fact, the size of any existing individual contact spot increases with force load, but at the same time new small spots are created to balance the force increment, which leaves the average unchanged. This behavior is induced by the exponential asperity height distribution assumed in the model.

The next step in the derivation is to make use of Preston's law. Since the force transmissions take place only over the asperity contact spots, the macro scale Preston's law $RR = K_p PV$ requires some careful modifications. Usually, the Preston coefficient K_p is taken as a constant containing all relevant effective material properties for polishing between two flat surfaces. For the polishing in the contact spot between an asperity and the wafer, Vasilev's [80] microscopic formulation of Preston's law can be adopted. That is, the removal caused by one contacting asperity is calculated as

$$\overline{RR_{asp}} = K_{asp}\,\overline{P_{asp}}V \tag{2.61}$$

where $\overline{P_{asp}}$ is the average real pressure under each asperity from Eq. 2.60, and K_{asp} is the microscopic Preston coefficient given by

$$K_{asp} = \frac{\pi\lambda}{\kappa_{asp}} K_p \tag{2.62}$$

Using Eq. 2.57, 2.60, 2.61 and 2.62, the local up area and down area removal rates can be calculated as

$$\begin{cases} RR_U = \overline{RR_{U,asp}} \dfrac{n_U}{A\rho} = \dfrac{e^{\frac{h}{\lambda}} \kappa_U \sqrt{\kappa_D}}{\kappa_{asp}\left(\sqrt{\kappa_U}(1-\rho)+e^{\frac{h}{\lambda}}\sqrt{\kappa_D}\rho\right)} \dfrac{F_T}{A} K_p V \\ RR_D = \overline{RR_{D,asp}} \dfrac{n_D}{A(1-\rho)} = \dfrac{\kappa_D \sqrt{\kappa_U}}{\kappa_{asp}\left(\sqrt{\kappa_U}(1-\rho)+e^{\frac{h}{\lambda}}\sqrt{\kappa_D}\rho\right)} \dfrac{F_T}{A} K_p V \end{cases} \tag{2.63}$$

where A is the total local area. Noticing that the local pressure is $p(x,y) = \frac{F_T}{A}$, Eq. 2.63 can be expressed as

$$\begin{cases} RR_U = p_U(x,y) K_p V = K_0 \dfrac{p_U(x,y)}{P_0} \\ RR_D = p_D(x,y) K_p V = K_0 \dfrac{p_D(x,y)}{P_0} \end{cases} \tag{2.64}$$

$$\begin{cases} p_U(x,y) = \dfrac{e^{\frac{h}{\lambda}} \kappa_U \sqrt{\kappa_D}}{\kappa_{asp}\left(\sqrt{\kappa_U}(1-\rho)+e^{\frac{h}{\lambda}}\sqrt{\kappa_D}\rho\right)} p(x,y) \\ p_D(x,y) = \dfrac{\kappa_D \sqrt{\kappa_U}}{\kappa_{asp}\left(\sqrt{\kappa_U}(1-\rho)+e^{\frac{h}{\lambda}}\sqrt{\kappa_D}\rho\right)} p(x,y) \end{cases} \tag{2.65}$$

where $K_0 = K_p P_0 V$ is the blanket removal rate under reference pressure P_0. Here $p_U(x,y)$ and $p_D(x,y)$ are the nominal local up area pressure and down area pressure. Based on Eq. 2.64 and 2.65, we see that the die-level pressure distribution is related to local material removal. Once the local pressure $p(x,y)$ is solved, the die-level wafer topography evolution can be estimated.

Referring Eq. 2.52 and the exponential asperity height distribution, the local pressure from asperity response is represented as

$$p(x,y) = k'\left[\frac{\rho}{\sqrt{\kappa_U}} + \frac{(1-\rho)}{\sqrt{\kappa_D}} e^{-\frac{h(x,y)}{\lambda}}\right] \cdot \lambda^{\frac{3}{2}} e^{-\frac{w(x,y)-z_u(x,y)}{\lambda}} \tag{2.66}$$

where $k' = \frac{E_a N\sqrt{\pi}}{A}$ is the effective spring constant, $z_u(x,y)$ is the wafer profile and $w(x,y)$ is the pad bulk profile. The pressure from pad bulk response remains the same as Eq. 2.18 in the original physical die-level model. Therefore, the pressure and pad bulk deflection are described by

$$\begin{cases} p(x,y) = k'\left[\frac{\rho}{\sqrt{\kappa_U}} + \frac{(1-\rho)}{\sqrt{\kappa_D}} e^{-\frac{h(x,y)}{\lambda}}\right] \cdot \lambda^{\frac{3}{2}} e^{-\frac{w(x,y)-z_u(x,y)}{\lambda}} \\ w(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(x-x', y-y') \cdot p(x',y') dx' dy' + w_0 \end{cases} \tag{2.67}$$

This problem is also subject to the boundary conditions of Eq. 2.21. Since Eq. 2.67 has a similar format as Eq. 2.29, it can be solved by the same computational approach discussed in subsection 2.2.3. Then time-stepped chip topography evolution is enabled by Eq. 2.64 and 2.65.

Comparing to the original die-level model, the extended model includes feature size by calculating curvatures of contact surfaces. When asperity size decreases or feature size increases significantly, the extended model reduces to the original model due to a weak size effect.

2.5.2 Simulation: pitch size effect in CMP

In this section, we compare the polishing simulation results from the extended die-particle-level model and the original die-level model. The chip layout and initial topography of a die on an SKW7-2 wafer (Figure 2.20) is used. The applied reference pressure is assumed to be 5 psi. Model parameters are listed in Table 2.1.

Table 2.1: Parameters used in the pitch size effect simulations.

Parameter	Extended die-particle-level model	Original die-level model
Blanket removal rate K_0 (nm/min)	**200**	**200**
Pad bulk reduced modulus E_0^* (MPa)	**320**	**320**
Characteristic asperity height λ (nm)	**100**	**100**
Asperity radius of curvature R_{asp} (μm)	**60**	
Feature shape factor α	**10**	

Simulated step height evolutions are compared in Figure 2.39. The extended die-particle-level model offers the capability to capture the feature size dependence, as shown in Figure 2.39(b). Small features (10 μm pitch) are planarized very fast, because the down areas of small features are not touched by the asperities. The original die-level model only considers the pattern density effect, so the predictions for various pitch sizes do not have significant differences, as shown in Figure 2.39(c).

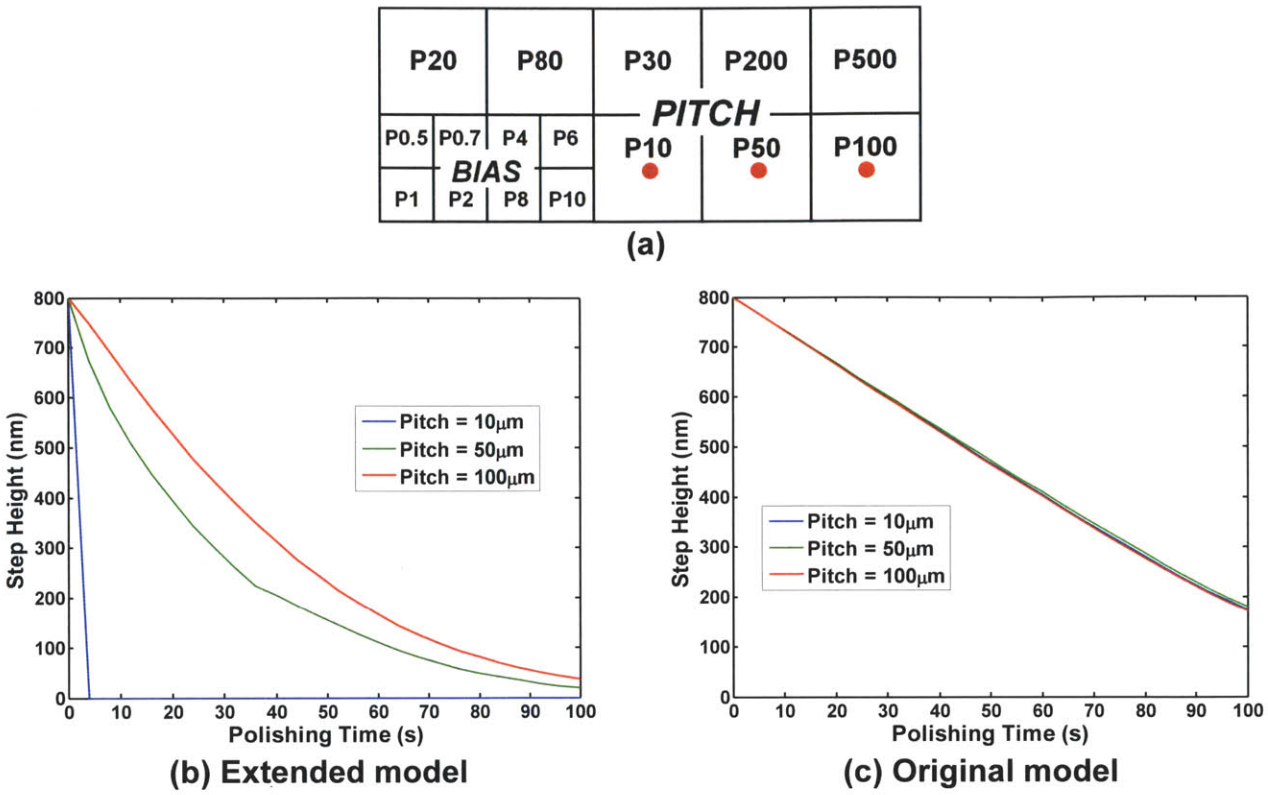


Figure 2.39: Pitch size effect in CMP: (a) Monitor sites in PITCH arrays (50% pattern density) of MIT standard layout; (b) step height evolution simulated by the extended die-particle-level model; (c) step height evolution simulated by the original die-level model.

We can investigate the asperity shape effect by varying the asperity radius of curvatures in the simulation. Figure 2.40 compares the step height reductions in the 50% density area for different asperity radius. Smaller asperity radius results in slower planarization, because the down area is easier to be touched by smaller asperities.

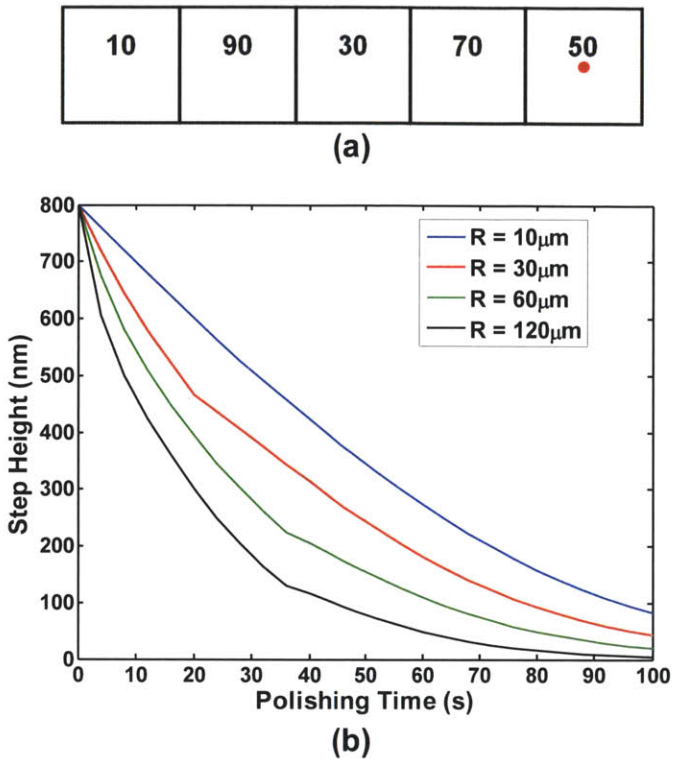


Figure 2.40: Asperity radius of curvature effect on planarization: (a) Monitor site in 50% STEP array of a single die; (b) Step height evolutions.

2.5.3 Summary of die-particle-level model integration

An extended physical die-particle-level model is derived by integrating particle-level and die-level models together. The extended die-level model includes pad modulus, asperity shape and asperity height distribution. It combines both pattern density and feature size effects, and is able to explain the rapid planarization of features with small spacing between those features.

2.6 Summary

In this chapter, three physical CMP models are developed at the wafer-level, die-level and particle-level, separately. The wafer-level model investigates the CMP tool effects on wafer-level pressure non-uniformity. CMP pad thickness, retaining ring size

and retaining ring reference pressure are strong factors related to wafer-level non-uniformity. The die-level CMP model is developed to study die-level non-uniformity of polishing result, where the pattern density dependence is captured. Pad properties including pad bulk modulus and pad asperity height distribution are related to planarization performance. The particle-level model focuses on the contact mechanism between pad asperities and the wafer. Pad-wafer contact percentage can be predicted. A modeling priority is the die-level non-uniformity of polishing, which is a major concern of both layout designers and process engineers. Therefore, two model integration approaches are proposed to connect the die-level model to the wafer-level and particle-level, so that CMP system impacts on die-level uniformity and feature size dependence are considered. Figure 2.41 shows the overall modeling framework of this chapter. The basic physical models and extended models relate CMP tool and process factors to chip-scale non-uniformity. In the reverse direction, if we have a non-uniform polishing result, model fitting and model parameter extraction can help to find the inducing factors.

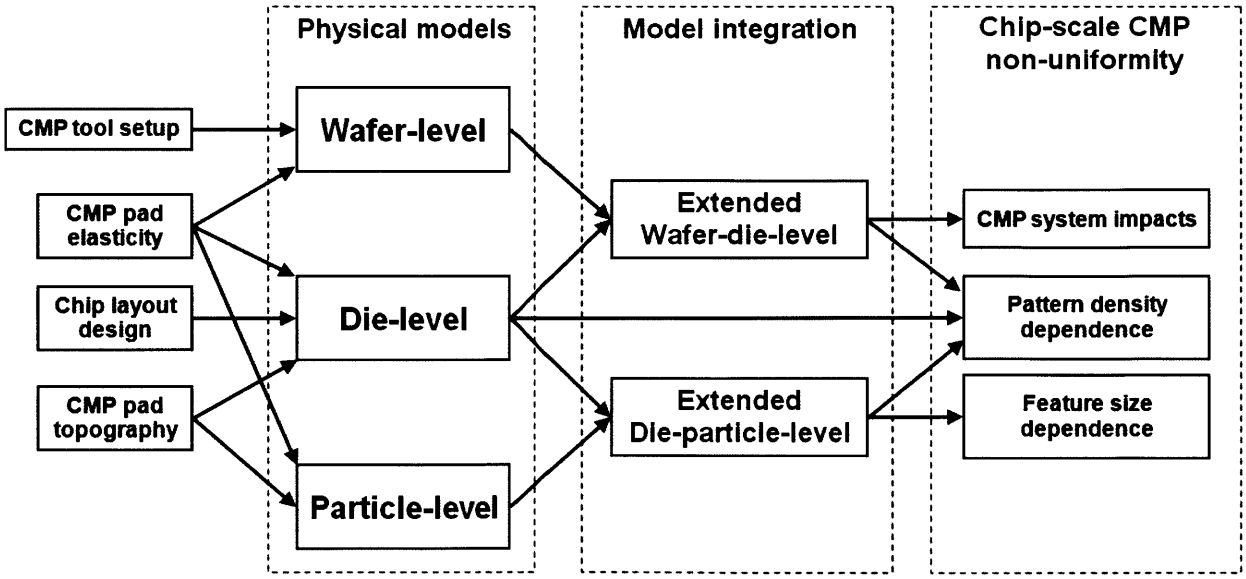


Figure 2.41: Summary of physical CMP modeling.

3 Applications of physical CMP models

Physical CMP models can be applied in many ways, such as verifying chip layout design, testing process parameters, evaluating CMP consumable properties and optimizing CMP processes. Compared to exploring these questions with polishing experiments, the CMP models have advantages in reducing time and cost. In this chapter, CMP models are applied to a number of practical CMP process problems. Our model applications are focused on, but not limited to chip-scale non-uniformity. Section 3.1 explains the general methodology of applying CMP models. Section 3.2 utilizes the physical die-level model to characterize CMP pad properties. Section 3.3 introduces the CMP endpoint variation analysis enabled by the model simulation. Section 3.4 evaluates wafer-level non-uniformity impact on die-level non-uniformity using the extended wafer-die-level model. Section 3.5 analyzes the pitch size effect in the oxide removal stage of STI CMP.

3.1 Methodology of applying physical CMP models

A physical CMP model can capture some main factors of a specific CMP process. Once the model parameters are extracted from experimental data, the model is calibrated and ready to use in process simulation. The general model application methodology has four steps:

1) Design of the experiment: Test wafers/layout patterns are designed to assess the polishing performance and enable the study of target factors in a CMP process. Process parameters are selected to explore the target polishing requirements.

2) Polishing tests: Test wafers are polished under specified processes. Wafer/die topography including film thickness and step height is measured before and after

polishing. Different polishing time splits are preferred when dynamic evolution of topography needs to be captured.

3) Model parameter extraction: The experiment data is fit against the CMP model. A set of optimized model parameters are chosen to minimize the fitting error, i.e., miss-match between model calculation and experiment result. The model is "calibrated" once the optimized model parameters are extracted.

4) Model simulation: With the calibrated model, new wafer/layout designs are taken as model input for the CMP process simulation and the polishing result can be predicted.

CMP model fitting and prediction are important in high yield semiconductor manufacturing. Without the assistance of CMP models, the manufacturing is a single stream from layout design to fabrication: the layout designer follows a set of design rules; the process engineers tune the process parameters to obtain acceptable polishing results and yield. As the chip design becomes more and more complex and the process control is more and more challenging, the design-to-fabrication single stream cannot guarantee a successful high yield manufacturing, even after many efforts from both design and process ends.

When a CMP model is applied, a system with feedback can be built for design and fabrication as shown in Figure 3.1. A precisely calibrated CMP model is a key component of the system. A chip layout is provided by the designers, while process engineers suggest possible process parameters. Both the layout design and the process parameters are sent to the calibrated CMP model as inputs to predict the polishing results. The model simulation will be iterated multiple times with changing process parameters or layout design, until the predicted results meet the manufacturing requirements. Then the model proved design and process are tested in the fab. If the test results are satisfactory, the design and process can be considered for production. If the fab test results do not meet the requirements or disagree with the simulation results, the polishing data will be used to recalibrate the CMP model until they agree. In this way, the fab test helps to tune and upgrade the model as the process improves or changes over time.

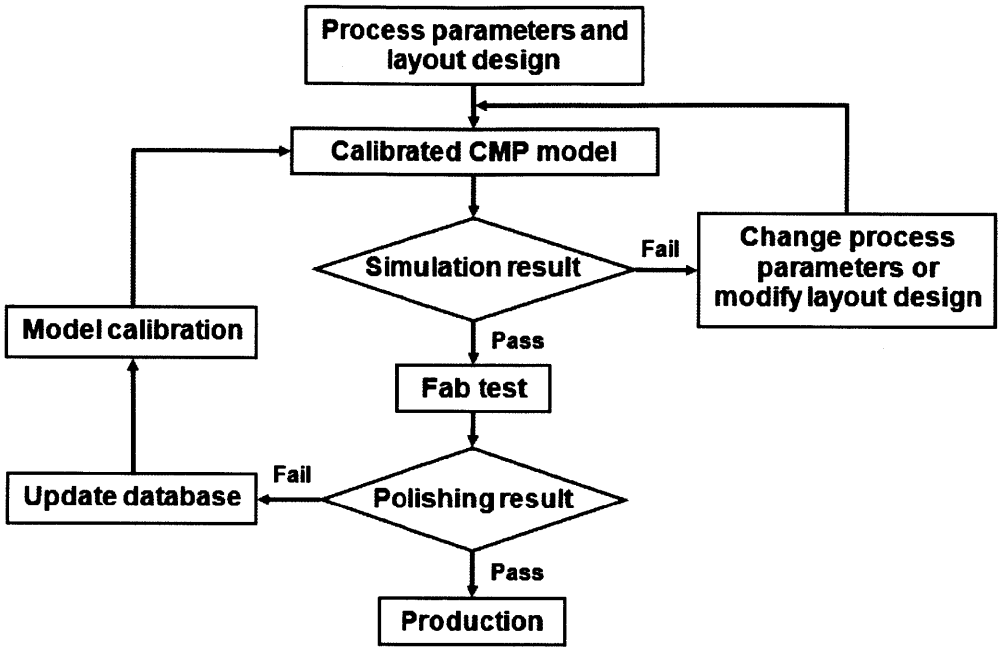


Figure 3.1: IC design and manufacturing with assistance of CMP model.

CMP model assistance reduces both time and cost for manufacturing development. Model simulation requires less time than a fab test experiment. Fewer fab tests are needed compared to single stream development, when there are complex dependencies between the layout design and CMP fabrication process. Even in cases where fab test results fail to meet requirements and disagree with the model, we still gain data to upgrade the model and drive the whole system toward more accuracy and robustness.

3.2 Study on stiffness and conditioning effects of CMP pad based on physical die-level CMP model

Understanding the relationships between CMP pad properties and planarization performance is important for both IC process engineers and CMP pad vendors to meet fabrication requirements. Physical model parameters are good abstractions of pad

properties, and these CMP pad properties can be studied by extracting physical model parameters.

There are three key parameters in the physical die-level CMP model: blanket removal rate K_0, effective pad bulk modulus E_0^*, and characteristic asperity height λ. The blanket removal rate is affected by many CMP tool, consumable, and process parameters, such as the CMP system reference pressure. The effective modulus is related to properties of the pad bulk, and is hypothesized to most strongly impact within die uniformity and layout pattern density effects, resulting from long range pad bending due to differential removal rates in different die pattern density regions. The characteristic asperity height reflects the distribution of pad asperity heights, and is hypothesized to most strongly impact the feature scale step height reduction. The pad property related parameters, E_0^* and λ, are independent from each other; in the model we assume that the CMP pad can be divided into two parts, pad bulk and asperities. E_0^* is only affected by pad bulk stiffness, while λ is only affected by the pad surface asperity height distribution. Thus, extracting these two independent parameters from experimental data enables investigation of pad bulk and surface properties separately.

In this section, specific CMP pad properties are related to the physical CMP model parameters, enabling us to clarify how pad bulk and surface properties affect within-die planarization. Effects of pad stiffness and conditioning disk diamond shape are investigated so as to intentionally modify pad bulk and surface properties, and fits of the model to experimental data from the polishing of patterned wafers are used to relate planarization model parameters to pad properties.

3.2.1 Experimental method

To evaluate the effect of pad stiffness, three water soluble particle (WSP) pads (produced by JSR Corporation) were engineered to have different stiffnesses (low, standard and high), and used in patterned wafer polishing. In the conditioning effect study, three conditioning disks with different diamond shapes (sharp, standard and

blocky) were applied on JSR standard pads. These diamond shapes are expected to "cut" into the pad surface differently, generating different pad asperity surface structures.

SKW7-2 oxide wafers patterned with an MIT CMP test layout were polished using JSR WSP pads. We applied two different sets of experimental conditions, to emphasize either the pad bulk dependence or the pad surface and asperity structure dependence, while the CMP tool recipe was kept the same, as listed in Table 3.1. In the first set of experiments, we polished wafers using pads with different stiffnesses (low, standard and high) while we used the same conditioning disk. In the second set of experiments, we polished wafers using a standard pad while we applied three different conditioning disks (sharp, standard and blocky). The first set investigates pad stiffness effect, and the second set verifies pad conditioning effects. Each test wafer was polished under a reference pressure of 5 psi for different time intervals accumulative to more than two minutes: 0, 20, 40, 70, 90, 110 and 130 seconds.

Table 3.1: Experimental conditions for pad property studies.

	Pad stiffness effect	**Conditioning effect**
Polishing machine	**MIRRA/MESA (AMAT)**	
Polishing pad	**JSR low stiffness** **JSR standard stiffness** **JSR high stiffness**	**JSR standard**
Conditioning disk	**Mitsubishi 325 grit**	**Read sharp** **Read standard** **Read blocky**
Slurry	**Cabot SS25 50% diluted**	
Platen speed (rpm)	**62**	
Polishing head speed (rpm)	**56**	
Conditioning head speed (rpm)	**60**	
Wafer reference pressure (psi)	**5**	
Retaining ring reference pressure (psi)	**6**	
Conditioning head down force (lbf)	**4**	
Slurry flow rate (ml/min)	**150**	
Pad break-in time (min)	**10**	

Optical measurement of film thickness and profilometry measurement of step height were performed after each polishing interval. On each wafer, we select a center die, a middle die and an edge die to test, as shown in Figure 3.2. Figure 2.20 shows the pattern type and local pattern density within each die of the SKW7-2 wafer. For each testing die, oxide thickness and step height are measured on five points in the STEP blocks with local pattern densities of 30%, 70% and 50% respectively. Oxide thickness data is used to extract the necessary model parameters and simulate polishing performance of each pad using the physical die-level CMP model. We extract model parameters by fitting to experimental data and minimizing the fitting error from time split experiment data. The hypothesized pad property effects are seen clearly in the pad parameter and model simulation results, as discussed below.

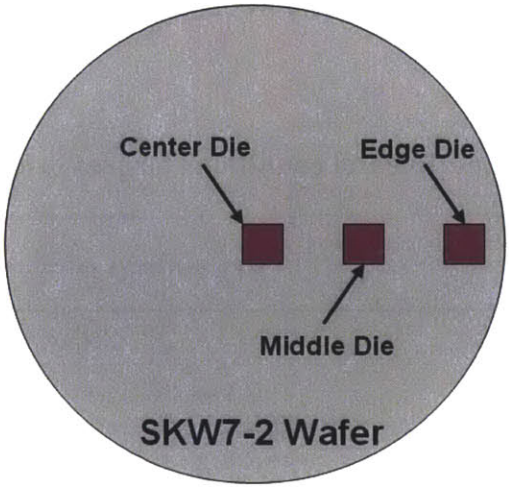


Figure 3.2: Measured die positions on an SKW7-2 dielectric polishing test wafer.

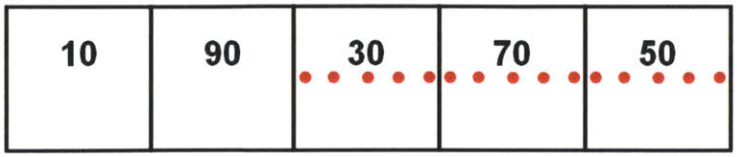


Figure 3.3: Measurement site positions in the STEP blocks in a die of SKW7-2 dielectric polishing test wafer.

3.2.2 CMP Pad Stiffness Effect

JSR pads with different stiffnesses (low, standard and high) were tested by polishing experiments and model data fitting. Based on the physical model, fitting results were hypothesized to have different effective moduli, but similar asperity heights and blanket removal rates. Figure 3.4 compares the extracted model parameters. As expected, effective modulus varies strongly corresponding to pad bulk stiffness. In addition, characteristic asperity height and blanket removal rate do not change substantially. This is consistent with an expectation that conditioning (which was the same for all JSR pad types used here) is the dominant factor in determining pad asperity height and contact properties.

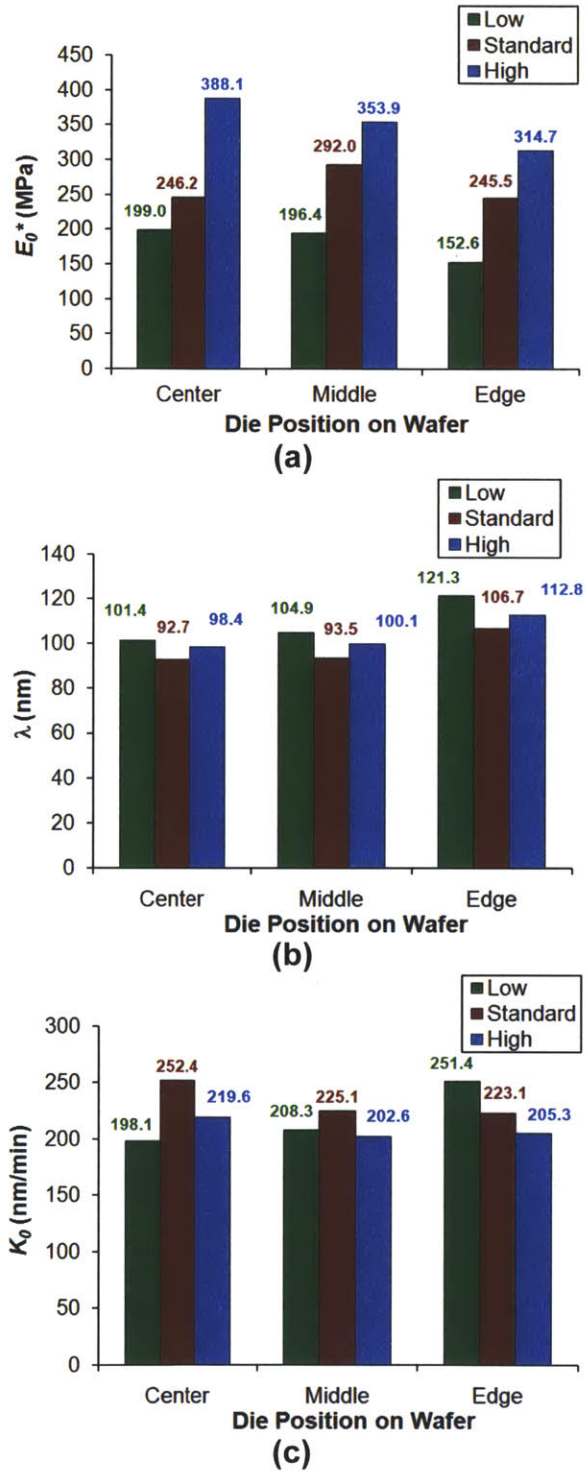


Figure 3.4: Extracted model parameters for different pad stiffnesses (low, standard, and high): (a) Effective modulus. (b) Characteristic asperity height. (c) Blanket removal rate.

In the following discussion, we only focus on the middle die, and do not consider in detail the across-wafer variation observed in the measured results. Our model application in this section is die-level only, and does not seek to account for CMP tool design or process non-uniformities in pressure, velocity, pad microstructure, conditioning profile, or other parameters across the wafer.

