Modeling of Planarization Technologies

by

Daniel Truque

Bachelor of Science, Louisiana State University, May 2005

Submitted to the Department of Electrical Engineering and Computer Science

in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Electrical Engineering and Computer Science

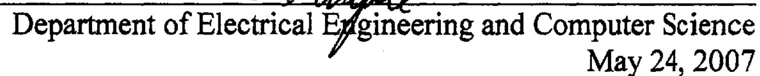
at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 24, 2007

[June 2007]

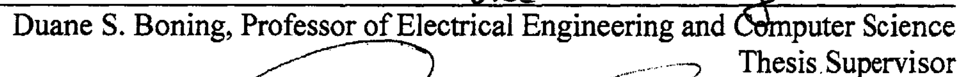
© Copyright Massachusetts Institute of Technology. All rights reserved.

Author



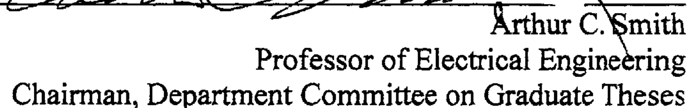
Department of Electrical Engineering and Computer Science
May 24, 2007

Certified by

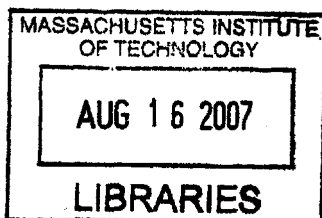


Duane S. Boning, Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by



Arthur C. Smith
Professor of Electrical Engineering
Chairman, Department Committee on Graduate Theses



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

AUG 16 2007

LIBRARIES

ARCHIVES

Modeling of Planarization Technologies

By
Daniel Truque

Submitted to the Department of Electrical Engineering and Computer Science on May 24, 2007, in partial fulfillment of the requirements for the degree of Master of Science

Abstract

The need for better planarity becomes more critical in semiconductor manufacturing as dimensions and tolerance margins keep shrinking. The purpose of this thesis is to understand and model new technologies for the planarization of integrated circuits. Two technologies and models are explored: characterization and modeling of a novel polishing pad, and a dynamic wafer level physical model for electrochemical-mechanical polishing (ECMP). The novel pad contains water soluble particles that dissolve in the slurry when they reach the surface. Different pads, with varying particle concentrations and sizes, are tested by polishing test wafers to extract the necessary model parameters to model each pad's polishing performance. The effect on chip uniformity and step height planarization are studied. The simulations and analysis enable understanding and comparison of pad design decisions, to achieve the best tradeoff in the performance metrics considered. In the second planarization technology, wafer scale uniformity effects are studied in an emerging copper polishing alternative, ECMP. The proposed dynamic wafer level physical ECMP model is able to capture the physics behind the ECMP process, based on modeling of electrical current flowing through the copper thin film on the wafer surface and the ECMP electrolyte solution. The model is able to fit zonal relative removal rates with root mean square error less than 7%, compared to existing empirical models and experiments.

Thesis Supervisor: Duane S. Boning

Title: Professor of Electrical Engineering and Computer Science

In the pursuit of knowledge...

I would like to thank my advisor Duane for giving me the opportunity to be part of his group during these two years at MIT in which I have learned so much. I also want to thank Xiaolin for the guidance he provided me into the world of CMP and the endless quest of mastering Matlab. I also want to thank all our group members: Ajay, Daihyun, Ed, Hayden, Hong, Karthik, Nigel, and Zhipeng for insightful discussions, endless conversations and sharing the donuts of knowledge every week at our group meetings... Always get the filled ones!

I want to thank my Sidekick KGs for their friendship and support during these unforgettable years at MIT. It all started with some karaoke at the Thirsty Ear, followed by uncountable ways of procrastinating, but we learned our main lesson: the more you know, the less you know because you discover how much more is out there to be learned, invented and discovered. Thank you Koshita for a wonderful year together and the ones that might come...