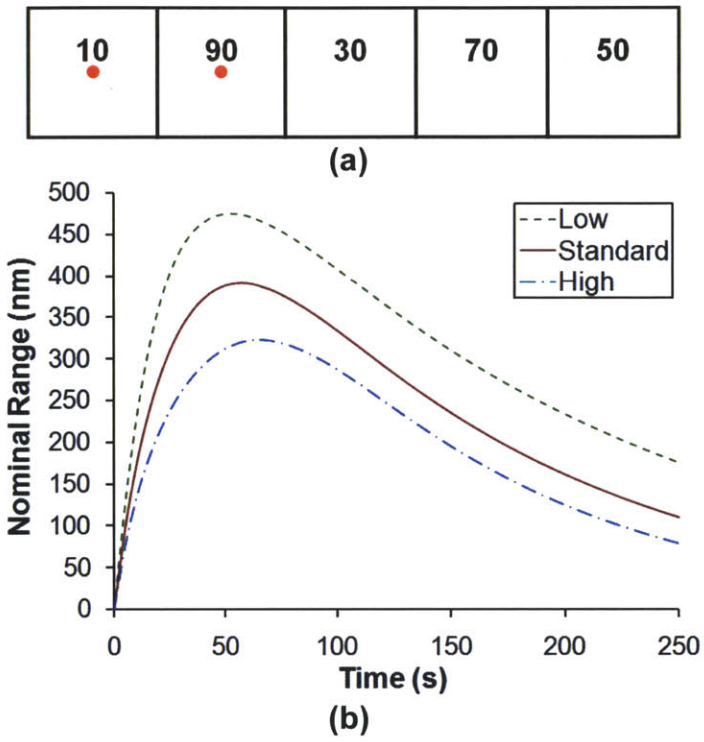


Figure 3.5: Simulated middle die nominal range evolution for different pad stiffnesses: (a) Nominal range monitor sites; (b) Nominal range evolution.

A useful output from our model to reflect the main within-die uniformity effect of pad stiffness is the "nominal range," which is defined as the difference between the up area oxide thickness of the 90% pattern-density area and that of the 10% pattern-density area. Within-die non-uniformity results from faster removal rates in low pattern density regions on the die compared to high pattern density areas: the local pad force is applied to a smaller number of up features in the low pattern density region, generating much higher local pressure and resulting in a large "recessed" or "eroded" region on the die. Across

lateral distances of several millimeters, this within-die non-uniformity can be substantial, with oxide thickness differences of several hundred nanometers. To combat this effect, pads with higher stiffness are of interest, as they are subject to less long-range bending. Across the oxide step structures, hard pads deflect less into recessed low pattern density areas, reducing the regional polish rates compared to high pattern density regions, resulting in a smaller final nominal range, as shown in Figure 3.5. In this figure, the creation of within-die non-uniformity in the first 50-70 seconds due to pattern density removal rate dependence is seen, with subsequent reduction as polishing progresses. During all times the stiffer pads create less nominal thickness range between 90% and 10% pattern density regions.

We can also understand the pad stiffness effect versus asperity height effect by comparing step-height evolution in the polishing process. Figure 3.6(a) shows the data fitting results of step-height vs. time on the center point of a middle die in the 50% pattern density array. We see that the high stiffness pad maintains a linear step-height reduction region to a comparable remaining step height (at which point an exponential decay in time occurs) across all three pads, corresponding to our expectation that bulk pad stiffness is not the primary factor in individual feature step height removal or down area polish. Comparing the step-height evolution at center (Figure 3.6(a)), left edge (Figure 3.6(b)) and right edge (Figure 3.6(c)) points of the 50% pattern density array on the middle die, however, we see the pattern density dependence of polishing as expected. As indicated by the red dashed lines, step height reaches to 200nm remaining height more slowly (at about 95s) at the left edge point (next to 70% pattern density array) and more quickly (at about 66s) at the right edge point (next to 10% pattern density array). This is the result of the long-range pad bending, which results in different "effective" pattern density at different points within the same local pattern density region, due to spatial "averaging" of applied pad pressure across the die.

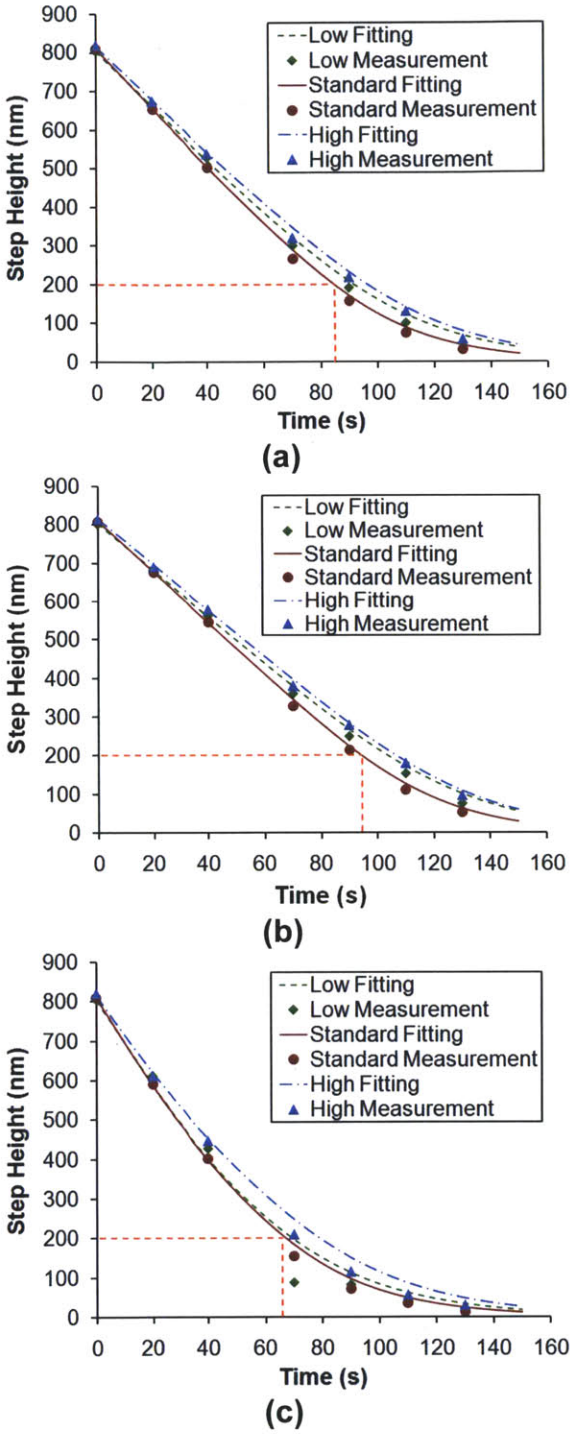


Figure 3.6: Middle die 50% pattern density array step height evolution for different pad stiffnesses (low, standard, and high): (a) Center point. (b) Left edge point (next to 70% pattern density array in Figure 3.3). (c) Right edge point (next to 10% pattern density array in Figure 3.3).

3.2.3 CMP Pad Conditioning Effect

JSR standard pads with different conditioning disk diamond type (blocky, standard and sharp) were tested. Since we used the same type of pad, extracted effective modulus is similar for each conditioning disk (Figure 3.7(a)), as expected. Especially for the middle die, the modulus differences are very small (less than 2.5%). But the characteristic asperity height is consistently different (Figure 3.7(b)); the characteristic asperity height varies corresponding to diamond shape. The blanket removal rate does not change substantially (Figure 3.7(c)).

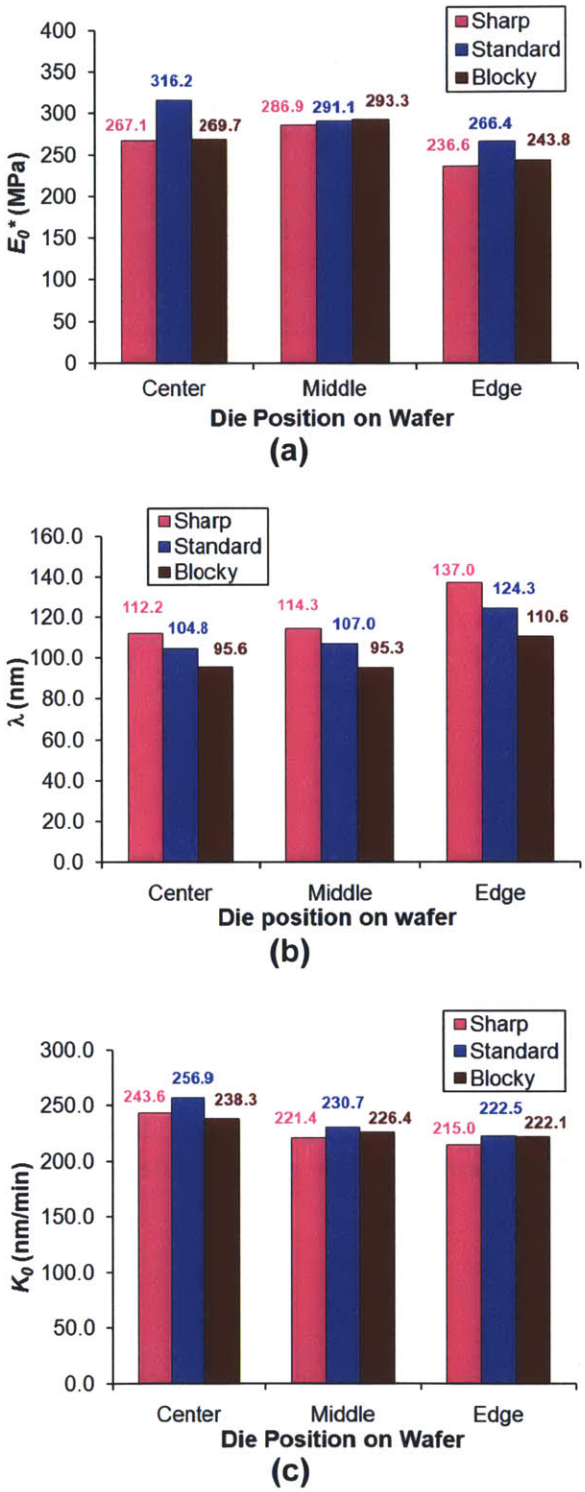


Figure 3.7: Extracted model parameters for different conditioning disk diamond shapes: (a) Effective modulus. (b) Characteristic asperity height. (c) Blanket removal rate.

In the conditioning test, the simulated nominal range (or within-die oxide thickness uniformity) is not impacted by characteristic asperity height significantly, and thus the range evolutions are similar for different conditioning disks (Figure 3.8).

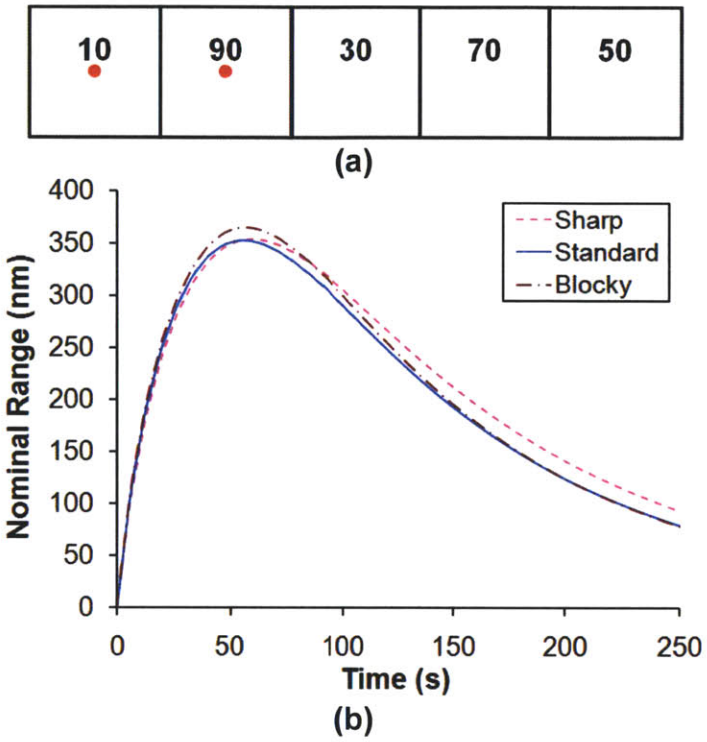


Figure 3.8: Simulated middle die nominal range evolution for different conditioning disk diamond shapes: (a) Nominal range monitor sites; (b) Nominal range evolution.

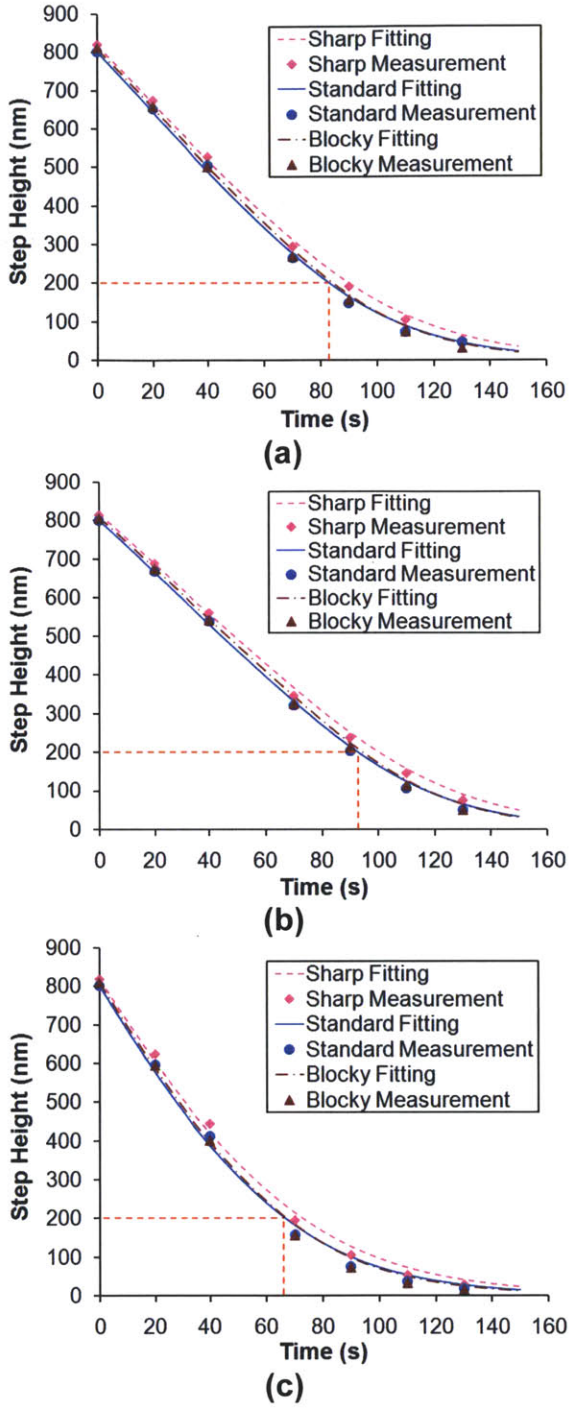


Figure 3.9: Middle die 50% pattern density array step height evolution for different conditioning disks: (a) Center point. (b) Left edge point (next to 70% pattern density array in Figure 3.3). (c) Right edge point (next to 10% pattern density array in Figure 3.3).

However, there is a difference in step-height evolution. Figure 3.9 is the data fitting result of step-height vs. time of the middle die 50% pattern density array. Pattern density dependence of polish is also verified by comparing the step-height evolution at center (Figure 3.9(a)), left edge (Figure 3.9(b)) and right edge (Figure 3.9(c)) points of the 50% pattern density array on the middle die. As indicated by the red dashed lines, step height reaches to 200nm remaining height more slowly (at about 92s) at the left edge point (next to 70% pattern density array) and more quickly (at about 65s) at the right edge point (next to 10% pattern density array).

In Figure 3.9, the sharp diamond disk data goes to a non-linear height-reduction region slightly earlier (at larger remaining step height) than the other disks. This transition from linear step height reduction in time occurs when asperities start to polish both up and down feature areas rather than just up areas; we thus expect this transition to depend on pad asperity height. A good way to understand the linear to non-linear transition in step height is shown in Figure 3.10. In the beginning of the process, down area removal amount is zero. When the step height reduction goes to non-linear, down area removal amount starts to increase above zero. In Figure 3.10, the sharp disk curve goes to above zero at the largest remaining step height among the three disks, while blocky disk curve goes to above zero at the lowest remaining step height. The characteristic asperity heights vary about 15% for the three conditioning disks, so we see a clear relationship verifying the physical model sensitivity to the pad asperity height.

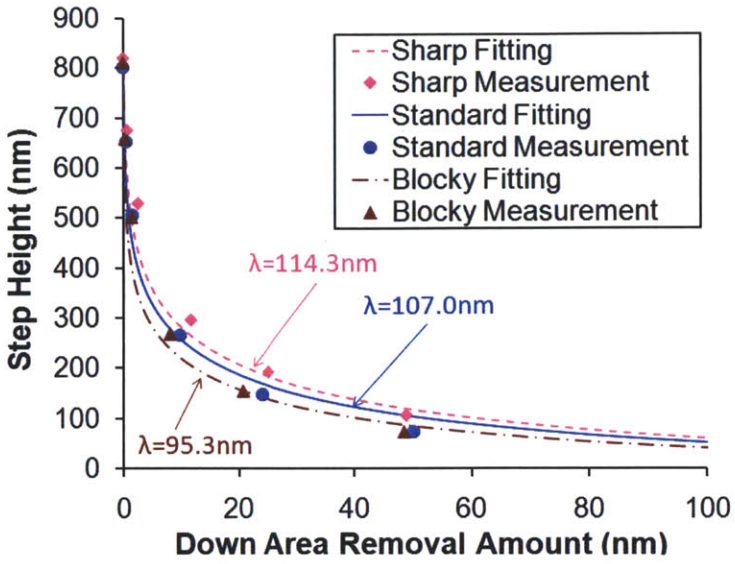


Figure 3.10: Middle die 50% pattern density array center point step height versus down area removal amount for different conditioning disk diamond shapes. Taking the down area removal amount at 1nm as the linear to non-linear transition point, the remaining step heights of blocky, standard and sharp disks at the transition point are 436.0nm, 497.7nm, and 533.5nm, respectively.

3.2.4 Conclusion

In this study, the physical die-level CMP model is applied to pad property and polish process analysis. The experimental results confirm the decomposition of the model between pad bulk and pad asperity components. We are able to relate pad stiffness and conditioning disk effects to model parameters and data fitting results. Quantitative relationships between pad properties and model parameters are established, which can be used to evaluate pad performance and assist in pad design and optimization. Model prediction is the consistent with our expectation: higher pad stiffness gives better within-die uniformity, and a conditioning disk with blocky diamonds results in a tighter asperity height distribution and keeps the linear height-reduction polishing region longer. This physical CMP model simulates pattern density dependence and within-die uniformity correctly.

3.3 CMP process endpoint analysis

Inter-level dielectric CMP is used to planarize the wafer surface, i.e., to remove topography resulting from previous patterning and deposition processes. In practice, perfect planarization is difficult to achieve due to the pattern density differences on the die, where planarization of low density areas is faster than that of high density areas. Generally, the CMP process must be stopped once a certain endpoint criteria is fulfilled. In one approach, the polishing should stop when the original step-height has been reduced to some value that is "small enough" to meet product requirements; we call this a step-height target strategy. Another commonly used approach is a thickness target strategy, in which the process should stop when the up area thickness has been reduced to a specified value determined by product requirements. Since the physical die-level CMP model has the capability to simulate the whole chip topography evolution during the polishing process, the process endpoint variation can be analyzed using the model simulation with extracted physical model parameters and chip layout. This section performs whole chip simulations and demonstrates endpoint variation analysis under different strategies.

3.3.1 Full chip simulation of topography evolution

To understand the die-level uniformity and impact on process endpoint, full-chip simulations (using the middle die extracted model parameters from Figure 3.9) are run for pad conditioning effects using the JSR standard pad. The layout design is the SKW7-2 test pattern as shown in Figure 2.20. Figure 3.11 and Figure 3.12 show up area thickness and step height evolutions respectively. We can see up that the area thickness becomes uniform between 120s and 180s. Step height becomes uniform around 120s. If different endpoint strategies are considered, the endpoint time needs to be selected correspondingly, as discussed in the following subsections.

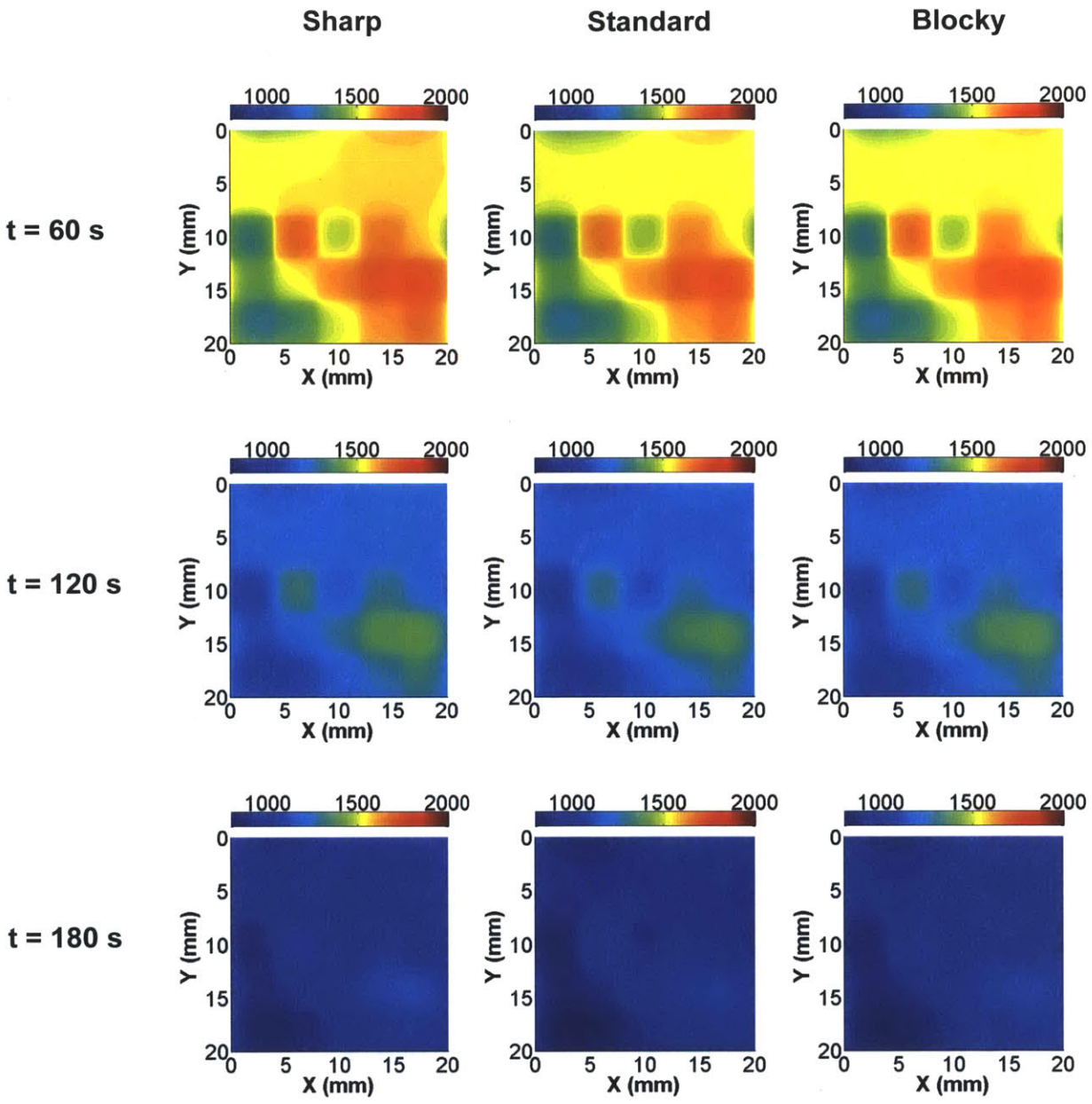


Figure 3.11: Full chip up area thickness (nm) evolutions of the SKW7-2 test pattern polished by a JSR standard pad with conditioning of different disks: sharp, standard and blocky.

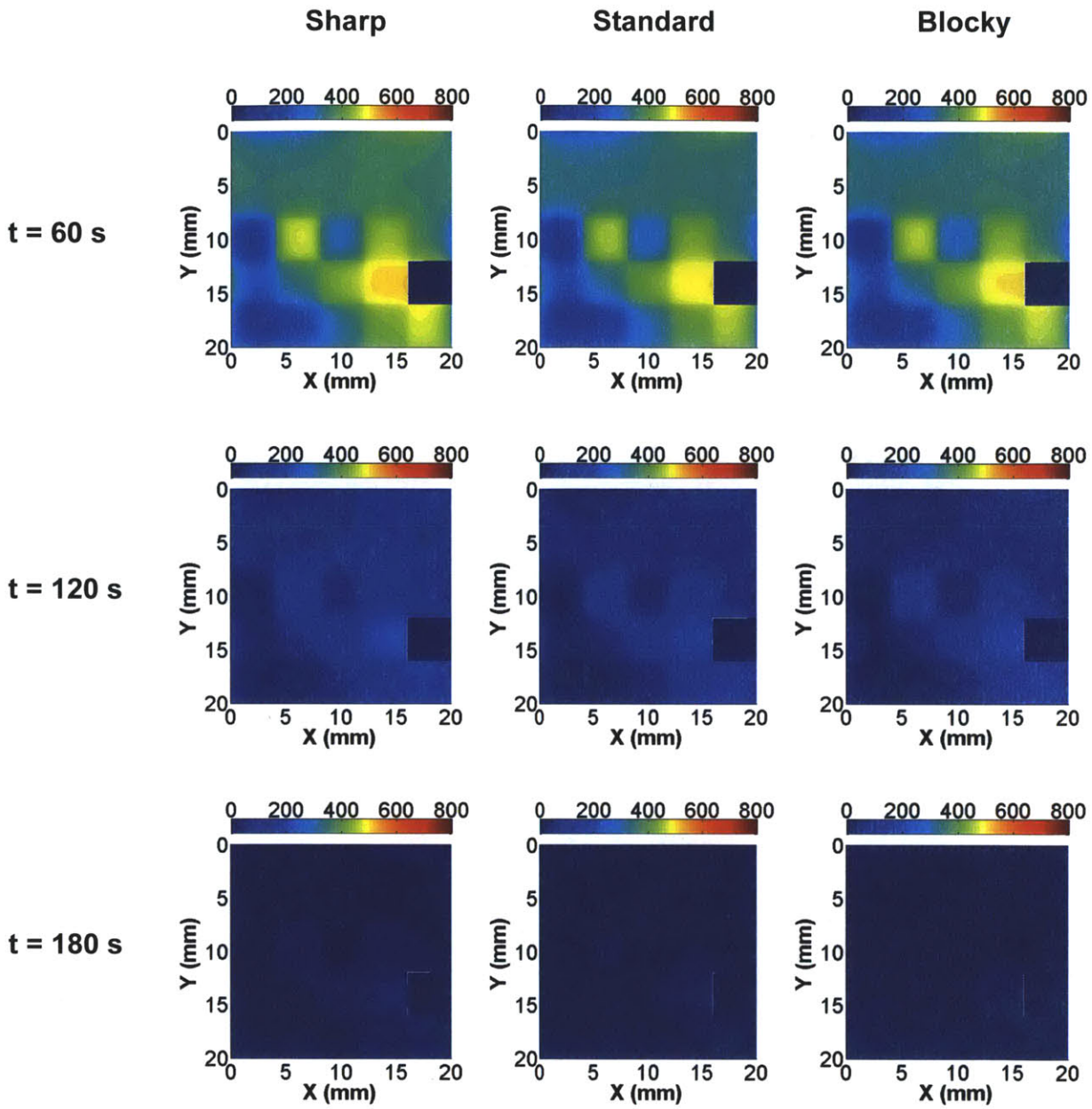


Figure 3.12: Full chip step height (nm) evolutions of the SKW7-2 test pattern polished by a JSR standard pad with conditioning of different disks: sharp, standard and blocky. The step height in the 100% density array is defined as our zero reference.

3.3.2 Step height target strategy

A step height target strategy is usually preferred when local planarization is the primary concern. In this study, the endpoint is defined as occurring when a step height target of 100nm at the 50% pattern density area is reached, i.e., the polishing stops when the remaining step height is 100nm at the center point of the 50% pattern density area. Table 3.2 lists the simulation results for the different conditioning diamond shapes. We can see the differences in final within-die and step-height uniformity in this table. When the 50% pattern density area reaches the endpoint, the 10% pattern density area is over polished and the 90% pattern density area still has a remaining step-height substantially larger than the endpoint target. From the endpoint time, we see that the standard pad with blocky conditioning disk planarizes faster. Table 3.2 also tells us that within-die oxide thickness uniformity is not substantially impacted by conditioning disks. Both nominal range (difference between the up area oxide thickness of the 90% pattern-density area and that of the 10% pattern-density area) and full chip range (difference between maximum up area thickness and minimum up area thickness across the entire chip) are similar for the three different disks.

Table 3.2: Simulation results of step height target strategy for pad conditioning effects on JSR standard pad.

Conditioning disk diamond shape		Sharp	Standard	Blocky
Initial step height (nm)		800	800	800
Initial up area thickness (nm)		2000	2000	2000
Endpoint time (s)		114	107	105
Remaining step height (nm) (100% density area excluded)	Center point of 10% pattern density area	36.5	34.5	30.7
	Center point of 50% pattern density area	**100 (endpoint)**		
	Center point of 90% pattern density area	186.7	192.6	205.6
	Maximum	259.5	267.9	284.3
	Minimum	29.5	27.4	23.7
Nominal range (nm)		272.6	275.9	285.0
Full chip range (nm)		385.8	389.5	398.7
Up area thickness (nm)	Center point of 10% pattern density area	1089.5	1097.6	1109.0
	Center point of 50% pattern density area	1237.2	1246.2	1261.1
	Center point of 90% pattern density area	1362.1	1373.4	1394.0
	Maximum	1447.3	1458.8	1479.5
	Minimum	1061.4	1069.4	1080.7

3.3.3 Film thickness target strategy

A film thickness target strategy is usually preferred when global planarization is the primary concern. The endpoint of the thickness target strategy is defined as occurring when an up area thickness of 1000nm at the 50% pattern density area is reached, i.e., the process stops when the remaining up area thickness is 1000nm at the center point of the 50% pattern density area. Table 3.3 lists the simulation results of this thickness strategy for the different diamond conditioning disks. We see a strong pattern density dependent within-die non-uniformity in these results, where the up area thickness and remaining

step heights vary in different pattern density regions at the end point. The planarization efficiency with the blocky disk is higher than with the other two conditioning disk types, because at the endpoint the remaining step-height of the blocky disk is less than with the other two.

Table 3.3: Simulation results of up area thickness target strategy for pad conditioning effects on JSR standard pad.