I would like to thank my family for their support throughout my years in school, which amounts to my whole life up to now! I was very fortunate to grow up with two sisters who molded my character as we all matured together though life. I want to thank my aunt for filling the role of a mother during my studies abroad. I would like to offer a special thanks to my mom for always pushing me past my limits to do my best and overachieve, because I have arrived to where I am thanks to her. Also I would like to thank my grandmother for her prayers which have helped me out and kept me afloat during hard times. I would also like to thank my grandfather for being a role model for me, showing me how rewarding it is being righteous and helping others, and how confidence and hard work always pays off. Finally, I would like to express my admiration for my dad for his perseverance in all aspects of life, showing me by example how to learn and adapt to whatever life might throw at you. Thank you all,

Daniel Truque

May 2007

Table of Contents

Chapter 1 Introduction and Background ... 11
1.1 Background ... 11
1.2 Pad Design and Modeling ... 13
1.2.2 Previous pad engineering ... 14
1.2.2 Traditional polishing pads ... 15
1.2.2 Experimental pad ... 16
1.3 ECMP ... 17
1.3.1 Electrochemistry of ECMP ... 18
1.3.2 ECMP implementations ... 18
1.4 Summary ... 20

Chapter 2 Experimental Pad Characterization ... 23
2.1 Polishing performance metrics ... 23
2.2 Conventional oxide CMP PDSH model description ... 26
2.3 Experimental ... 28
2.3.1 Setup and methodology ... 28
2.3.2 Ramp-up time ... 30
2.3.3 Preliminary Results ... 31
2.3.4 Model parameter sensitivity analysis ... 35
2.3.5 Variable WSP concentration experiments ... 38
2.3.6 Variable WSP size experiments ... 55
2.4 Physical modeling ... 65
2.5 Summary ... 67

Chapter 3 ECMP Model and Simulations ... 69
3.1 Empirical modeling ... 69
3.2 Physical Model Description ... 71
3.3 Experimental Data Fitting ... 75
3.4 Physical Model Optimization and Application ... 77
3.5 Model Possible Enhancements ... 78

Chapter 4 Conclusion ... 79

References 81

Table of Figures

Figure 1: Standard orbital CMP setup with pad conditioner [3] (left) and an assortment of pad conditioners made from stainless steel with brazed diamonds [2] (right). 13

Figure 2: Layered pad design [20]. 15

Figure 3: Four main pad types according to L. Cook [23]; Types 1-4 from left to right. 16

Figure 4: JSR Micro's pad with water soluble particles. 16

Figure 5: Basic hybrid ECMP setup. 19

Figure 6: Oxide CMP pre and post polish oxide thickness variations. 23

Figure 7: STI CMP dishing and erosion. 24

Figure 8: Performance metrics (Above – Time to target step height; Below - Ending Range). 25

Figure 9: Raw pattern density (left) and effective pattern density (right) across test die.26

Figure 10: Contact height h_c. 26

Figure 11: Exponential step height - pattern density model. 27

Figure 12: Test die pattern. 29

Figure 13: Experimental polishing conditions. 30

Figure 14: Ramp-up time and multiple ramp-up times effect on measurements. 31

Figure 15: Initial experimental results for JSR WSP vs. IC1000 pad. 32

Figure 16: Pad asperities. 33

Figure 17: Model predictions for different pattern densities for the JSR pad (right) and the IC1000 pad (left). 34

Figure 18: Step height evolution between IC1000 pad (left) and JSR WSP pad (right). 35

Figure 19: Model parameter error sensitivity analysis. 36

Figure 20: *PL* influence on step height evolution for different pattern densities (profilometry and optical measurements shown where available). 37

Figure 21: τ influence on step height evolution for different pattern densities. (profilometry and optical measurements shown where available). 38

Figure 22: SEM images of pad conditioners: #125 grit (left) and #325 grit (right) [31]. 39

Figure 23: #100 grit parameters normalized to #325 grit parameters. 41

Figure 24: Model predicted blanket removal rate vs. given data on blanket removal rate. 42