Conditioning disk diamond shape		Sharp	Standard	Blocky
Initial step height (nm)		800	800	800
Initial up area thickness (nm)		2000	2000	2000
Endpoint time (s)		170	163	165
Remaining step height (nm) (100% density area excluded)	Center point of 10% pattern density area	8.1	6.5	4.4
	Center point of 50% pattern density area	18.0	15.0	10.9
	Center point of 90% pattern density area	33.9	29.2	23.2
	Maximum	65.3	59.7	53.0
	Minimum	6.4	5.0	3.3
Nominal range (nm)		176.6	172.0	168.4
Full chip range (nm)		295.4	291.7	289.7
Up area thickness (nm)	Center point of 10% pattern density area	903.0	905.2	910.3
	Center point of 50% pattern density area	**1000 (endpoint)**		
	Center point of 90% pattern density area	1079.6	1077.2	1078.7
	Maximum	1171.1	1169.5	1172.4
	Minimum	875.7	877.8	882.7

3.3.4 Conclusion

Full chip topography evolution in CMP is simulated using the calibrated physical die-level CMP model. Endpoint is predicted according to two different process target strategies. Full chip simulation enables one to evaluate the within-die non-uniformity by identifying the critical results including the nominal range and remaining step height.

Different process inputs can be verified to see if the polishing results meet the process requirements.

3.4 Wafer-level pressure non-uniformity evaluation

In Sections 3.2 and 3.3, the die-level physical model is applied to understand the pad property effects and process endpoints. Those studies focus on the middle die from a wafer: wafer-level non-uniformity is not considered. However, the extracted model parameters from center die, middle die and edge die are different, as shown in Figure 3.4 and Figure 3.7. These differences may be induced by substantial wafer-level non-uniformity.

This section examines the wafer-level non-uniformity using the extended wafer-die-level model. There are many sources of wafer-level non-uniformity in CMP, such as pressure distribution, relative velocity mismatch, slurry delivery, CMP pad glazing and non-uniform beginning profile of input wafer. Our model focus on non-uniform pressure distribution caused by CMP pad properties and polishing tool design; thus a successful model fitting of wafer-level experiment data is not guaranteed. If the main source of wafer-level non-uniformity is pressure, the model can capture this effect; otherwise, the model will not fit when the main source of non-uniformity is due to other factors. If we verify the pressure to be the main factor impacting wafer-level non-uniformity, optimized process parameters may be suggested.

3.4.1 Wafer-level experiment data fitting

Recall the patterned wafer polishing experiment in Section 3.2, where we measured three dies on the wafer after each polishing interval. Each die is fitted by the physical die-level model separately to extract model parameters. We get three sets of parameters for each wafer in Section 3.2; these parameters have substantial differences, suggesting that wafer-level non-uniformity exists. Since we know the die positions on the wafer, the integrated wafer-die-level model can be employed to fit the three dies all together, as

discussed in Section 2.4. Only one set of model parameters, including both wafer- and die-levels parameters, will be extracted from each wafer.

As the integrated model has limitations to cover multiple non-uniformity factors, the model may not fit the experiment data well. Here we can check fitting error to see if pressure is the main factor affecting wafer-level non-uniformity. The fitting error is defined as the minimized root mean squared (RMS) error between model estimation and polishing data. If the fitting error is fairly high, the extracted parameters are not acceptable, indicating that the pressure distribution is not recognized as the main source of wafer-level non-uniformity.

Table 3.4 lists the extracted model parameters and fitting errors for polishing data using different pads. The fitting errors for standard and high stiffness pads are small enough to suggest that pressure variation may be an important factor driving wafer-level non-uniformity. In the following, we focus on the standard pad, as it has the smallest fitting error.

Table 3.4: Wafer-level extracted model parameters and fitting errors for SKW7-2 wafer polishing with different pad stiffness (low, standard and high).

Parameter	Low	Standard	High
Wafer-level blanket removal rate K_0 (nm/min)	**202.4**	**224.3**	**201.0**
Pad bulk effective modulus E_0^* (MPa)	**154.6**	**271.2**	**362.1**
Characteristic asperity height λ (nm)	**100.8**	**97.4**	**102.7**
Fitting error (nm)	**57.6**	**33.4**	**39.6**

Since the standard pad is a possible good fit, we compare the model calculation and experiment data as shown in Figure 3.13. The whole wafer step-height evolution is fitted by the integrated model. Wafer-level non-uniformity is observed at each monitor site. The model captures the main trend: the center die has a fast step height reduction than the other dies. This trend could result from the wafer-level pressure distribution. The

extracted wafer-level pressure non-uniformity is shown in Figure 3.14: the pressure in the wafer center is higher than in the other regions, which causes the center die to planarize faster. The extracted wafer-level pressure non-uniformity is large, which agrees with the model simulation results in Section 2.1.3. This wafer-level pressure non-uniformity may be induced by non-optimized process parameters, including pad thickness, retaining ring gap and retaining ring reference pressure.

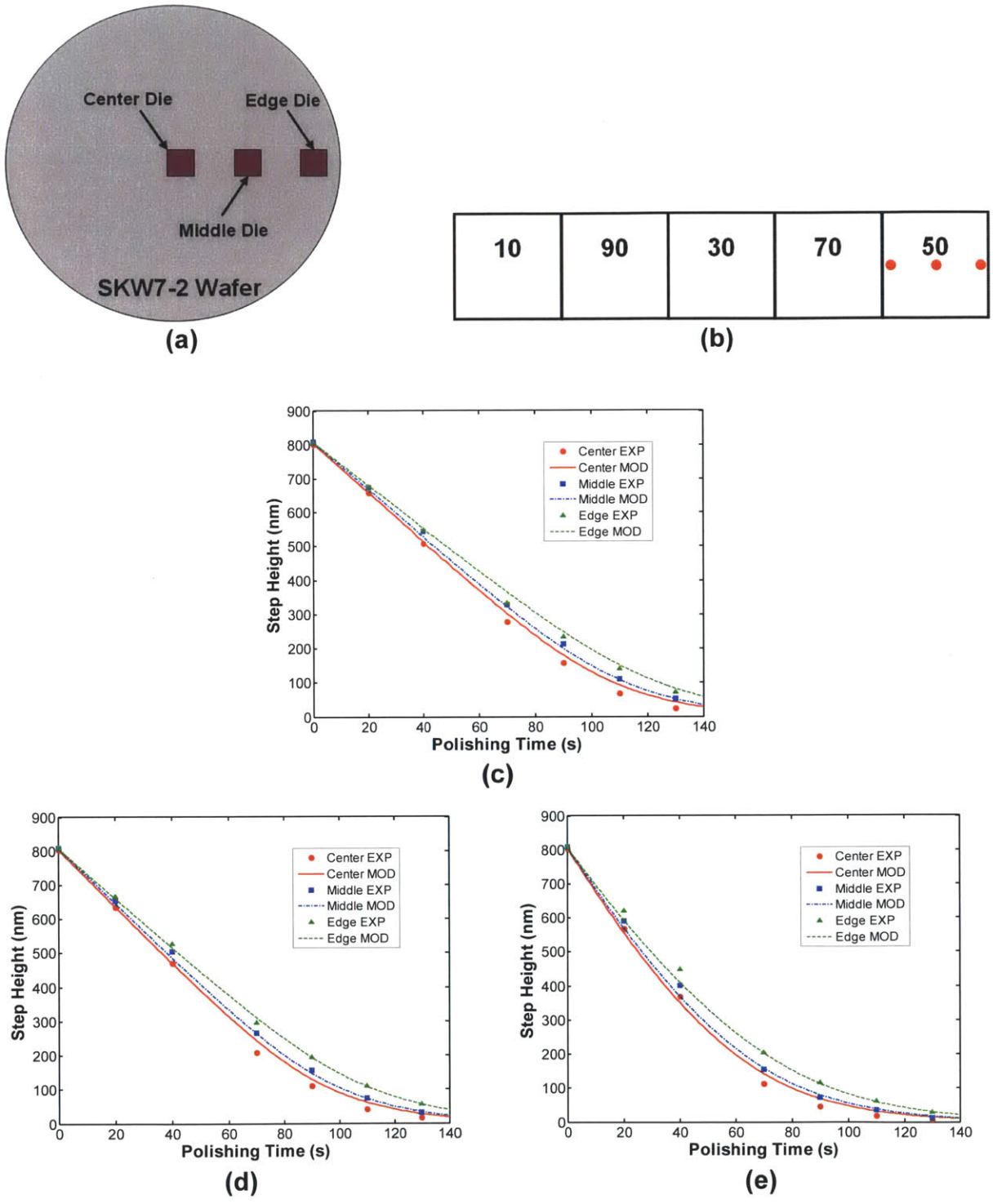


Figure 3.13: Wafer-level step height evolution: (a) Die positions; (b) Monitor sites in 50% pattern density array of each die; (c) Center monitor site in 50% pattern density array; (d) Left monitor site (next to 70% pattern density array); (e) Right monitor site (next to 10% pattern density array). Here "EXP" indicates experimental data, and "MOD" is the fitted model.

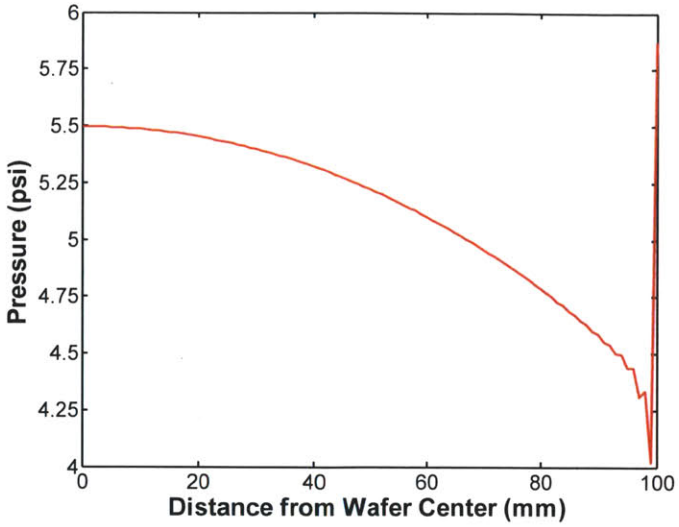


Figure 3.14: Extracted wafer-level pressure distribution for JSR pad with standard stiffness.

3.4.2 Wafer-level non-uniformity impact on die-level non-uniformity

In this section, we discuss the pressure distribution impact on die-level non-uniformity. The die-level up area oxide thickness non-uniformity can be reflected by the nominal range (the difference between the up area oxide thickness of the 90% pattern-density area and that of the 10% pattern-density area). As shown in Figure 3.15, all three dies have similar "final" nominal ranges when the polishing time is longer than 100 seconds. Thus the wafer-level pressure non-uniformity does not introduce strong die-level final up area oxide thickness non-uniformity; rather, it only changes the within-die material removal speed simultaneously across all pattern regions in the die. The within-die up area thickness non-uniformity is primarily controlled by the pad modulus, which does not change for different dies across the wafer.

To compare within-die local non-uniformity, we define a parameter called the step range as the difference between the maximum step height and the minimum step height in a die. Smaller step range means more uniform local planarization. Figure 3.16 shows that

the center die achieves lower step range than the others. Higher wafer center pressure improves the center die local planarization uniformity.

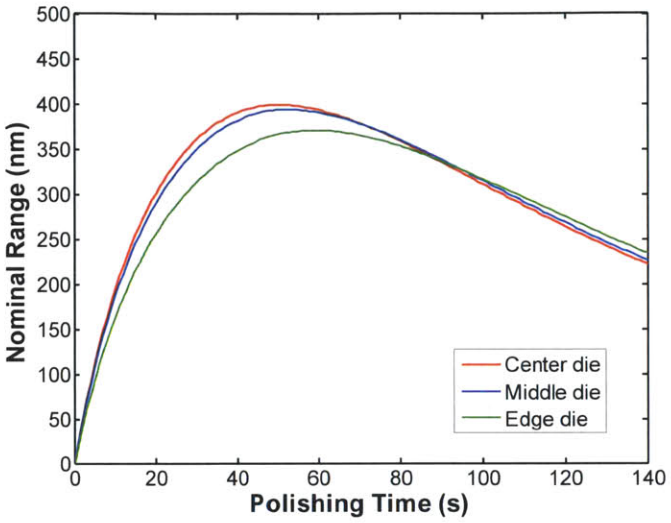


Figure 3.15: Simulated nominal range evolutions for different die positions from an SKW7-2 wafer polished by JSR standard pad.

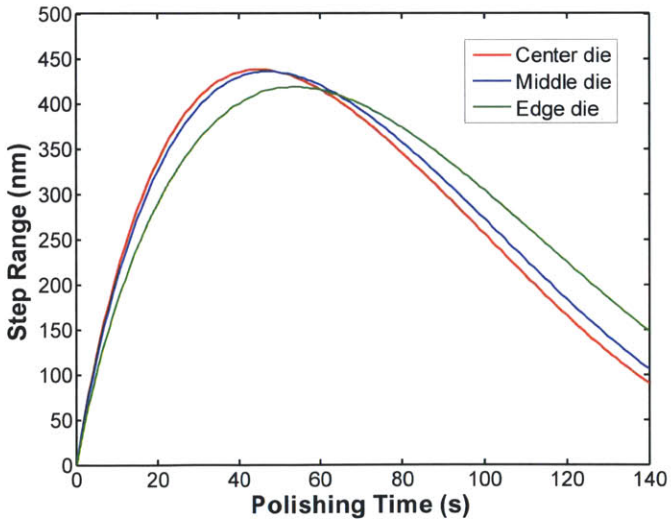


Figure 3.16: Simulated step range evolutions for different die positions from an SKW7-2 wafer polished by JSR standard pad.

3.4.3 Conclusion

The polishing data from different dies on a wafer is fitted by the integrated wafer-die-level model. Wafer-level pressure non-uniformity impact on die-level non-uniformity is verified: the wafer-level pressure distribution does not impact the up area thickness uniformity within different die on the wafer, but does significantly affect the final within-die local planarization uniformity for different die on the wafer.

3.5 Pitch size effect evaluation

In oxide CMP, the layout pitch size is also a source of within-die variation. In the same pattern density region, the planarization speed varies depending on the pitch size (or equivalently, depending on the feature size and spacing between features, as shown in Figure 3.17). Layout designers not only have design rules for pattern density, but also for pitch or feature size. This section examines the pitch size dependence in CMP.

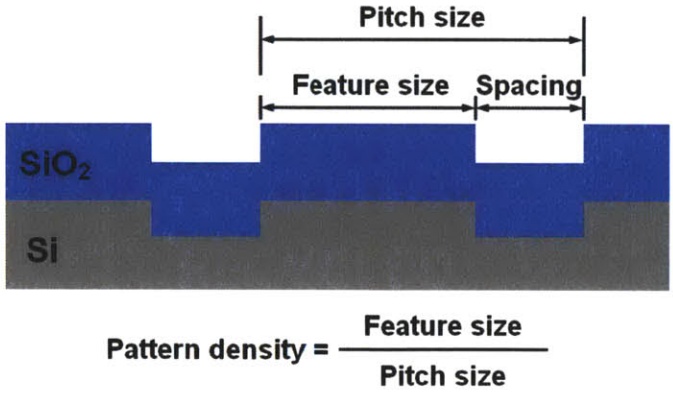


Figure 3.17: Definitions of pattern density and pitch size.

3.5.1 Polishing experiment

The experiment involves polishing four 200 mm wafers patterned with the MIT standard STI mask (Appendix A). All wafers began with 9 nm of a thermally grown pad oxide on a p-type silicon substrate, followed by a 119 nm silicon nitride deposition. Wafers were patterned and etched to obtain a trench depth of 500 nm (initial step height),

then subjected to a sidewall oxide layer growth of 25 nm using dry oxidation. This was followed by deposition of TEOS oxide for trench fill of 575 nm (initial up area thickness). The four wafers were polished with time splits of 48, 72, 96 and 120 seconds separately. The CMP process was run on an AMAT Mirra polisher with a reference pressure of 1.5 psi, using IC1000 polishing pad and silica based slurry.

Ten optical film thickness measurements (including up area and down area) were performed on the middle die of each wafer, as shown in Figure 3.18. We have three monitor sites in 50% pattern density regions, which enable us to consider different pitch sizes at the same pattern density. The full chip pattern density and pitch size arrangements are illustrated in Figure 3.19. We see that the pitch size arrangement is more complicated than the pattern density arrangement, which is expected to induce within-die variation.

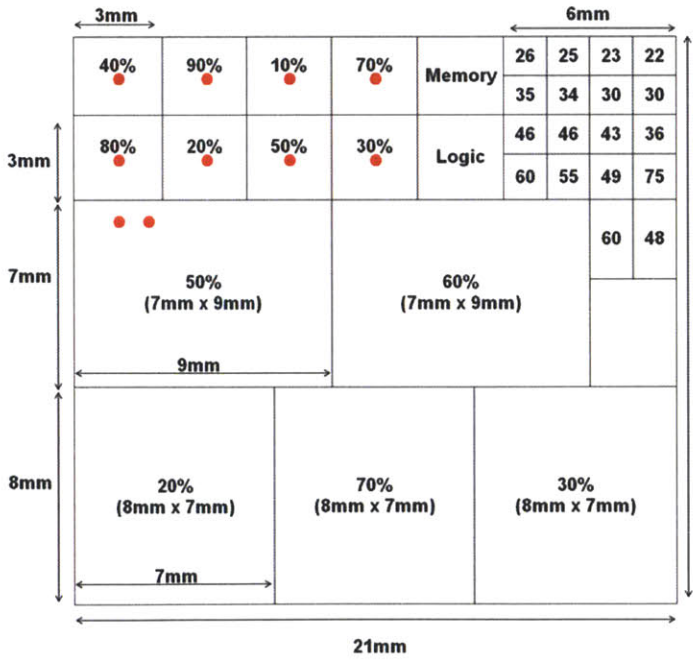


Figure 3.18: Measurement sites on MIT standard STI CMP test layout.

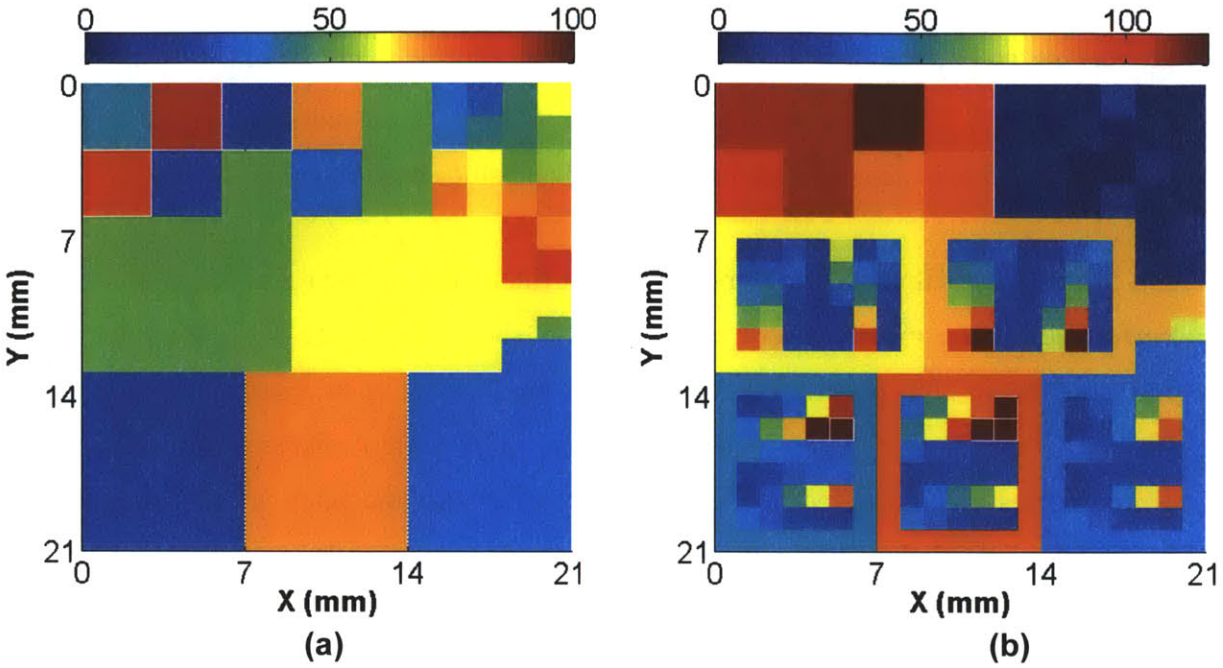


Figure 3.19: Pattern density and pitch size specifications of MIT standard STI CMP layout: (a) Pattern density (%) map; (b) Pitch size (μm) map.

3.5.2 Model fitting

Oxide removal amount data is used to fit the extended die-particle-level CMP model. The model parameters are extracted by minimizing the root mean squared (RMS) error between polishing data and model estimations, as listed in Table 3.5. The fitting error is 18.4 nm.

Table 3.5: Extracted model parameters for oxide polishing in STI CMP experiment.

Parameter	Value
Blanket removal rate K_0 (nm/min)	**99.4**
Pad bulk reduced modulus E_0^* (MPa)	**447.9**
Characteristic asperity height λ (nm)	**137.5**
Asperity radius of curvature R_{asp} (μm)	**15.0**
Feature shape factor α	**2.52**

Figure 3.20 compares the oxide removal amounts of different pitch size in 50% density regions. Both up area and down area have pitch size dependence. As expected, up area removal is faster for smaller pitch size, while down area removal is slower for smaller pitch size. This is because in the early stage of polishing, asperities can only touch the up area of small features. The extended die-particle level model has the ability to capture this pitch size dependence.

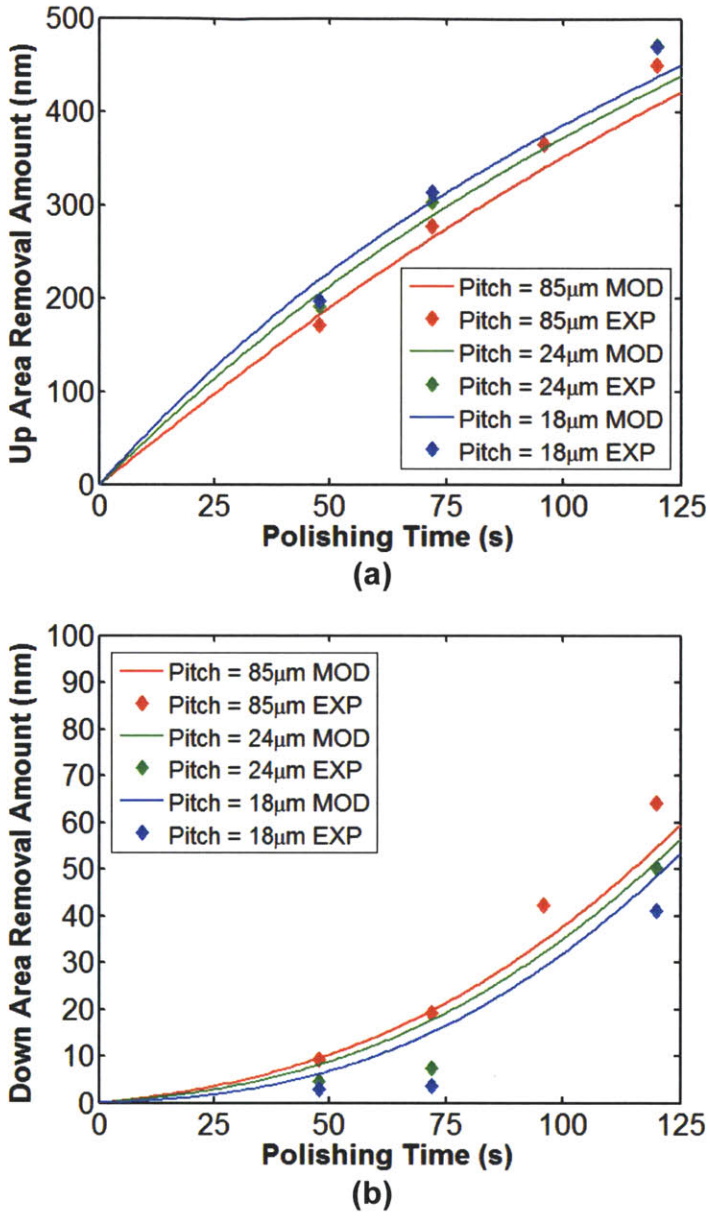


Figure 3.20: Oxide removal amount in 50% pattern density region: (a) Up area removal amount. (b) Down area removal amount. Here "EXP" indicates experimental data, and "MOD" is the fitted model.

3.5.3 Pitch size effect on within-die non-uniformity

Since the pitch size effect does exist and the calibrated model is able to make predictions based on this, the pitch size effect can be studied for the whole chip or in the same density region.

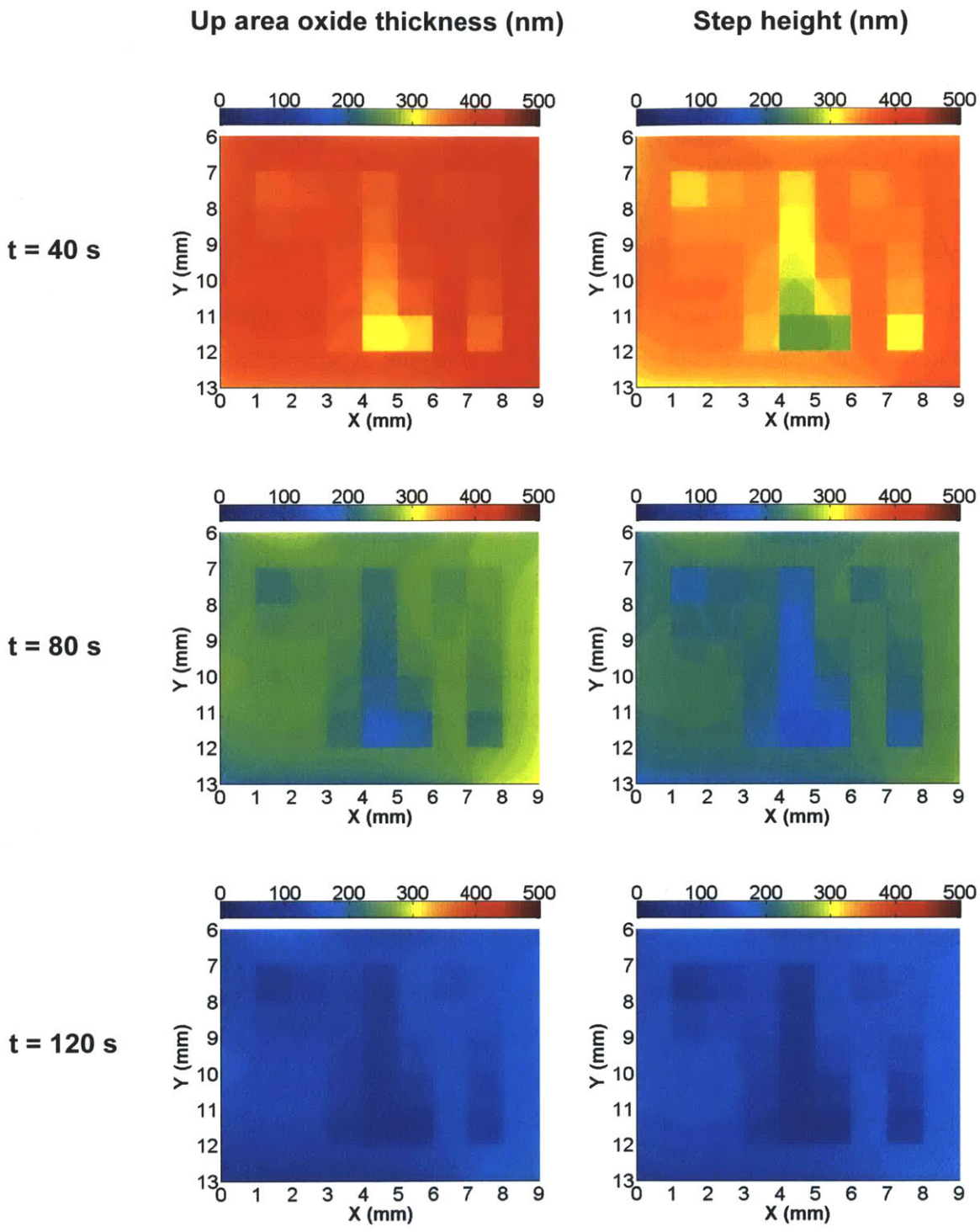


Figure 3.21: Up area thickness and step height evolutions of the 50% density region in the MIT standard STI CMP layout. Non-uniformity is induced by the different pitch sizes arranged in this region.

Figure 3.21 shows the simulated up area thickness and step height evolutions in the 50% density region. The non-uniformity can be attributed to the pitch size, referring to the pitch size map in Figure 3.19(b). The non-uniformity is strongly affected by the small features in the middle column.

3.5.4 Conclusion

The extended die-particle-level model is applied to fit oxide removal data of STI CMP. Pitch size dependence is observed from the polishing experiment data and captured by the model. Simulation suggests that even in a constant pattern density region, non-uniformity caused by different pitch sizes is also a concern of layout design.

3.6 Summary

In this chapter, several model applications in oxide CMP have been presented, focusing on die-level non-uniformity. The general methodology of applying the physical die-level model can be used for candidate product layout designs, to analyze potential die-level non-uniformity concerns.

CMP pad stiffness and conditioning effects are studied by extracting and comparing die-level physical model parameters from polishing data. Quantitative relationships between pad properties and model parameters are established and applied in the model prediction. Within-die non-uniformity at process endpoint is analyzed under different endpoint strategies based on simulation results of the calibrated model.

Integrated models are applied to verify and consider more substantial impacts on within-die non-uniformity. The extended wafer-die-level model fitting combines pattern density effects and wafer-level pressure variation. The within-die non-uniformity of local planarization is shown to be a function of die position on the wafer due to the wafer-level non-uniform pressure distribution. The extended die-particle-level model is shown to account for both pattern density and pitch size effects.

4 Physical characterization of CMP pad properties

4.1 Introduction

In the CMP process, pad modulus and asperity height distribution are two important properties that affect planarization results. The pad modulus is primarily determined by the pad material, while the asperity height distribution depends both on the pad material and the conditioning process. The coefficients of these two pad properties, pad modulus and characteristic asperity height, are employed as CMP model parameters in previous chapters of this thesis and many other CMP modeling works [18, 27] to understand polishing performance and the interaction between the pad and the wafer. Physical measurements of pad modulus and asperity height are therefore valuable for improving model fitting and for verifying underlying assumptions.

Many measurement approaches have been developed to test CMP pad mechanical properties [93-95], especially the pad modulus, based on tensile test, compression test or dynamic mechanical analysis. However, polishing interactions occur between pad asperities and the wafer surface at the micrometer scale, and traditional macro-scale physical measurement approaches have limitations in their ability to capture such details. Recently, nanoindentation has been used in pad mechanical property tests with nanometer scale contact control precision and force sensitivity. The details of traditional porous pad surface mechanical behaviors, such as local particle-level contact, cell bending, long range pad asperity contact and the bulk response can be examined by nanoindentation [96, 97]. New and broken-in pads can be recognized by comparing the nanoindenation results of pad surfaces [29].

Characterization of the polishing pad surface topography assists in evaluating the material removal ability, slurry holding capacity, and the lifetime of the pad. Various metrology approaches have been utilized to study CMP pad surfaces, including interferometry [98], confocal microscopy [99], and stylus mircoprofilometry [100].

Quantitative correlations can be found between pad surface topography and conditioning effects.

In this chapter, a comprehensive study is presented to relate CMP pad surface property measurements and pad aging effects investigation. Section 4.2 introduces CMP pad nanoindentation, which measures the pad asperity modulus. Section 4.3 explains an approach using pad surface profilometry for characteristic asperity height measurement. Pad aging effects and spatial dependence of pad wear are evaluated in Sections 4.4 and 4.5, respectively, using these measurement approaches.

4.2 Nanoindentation of CMP pad

4.2.1 Background of nanoindentation

The principal goal of nanoindentation testing is to extract elastic modulus and hardness of the specimen material from experimental readings of indenter load and depth of penetration [101]. Generally in nanoindentation, we use the term "load" to indicate the force applied to the indenter, and "force" to indicate the force measured by the force sensor. In some indentation instruments the load is recorded, while in others (which have a force sensor) the force is measured. Ideally load and force would be identical.

In a typical test, force and depth of penetration are recorded as the load is applied from zero to some maximum and then from maximum force back to zero, i.e., over a complete loading and unloading cycle. The depth of penetration together with the known geometry of the indenter provides an indirect measure of the area of contact at full load, from which the mean contact pressure, and thus hardness, may be estimated. When load is removed from the indenter, the material attempts to regain its original shape, but is prevented from doing so because of plastic deformation. However, there is some degree of recovery due to the relaxation of elastic strains within the material. An analysis of the initial portion of this elastic unloading response gives an estimate of the elastic modulus of the indented material.

A typical load (force) vs. displacement curve in a nanoindenation test is illustrated in Figure 4.1. The test is performed with maximum load P_{max} and maximum depth beneath the specimen free surface h_{max}. During the loading of the indenter, the material undergoes both elastic and plastic deformation. The elastic part of the unloading curve (typically the beginning section of unloading near the maximum load point) provides information about the elastic properties of the tested materials. Several approaches to analyze nanoindentation data have been developed over several decades [102-104]. An often-used approach is known as the Oliver-Pharr method [105].

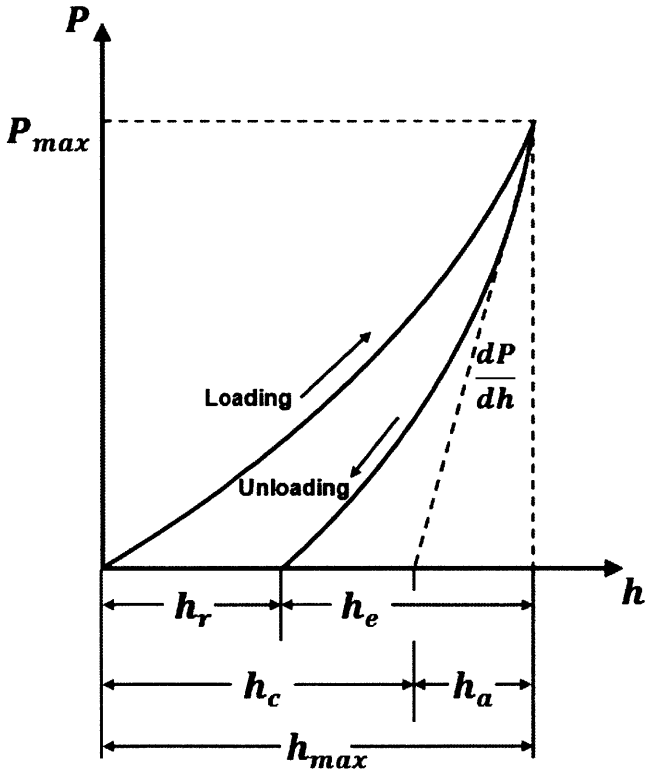


Figure 4.1: Schematic diagram of a typical load vs. displacement curve in nanoindenation. The contact depth h_c and slope of the elastic unloading $\frac{dP}{dh}$ allow specimen modulus and hardness to be calculated. Here h_r is the depth of the residual impression, h_e is the displacement associated with the elastic recovery during unloading, and h_a is the displacement from the edge of the contact to the specimen surface at full load.

The Oliver-Pharr method is adopted in the CMP pad nanoindentation test of this thesis. The contact depth h_c and the contact stiffness $S = \frac{dP}{dh}$ are used to calculate the reduce modulus of measured samples. The relation between reduced modulus and contact stiffness can be written as

$$E^* = \frac{\sqrt{\pi}}{2\sqrt{A}} S \tag{4.1}$$

The reduced modulus is given by

$$\frac{1}{E^*} = \frac{1-\nu_s^{\,2}}{E_s} + \frac{1-\nu_i^{\,2}}{E_i} \tag{4.2}$$

where E_s and ν_s are Young's modulus and Poisson's ratio for the sample, and E_i and ν_i are the same parameters for the indenter. A is the projected contact area, which depends on the indenter type. Four basic indenters are illustrated in Figure 4.2. Once the indenter type is specified, the contact area A can be estimated by the indenter geometry parameters listed in Table 4.1.

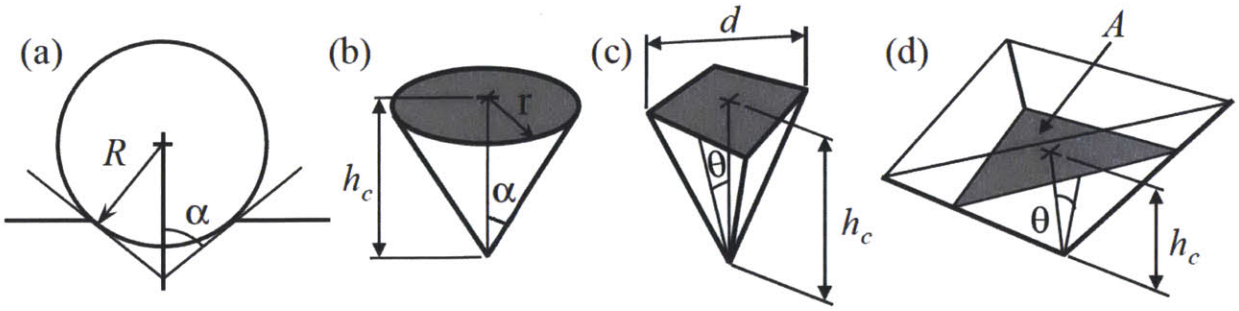
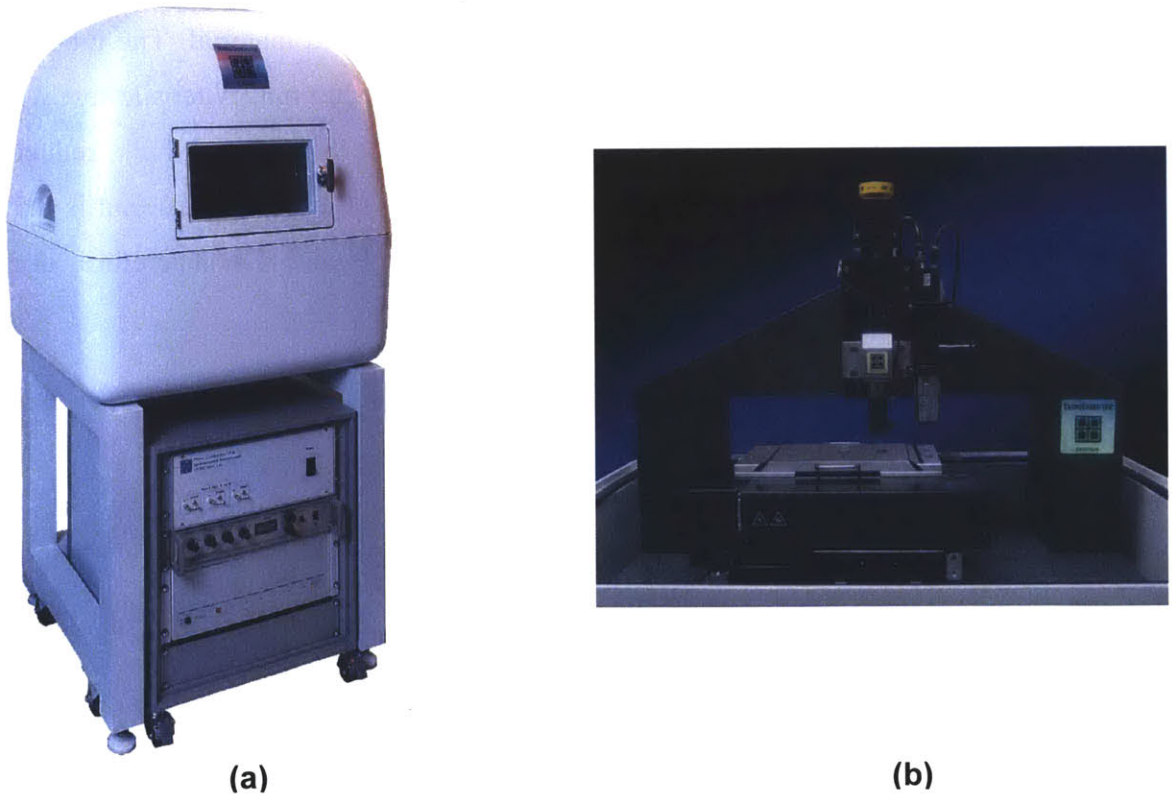


Figure 4.2: Indenter shape and geometry parameters [101]: (a) spherical, (b) conical, (c) Vickers, and (d) Berkovich.

Table 4.1: Geometry parameters for various types of indenters.

Indenter type	Contact area A	Semi-angle θ (deg)	Effective cone angle α (deg)
Spherical	$A \approx 2\pi R h_c$	N/A	N/A
Conical	$A = \pi h_c^2 \tan^2 \alpha$	α	α
Vickers	$A = 4h_c^2 \tan^2 \theta$	68	70.3
Berkovich	$A = 3\sqrt{3} h_c^2 \tan^2 \theta$	65.3	70.3

In our study, the nanoindenter is a Hysitron TriboIndenter model TI 900, as shown in Figure 4.3. The indenter is a Hysitron standard diamond Berkovich tip. With an average radius of curvature of about 150 nm, the Berkovich tip is primarily used for bulk materials and thin films greater than 100 nm thick, and is recommended in hard polymer material tests [107], which fits our test need.



(a) (b)

Figure 4.3: Hysitron TI 900 TriboIndenter [106]: (a) Chamber and controller; (b) Sensor and optical microscope in the chamber.

4.2.2 Indentation depth dependence

In polymeric material nanoindentation, depth dependence of modulus is observed [108], i.e., measured reduced modulus near the material surface is higher than that of the bulk. A recent nanoindentation study has determined the elastic moduli over a range from 5 to 200 nanometers from the free surface of various polymeric materials [109]. The surface modulus can exceed that of the bulk by up to 200%, independent of processing scheme, macromolecular structural characteristics and relative humidity. The enhanced stiffness or modulus trend is intrinsic and can be attributed to the contact stress-induced formation of a mechanically unique confined interphase at the contact region of the polymer surface and the indenter tip [109]. Therefore, the contact-induced stiffing may dominate the surface deformation in the nanometer scale.

CMP pads are usually made from polymeric materials [81], especially polyurethane. Indentation depth dependence is observed. Figure 4.4 shows the measured results of reduced modulus vs. indentation depth for Cabot 42D solid pad material. The modulus becomes fairly constant when the indentation depth is over 200 nm. A tensile test with TA Instruments 2980 Dynamic Mechanical Analyzer shows that the bulk reduced modulus of this material is 30.0 MPa (assuming Poisson's ratio of this material is 0.5), which is very close to the deep indentation modulus in Figure 4.4. Therefore, deep indentation results approach the bulk modulus.

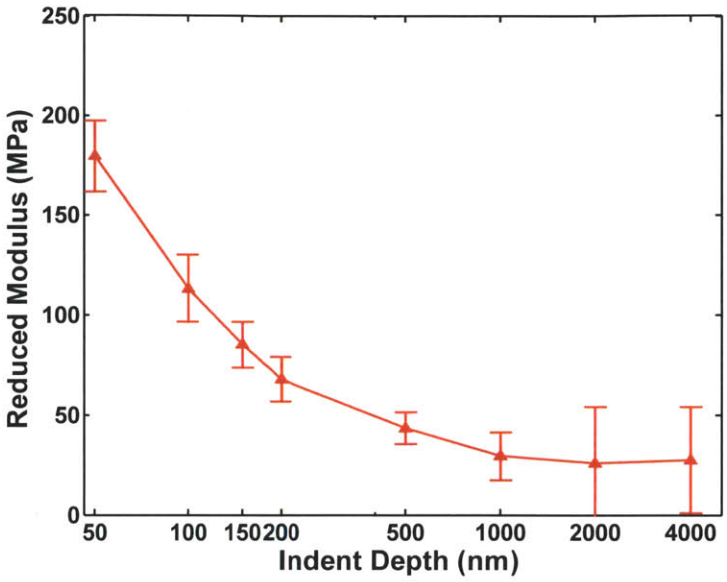


Figure 4.4: Reduced modulus vs. indentation depth for Cabot 42D solid pad material. Error bars represent one standard deviation of replicated measurements.

4.2.3 Solid contact between indenter and asperities

When we measure a conditioned CMP pad, asperity shape is a concern affecting our ability to perform successful indentation. We need to obtain the modulus from large asperities in order to avoid measurement artifacts. To capture a single large asperity, the indents are made on the pad sample surface within a 40 by 40 μm area at 10 μm intervals, shown in Figure 4.5. Within the indent test pattern, two typical contacts between indenter tip and pad asperity can be identified by the shape of the testing curve. When the tip is sliding on the asperity, there are abrupt changes in the testing curve, as shown in Figure 4.6(a). If the tip and asperity are in solid contact, the testing curve is continuous as shown in Figure 4.6(b). Only solid contact test curves are used to extract asperity modulus, following the standard data analysis method in Subsection 4.2.1.

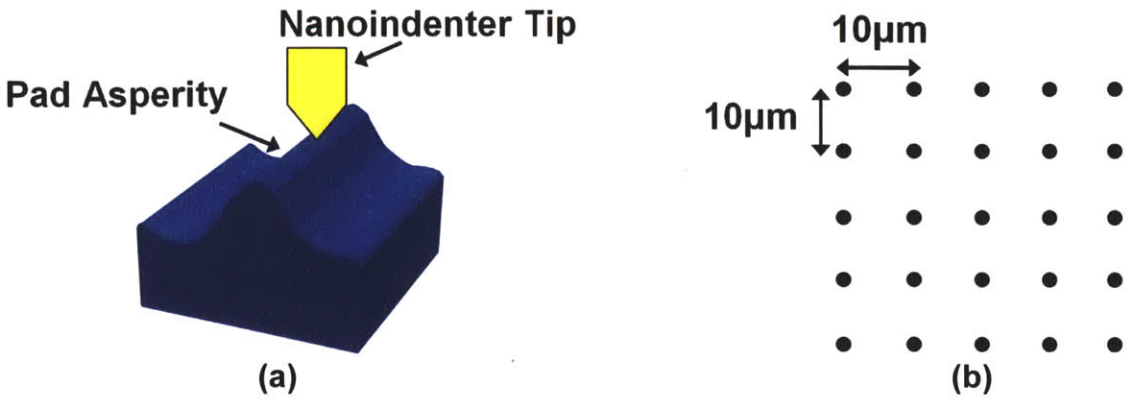


Figure 4.5: Conditioned pad sample nanoindentation: (a) asperity top indentation; (b) test pattern of indents.

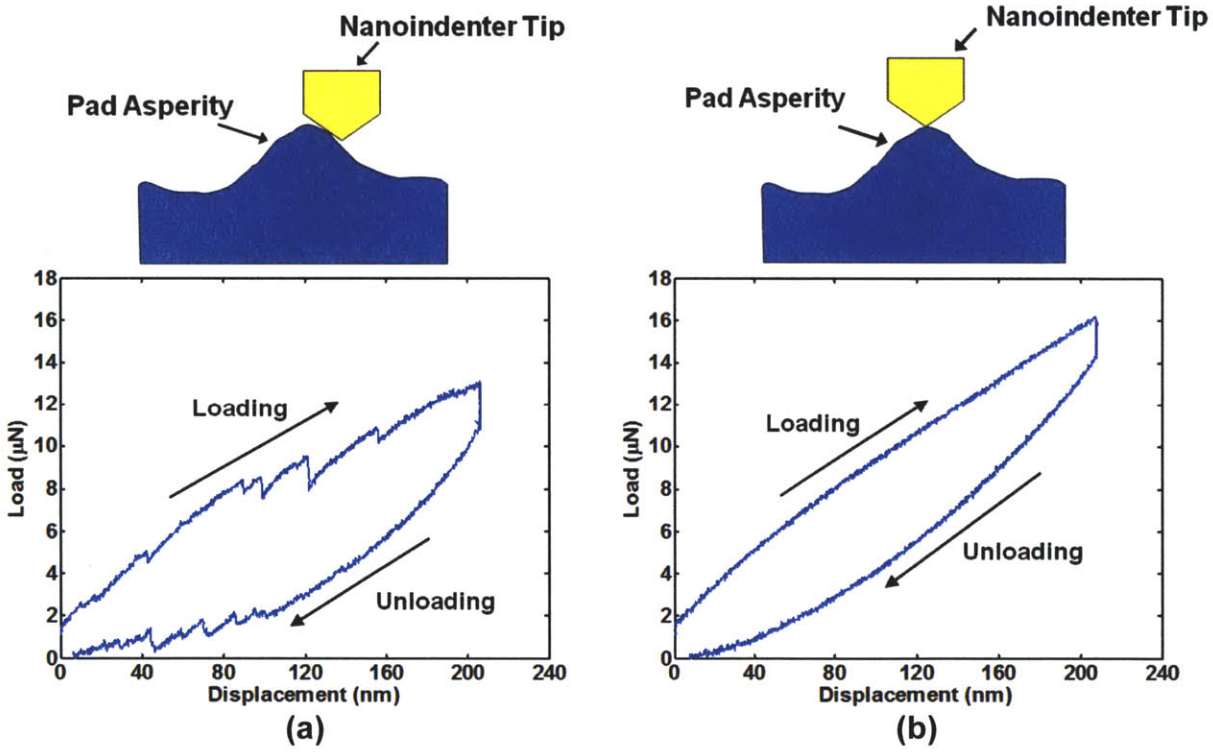


Figure 4.6: Testing curves of conditioned pad sample nanoindentation: (a) indenter tip sliding on asperity top; (b) solid contact between indenter tip and asperity.

4.3 Profilometry of CMP pad surface

4.3.1 Method of pad surface profilometry

A certain level of surface roughness is required on the pad surface to transport slurry and provide a sufficient number of asperities to make contact with the wafer. Surface profilomtery analysis has been used in evaluating pad polishing ability and conditioning effects [110].

A traditional method of pad surface measurement is stylus contact profiling. In this technique, a stylus passes over the surface going up and down according to the surface texture. A transducer converts this movement into surface topography information. Figure 4.7(a) shows a cross-section of a polishing pad, with an approximation of the surface profile as "seen" by a stylus. This illustrates the possible inadequacy of the stylus technique to accurately reflect the nature of the surface, resulting from stylus geometry interference. Reducing the stylus diameter has little effect on the surface profile scanned unless the stylus angle is significantly reduced.

Recently, the use of white light interferometry has become a popular non-contact method of gaining surface topography data. White light is split in a special objective lens, where part of the light travels to the pad sample, and the remainder is directed to a reference mirror. When the two parts recombine, interference fringes appear at the point of focus. Fourier transform algorithms convert the data into surface topography information. A disadvantage of interferometry can be the effect of the shape of some of the surface asperities on the pad surface. If the shape of the asperity hides a part of the surface from the light source, its actual shape will not be seen by the microscope as displayed in Figure 4.7(b). Another influencing factor is the vertical scan range. If the scan range is too small, the microscope will not acquire data for the lower areas of the pad surface, such as within pad pores. The intensity of the light reflected from the pad surface may also be a problem to obtain an accurate profile, if reflected light intensity is lower than the intensity threshold of the interferometer.

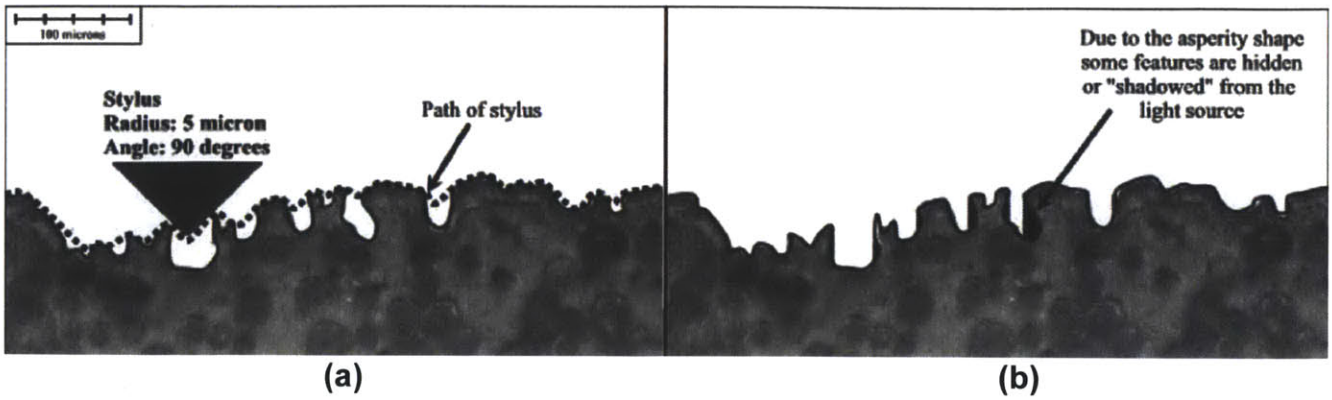


Figure 4.7: Contact and non-contact methods of CMP pad surface profilometry [110]: (a) Path of stylus over actual pad surface; (b) the pad surface as 'seen' by the white light interferometer.

In our pad surface study, we are interested in the "effective" asperities which result in a strong force response. Figure 4.8 illustrates three types of asperity-wafer contact in CMP. Solid contacts of fully supported asperities are assumed to make the largest contribution in material removal, but are very rare in number, as shown in Figure 4.8(a). Applied force produces a small deflection and a small contact area on a fully supported asperity. Less well supported (Figure 4.8(b)) and poorly supported (Figure 4.8(c)) summit contacts are assumed to be more common on the porous pad surface. These two types are easy to deflect, but have little contribution in material removal, due to limited force application. Stylus profilometry requires a small vertical force (usually several μN) applied on the pad surface during the surface scan. The force will deflect less supported and poorly supported asperities. The stylus profilometer thus only "sees" fully supported asperities rather than overall topography. The stylus profilometer is thus preferred in our work, as the "effective" topography obtained from stylus profilometry can be used to extract characteristic asperity height of fully supported asperities.

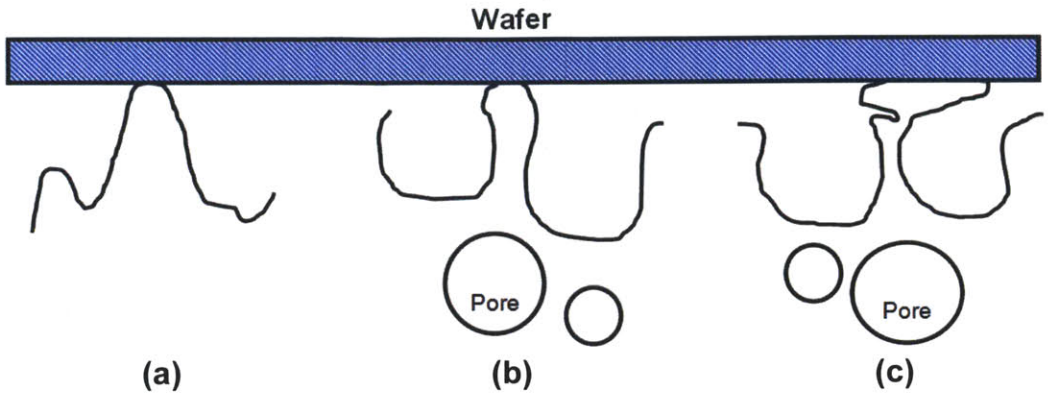


Figure 4.8: Asperity-wafer contact summit types in CMP: (a) fully supported; (b) less well supported; (c) poorly supported.

4.3.2 Characteristic asperity height

This subsection explains the procedure for extracting characteristic asperity height. A Tencor P-16 stylus profilometer is used to scan the surface of a conditioned pad under applied normal force of 20 μN, and the measured region is 500 by 500 μm, as shown in Figure 4.9(a). We adopt the data analysis approach described by Sun [84]. The profilometer measures surface heights over the selected region relative to an arbitrary reference plane assumed by the tool software. The height range is divided into equal bins and a histogram is made of the number of times that the surface height falls into each bin. It is convenient to normalize the histogram by dividing by the total number of data points. The histogram then becomes a probability density function (PDF), as shown in Figure 4.9(b). The area under the PDF in any chosen height range gives the probability of finding a point on the surface within that range. Since the reference plane used by the profilometer is arbitrary, it is also common to shift the mean height to zero. This facilitates the comparison of PDFs from different pad surfaces or different parts of the same surface. Heights to the right of the mean, or the positive side, correspond to the asperities that might contact the wafer under different load conditions. When the asperity summits have exponentially distributed heights, then the right hand tail of the PDF will

be linear on a log plot and can be characterized by a decay length λ (i.e., the distance over which the tail drops by a factor of $\frac{1}{e}$), which is called the characteristic asperity height. When λ is small, there is a narrow range of surface heights and the surface appears smooth. When λ is large, there is a wider range of surface heights, so the surface appears to be less smooth.

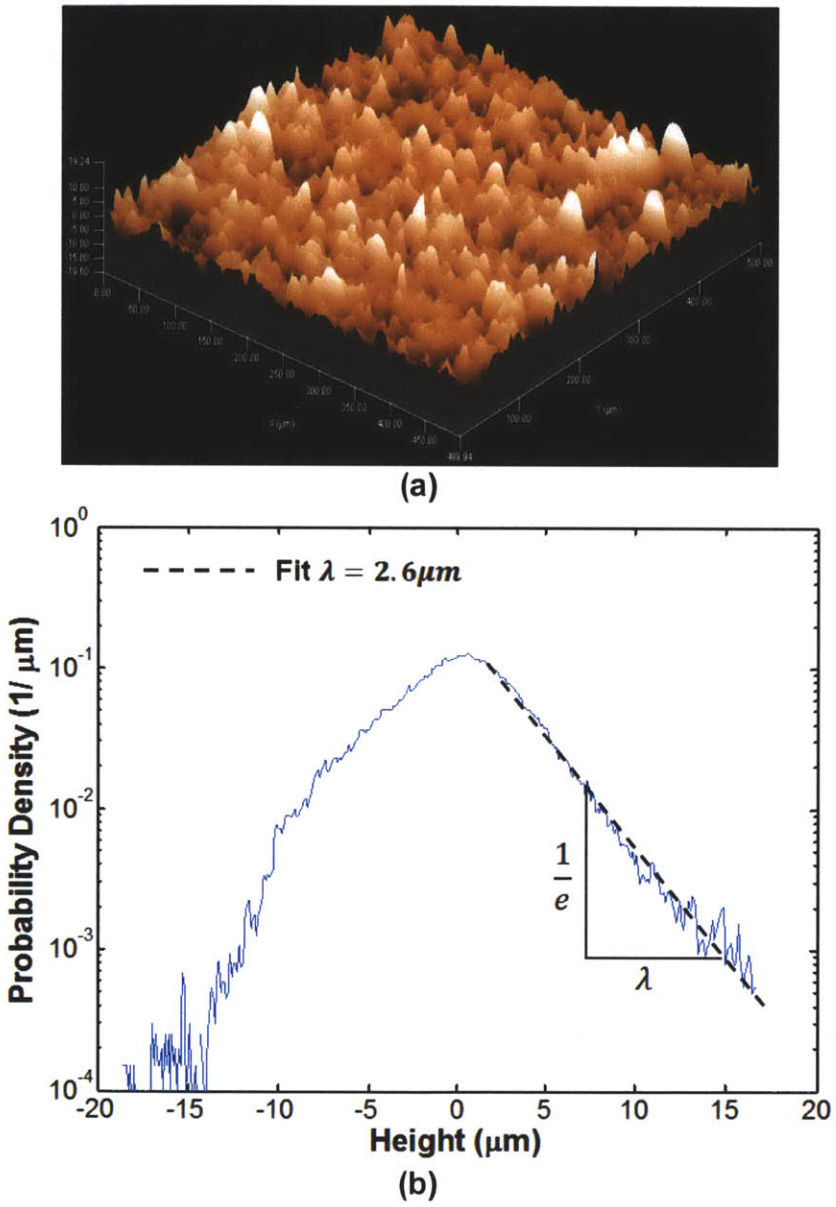


Figure 4.9: Extraction of characteristic asperity height: (a) Profilometry data. (b) Probability density fitting of exponentially distributed heights.

Characteristic asperity height λ is a measure closely related to the tall asperities, which actually contact the wafer surface. Compared to conventional surface roughness, λ is more sensitive to small changes in the high tail of the profile range or large deviations from the mean height.

4.4 Characterization of CMP pad aging effects

Pad aging is an important factor in CMP, as typical processes suffer lot-to-lot, or even wafer-to-wafer, removal rate decay due to aging [111]. The effectiveness of pad conditioning is impacted by pad aging [112], potentially making the pad surface properties and the resulting pad-wafer contact inconsistent in the latter stages of pad usage. This is compounded by the fact that in-situ measurement of changes in pad surface properties as well as pad-wafer contact events is difficult [113], so that process control monitoring of this degradation is challenging. The relationship between pad property changes and pad aging is not well understood.

In this section, we investigate CMP pad surface properties and evaluate pad aging effects by ex-situ measurements. Pad asperity modulus and height are measured by nanoindentation and microprofilometry, respectively. Pad-wafer contact fraction is measured by laser confocal microscopy [85]. To cause aging, a CMP pad is used to polish blanket PETEOS wafers, with in-situ conditioning for a total of 16 hours. At different stages of this marathon test, physical measurements are performed at the same location on the pad and the measured results are compared.