Figure 25: Planarization length larger than the test density region. 43

Figure 26: New sites near the edge to improve model fitting. 44

Figure 27: Blanket removal rate across wafer locations for the second experimental set.45
Figure 28: Blanket removal rates for all grit and WSP concentrations tested................ 46
Figure 29: Planarization length across wafer locations for the second experimental set. 47
Figure 30: Planarization length for all grit and WSP concentrations tested................... 47
Figure 31: τ values across wafer locations for the second experimental set. 48
Figure 32: τ values for all grit and WSP concentrations tested. 49
Figure 33: Model error across wafer locations for the second experimental set............. 50
Figure 34: Model error for all grit and WSP concentrations tested................................ 50
Figure 35: Model step height and thickness evolution plots. The chip-scale plots at top correspond to t = 200 sec. 52
Figure 36: Time to reach step height target for #325 grit conditioned pads................... 53
Figure 37: Ending thickness ranges for #325 grit conditioned pads. 54
Figure 38: Blanket removal rate across wafer locations for the third experimental set. . 56
Figure 39: Blanket removal rate for different WSP sizes... 57
Figure 40: Suggested relationship between WSP size and blanket removal rate K. 58
Figure 41: Planarization length across the wafer for the third experimental set. 59
Figure 42: Planarization length for different WSP sizes. 59
Figure 43: Suggested relationship between WSP size and PL. 60
Figure 44: τ across wafer for the third experimental set.. 61
Figure 45: τ values for different WSP sizes. 61
Figure 46: Suggested relationship between WSP size and τ. 62
Figure 47: Fitting error across the wafer for the third experimental set.......................... 63
Figure 48: Fitting error for the third experimental set... 63
Figure 49: Time to reach step height target (200 Å) for different WSP sizes. 64
Figure 50: Ending thickness ranges for different WSP sizes. 65
Figure 51: Exponential distribution of conditioned pad asperities [30]. 67
Figure 52: Empirical model vs. Applied Materials model. 70
Figure 54: Normalized amount removed fit and RMS zone error table.......................... 75
Figure 55: Model output: voltage drop, current distribution, removal rates and final thickness. 76

Chapter 1 *Introduction and Background*

This first chapter will give a brief background on chemical mechanical polishing (CMP), polishing pads and electrochemical mechanical polishing (ECMP). It is composed of two sections. The first section will discuss pad design and modeling, going into detail about existing modeling techniques, polishing pads, and the experimental pad we examine in this research. The second section will focus on ECMP and explain its electrochemistry and how it is applied on new ECMP tools. It also explains the hybrid ECMP setup from which previous data was acquired and how this can be used to create a model for wafer-level removal rates and uniformity in ECMP.

1.1 Background

The need for better planarity becomes more critical in semiconductor manufacturing as dimensions and tolerance margins keep shrinking. Over the past decade, chemical mechanical polishing (CMP) has developed to become an established process widely used in the semiconductor industry. There are three main types of CMP performed during IC manufacturing. Oxide CMP is used to planarize silicon oxide in front-end processing and back-end processing, since oxide is widely used throughout the structure to provide electric isolation. Shallow Trench Isolation (STI) CMP is performed during the front end before the transistors are even laid down, to provide isolation between devices. STI CMP differs from oxide CMP used in interconnect formation in that two materials (silicon nitride and silicon oxide) are polished instead of only one. Copper CMP is performed during back-end processing, where copper is used for interconnects. This last is very similar to STI CMP in that multiple materials (metal, metal barrier and dielectric) are being polished simultaneously. The big difference lies in that copper CMP may be performed while also applying a voltage to the wafer during polishing, a process known as electrochemical mechanical polishing (ECMP).

There are still a number of challenges to overcome, especially as industry moves towards sub-45 nm technology nodes and new materials are introduced with dimensions reaching critical fundamental limits. The need for a deeper understanding of the existing and developing planarization technologies is essential to move forward. There are

currently various approaches to CMP, ranging from rotational tools by Applied Materials to orbital tools by Novellus, but all of them have one vital factor in common: the polishing pad. Rethinking conventional CMP from the point of view of engineering an ideal pad can benefit most of the existing CMP tool configurations. Modeling relatively new processes such as electrochemical mechanical polishing (ECMP) will help us to understand the different mechanisms by which material removal occurs in these novel processes.