4.4.1 Polishing experiment and pad sample collection

A 31-inch commercially available JSR water soluble particle (WSP) pad was used to polish 300-mm blanket PETEOS wafers on an Araca APD-800 polisher. The wafer polishing pressure was kept at 4 psi with the carrier rotating at 25 rpm. The pad was in-situ conditioned using a 3M A98-AF diamond disk at 8 lbf. The conditioning head

rotated at 95 rpm and swept at 10 times per minute with a standard sinusoidal sweep schedule. Polishing time for each wafer was kept to 1 minute. Removal rate was measured on each wafer after polishing.

At three different polishing/conditioning stages, zero (i.e., just after break-in), 8 and 16 hours, a square sample was extracted at 23 cm from the center of the polishing pad (Figure 4.10). The sample size is 2.5 by 2.5 cm. Each sample was rinsed with deionized (DI) water to remove any slurry residues from the pad surface. The sample was then left to air-dry. The collected samples were used to perform pad property tests as described in Section 4.2 and Section 4.3.

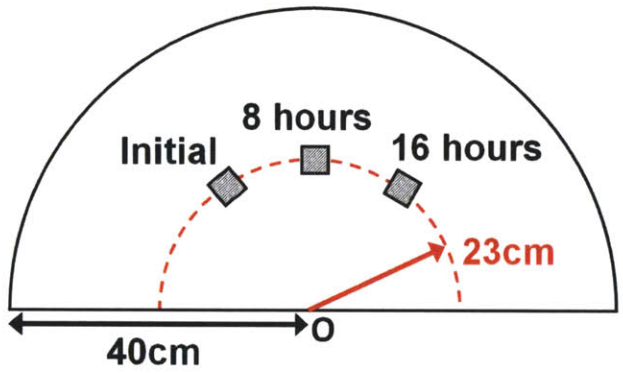


Figure 4.10: Sample collection from the polishing pad at different usage stages.

4.4.2 Pad aging results

Figure 4.11 shows the asperity modulus measured at different pad aging stages. Error bars indicate one standard deviation based on solid contact tests achieved within a test pattern in nanoindentation. Results show that pad asperity modulus is independent of pad age across the entire polishing/conditioning process (i.e., 16 hours). This conclusion is made in light of the fact that confidence intervals overlap at the same indentation depth. Figure 4.11 also shows that there is a strong depth dependence of the asperity modulus whereby modulus decreases as indentation depth is increased.

Figure 4.12 shows the characteristic asperity height extracted from pad sample surface scans at different pad aging time (error bars indicate one standard deviation based on the characteristic asperity heights extracted from two scanned regions). No time dependence of characteristic asperity height is found.

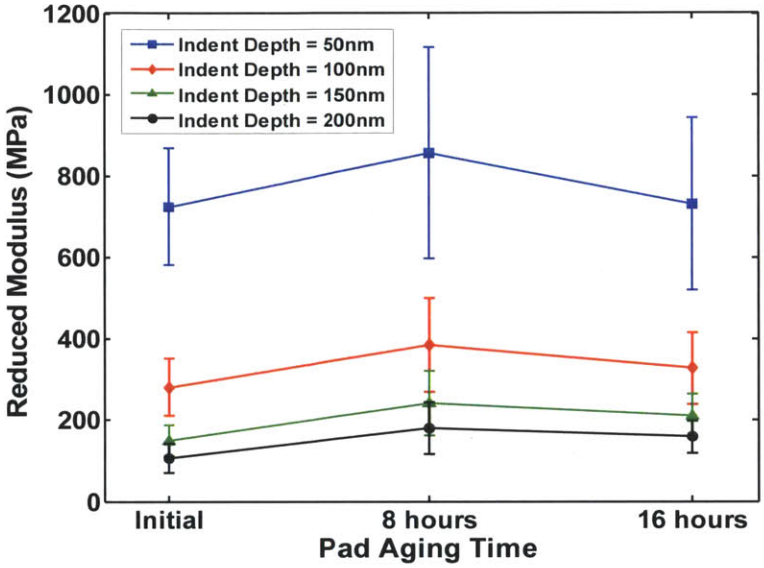


Figure 4.11: Pad asperity modulus vs. pad aging time. Error bars represent one standard deviation across multiple indentation measurements in a 40 μm by 40 μm indentation test region.

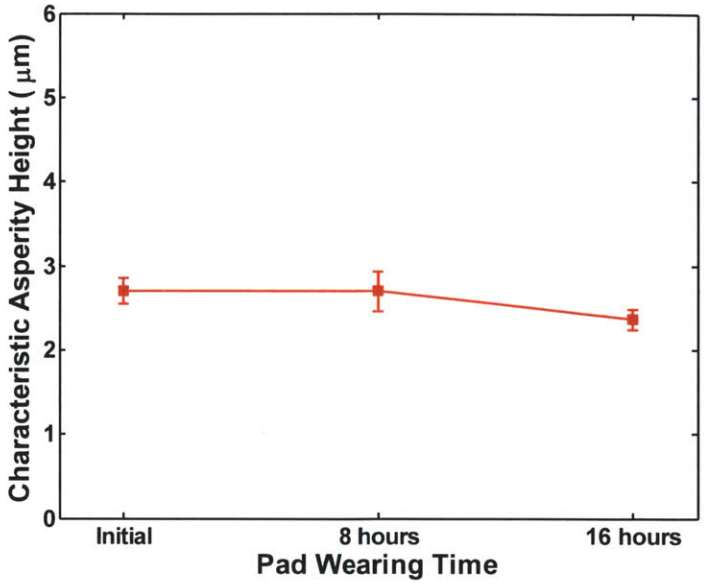


Figure 4.12: Characteristic asperity height vs. pad aging time. Error bars represent one standard deviation.

The pad groove depth can be measured using the microscope and the positioning control system on the Hysitron TriboIndenter. This is done by first focusing the microscope at the top of a stripe and then again at the bottom of the groove, as illustrated in Figure 4.13. The positioning control system records the change in distance between the two focus steps thus providing an accurate value for the groove depth.

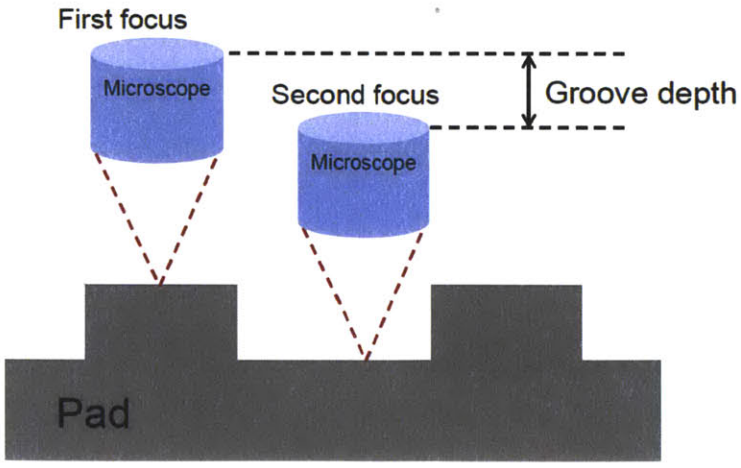


Figure 4.13: Pad groove depth measurement with positioning control system on nanoindenter.

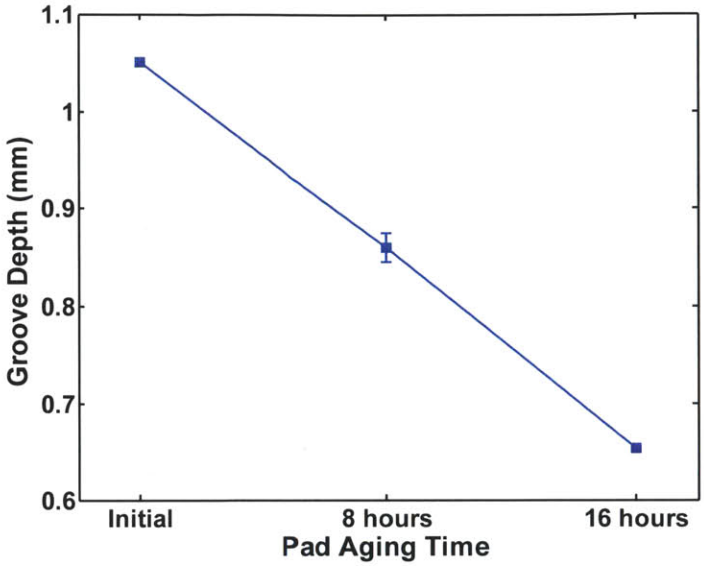


Figure 4.14: Pad groove depth vs. pad aging time. Error bars represent one standard deviation.

Figure 4.14 shows the measured groove depth at different pad ages (error bars indicate one standard deviation based on three depths measured at each age). As expected, pad groove depth decreases linearly with polishing/conditioning time due to uniform pad wear as pad material is removed by conditioning during the process.

Pad surface contact area is measured with a Zeiss LSM 510 Meta NLO laser confocal microscope. The procedure is described in detail by Sun [85]. All measurements are performed at a reference pressure of 4 psi. A series of ten contiguous, non overlapping 450 by 450 μm images are taken. Within a recorded image, the contact percentage is determined by dividing the contact area by the total area of the image. No statistically significant time dependence of contact percentage is observed, as shown in Figure 4.15.

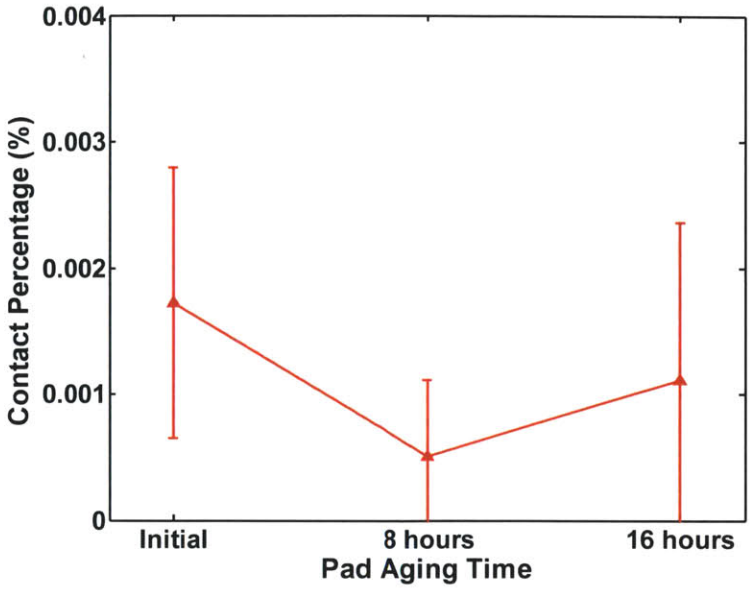


Figure 4.15: Contact percentage vs. pad aging time. Error bars represent one standard deviation.

PETEOS removal rates are compared in Figure 4.16 at three pad ages (error bars indicate one standard deviation based on removal rates measured on five monitor wafers at each age). Results indicate that removal rate is relatively stable during the entire 16 hours of pad use.

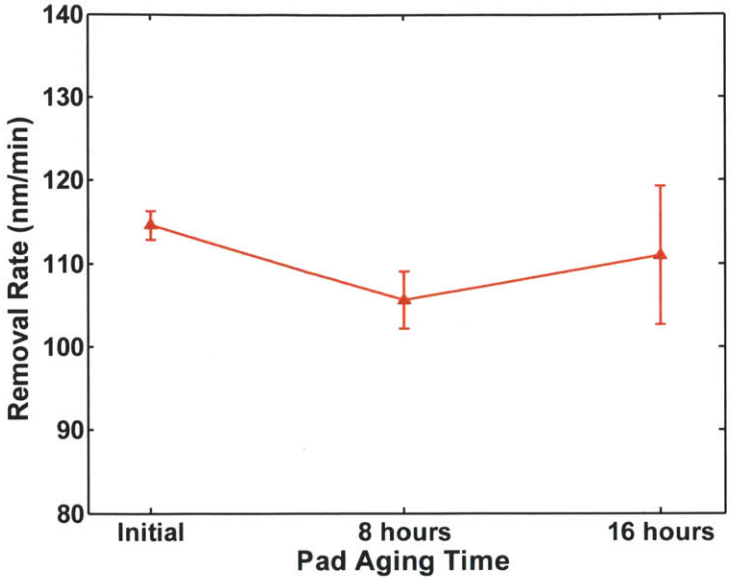


Figure 4.16: TEOS removal rate vs. pad aging time. Error bars represent one standard deviation.

4.4.3 Conclusion

Multiple measurement techniques, including nanoindentation, microprofilometry and laser confocal microscopy are used to characterize CMP pad properties and pad aging effects. Asperity modulus and asperity height distribution are approximately uniform during the whole polishing/conditioning time. Pad groove depth indicates uniform pad wear throughout the polishing/conditioning time; groove depth decreases from 1.06 to 0.66 mm, showing substantial pad wear over this time. The pad-wafer contact area percentage is consistent during the pad usage time. PETEOS removal rate is uniform during the whole polishing time up to 16 hours. The pad conditioning in the polishing experiment thus achieves consistent pad surface properties for stable polishing results, even though substantial pad material is removed by conditioning in the process.

4.5 Spatial variation of pad aging

This section investigates the spatial dependence of pad wearing. After 16 hours polishing/conditioning, eight square samples were extracted from the different locations on the polishing pad as illustrated in Figure 4.17. The sample size is 1.5 by 1.5 cm. Each sample was rinsed with deionized (DI) water to remove any slurry residues from the pad surface, then left to air-dry. The collected samples were used to perform pad property tests as described in Section 4.2 and Section 4.3.

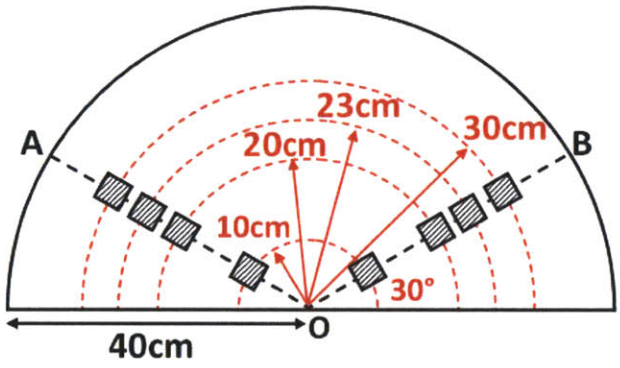


Figure 4.17: Spatial sample collection from the polishing pad after 16 hours polishing/conditioning.

Figure 4.18 compares the asperity modulus measured at different locations on the pad. Error bars indicate one standard deviation based on solid contact tests achieved within a test pattern in nanoindentation. Since confidence intervals overlap at the same indentation depth, the results indicate that pad asperity modulus is approximately independent of pad locations in both OA direction and OB direction.

Figure 4.19 shows the characteristic asperity height extracted from pad sample surface scans at different locations on the pad (error bars indicate one standard deviation

based on the characteristic asperity heights extracted from two scanned regions). No statistically significant spatial dependence of characteristic asperity height is observed.

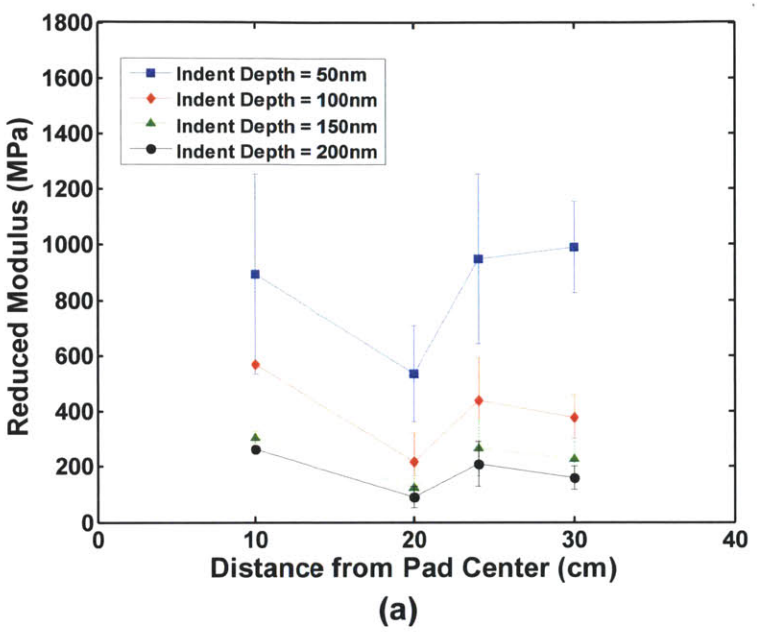
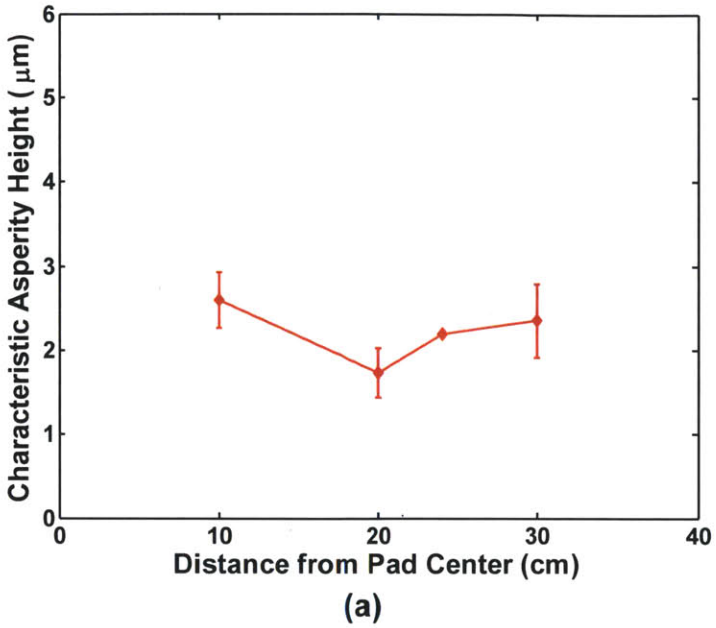


Figure 4.18: Pad asperity modulus vs. sample locations. Error bars represent one standard deviation: (a) OA direction in Figure 4.17. (b) OB direction in Figure 4.17.



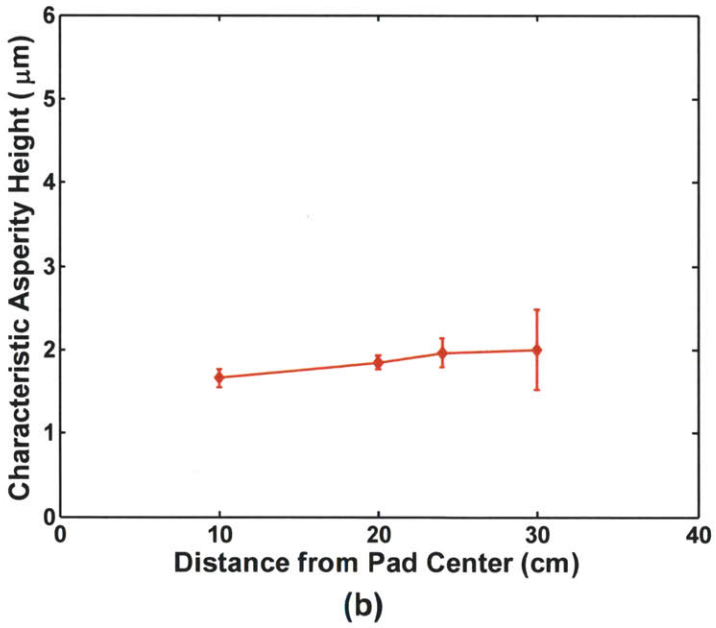


Figure 4.19: Characteristic asperity height vs. sample locations. Error bars represent one standard deviation: (a) OA direction in Figure 4.17. (b) OB direction in Figure 4.17.

Groove depth has a strong radial dependence, as shown in Figure 4.20 (error bars indicate one standard deviation based on three depths measured at each radius). More pad wear occurs near the pad center area so that less groove depth remains after 16 hours

polishing and conditioning. This is caused by the non-optimized pad conditioning recipe in the experiment.

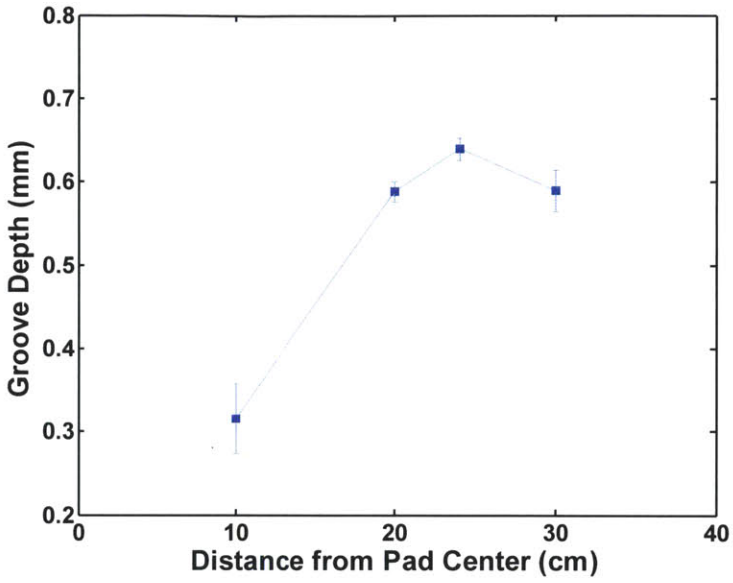


Figure 4.20: Groove depth vs. locations on the pad after 16 hours polishing/condition. Error bars represent one standard deviation.

The spatial results of modulus, characteristic asperity height and groove depth indicate that although the non-optimized conditioning makes pad wear non-uniform, conditioning still succeeds in keeping the surface properties approximately uniform across the whole pad.

4.6 Comments on physical characterization and model characterization of pad properties

In Chapter 3, we characterized CMP pad properties by fitting polishing data and extracting physical model parameters. In this chapter, pad properties are physically measured. In this section, we compare and discuss pad properties as evaluated by physical characterization and model characterization.

The measured pad surface modulus has strong indentation depth dependence. However, in model fitting, only a single modulus is extracted. Figure 4.21 shows the depth dependence of modulus measured on a conditioned IC1000 pad. The modulus decreases from above 800 MPa at the surface to below 250 MPa in the bulk. The model extracted modulus is 447.9 MPa for an IC1000 pad in Section 3.5, shown as the dashed line in Figure 4.21. The model extracted modulus falls between the shallow indent modulus and the deep indent modulus.

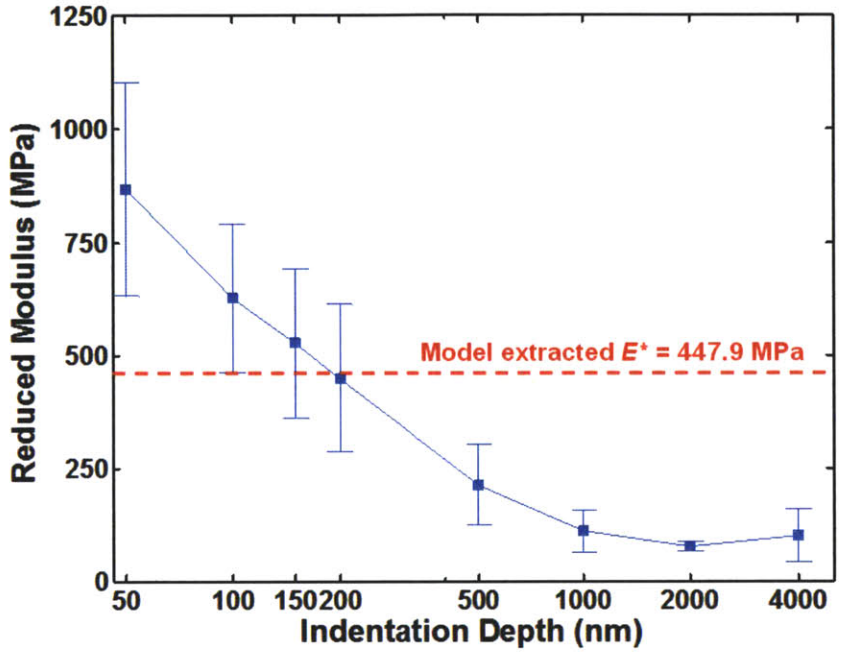


Figure 4.21: Reduced modulus vs. indentation depth for IC1000 CMP pad.

The measured modulus of a porous CMP pad by nanoindentation contains multiple-scale pad surface force response effects [97]. When the indentation depth is less than 200 nm, the indenter tip settles over a single asperity. The measured modulus in this depth region is single asperity dominant. When the indentation depth is greater than 500 nm, the single cell wall bending becomes more pronounced. So the measured modulus is local cell structure dominant. On the other hand, the die-level model extracted pad modulus is close to the macro scale pad bulk modulus, 300 ~ 500 MPa [93, 94]. This is due to the fact that pad long range bending is around a few millimeters in chip-scale. Therefore, physical characterization evaluates the pad surface asperity modulus, and model

characterization evaluates the pad bulk modulus. To understand more details about measured pad bulk modulus and model extracted pad bulk modulus, microindentation or nanoindentation with a flat punch tip in millimeter scale [96, 97, 114] could be considered in future work .

The model extracted characteristic asperity height λ (about 100 nm in Section 3.2) is much smaller than the λ values we typically measured from pad surface topography through stylus profilometry (2~3 μm in Section 4.4). The large difference is due to our die-level model assumption: pad surface is constantly in contact with the wafer surface. As learned from pad-wafer contact percentage in laser confocal microscopy (Figure 4.15), the contact area is less than 0.01% of the total wafer area. The die-level model assumption does not consider the actual contact percentage between pad and wafer. In the chip scale or feature scale dimension, the layout pattern size is about 50 μm and the step height is about 2 μm in the experiment discussed in Section 3.2. The feature line or space may only "see" the micro/nano topography on the head of an asperity. The die-level model assumes that wafer features are affected only by the interacting surface of the pad. Thus, the model fitting only takes into account the nano topography of asperity heads to extract λ, and the extracted λ is based only on the asperity top roughness, which is at a nanometer scale rather than micrometer scale of overall pad asperity heights.

The model extracted characteristic asperity height also reflects an important pad surface property, as discussed in Section 3.2: it tells the planarization ability of the pad. Specifically, a small extracted characteristic asperity height means higher planarization efficiency, such that fewer pad asperity tips are able to contact wafer surface down areas. However, some important pad topography information is missed by the extracted die-level model λ, as the large scale pad roughness may affect slurry holding capacity and pad-wafer contact. Thus, both model and physical measurement improvements should be considered in future work. In addition, statistical assumptions about pad-wafer contact could be added to the die-level model. Atomic force microscopy (AFM) could be considered to obtain nano-scale topography on asperity heads.

4.7 Summary

In this chapter, we characterize CMP pad surface properties by ex-situ physical measurements. Pad asperity modulus and characteristic asperity height are measured by nanoindentation and microprofilometry, respectively.

Pad aging effects are investigated by comparing physical measurement results at different pad usage stages. To cause aging/wearing, a CMP pad is used to polish blanket TEOS wafers, with in-situ conditioning for a total of 16 hours. At different stages of the test, physical measurements are performed at the same location on the pad. In general, conditioning succeeds in keeping the pad surface properties consistent to perform polishing. Asperity modulus and asperity height distribution are approximately consistent during the whole polishing/conditioning time. But pad groove depth significantly decrease in the process, confirming substantial pad wear.

After 16 hours polishing/conditioning, spatial variation of pad aging/wearing is verified by comparing pad properties at different locations on the pad. Pad asperity modulus and characteristic asperity height are approximately uniform across the whole pad. Groove depth has a strong radial dependence, since the non-optimized conditioning recipe results in more wear in the center region of the pad.

5 Modeling for CMP with pad-in-a-bottle (PIB)

This chapter explores physical models for CMP with an alternative approach, pad-in-a-bottle (PIB). Section 5.1 first reviews the current understanding of the CMP process, and then introduces a new CMP technology using pad-in-a-bottle. Section 5.2 proposes two modeling approaches and discusses potential challenges in applying pad-in-a-bottle on the basis of these models.

5.1 Introduction to pad-in-a-bottle

5.1.1 The nature of CMP

The laser confocal microscopy measurement (in Section 4.4) shows that the contact area between pad and wafer is extremely small (less than 0.01%) under the reference pressure of CMP. Only the tallest asperities are in contact with the wafer surface. The pressure within the contact spots is concentrated and extremely high. Polishing is believed to be driven by the material removal under these microscopic contacts.

Figure 5.1 illustrates the relation between random material removal events and polishing results. During the CMP process, random pad asperity contacts create multiple single removal events on the wafer surface at different moments. As these material removal events are accumulated, wafer surface material is reduced; that is to say, polishing occurs.

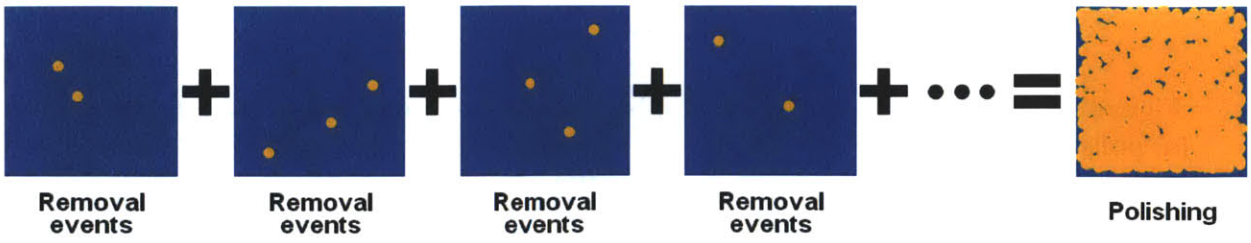


Figure 5.1: The nature of CMP: polishing is the accumulated result of many random material removal events.

5.1.2 What is pad-in-a-bottle

To achieve polishing, the key task is to generate large numbers of these single removal events. In conventional CMP, the contacts are provided by pad asperities, as illustrated in Figure 5.2(a). The pad surface requires conditioning with a diamond disk to create and maintain asperities by cutting the pad.

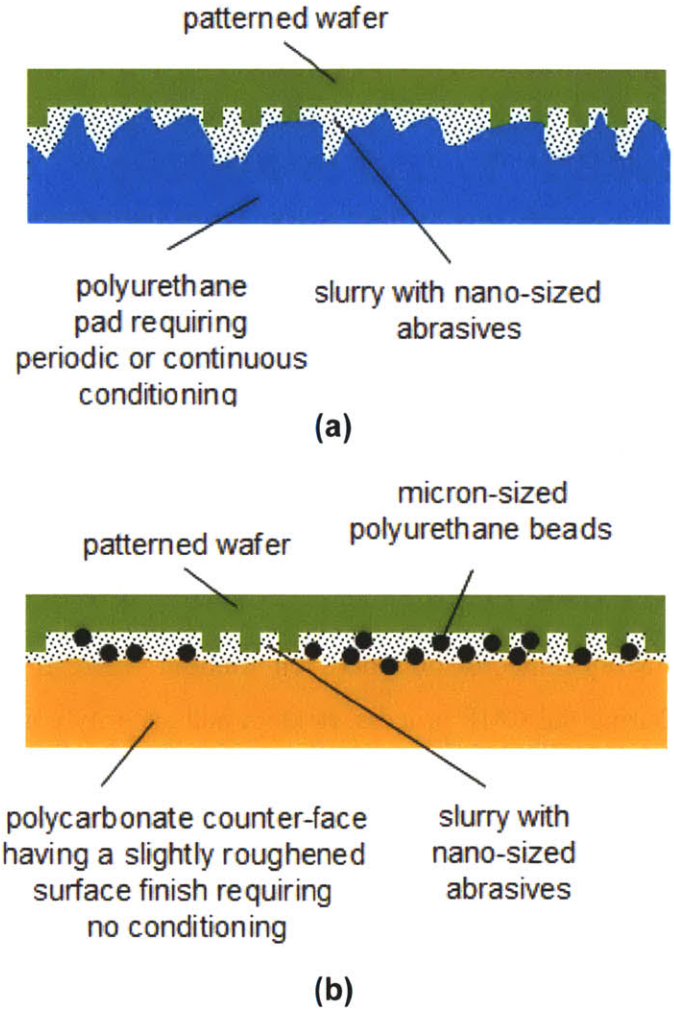


Figure 5.2: CMP with conventional polishing pad and novel pad-in-a-bottle [116]: (a) polishing with polyurethane pad; (b) polishing with polyurethane beads as pad-in-a-bottle.

Instead of using randomly created pad asperities, a new idea has been proposed that contacts can be made with controlled "pad particles" included within the slurry; this is referred as the "pad-in-a-bottle" approach [115]. Generally, pad-in-a-bottle consists of suspended polymer beads (preferred but not limited to polyurethane) mixed into slurries, as shown in Figure 5.2(b). A low-cost counterface (for example, polycarbonate) is applied instead of the pad in CMP. The counterface has a certain roughness to catch the polymeric beads and prevent their rolling motion. The polymeric beads are pressed onto the wafer surface and contacts can be achieved to generate material removal events.

5.1.3 Why use pad-in-a-bottle

Pad-in-a-bottle has a potential intrinsic advantage compared to a conventional pad: controllability. In conventional CMP, pad asperities have irregular shapes and wear properties, as shown in Figure 5.3(a). The uncontrolled shapes cause variations in CMP performance and are believed to contribute to defects. Pad-in-a-bottle offers predictable and controllable contacts. As shown in Figure 5.3(b), a single layer of monosized polymer particles can give very uniform contact conditions including height, curvature and density. This should substantially reduce process variability since we will have almost identical material removal events driven by the polymer beads.

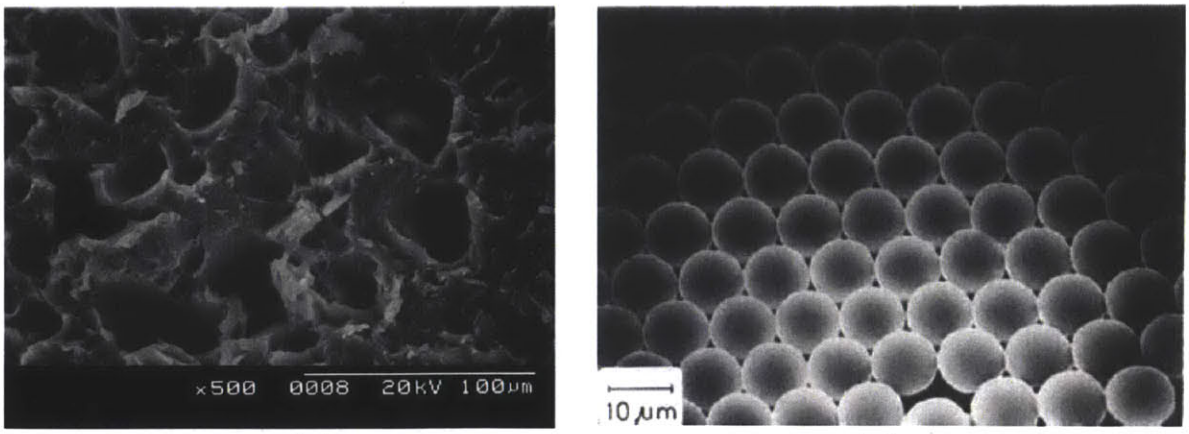


Figure 5.3: Asperities vs. monosized polymer particles: (a) Surface of conditioned IC1400 pad [117]; (b) Single layer of monosized 10 µm polystyrene particles [118].

Pad-in-a-bottle also has the potential to reduce the cost compared to conventional pad based CMP. A conventional polishing pad typically costs a few hundred dollars, and the pad life is usually only two or three days of continuous use in high volume manufacturing. A disposed pad still contains about two thirds of the pad material. Changing a pad can take several hours, including installing, break-in and monitor wafer test, until the new pad is ready to use in the production line. Both pad and tool down time costs are significant. Conditioning disks are another cost concern; these are usually several hundred dollars each and have a similar life time as a pad. Applying pad-in-a-bottle has the potential to eliminate the use of pad and conditioning disk. Instead, fresh polymer beads are delivered on the counter-face with the slurry continuously so they can generate the polishing action. Since polymer beads will be crushed or deformed by the wafer, it is necessary to remove used beads from the counterface. This can be done by a soft brush gently sweeping on the counterface. For pad-in-a-bottle, less substantial wear is expected to occur on the counterface, and little wear should occur in the soft brush. Thus the polishing tool should be easier to maintain, and counterface and brush changing is required less frequently.

Pad-in-a-bottle also has important environmental benefits. It reduces polymer material usage, because of no pad disposal. The polymer beads may be able to be extracted or recycled from waste stream.

5.1.4 Current progress of pad-in-a-bottle

Pad-in-a-bottle is still in a very early stage of development. Only a few published works have sought to demonstrate the concept in blanket silicon wafer polishing. Lu [119] polished 100 mm silicon wafers with benzoguanamine (BG) particles as micro pads and with a glass plate as counterface. Similar removal rate and roughness were achieved as in polishing with a conventional pad, as shown in Figure 5.4. Xu [120] used polystyrene (PS) particles in blanket silicon wafer polishing. Significant removal rate and scratch reduction were observed, shown in Figure 5.5. However, detailed force response of polymer particles has not been studied. The relationship between removal rate and applied pressure is not understood. Successful dielectric polishing with the concept of

pad-in-a-bottle has not yet been reported. Current efforts are focused on polymer beads and counterface selection for enabling oxide polishing [116].

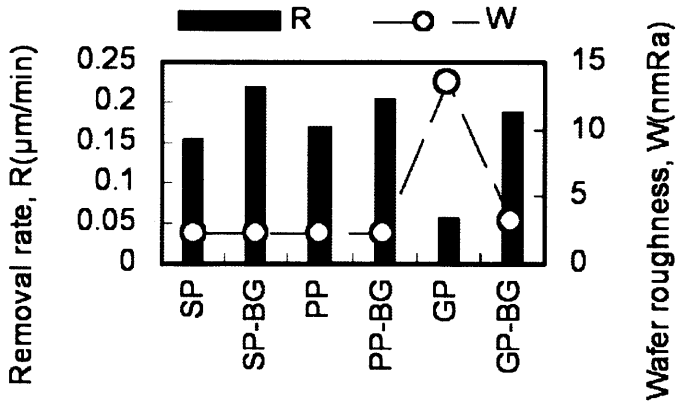


Figure 5.4: Polishing results of different methods [119]. SP: solid pad. SP-BG: solid pad and BG particles. PP: porous pad. PP-BG: porous pad and BG particles. GP: glass plate. GP-BG: glass plate and BG particles.

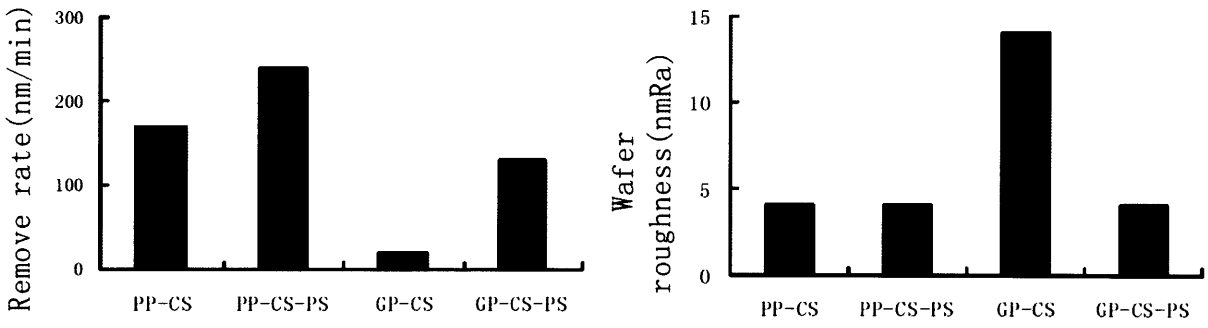


Figure 5.5: Effect of polymer particle on polishing [120]. PP-CS: polyurethane pad and colloidal silica slurry. PP-CS-PS: polyurethane pad, colloidal silica slurry and polystyrene particles. GP-CS: glass plate and colloidal silica slurry. GP-CS-PS: glass plate, colloidal silica slurry and polystyrene particles.

5.2 Modeling for mechanical response of pad-in-a-bottle

Physical models can help to understand the behavior of pad-in-a-bottle and provide some guidance for experimentation. This section analyzes the force response of pad-in-a-bottle, and hypothesizes a correlation between material removal rate and applied pressure. In Section 2.3, the particle-level model assumes that every asperity has a spherical head and utilizes Hertz contact theory. For pad-in-a-bottle, the polymer beads are assumed to be nearly ideal spherical in shape and size, so that a particle-level modeling approach can be followed.

Two cases of polymer bead formation are considered in our modeling work, bead packing and bead stacking, illustrated in Figure 5.6. In both cases, we assume that the polymer beads are ideally monosized. No rolling motion of polymer beads occurs. Counterface roughness is much smaller than bead size so that the counterface can be treated as a flat plane. Slurry abrasive particles are densely adsorbed on the surface of the polymer beads, and the abrasive size is also much smaller than the polymer bead size. When a polymer bead comes in contact with the wafer surface, abrasives are pressed between them and material removal occurs under relative motion.

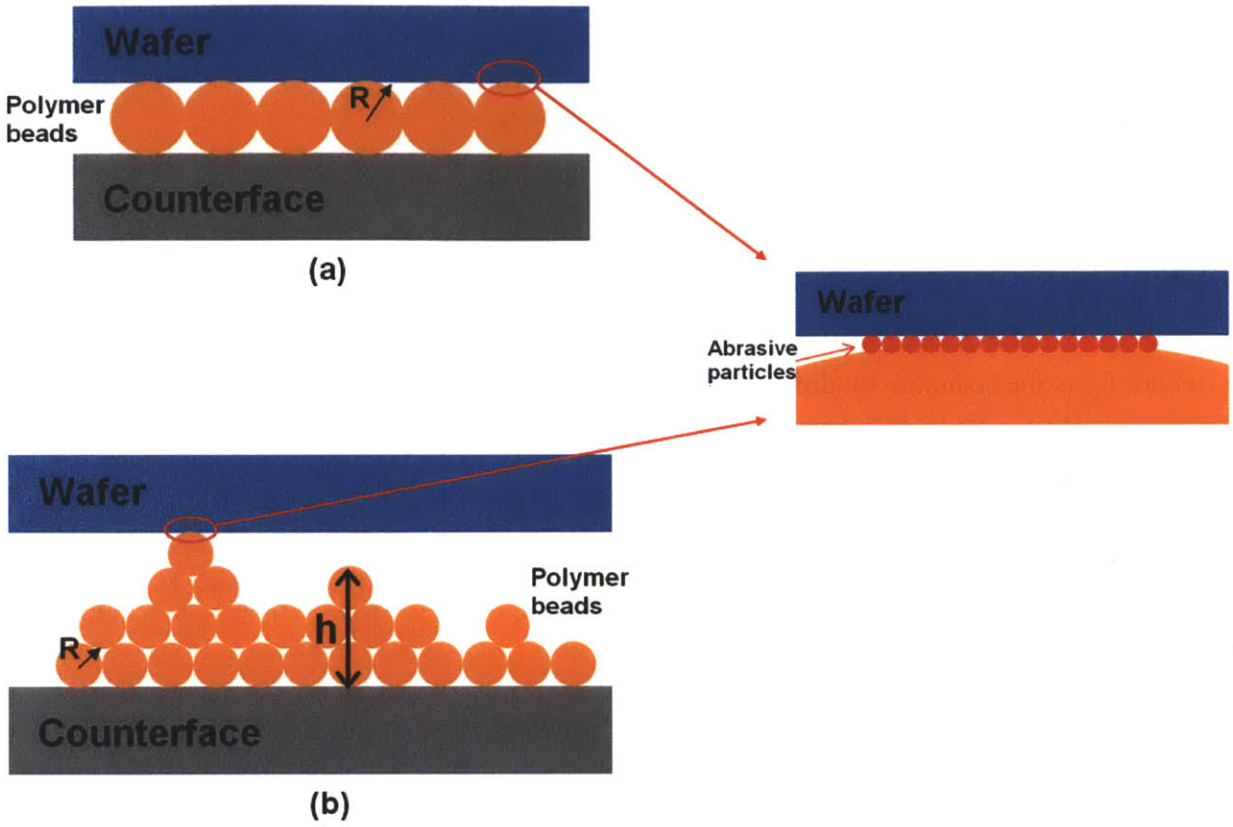


Figure 5.6: Model approaches for pad-in-a-bottle: (a) Bead packing; (b) Bead stacking.

5.2.1 Bead packing

The bead packing case describes the behavior of a monolayered set of polymer beads. A single layer of monosized beads are spread on the counterface, as shown in Figure 5.6(a). The beads are densely packed, i.e., there is no gap between any neighboring beads. All of the beads with the same radius R bear the force from the wafer identically, with a compressed deflection of δ. Using the Hertz contact results, the single bead contact area $a(\delta)$, the single bead force load $L(\delta)$ and the average pressure $\bar{p}(\delta)$ within the single contact spot can be expressed as [58]:

$$\begin{cases} a(\delta) = \pi R \delta \\ L(\delta) = \dfrac{4}{3} E_p R^{\frac{1}{2}} \delta^{\frac{3}{2}} \\ \bar{p}(\delta) = \dfrac{L(\delta)}{a(\delta)} = \dfrac{4E_p}{3\pi} R^{-\frac{1}{2}} \delta^{\frac{1}{2}} \end{cases} \tag{5.1}$$

where E_p is the reduced modulus of the polymer beads. We assume that the total number of beads in contact is N. The total area occupied by N densely packed beads is $A_0 = 4NR^2$. Then the total force can be calculated as

$$F(\delta) = NL(\delta) = P_0 A_0 \tag{5.2}$$

where P_0 is the applied reference pressure. Combining Eq. 5.1 and 5.2, we can solve for the average pressure $\bar{p}$ within a contact spot as

$$\bar{p}(P_0) = \frac{4}{\pi}\left(\frac{E_p}{3}\right)^{\frac{2}{3}} P_0^{\frac{1}{3}} \tag{5.3}$$

This average pressure can be applied in the microscopic formulation of Preston's law [80] to correlate pressure and removal rate as in the model derivation in Section 2.5.

However, the average spot pressure $\bar{p}(\delta)$ is determined by the applied pressure P_0. This is different from the pad asperity response. Recalling the average spot pressure of pad asperity response in Eq. 2.60, in that case we found that the pressure only depends on asperity modulus, asperity curvature and asperity height. The difference between the standard pad case and the pad-in-a-bottle case in Eq. 5.3 comes from our model assumption. In our bead packing model for pad-in-a-bottle, we have contact spots everywhere. But in the asperity model, we have only a few contact spots due to the distribution of asperity height. So the average spot pressure from monodisperse beads may be much lower than what we find from asperity contact. This may cause a serious

problem in CMP: we may not overcome a threshold pressure required to initiate polishing action. Before we proceed to the next step of derivation, the threshold pressure problem needs to be considered in more detail.

A threshold pressure P_{th} is observed in blanket wafer polishing for many materials [121-124]. When the applied reference pressure is lower than the threshold pressure, no material removal is achieved. If the reference pressure exceeds the threshold pressure, the material is removed and the removal rate follows the trend of Preston's law [23]. A mechanism for threshold pressure is suggested by Zhao and Shi [125]. At a microscopic scale, a critical value of pressure p_c has to be overcome on each asperity contact spot before material removal can happen. This critical value is observed as a threshold pressure in polishing. Based on this mechanism, we can estimate the critical pressure p_c for oxide polishing. The threshold pressure observed in oxide polishing with Rohm and Haas IC pad and silica slurry is about 1 psi [121]. The measured pad-wafer contact indicates that the contact percentage of an IC pad is about 0.02% under 1 psi [85]. To establish a lower bound on threshold pressure, we assume that the contact percentage is at most 0.1% under 1 psi. Then, the average local pressure within the contact spots is 1000 psi (6.9 MPa), dividing 1 psi by 0.1%. An equivalent asperity contact pressure of p_c = 6.9 MPa might be a difficult challenge to achieve in the case of bead packing. Assume we use polymer beads with E_p = 100 MPa. Letting $\bar{p} = p_c$ in Eq. 5.3, we get P_{th} = 20.7 psi (0.14 MPa). This is a very high pressure, not usually considered or used in CMP. It is possible that the bead packing formation provides too much contact area and results in too low a value for local contact pressure. The monolayer bead distribution may not be a good option for pad-in-a-bottle.

Once the critical pressure is derived, we can utilize microscopic formulation of Preston's law [80] to calculate the removal rate of each bead as

$$\overline{RR_p} = K'V\left(\overline{p} - p_c\right) \tag{5.4}$$

where K' is the microscopic Preston constant and V is the relative velocity. The removal rate on a blanket wafer is estimated as

$$RR = \overline{RR_p} \frac{N}{A_0} \tag{5.5}$$

The relation between critical pressure and threshold pressure is expressed as

$$p_c = \frac{4}{\pi}\left(\frac{E_p}{3}\right)^{\frac{2}{3}} P_{th}^{\frac{1}{3}} \tag{5.6}$$

Combining Eq. 5.3 through 5.6, we have

$$RR = \frac{1}{\pi R^2}\left(\frac{E_p}{3}\right)^{\frac{2}{3}} K'V\left(P_0^{\frac{1}{3}} - P_{th}^{\frac{1}{3}}\right) \tag{5.7}$$

As estimated above, the macroscopic applied threshold pressure, P_{th}, could be as high as 20 psi. Material removal is not easy to achieve using a single layer of monosized polymer beads.

The result of a single layer of densely packed beads can be easily generalized to multiple layers of densely packed beads, with the same result. We can just take the bottom plane of the top layer of beads as the reference plane; the derivation and expressions are exactly the same mathematically, with the same behavioral result and dependencies.

5.2.2 Bead stacking

The bead stacking case describes the behavior of randomly aggregated polymer beads. The friction between neighboring beads is assumed to be strong enough so that no relative motion occurs in the configuration, shown in Figure 5.6(b). The stacking height

distribution is an important consideration: only some tall peaks are in contact with the wafer. This case is very similar to the pad asperity response in conventional pad CMP.

The top elastic deformation of a single active stacked peak with height h is the same as that of a single bead in Eq. 5.1, because the contact is made by a single bead. We assume a stacking height distribution or probability height density function $\phi(h)$. If the distance between the wafer and the counterface is d, the peaks with height larger than d will be in contact with the wafer surface. The number of peaks in contact is

$$n = N\int_{d}^{\infty}\phi(h)\,dh \tag{5.8}$$

where N is the total number of active peaks. For the peaks with height $h > d$, the deformation is $\delta = h - d$. The total peak contact area is

$$A = N\int_{d}^{\infty}a(h-d)\phi(h)\,dh \tag{5.9}$$

For an applied force of F_0, the distance d can be obtained as

$$F_0 = N\int_{d}^{\infty}L(h-d)\phi(h)\,dh \tag{5.10}$$

To simplify the mathematics for the derivation, we consider an example assuming an exponential peak height distribution

$$\phi(h)=\frac{1}{\lambda}e^{-\frac{h}{\lambda}} \tag{5.11}$$

where λ is defined as the characteristic peak height. Then the number of peaks in contact n, total contact area A, and the applied force F_0 can be determined as

$$\begin{cases} n = Ne^{-\frac{d}{\lambda}} \\ A = N\pi R\lambda e^{-\frac{d}{\lambda}} \\ F_0 = E_p N\sqrt{\pi R\lambda^3}e^{-\frac{d}{\lambda}} \end{cases} \tag{5.12}$$

We can solve for the average pressure within a contact spot, obtaining

$$\bar{p} = \frac{F_0}{A} = \frac{E_p\sqrt{\lambda}}{\sqrt{\pi R}} \tag{5.13}$$

This value is independent of applied reference pressure or force, which is similar to the result for conventional CMP with asperity contact, i.e., the dominant factor to obtain more contact area is to create more contact spots. However, this value for average pressure is fairly high if we choose hard polymer beads, and the microscopic critical pressure may be overcome easily. Take the oxide polishing critical pressure as we estimated in Subsection 5.2.1, p_c = 6.9 MPa. If we use beads with E_p = 100 MPa and R = 10 μm, and assume a characteristic peak height of λ = 100 μm, we find $\bar{p}$ = 178.6 MPa, which is significantly higher than p_c. In this case, the removal rate of each bead is estimated as

$$\overline{RR_p} = K'V\bar{p} \tag{5.14}$$

From Eq. 5.12, we can find the number of bead peaks as

$$n = \frac{F_0}{E_p\sqrt{\pi R\lambda^3}} \tag{5.15}$$

Considering reference pressure as $P_0 = \frac{F_0}{A_0}$, the removal rate on a blanket wafer is estimated as

$$RR = \overline{RR_p}\frac{N}{A_0} = \frac{1}{\pi R\lambda}K'VP_0 \tag{5.16}$$

The relation is linearly dependent on applied reference pressure. We also notice that the removal rate is independent of polymer bead modulus. This is due to the assumption of exponential peak height distribution, in which case the removal is dominated by the height distribution.

5.2.3 Remarks on model trend

Two possible bead formations are modeled in the above sections. In the bead packing case, polymer beads are uniformly dispersed. Contact spots are available everywhere on the wafer. But the bearing load of each bead contact is too low to overcome the polishing critical pressure, and material removal is not likely to happen unless extremely high pressure is applied or extremely rigid beads are chosen. In the bead stacking case, polymer beads are randomly aggregated, and only a few peak spots are in contact with the wafer. The bearing load of each spot is very high, and thus the critical pressure is easily overcome and material removal is achieved by the peak spots. Thus, the removal in the bead stacking case is contact event driven, similar to that in conventional CMP.

Comparison of bead packing and bead stacking provides us two important points to understand the nature of polishing: First, polishing is removal event driven; second, a single removal event is contact pressure driven. In bead packing, we create contact events constantly, but these may not be sufficient to create removal events. Every contact spot pressure is too low to achieve material removal. In bead stacking, we randomly create a few contacts, but every contact is a successful removal event because of the high localized contact pressure.

5.3 Outlook

The hypothesized physical models tell us that perfectly uniform controlled polymer bead size and shape may not assist polishing. In future experiments or applications of pad-in-a-bottle, monodispersed beads may be less preferred. In practice, a certain bead size distribution or shape variation may be needed. Aggregation of polymer beads might be controlled by surfactant or additives in the slurry. The selected polymer beads should be superiorly chemical and heat resistant.

The selection of the polishing counterface is also important. The counterface needs sufficient hardness to resist wear. It should have sufficient hydrophilicity, so that slurry and beads achieve traction. Some surface texture has to be made on the counterface to prevent rolling motion of the polymer beads, but the roughness needs to be low enough to prevent direct contact with the wafer.

Since the pad-in-a-bottle has polymer beads mixed into the slurry during polishing, slurry design should consider the following issues. First, polymer beads and abrasives can both be charged in the background chemical environment; we need the abrasives to adsorb onto the surface of the polymer beads. Second, the slurry abrasive particle concentration needs to be high enough to coat the polymer bead surfaces, but low enough to prevent rapid abrasive-bead agglomeration.

Since pad-in-a-bottle is in the earliest stages of development, polishing experiments and applications will first focus on blanket wafers. Removal rate, roughness and defect rate should be considered in this stage. Once successful blanket wafer polishing is achieved, planarization of patterned wafers with pad-in-a-bottle can be explored and physical models to predict chip-scale non-uniformity need to be proposed in the future.

6 Conclusions and future work

6.1 Conclusions

This thesis has proposed an advanced methodology for physical modeling and characterization to understand and simulate the chemical mechanical polishing (CMP) process in integrated circuit (IC) manufacturing. We contribute a multiple-level systematic approach to characterize the CMP die-level non-uniformity induced by layout design, pressure distribution and CMP pad properties.

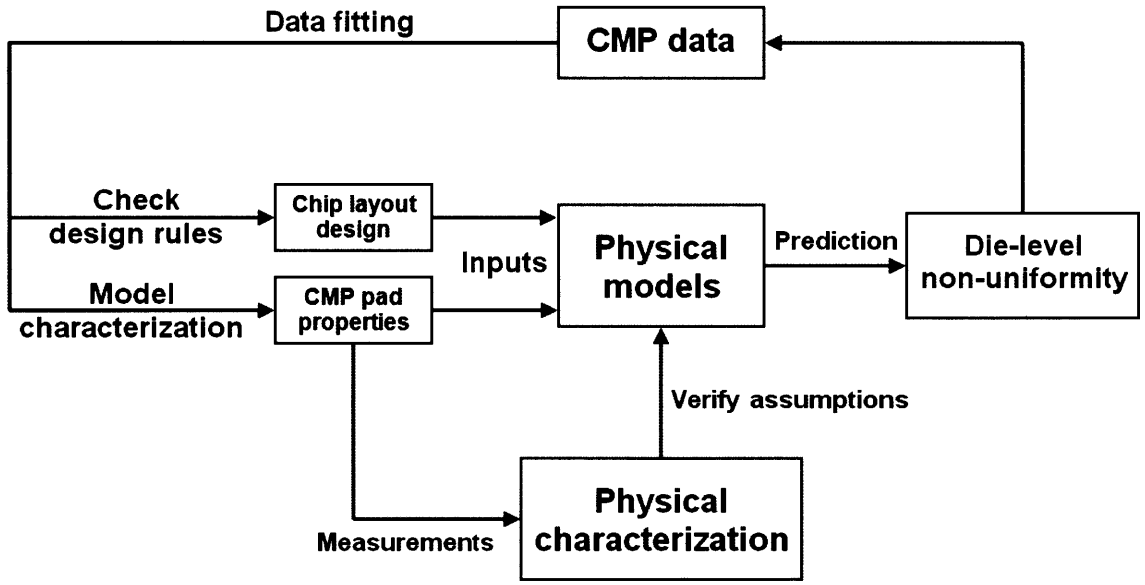


Figure 6.1: Summary of the methodology of CMP modeling and physical characterization.

Figure 6.1 illustrates our systematic methodology for CMP modeling and physical characterization. Physical CMP models are the core of this system. Model fitting of experimental data not only helps to generate simulations that can be used to check layout design rules, but also extracts model parameters with physical meanings. CMP pad

properties can be related to and studied through these model parameters. Physical characterization approaches are developed to directly measure CMP pad properties. The measured results are used to verify model assumptions.

A key contribution of this thesis is the development of a physical model system, including three single-level models and two multiple-level model integration approaches. Three physical CMP models are developed, at the wafer-level, die-level and particle-level. The wafer-level model investigates CMP tool effects on wafer-level pressure non-uniformity. CMP pad thickness, retaining ring size and retaining ring reference pressure are found to be strong factors impacting wafer-level non-uniformity. The die-level CMP model is developed to study die-level non-uniformity of polishing. Pattern density dependence is captured by the die-level model. Pad properties including pad bulk modulus and pad asperity height distribution are related to planarization performance. The particle-level model focuses on the contact mechanism between pad asperities and the wafer. Pad-wafer contact percentage is predicted using the model.

Two extended models are proposed to connect the die-level model to wafer-level and particle-level models, so that CMP system impacts on die-level uniformity and feature size dependence are considered. The wafer-die-level model is accomplished by taking the wafer-level pressure non-uniformity as the pressure boundary condition for die-level simulation. The effect of wafer-level pressure spatial non-uniformity on die-level planarization is estimated. The extended physical die-particle-level model is derived by integrating the particle-level and die-level models together and considering the curvature of the contacting surfaces between pad asperities and chip features. It combines both pattern density and feature size effects.

As a major motivation is to help both layout designers and process engineers, the modeling priority of this thesis lies in die-level non-uniformity of polishing. An important advantage of the proposed physical modeling system is that the previous empirical parameter, planarization length (PL), is no longer used as a model parameter. Instead, physical meanings are attributed to two key model parameters, pad effective modulus E_0^* and characteristic asperity height λ. The pad effective modulus is related to

properties of the pad bulk, and is shown to most strongly impact within-die uniformity and layout pattern density effects resulting from long range pad bending. The characteristic asperity height reflects the distribution of pad asperity heights, and is shown to most strongly impact the feature scale step height reduction efficiency.

Physical CMP models are applied to a number of practical engineering problems in CMP. CMP pad stiffness and conditioning effects are studied by extracting physical model parameters from polishing data. Quantitative relationships between pad properties and model parameters are established and applied in model prediction. Higher pad stiffness is shown to achieve better within-die uniformity. A conditioning disk with blocky diamonds is seen to achieve higher local planarization efficiency. Within-die non-uniformity at process endpoint is analyzed under different endpoint strategies based on simulation results using a calibrated model. An extended wafer-die-level model is used to fit polishing data from multiple chips from a wafer, combining pattern density effects and wafer-level pressure variation. The within-die non-uniformity of local planarization is verified to be a function of die position on the wafer due to pressure distribution non-uniformity. The extended die-particle-level model is fitted to oxide stage polishing in an STI process, and accounts for both pattern density and pitch size effects. Within the same pattern density area, non-uniformity occurs due to pitch size variation, and this effect is captured by the extended model.

Another key contribution of this thesis is the development of a set of physical measurement approaches to characterize CMP pad properties. Pad asperity modulus and characteristic asperity height are measured by nanoindentation and microprofilometry, respectively. The measurements focus on the active asperities, i.e., potential solid contact sites between pad and wafer. Floppy asperities can be recognized from nanoindentation testing curves. Stylus profilometry suppresses floppy asperities response in topography scans, enabling extraction of polishing-relevant pad asperity properties.

Pad aging effect is investigated by comparing physical measurement results at different pad usage stages. In-situ conditioning is found to keep pad surface properties (asperity modulus and asperity height distribution) consistent to perform polishing up to

16 hours, even in the face of substantial pad wear. As a verification of model assumptions, the physical characterization indicates that our model parameters modulus E_0^* and characteristic asperity height λ remain constant during the pad aging experiment.

Finally, we have applied the physical particle-level modeling approach to explore an alternative CMP process, a pad-in-a-bottle approach, which consists of suspended polymer beads mixed in slurries. Force responses of two possible bead formations are modeled, bead packing and bead stacking. Considering critical pressure in CMP, physical model predictions suggest that material removal is easier to obtaining in a bead stacking case. In future development of pad-in-a-bottle technology, a polymer bead size distribution and shape variations are suggested to be needed in achieve practical polishing.

6.2 Future work

This thesis has contributed to physical modeling and characterization of CMP. Given the complexity of the process and consumables, many interesting directions of future research remain, in which our general modeling and characterization approaches could be extended.

Dynamic effects need to be considered in the wafer-level model and experimental data fitting. In Subsection 2.1.7, the non-centered wafer position effect is studied by calculating a time-averaged wafer-level pressure distribution. The dynamic wafer position may be affected by many factors, such as platen speed, wafer carrier down force and retaining ring design. Dynamic pressure distribution affected by slurry fluid flow could be considered. To help understanding the details and model fitting, some in-situ test approaches need to be developed to observe the wafer dynamic positions and the dynamic fluid pressure during polishing.

Currently, we only consider the statistical distribution of asperity height in the die-level CMP model. In the further development of extended die-level models, statistical

distribution of asperity size could also be introduced, so that the feature size dependence can be captured more precisely.

In further advanced CMP particle-level models, statistical assumptions about asperity-wafer contact events could be considered, including asperity-abrasive-wafer three body contact. Defect rate could be studied, potentially related to contact event probability and abrasive size distribution.

Further CMP pad characterization approaches could be proposed and applied. In Section 4.4, the CMP pad nanoindentation is performed in a dry environment. Further CMP pad mechanical property tests should consider wet environment nanoindentation, potentially with custom designed fluid cell indenter tips [107]. The pad modulus could then be measured while the pad sample is submersed within slurry, by which we could study the pad response under conditions more similar to those during polishing. AFM imaging is a possible way to study nano-scale topography of asperity tops, and could help differentiate between pad asperity height and asperity head roughness effects in CMP. In-situ characterization of asperity-wafer contact events is a further direction for future study, possibly enabled by MEMS based sensors [113].

Pad-in-a-bottle is a developing novel CMP technology, and modeling for pad-in-a-bottle is a new area to explore. Three-body contact has to be included. Slurry particle agglomeration models could be integrated in a force response model to understand polymer beads, slurry and wafer surface interactions. Physical models of pad-in-a-bottle could also bring benefits in improved understanding of polymer particle assisted CMP, in which both traditional polishing pad and polymer particles acting as micro pads are used.

A. MIT STI CMP test mask

An MIT STI CMP test mask is designed for characterizing and modeling pattern dependent variation in CMP processes for shallow trench isolation (STI) [18]. The mask contains a 21 mm by 21 mm die. The die layout plan is shown in Figure A.1. The structures are grouped into five categories: 1) Long-range pattern density dependence structures; 2) Bias structures; 3) Edge-acceleration effect structures; 4) Dishing and Erosion structures, and L-shape and X-shape structures; and 5) Product cells provided by National Semiconductor.

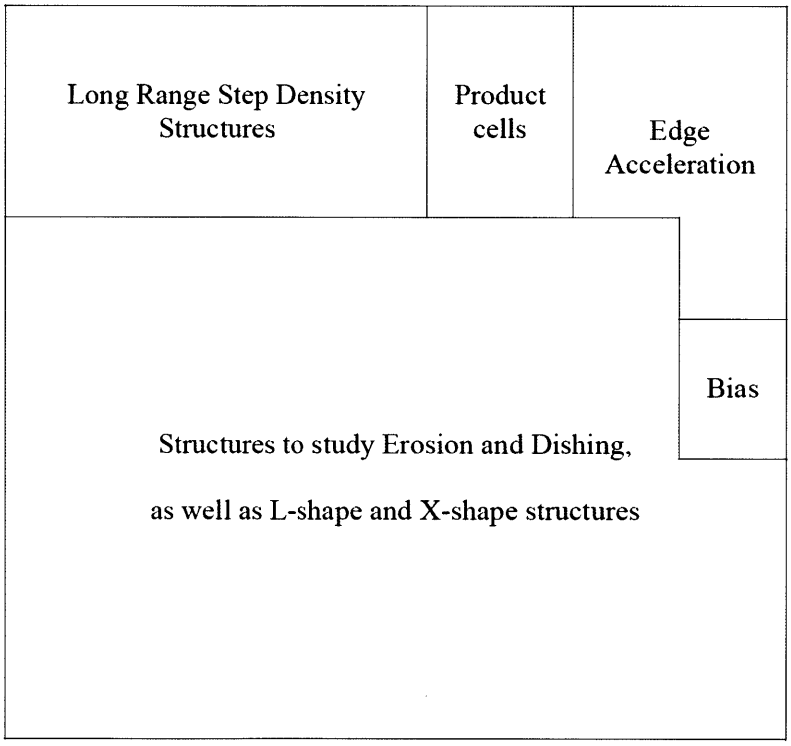


Figure A.1: Layout plan of the MIT STI CMP test mask.

The pattern density arrangement is illustrated in Figure A.2. The number in each block denotes the active density of the region, for example, 50 means 50% density, where active density is defined as the ratio of the active area (covered by silicon nitride) to total area.

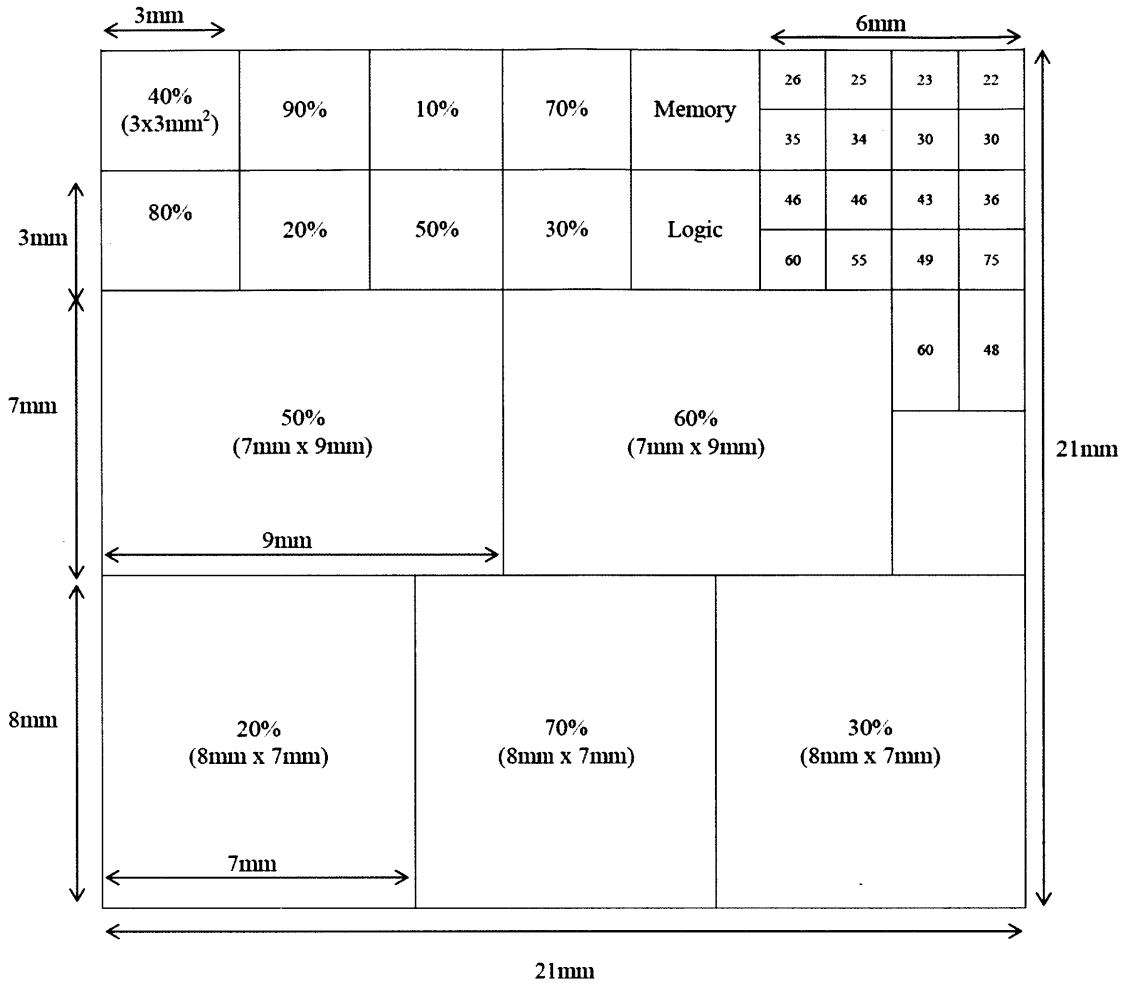


Figure A.2: Pattern density arrangement of the MIT STI CMP test mask.

A.1. Long-range pattern density (LRPD) dependence structures

This region is designed to study long-range pattern density dependency. It is composed of eight 3 mm by 3 mm cells. Each cell has repeated rectangular structures. The density of cells ranges from 10% to 90%. The density cells are placed in random layout arrangements to achieve a good contrast of low and high densities, as shown in Figure A.3.

LRPD_40	LRPD_90	LRPD_10	LRPD_70
LRPD_80	LRPD_20	LRPD_50	LRPD_30

Figure A.3: Arrangement of long-range pattern density structures.

Table A.1 lists the geometry parameters of the long-range pattern density structures. The definitions of L, W and S are explained in Section A.5.

Table A.1: Geometry parameters of the long-range pattern density structures.

Name	L (μm)	W (μm)	S (μm)	Active Density (%)
LRPD_10	40	40	80	10
LRPD_20	50	50	60	20
LRPD_30	50	50	40	30
LRPD_40	70	70	40	40
LRPD_50	60	60	25	50
LRPD_70	80	80	15	70
LRPD_80	90	90	10	80
LRPD_90	100	500	10	90

A.2. Bias structures

This region is composed of fifty 20 μm by 20 μm cells, where each cell contains repeated rectangular structures. These structures are designed to characterize deposition bias, and feature size from 0.25 μm to 5 μm. The descriptions of the structures are listed in Table A.2.

Table A.2: Geometry parameters of the bias structures.

Name	L (μm)	W (μm)	S (μm)	Active Density (%)	Name	L(μm)	W(μm)	S(μm)	Active Density (%)
BI_01	0.5	0.5	5	0.23	BI_26	1	0.5	0.75	22.22
BI_02	0.75	0.5	5	0.43	BI_27	0.75	0.5	0.5	33.33
BI_03	0.5	0.5	2	1.23	BI_28	2	0.75	2	10.00
BI_04	1	0.5	5	0.79	BI_29	2	0.5	0.75	26.67
BI_05	0.5	0.5	1	4.00	BI_30	5	1	5	8.33
BI_06	0.75	0.5	2	2.22	BI_31	2	2	5	8.16
BI_07	2	0.5	5	1.36	BI_32	1	1	2	11.11
BI_08	0.75	0.75	5	0.83	BI_33	1	0.75	1	16.67
BI_09	1	0.5	2	3.70	BI_34	0.75	0.75	0.75	25.00
BI_10	0.5	0.5	0.75	11.11	BI_35	5	0.75	2	14.29
BI_11	0.75	0.5	1	6.67	BI_36	5	0.5	0.75	30.30
BI_12	0.75	0.75	2	4.00	BI_37	1	0.5	0.5	40.00
BI_13	2	0.5	2	5.56	BI_38	2	0.75	1	22.22
BI_14	1	0.75	5	1.52	BI_39	2	0.5	0.5	44.44
BI_15	1	1	5	2.78	BI_40	2	1	2	16.67
BI_16	0.5	0.5	0.5	25.00	BI_41	0.75	0.75	0.5	44.44
BI_17	1	0.5	1	10.00	BI_42	5	2	5	14.29
BI_18	5	0.5	2	7.94	BI_43	1	0.75	0.75	33.33
BI_19	0.75	0.5	0.75	16.67	BI_44	5	0.75	1	27.78
BI_20	2	0.75	5	2.60	BI_45	1	1	1	25.00
BI_21	1	0.75	2	6.67	BI_46	5	1	2	23.81
BI_22	0.75	0.75	1	11.11	BI_47	2	0.75	0.75	40.00
BI_23	2	0.5	1	13.33	BI_48	2	2	2	25.00
BI_24	2	1	5	4.76	BI_49	1	0.75	0.5	53.33
BI_25	5	0.5	1	16.67	BI_50	5	0.75	0.75	45.45

These bias structures have very low densities as indicated Table A.2. To keep them from affecting nearby structures, rectangular buffer structures are added around them.

The bias structures are aligned vertically and surrounded by buffer structures, as shown in Figure A.4.

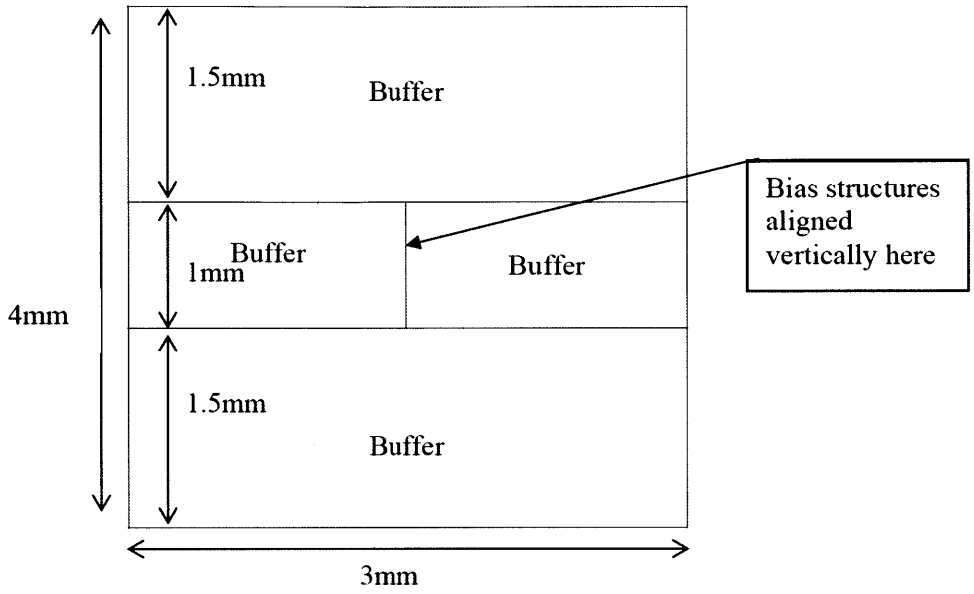


Figure A.4: Arrangement of bias structures and surrounding butter structures.

A.3. Edge-acceleration effect structures

This region is composed of twenty 1.5 mm by 1.5 mm cells, where each cell contains repeated rectangular structures. These structures are designed to study edge-accelerating effect, and feature size from 0.25 μm to 5 μm. The descriptions of the structures are listed in Table A.3.

Table A.3: Geometry parameters of the edge-acceleration structures.

Name	L (μm)	W (μm)	S (μm)	Active Density (%)
EA01	2	1	1	33.33
EA02	5	5	5	25.00
EA03	1	1	0.75	44.44
EA04	2	0.75	0.5	59.26
EA05	5	2	2	35.71
EA06	5	1	1	41.67
EA07	2	2	1	44.44
EA08	2	1	0.75	53.33
EA09	1	1	0.5	64.00
EA10	5	1	0.75	60.61
EA11	5	5	2	51.02
EA12	5	2	1	55.56
EA13	2	1	0.5	71.11
EA14	2	2	0.75	64.00
EA15	3	2	0.75	68.57
EA16	5	2	0.75	72.73
EA17	2	2	0.5	79.01
EA18	5	5	1	69.44
EA19	4	4	0.75	79.01
EA20	5	5	0.75	82.64

A.4. Dishing and erosion structures

The dishing and erosion region is divided into 5 sub-regions with different densities. Each sub-region has a 1 mm buffer region around the border.

Three sub-regions are arranged as 8 mm by 7 mm, and their densities are 20%, 30% and 70%, respectively. The inside 6 mm by 5 mm area is composed of thirty 1 mm by 1 mm cells. These 30 cells are grouped as 6 classes: fixed active area (A), fixed ratio of L:W:S (R), fixed S (S), fixed L (F), fixed W (W) and L-shape (L). The structure arrangement is illustrated in Figure A.5. The structure geometry parameters are listed in Table A.4, Table A.5 and Table A.6.

B	B	B	B	B	B	B
B	R1	R2	R3	R4	R5	B
B	L1	L2	L3	L4	L5	B
B	S1	S2	S3	S4	S5	B
B	W1	W2	W3	W4	W5	B
B	F1	F2	F3	F4	F5	B
B	A1	A2	A3	A4	A5	B
B	B	B	B	B	B	B

Figure A.5: Arrangement of dishing and erosion structures in 8 mm by 7 mm sub-region.

Table A.4: Geometry parameters of the structures in 20% density dishing and erosion sub-region.

	Name	S (μm)	L (μm)	W (μm)	Active Density (%)
Buffer	D2B	25	20	21	20
Fix S	D2S1	25	20	21	20
	D2S2	25	25	17	20
	D2S3	25	37	13	20
	D2S4	25	61	10	20
	D2S5	25	85	9	20
Fix Ratio	D2R1	9.7	8.3	8.3	21
	D2R2	14.7	12.3	12.3	21
	D2R3	19.7	16.3	16.3	21
	D2R4	39.7	32.3	32.3	20
	D2R5	59.7	48.3	48.3	20
Fix Area	D2A1	22	18	18	20
	D2A2	20.8	29.2	11.2	20
	D2A3	17.8	46.2	7.2	21
	D2A4	14.8	65.2	5.2	21
	D2A5	12.1	87.9	3.9	21
Fix L	D2F1	13	40	5	21
	D2F2	21.7	40	10	20
	D2F3	34	40	20	20
	D2F4	42	40	30	20
	D2F5	49	40	40	20
Fix W	D2W1	12.5	11	10	21
	D2W2	16.7	20	10	20
	D2W3	21.7	40	10	20
	D2W4	24.7	60	10	20
	D2W5	27	80	10	20
L-Shape	D2L1	9.7	30.3	6.3	21
	D2L2	14.7	45.3	9.3	21
	D2L3	19.7	60.3	12.3	21
	D2L4	29.7	90.3	18.3	21
	D2L5	39.7	120.3	24.3	21

Table A.5: Geometry parameters of the structures in 30% density dishing and erosion sub-region.

	NAME	S (μm)	L (μm)	W (μm)	Active Density (%)
Buffer	D3B	16	20	20	31
Fix S	D3S1	16	20	20	31
	D3S2	16	29	15	31
	D3S3	16	44	12	31
	D3S4	16	64	10	31
	D3S5	16	84	9.5	31
Fix Ratio	D3R1	8.7	11.3	11.3	32
	D3R2	13.2	16.8	16.8	31
	D3R3	17.7	22.3	22.3	31
	D3R4	26.7	33.3	33.3	31
	D3R5	35.7	44.3	44.3	31
Fix Area	D3A1	17.8	22.2	22.2	31
	D3A2	17	33	15	31
	D3A3	15	49	10	31
	D3A4	12.6	67.4	7.4	31
	D3A5	10.4	89.6	5.6	31
Fix L	D3F1	8.5	50	5	32
	D3F2	15	50	10	31
	D3F3	24	50	20	31
	D3F4	36	50	40	31
	D3F5	40	50	50	31
Fix W	D3W1	8.2	11	10	31
	D3W2	11	21	10	31
	D3W3	14	40	10	31
	D3W4	15.6	60	10	31
	D3W5	17	80	10	31
L-Shape	D3L1	5.7	20.3	6.3	32
	D3L2	8.7	30.3	9.3	31
	D3L3	11.7	40.3	12.3	31
	D3L4	17.7	60.3	18.3	31
	D3L5	23.7	80.3	24.3	31

Table A.6: Geometry parameters of the structures in 70% density dishing and erosion sub-region.

	NAME	S (μm)	L (μm)	W (μm)	Active Density (%)
Buffer	D7B	14.7	80.3	75.3	71
Fix S	D7S1	6	34	30	71
	D7S2	6	44	25	71
	D7S3	6	54	23	71
	D7S4	6	74	20	71
	D7S5	6	104	18	71
Fix Ratio	D7R1	5.7	32.3	30.3	72
	D7R2	8.7	48.3	45.3	71
	D7R3	11.7	64.3	60.3	71
	D7R4	14.7	80.3	75.3	71
	D7R5	23.7	128.3	120.3	71
Fix Area	D7A1	6.3	33.7	33.7	71
	D7A2	6.1	43.9	25.9	71
	D7A3	5.5	58.5	19.5	71
	D7A4	4.8	75.2	15.2	71
	D7A5	4	96	12	72
Fix L	D7F1	5.7	60	20	71
	D7F2	7.6	60	30	71
	D7F3	9.2	60	40	70
	D7F4	10.5	60	50	70
	D7F5	11.5	60	60	70
Fix W	D7W1	3.6	20	20	72
	D7W2	4.4	30	20	71
	D7W3	5	40	20	71
	D7W4	5.7	60	20	71
	D7W5	6.2	80	20	71
L-Shape	D7L1	5.7	42.3	30.3	71
	D7L2	8.7	63.3	45.3	71
	D7L3	11.7	84.3	60.3	71
	D7L4	17.7	126.3	90.3	71
	D7L5	23.7	168.3	120.3	71

The other two are arranged as 7 mm by 9 mm, and their densities are 50% and 60%, respectively. The inside 5 mm by 7 mm area is composed of thirty-five 1 mm by 1 mm cells. These 35 cells are grouped as 7 classes: fixed active area (A), fixed ratio of L:W:S (R), fixed S (S), fixed L (F), fixed W (W), L-shape (L) and X-shape (X). Their arrangement is illustrated in Figure A.6. The structure geometry parameters are listed in Table A.7 and Table A.8.

B	B	B	B	B	B	B	B	B
B	R1	L1	S1	W1	F1	X1	A1	B
B	R2	L2	S2	W2	F2	X2	A2	B
B	R3	L3	S3	W3	F3	X3	A3	B
B	R4	L4	S4	W4	F4	X4	A4	B
B	R5	L5	S5	W5	F5	X5	A5	B
B	B	B	B	B	B	B	B	B

Figure A.6: Arrangement of dishing and erosion structures in 7 mm by 9 mm sub-region.

Table A.7: Geometry parameters of the structures in 50% density dishing and erosion sub-region.

	NAME	S (μm)	L (μm)	W (μm)	Active Density (%)
Buffer	D5B	20	50	50	51
Fix S	D5S1	10	30	21	51
	D5S2	10	25	25	51
	D5S3	10	40	18	51
	D5S4	10	50	16	51
	D5S5	10	90	13	51
Fix Ratio	D5R1	5.7	18.3	12.3	52
	D5R2	9.7	30.3	20.3	51
	D5R3	14.7	45.3	30.3	51
	D5R4	19.7	60.3	40.3	51
	D5R5	29.7	90.3	60.3	50
Fix Area	D5A1	11.5	28.5	28.5	51
	D5A2	11	39	21	51
	D5A3	9.9	54.1	15.1	51
	D5A4	8.5	71.5	11.5	51
	D5A5	7	93	9	52
Fix L	D5F1	20	50	50	51
	D5F2	18	50	40	51
	D5F3	11.8	50	20	51
	D5F4	8.1	50	12	51
	D5F5	5.8	50	8	52
Fix W	D5W1	7.7	80	10	52
	D5W2	7.3	60	10	52
	D5W3	6.6	40	10	52
	D5W4	5.2	20	10	52
	D5W5	3.7	10	10	53
L-Shape	D5L1	5.7	18.3	12.3	52
	D5L2	7.7	24.3	16.3	51
	D5L3	11.7	36.3	24.3	51
	D5L4	15.7	48.3	32.3	51
	D5L5	19.7	60.3	40.3	51
X-Shape	D5X1	5.7	20.3	12.3	51
	D5X2	8.7	30.3	18.3	51
	D5X3	11.7	40.3	24.3	51
	D5X4	17.7	60.3	36.3	50
	D5X5	23.7	80.3	48.3	50

Table A.8: Geometry parameters of the structures in 60% density dishing and erosion sub-region.

	NAME	S (μm)	L (μm)	W (μm)	Active Density (%)
Buffer	D6B	19.7	80.3	60.3	61
Fix S	D6S1	7	25	25	61
	D6S2	7	28	23	61
	D6S3	7	33	20	61
	D6S4	7	44	17	61
	D6S5	7	60	15	61
Fix Ratio	D6R1	7.7	32.3	24.3	61
	D6R2	11.7	48.3	36.3	61
	D6R3	15.7	64.3	48.3	61
	D6R4	19.7	80.3	60.3	61
	D6R5	23.7	96.3	72.3	60
Fix Area	D6A1	8.8	31.2	31.2	61
	D6A2	8.5	41.5	23.5	61
	D6A3	7.7	56.3	17.3	61
	D6A4	6.6	73.4	13.4	61
	D6A5	5.5	94.5	10.5	62
Fix L	D6F1	4.8	55	10	62
	D6F2	6.8	55	15	61
	D6F3	8.5	55	20	61
	D6F4	13.3	55	40	60
	D6F5	15.8	55	55	60
Fix W	D6W1	2.5	10	10	64
	D6W2	3.6	20	10	62
	D6W3	4.5	40	10	62
	D6W4	4.9	60	10	62
	D6W5	5.2	80	10	62
L-Shape	D6L1	5.7	24.3	18.3	62
	D6L2	8.7	36.3	27.3	61
	D6L3	11.7	48.3	36.3	61
	D6L4	17.7	72.3	54.3	61
	D6L5	23.7	96.3	72.3	60
X-Shape	D6X1	5.7	24.3	18.3	62
	D6X2	8.7	36.3	27.3	61
	D6X3	11.7	48.3	36.3	61
	D6X4	17.7	72.3	54.3	61
	D6X5	23.7	96.3	72.3	60

A.5. Geometry parameters of pattern structures

Geometry parameters L, W and S of rectangular structures are defined in Figure A.7. L is defined not smaller than W. The active density is defined as

$$\rho = \frac{LW}{(L+S)(W+S)} \tag{A.1}$$

The pitch size is defined as

$$pitch = W + S \tag{A.2}$$

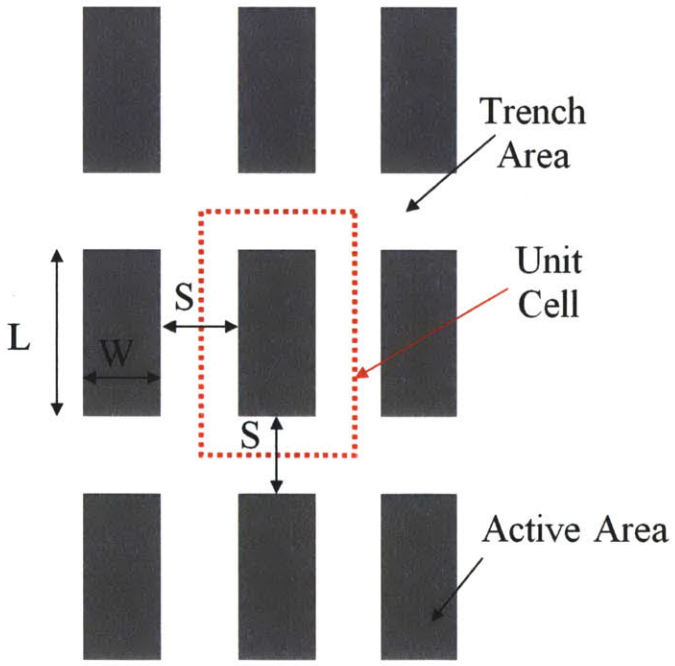


Figure A.7: Geometry parameters of rectangular structures.

Geometry parameters L, W and S of X-shape structures are defined in Figure A.8. The active density is defined as

$$\rho = \frac{2WL - W^2}{(L+S)^2} \tag{A.3}$$

The pitch size is defined as

$$pitch = L + S \tag{A.4}$$

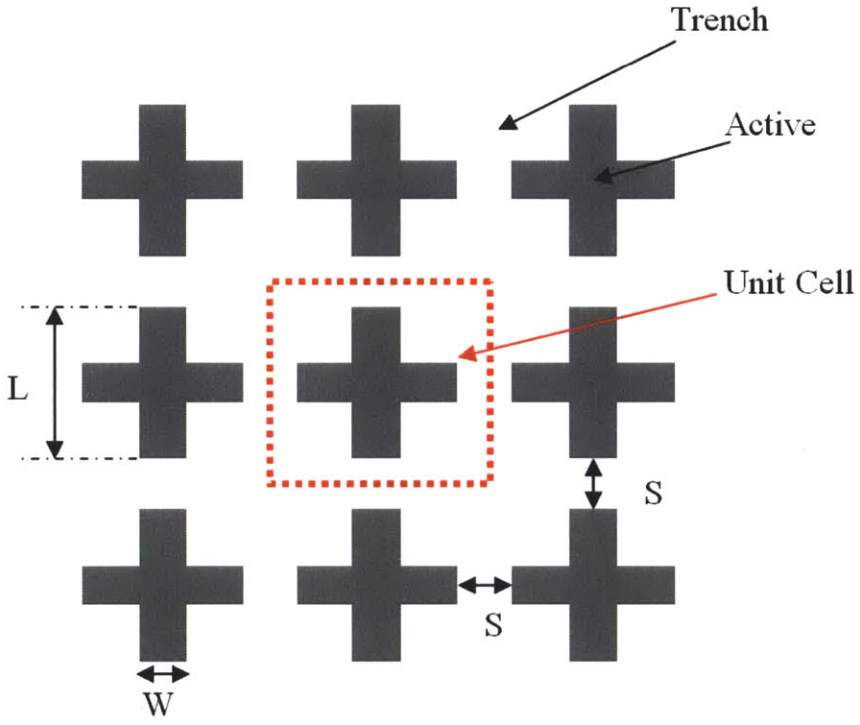


Figure A.8: Geometry parameters of X-shape structures.

Geometry parameters L, W and S of L-shape structures are defined in Figure A.9. The definitions of active density and pitch size are the same as Eq. A.3 and Eq. A.4.

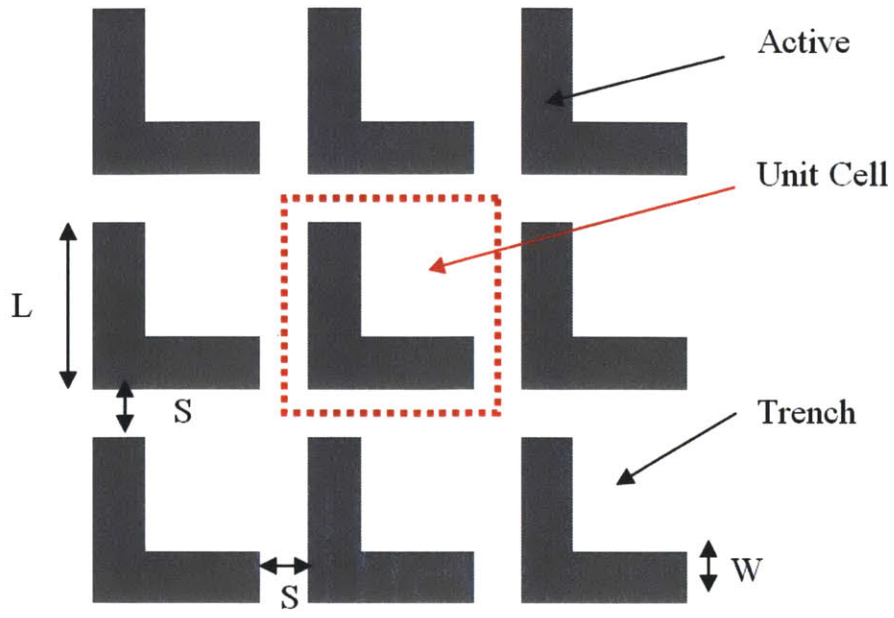


Figure A.9: Geometry parameters of L-shape structures.

References

1. J. M. Steigerwald, S. P. Murarka and R. J. Gutmann, *Chemical Mechanical Planarization of Microelectronic Materials*, New York: John Wiley & Sons, 1997.

2. D. Evans, "The Future of CMP," *MRS Bulletin*, vol. 27, pp. 779-783, 2002.

3. E. Mendel, "Polishing of Silicon," *Solid State Technology*, vol. 10, pp. 27-39, 1967.

4. B. Davari, C.W. Koburger, R. Schulz, J.D. Warnock, T. Furukawa, M. Jost, Y. Taur, W.G. Schwittek, J.K. DeBrosse, M.L. Kerbaugh and J.L. Mauer, "A New Planarization Technique, Using A Combination of RIE and Chemical Mechanical Polish (CMP)," *IEDM '89. Technical Digest*, pp.61-64, 1989.

5. D. Moy, M. Schadt, C. K. Hu, F. Kaufman, A. K. Ray, N. Mazzeo, E. Baran and D. J. Pearson, "A Two-level Metal Fully Planarized Interconnect Structure Implemented on a 64 kb CMOS SRAM," *Proc. 6th International IEEE VLSI Multilevel Interconnection Conference*, pp. 26-32, 1989.

6. International Technology Roadmap for Semiconductors, 2009.

7. International Technology Roadmap for Semiconductors, 2011.

8. B. Lee, "Modeling of Chemical Mechanical Polishing for Shallow Trench Isolation," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, 2002.

9. D. Edelstein, J. Heidenreich, R. Goldblatt, W. Cote, C. Uzoh, N. Lustig, P. Roper, T. McDevitt, W. Motsiff, A. Simon, J. Dukovic, R. Wachnik, H. Rathore, R. Schulz, L. Su, S. Luce and J. Slattery, "Full Copper Wiring in a Sub-0.25 μm CMOS ULSI Technology," *IEDM '97. Technical Digest*, pp. 773-776, 1997.

10. P. Wrschka, J. Hernandez, G. S. Oehrlein and J. King, "Chemical Mechanical Planarization of Copper Damascene Structures," *Journal of the Electrochemical Society*, vol. 147, pp. 706-712, 2000.

11. J. Lai, Nannaji Saka and Jung-Hoon Chun, "Evolution of Copper-Oxide Damascene Structures in Chemical Mechanical Polishing," *Journal of the Electrochemical Society*, vol. 149, pp. G31-G40, 2002.

12. K. T. Turner, "Wafer Bonding: Mechanics-Based Models and Experiments," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, 2004.

13. D. L. Hetherington and J. J. Sniegowski, Improved Polysilicon Surface-"Micromachined Micromirror Devices Using Chemical-Mechanical Polishing," *Proc. SPIE*, vol. 3440, pp. 148–153, 1998.

14. R. Nasby, J. Sniegowski, J. Smith, S. Montague, C. Barron, W. Eaton, P. McWhorter, D. Hetherington, C. Apblett and J. Fleming, "Application of Chemical-mechanical Polishing to Planarization of Surface-micromachined Devices," *Proc. Solid State Sensor and Actuator Workshop*, pp. 48–53, 1996.

15. C. Kourouklis, T. Kohlmeier, and H.H. Gatzen, "The Application of Chemical-mechanical Polishing for Planarizing a Su-8/permalloy Combination Used in MEMS Devices," *Sensors and Actuators A Physical*, vol. 106, pp. 268-271, 2003.

16. J. Fleming and C. Barron, "Novel Silicon Fabrication Process for High-Aspectratio Micromachined Parts," *Proc. SPIE*, vol. 2639, pp. 185-190, 1995.

17. S. Lin, J.G. Fleming and E. Chow, "Two- and Three-dimensional Photonic Crystals in III-V Semiconductors," *MRS Bulletin*, vol. 26, pp. 627–631, 2001.

18. X. Xie, "Physical Understanding and Modeling of Chemical Mechanical Planarization in Dielectric Materials," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, 2007.

19. R. Jairath, S. Chadda, E. Engdahl, W. Krussel, T. Mallon, K. Mishram, A. Pant and B. Withers, "Performance of Ontrak System's Linear Planarization Technology (LPT) for Dielectric CMP Processes," *Proc. 1997 International Conference on Chemical-Mechanical Polish (CMP) Planarization for ULSI Multilevel Interconnection (CMP-MIC)*, pp.194–201, 1997.

20. F. Sugimoto, Y. Arimoto and T. Ito, "Simultaneous Temperature Measurement of Wafers in Chemical Mechanical Polishing of Silicon Dioxide Layer," *Japanese Journal of Applied Physics*, vol. 34, pp. 6314-6320, 1995.

21. N. Kim, Y. Seo and W. Lee, "Temperature Effects of Pad Conditioning Process on Oxide CMP: Polishing Pad, Slurry Characteristics, and Surface Reactions," *Microelectronic Engineering*, vol. 83, pp. 362-370, 2006.

22. Y. Li, *Microelectronic Applications of Chemical Mechanical Planarization*, Hoboken: John Wiley & Sons, 2008.

23. F. Preston, "The Theory and Design of Plate Glass Polishing Machines," *Journal of the Society of Glass Technology*, vol. 11, pp. 214-256, 1927.

24. L. Hwee, S. Balakumar, S. Mahadevan, Zhou Mei Sheng, Alex See, M. Rahman and A. Senthilkumar, "Dishing and Nitride Erosion of STI-CMP for Different Integration Schemes," *Journal of Electronic Materials*, vol. 30, pp. 1478-1482, 2001.

25. P. Singer, "Copper CMP: Taking Aim at Dishing," *Semiconductor International*, Vol. 27, pp. 38-41, 2004.

26. H. Cai, "Modeling of Pattern Dependencies in the Fabrication of Multilevel Copper Metallization," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, 2007.

27. J. Luo and D. A. Dornfeld, *Integrated Modeling of Chemical Mechanical Planarization for Sub-Micron IC Fabrication*, Berlin: Springer, 2004.

28. J. Mello, "CMP Redefined," *SEMICON Taiwan 2011*, Taipei, Sep. 2011.

29. T. Eusner, "Multi-scale Scratching in Chemical-mechanical Polishing," Ph.D. thesis, Massachusetts Institute of Technology, 2010.

30. T. Eusner, "Nano-scale Scratching in Chemical-mechanical Polishing," Master's thesis, Massachusetts Institute of Technology, 2008.

31. T. Eusner, N. Saka and J. H. Chun, "Defect Reduction in Cu Chemical-mechanical Polishing," *Proc. 2010 International Symposium on Semiconductor Manufacturing (ISSM)*, pp.1-4, 2010.

32. J. Luo and D. A. Dornfeld, "Material Removal regions in Chemical Mechanical Planarization for Submicron Integrated Circuit Fabrication: Coupling Effects of Slurry Chemicals, Abrasive Size Distribution, and Wafer-pad Contact Area," *IEEE Transactions on Semiconductor Manufacturing*, vol. 16, pp. 45-56, 2003.

33. G. Fu, A. Chandra, S. Guha and G. Subhash, "A Plasticity-based Model of Material Removal in Chemical-mechanical Polishing (CMP)," *IEEE Transactions on Semiconductor Manufacturing*, vol. 14, pp. 406-417, 2001.

34. A. Maury, D. Ouma, D. Boning and J. Chung, "A Modification to Preston's Equation and Impact on Pattern Density Effect Modeling," *Advanced Metallization Conference (AMC)*, San Diego, CA, Oct. 1997.

35. P. Wrschka, J. Hernandez, Y. Hsu, T. S. Kuan, G. S. Oehrlein, H. J. Sun, D. A. Hansen, J. King and M. A. Fury, "Polishing Parameter Dependencies and Surface Oxidation of Chemical Mechanical Polishing of Al Thin Films," *Journal of the Electrochemical Society*, vol. 146, pp. 2689-2696, 1999.

36. C. W. Liu, B. T. Dai, W. T. Tseng and C. F. Yeh, "Modeling of the Wear Mechanism during Chemical-mechanical Polishing," *Journal of the Electrochemical Society*, vol. 143, pp. 716-721, 1996.

37. J. Xin, W. Cai and J. A. Tichy, "A Fundamental Model Proposed for Material Removal in Chemical–Mechanical Polishing," *Wear*, vol. 268, pp. 837–844, 2010.

38. B. Stine, D. Ouma, R. Divecha, D. Boning, J. Chung, D. L. Hetherington, I. Ali, G. Shinn, J. Clark, O. S. Nakagawa and S. Y. Oh, "A Closed-Form Analytic Model for ILD Thickness Variation in CMP Processes," *Proc. 1997 International Conference*

on Chemical-Mechanical Polish (CMP) Planarization for ULSI Multilevel Interconnection (CMP-MIC), pp. 266-273, 1997.

39. B. Stine, D. Ouma, R. Divecha, D. Boning, J. Chung, D. Hetherington, R. Harwood, S. Nakagawa and S. Y. Ho, "Rapid Characterization and Modeling of Pattern Dependent Variation in Chemical Mechanical Polishing," *IEEE Transactions on Semiconductor Manufacturing,* vol. 11, pp. 129-140, 1998.

40. 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 Transaction on Semiconductor Manufacturing*, vol. 15, pp. 232-244, 2002.

41. K. Noh, "Modeling of Dielectric Erosion and Copper Dishing in Copper Chemical-mechanical Polishing," Ph.D. thesis, Massachusetts Institute of Technology, 2005.

42. J. Lai, "Mechanics, Mechanisms, and Modeling of the Chemical Mechanical Polishing Process," Ph.D. thesis, Massachusetts Institute of Technology, 2001.

43. S. Bott, R. Rzehak, B. Vasilev, P. Kucher, and J. W. Bartha, "A CMP Model Including Global Distribution of Pressure," *IEEE Transactions on Semiconductor Manufacturing*, vol. 24, pp. 304-314, 2011.

44. D. Wang, J. Lee, K. Holland, T. Bibby, S. Beaudoin and T. Cale, "Von Mises Stress in Chemical-mechanical Polishing Processes," *Journal of the Electrochemical Society*, vol. 144, pp. 1121-1127, 1997.

45. G. Fu and A. Chandra, "A Model for Wafer Scale Variation of Removal Rate in Chemical Mechanical Polishing Based on Elastic Pad Deformation," *Journal of Electronic Materials*, vol. 30, pp. 400-408, 2001.

46. D. G. Thakurta, C. L. Borst, D. W. Schwendenman, R. J. Gutmann and W. N. Gill, "Three-dimensional Chemical Mechanical Planarization Slurry Flow Model Based on Lubrication Theory," *Journal of the Electrochemical Society*, vol. 148, pp. G207-G214, 2001.

47. D. G. Thakurta, C. L. Borst, D. W. Schwendenman, R. J. Gutmann and W. N. Gill, "Pad Porosity, Compressibility and Slurry Delivery Effects in Chemical-mechanical Planarization: Modeling and Experiments," *Thin Solid Films*, vol. 366, pp. 181-190, 2000.

48. D. White, J. Melvin and D. Boning, "Characterization and Modeling of Dynamic Thermal Behavior in CMP," *Journal of the Electrochemical Society*, vol. 150, pp. G271-G278, 2003.

49. K. D. Kopanski, "Analysis of Slurry Flow in Chemical-mechanical Polishing," Master's thesis, Massachusetts Institute of Technology, 2005.

50. A. Nishimoto, T. Smith, D. Ouma, E. Stuckey and D. Boning, "An In-situ Sensor for Reduced Consumable Usage through Control of CMP," *SRC TechCon'98*, Las Vegas, NV, Sept. 1998.

51. S. J. Shiu, C. C. Yu and S. H. Shen, "Multivariable Control of Multizone Chemical Mechanical Polishing," *Journal of Vacuum Science & Technology B*, vol. 22, pp. 1679-1687, 2004.

52. C. Mau, "Control of Wafer-scale Non-uniformity in Chemical-mechanical Planarization by Face-up Polishing," Master's thesis, Massachusetts Institute of Technology, 2008.

53. J. A. Greenwood and J. B. P. Williamson, "Contact of Nominally Flat Surfaces," *Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences*, vol. 295, pp. 300-319, 1966.

54. E. Tseng, C. Yi, and H. C. Chen, "A Mechanical Model for DRAM Dielectric CMP Process," *Proc. 1997 International Conference on Chemical-Mechanical Polish (CMP) Planarization for ULSI Multilevel Interconnection (CMP-MIC)*, p. 258-264, 1997.

55. T. H. Smith, S. J. Fang, D. S. Boning, G. B. Shinn, and J. A. Stefani, "A CMP Model Combining Density and Time Dependencies," *Proc. International Conference on Chemical-Mechanical Polish (CMP) Planarization for ULSI Multilevel Interconnection (CMP-MIC)*, pp. 97-104, 1999.

56. J. Grillaert, M. Meuris, E. Vrancken, N. Heylen, K. Devriendt, and M. Heyns, "Modeling Step Height Reduction and Local Removal Rates Based on Pad-substrate Interactions," *Proc. 1998 International Conference on Chemical-Mechanical Polish (CMP) Planarization for ULSI Multilevel Interconnection (CMP-MIC)*, p. 79-86, 1998.

57. J. P. Urbach and R. Rzehak, "A Novel Approach to Analyze and Model Feature Size Effects in CMP," *IEEE Transactions on Semiconductor Manufacturing*, vol. 22, pp. 566–571, 2009.

58. K. L. Johnson, *Contact Mechanics*, Cambridge: Cambridge University Press, 1985.

59. W. Chen, Q. Wang, Y. Liu, W. Chen, J. Cao, C. Xia, R. Talwar and R. Lederich, "Analysis and Convenient Formulas for Elasto-Plastic Contacts of Nominally Flat Surfaces: Average Gap, Contact Area Ratio, and Plastically Deformed Volume," *Tribology Letters*, vol. 28, pp. 27-38, 2007.

60. E. K. Dimitriadis, F. Horkay, J. Maresca, B. Kachar and R. Chadwick, "Determination of Elastic Moduli of Thin Layers of Soft Material Using the Atomic Force Microscope," *Biophysical Journal*, vol. 82, pp. 2798-2810, 2002.

61. T. C. O'Sullivan and R. B. King, "Sliding Contact Stress Field Due to a Spherical Indenter on a Layered Elastic Half-Space," *Journal of Tribology*, vol. 110, pp. 235-241, 1988.

62. 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, pp. 493-501, 1997.

63. 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," *Transactions of the ASME*, vol. 123, pp. 330-343, 2001.

64. A. E. H. Love, "The Stress Produced in a Semi-Infinite Solid by Pressure on Part of the Boundary," *Philosophical Transactions of the Royal Society A*, vol. 228, pp. 337-420, 1929.

65. H. Taylor, "Modeling and Controlling Topographical Nonuniformity in Thermoplastic Micro- and Nano-Embossing," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, 2009.

66. M. Webster and R. Sayles, "A Numerical Model for the Elastic Contact of Real Rough Surfaces," *Journal of Tribology*, vol. 108, pp. 314-321, 1986.

67. M. West and R. Sayles, "A 3-Dimensional Method of Studying 3-Body Contact Geometry and Stress on Real Rough Surfaces," *Proc. the 14th Leeds-Lyon Symposium on Tribology*, pp. 195-200, 1987.

68. H. A. van der Vorst, "Bi-CGSTAB: A Fast and Smoothly Converging Variant of Bi-CG for the Solution of Nonsymmetric Linear Systems," *Journal on Scientific Computing*, vol. 13, pp. 631-644, 1992.

69. D. Boning and X. Xie, "CMP at the Wafer Edge – Modeling the Interaction Between Wafer Edge Geometry and Polish Performance," *MRS Proceedings*, vol. 867, pp. W5.1.1-W5.1.11, 2005.

70. W. Fan, D. Boning, Y. Zhuang, Y. Sampurno, A. Philipossian, M. Moinpour and D. Hooper, "Characterization of CMP Pad Surface Properties and Aging Effects," *Proc. 2011 International Conference on Planarization/CMP Technology (ICPT)*, pp. 325-333, 2011.

71. C. Zhou, L. Shan, J. R. Hight, S. H. Ng and S. Danyluk, "Fluid Pressure and Its Effects on Chemical Mechanical Polishing," *Wear*, vol. 253, pp. 430-437, 2002.

72. L. J. Borucki, S. H. Ng and S. Danyluk, "Fluid Pressure and Pad Topography in Chemical Mechanical Polishing," *Journal of the Electrochemical Society*, vol. 152, pp. G391-G396, 2005.

73. C. F. Higgs, S. H. Ng, L. Borucki, I. Yoon and S. Danyluk, "A Mixed-Lubrication Approach to Prediction CMP Fluid Pressure Modeling and Experiments," *Journal of the Electrochemical Society*, vol. 152, pp. G193-G198, 2005.

74. S. Bott, R. Rzehak, B. Vasilev, P. Mucher and J. W. Bartha, "A CMP Model Including Global Distribution of Pressure," *IEEE Transactions on Semiconductor Manufacturing*, vol. 24, pp. 304-314, 2011.

75. L. Borucki, R. Zhuang, T. Sun, Y. Zhuang, A. Philipossian and D. Slutz, "Mechanical and Optical Analysis of Pad Surface Micro-texture Differences Caused by Conditioning," *presented at 2006 International Conference on Planarization Technology*, Foster City, CA, 2006.

76. W. Fan, D. Boning, L. Charns, H. Miyauchi, H. Tano and S. Tsujic, "Study on Stiffness and Conditioning Effects of CMP Pad Based on Physical Die-Level CMP Model," *Journal of the Electrochemical Society*, vol. 157, pp. H526-H533, 2010.

77. D. Ouma, "Modeling of Chemical Mechanical Polishing for Dielectric Planarization," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge 1999.

78. B. Stine, "A General Methodology for Assessing and Characterizing Variation in Semiconductor Manufacturing," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, 1998.

79. J. Johnson, "Modeling of Advanced Integrated Circuit Planarization Processes: Electrochemical-Mechanical Planarization (eCMP), STI CMP using Non-Conventional Slurries," Master's thesis, Massachusetts Institute of Technology, Cambridge, 2009.

80. B. Vasilev, R. Rzehak, S. Bott, P. Kucher and J. W. Bartha, "Greenwood-Williamson Model Combining Pattern-Density and Pattern-Size Effects in CMP," *IEEE Transactions On Semiconductor Manufacturing*, vol. 24, pp. 338-347, 2011.

81. M. R. Oliver, *Chemical-Mechanical Planarization of Semiconductor Materials*, Berlin: Springer, 2004.

82. C. T. Kelly, *Iterative Methods for Linear and Nonlinear Equations, Frontiers in Applied Mathematics, vol. 16*, Philadelphia: SIAM, 1995.

83. http://www.testwafer.com/PDF/skw7-2.pdf, last accessed in April 2012.

84. T. Sun, Y. Zhuang, L. Borucki and A. Philipossian, "Optical and Mechanical Characterization of Chemical Mechanical Planarization of Pad Surfaces," *Japanese Journal of Applied Physics*, vol. 49, p. 046501, 2010.

85. T. Sun, Y. Zhuang, L. Borucki and A. Philipossian, "Characterization of Pad-Wafer Contact in CMP Using Confocal Microscopy," *Japanese Journal of Applied Physics*, vol. 49, p. 066501, 2010.

86. T. Sun, L. Borucki, Y. Zhuang and A. Philipossian, "Investigating the Effect of Diamond Size and Conditioning Force on Chemical Mechanical Planarization Pad Topography," *Microelectronic Engineering*, vol. 87, pp. 553-559, 2010.

87. A. S. Lawing, "Pad Conditioning and Pad Surface Characterization in Oxide Chemical Mechanical Polishing," *MRS Proceedings*, vol. 732, p. I5.3.1, 2002.

88. J. A. Greenwood and J. B. P. Williamson, "Contact of Nominally Flat Surfaces," *Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences*, vol. 295, pp. 300-319, 1966.

89. D. Ouma, B. Stine, R. Divecha, D. Boning, J. Chung, G. Shinn, I. Ali and J. Clark, "Wafer-Scale Modeling of Pattern Effect in Oxide Chemical Mechanical Polishing," presented *at Manufacturing Yield, Reli-ability, and Failure Analysis Session, SPIE 1997 Symposium on Microelectronic Manufacturing*, Austin, TX, 1997.

90. D. Boning and W. Fan, "Characterization and Modeling of Pad Asperity Response in CMP," *MRS Proceedings*, vol. 1249, p. E05-04, 2010.

91. R. Rzehak, "Pitch-Dependence in Oxide CMP," *Proc. 2006 International Conference on Chemical-Mechanical Polish (CMP) Planarization for ULSI Multilevel Interconnection (CMP-MIC)*, p. 137-144, 2006.

92. J. P. Urbach and R. Rzehak, "A Novel Approach to Analyze and Model Feature Size Effects in CMP," *IEEE Transactions on Semiconductor Manufacturing*, vol. 22, pp. 566-571, 2009.

93. L. Charns, M. Sugiyama and A. Philipossian, "Mechanical Properties of Chemical Mechanical Polishing Pads Containing Water-soluble Particles," *Thin Solid Films*, vol. 485, pp. 188-193, 2005.

94. B. S. Kim, M. H. Tucker, J. D. Kelchner and S. P. Beaudoin, "Study on the Mechanical Properties of CMP Pads," *IEEE Transactions on Semiconductor Manufacturing*, vol. 21, pp. 454-463, 2008.

95. H. Lua, Y. Obeng and K.A. Richardson, "Applicability of Dynamic Mechanical Analysis for CMP Polyurethane Pad Studies," *Materials Characterization*, vol. 49, pp. 177-186, 2003.

96. S. D. Gouda, A. Bastawros and A. Chandra, "Multi-Scale Characterization of Pad Role on Material Removal Rate in CMP," *MRS Proceedings*, vol. 767, p. F2.3.1, 2003.

97. A. Bastawros and A. Chandra, "Role of Multi-Scale Polishing Pad Response on Evolution of Scratches during CMP," *Proc. 2007 International Conference on Chemical-Mechanical Polish (CMP) Planarization for ULSI Multilevel Interconnection (CMP-MIC)*, p. 248-257, 2007.

98. A. Lawing and C. Juras, "Pad Surface Analysis and Conditioning Effects: Implications on Process Design, Break-in Response and Next Generation Pad and Conditioning Platforms," *Proc. of 2007 International Conference on Planarization / CMP Technology (ICPT)*, pp. 1-5, 2007.

99. C. L. Elmufdi and G. P. Muldowney, "The Impact of Diamond Condition on Surface Contact in CMP pads," *MRS Proceedings*, vol. 991, p. 0991-C01-02, 2007.

100. J. Reilly, "Stylus Profiler Monitors Chemical Mechanical Planarization Performance," *Proc. 1994 Advanced Semiconductor Manufacturing Conference and Workshop*, p. 320-322, 1994.

101. A. C. Fischer-Cripps, *Nanoindentation, 2nd ed*, New York: Springer, 2004.

102. M. F. Doerner and W. D. Nix, "A Method for Interpreting the Data Depth-Sensing Indentation Instruments," *Journal of Materials Research*, vol. 1, pp. 601-609, 1986.

103. I. N. Sneddon, "The Relation between Load and Penetration in the Axisymmetric Boussinesq Problem for a Punch of Arbitrary Profile," *International Journal of Science Engineering*, vol. 3, pp. 47-57, 1965.

104. G. M. Pharr, W. C. Oliver and F. R. Brotzen, "On the Generality of the Relationship among Contact Stiffness, Contact area, and Elastic Modulus during Indentation," *Journal of Materials Research*, vol. 7, pp. 613-617, 1992.

105. W. C. Oliver and G. M. Pharr, "A New Improved Technique for Determining Hardness and Elastic Modulus Using Load and Sensing Indentation Experiments," *Journal of Materials Research*, vol. 7, pp. 1564-1582, 1992.

106. Hysitron Inc., *TI 900 TriboIndenter Information Sheet*, online document in http://www.hysitron.com/products/ti-series/ti-900-triboindenter.

107. Hysitron Inc., *Probe Selection Guide*, online document in http://www.hysitron.com/products/ti-series/ti-900-triboindenter.

108. B. J. Briscoe, L. Fiori and E. Pelillo, "Nano-Indentation of Polymeric Surfaces," *Journal of Physics D: Applied Physics*, vol. 31, pp. 2395-2405, 1998.

109. C. A. Tweedie, G. Constantinides, K. E. Lehman, D. J. Brill, G. S. Blackman and K. J. Van Vliet, "Enhanced Stiffness of Amorphous Polymer Surfaces under Confinement of Localized Contact Loads," *Advanced Materials*, vol. 19, pp. 2540-2546, 2007.

110. J. McGrath and C. Davis, "Polishing Pad Surface Characterization in Chemical Mechanical Planarization," *Journal of Material Processing Technology*, vol. 153-154, pp. 666-673, 2004.

111. S. Y. Kima, C. J. Park and Y. J. Seo, "Signal Analysis of the End Point Detection Method Based on Motor Current," *Microelectronic Engineering*, vol. 66, pp. 472-479, 2003.

112. P. L. Tso and S. Y. Ho, "Factors Influencing the Dressing Rate of Chemical Mechanical Polishing Pad Conditioning," *The International Journal of Advanced Manufacturing Technology*, vol. 33, pp. 720-724, 2007.

113. D. Gauthier, A. Mueller, R. D. White, V. Manno, C. Rogers, D. Hooper, S. Anjur, M. Moinpour, "Micromachined Lateral Force Sensors for Characterization of Microscale Surface Forces During Chemical Mechanical Polishing," *MRS Proceedings*, vol. 1085, p. T05-11, 2008.

114. Hysitron Inc., Hysitron Nanomechanical Testing Probe Selection Guide, online document in http://www.hysitron.com/products/options/indenter-tips/.

115. L. Borucki and Y. Sampurno, "Method for CMP using Pad in a Bottle," patent pending, WO2011/142764, Nov. 17, 2011.

116. A. Philipossian, "'Pad-in-a-Bottle': Planarization with Slurries Containing Suspended Polyurethane Beads," *presented at 2012 SRC-GRC/SEMATECH Engineering Research Center for Environmentally Benign Semiconductor Manufacturing Annual Review Meeting*, Tucson, AZ, 2012.

117. K. H. Park, H. J. Kim, O. M. Chang and H. D. Jeong, "Effects of Pad Properties on Material Removal in Chemical Mechanical Polishing," *Journal of Materials Processing Technology*, vol. 187–188, pp. 73–76, 2007.

118. J. Ugelstad, H. R. Mfutakamba, P. C. Mork, T. Ellingsen, A. Berge, R. Schmid, L. Holm, A. Jorgedal, F. K. Hansen and K. Nustad, "Preparation and Application of Monodisperse Polymer Particles," *Journal of Polymer Science: Polymer Symposium*, vol. 72, pp. 225-240, 1985.

119. Y. Lu, Y. Tani and K. Kawata, "Proposal of New Polishing Technology without Using a Polishing Pad," *CIRP Annals - Manufacturing Technology*, vol. 51, pp. 255–258, 2002.

120. X. F. Xu, B. X. Ma, F. Chen and W. Peng, "Study on Effect of Composite Particles in Polishing Process and Its Mechanism," *Advanced Materials Research*, vol. 24-25, pp. 155-159, 2007.

121. B. Zhao and F. G. Shi, "Threshold Pressure and Its Influence in Chemical Mechanical Polishing for IC Fabrication," *IEDM '98. Technical Digest*, pp. 341-344, 1998.

122. Y. Zhao and L. Chang, "A Micro-Contact and Wear Model for Chemical-Mechanical Polishing of Silicon Wafers," *Wear*, vol. 252, pp. 220-226, 2002.

123. X. Xia and G. Ahmadi, "Surface Removal Rate in Chemical-Mechanical Polishing, Particulate Science and Technology," vol. 20, pp. 187-196, 2002.

124. F. Liu and M. P. F. Sutcliffe, "Modeling of Delamination of Ultra Low-k Material during Chemical Mechanical Polishing," *Tribology Letters*, vol. 25, pp. 225-236, 2007.

125. B. Zhao and F. G. Shi, "Chemical Mechanical Polishing: Threshold Pressure and Mechanism," *Electrochemical and Solid-State Letters*, vol. 2, pp. 145-147, 1999.