Modeling is essential to understand the mechanisms that cause variation within a process. Variation is the deviation from intended or designed values for a structure or circuit parameter of concern [1]. There are different degrees of variation within a process, which can be grouped into three main categories: die level variation, wafer level variation and wafer to wafer variation. Die level refers to variations within a single die or chip. Layout pattern dependent variations in polishing rate, such as dishing and erosion, fall within this category. These are caused by pressure distribution variations over different regions on the chip, resulting from layout pattern density or feature size dependencies. Wafer level variation is observed on the wafer scale (e.g., across a 200 mm or 300 mm wafer diameter). Nanotopography [5], [10], [11], [12], [13], [14], [15], [16], being the vertical variation in the scale of nanometers across a few millimeters of lateral distance, is an example of wafer level variation. Wafer bow and thickness variations also fall within this category. These effects are manifested differently at the die level depending on the location of the die on the wafer. Wafer to wafer variation accounts for the differences between wafers. Examples include variations corresponding to thickness deviations of starting wafers, different wafer curvatures within a lot or batch of wafers, and film thickness deviations arising from polish rate drifts as the pad wears from polishing large numbers of wafers.

Pad wear is an important source of variation. To mitigate it, the pad must be conditioned, which means roughening it with diamond grit prevent glazing of the pad and *freshen* it, as pictured in Figure 1. This prevents particle agglomeration and surface smoothening that can cause polishing variations. The conditioner is usually a small disc made of stainless steel brazed with diamond grit. It usually sits at the opposite side of the pad from where the wafer is contacting the pad. In this fashion, the pad can be

conditioned while polishing. When the pad is conditioned in between wafer polishes, the process is said to use ex-situ conditioning, while if the pad is continuously polished as the wafer is polished then it is referred to as in-situ conditioning.

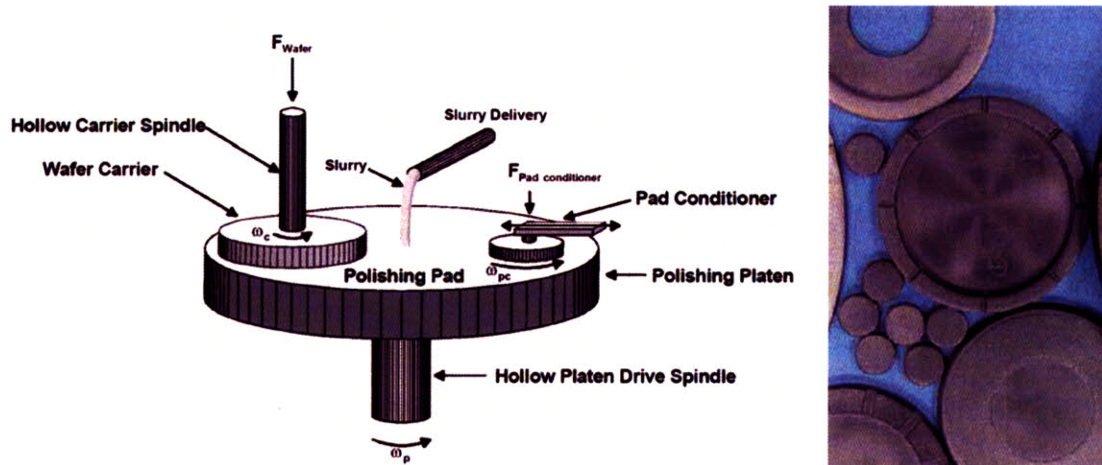


Figure 1: Standard orbital CMP setup with pad conditioner [3] (left) and an assortment of pad conditioners made from stainless steel with brazed diamonds [2] (right).

1.2 Pad Design and Modeling

The polishing pad plays a major role in the polishing, though its contribution is hard to understand or model. Our group has previously developed CMP models based on wafer pattern density and step height dependencies (PDSH model) [4], [5], [6], [7], [8], [9]. The parameters for this model were empirical and lacked physical meaning, though they are able to give an extensive characterization of the polishing performance of a given CMP process. Relating these parameters to the physical properties of the pad can enable the pad manufacturers to engineer an optimal pad based on modeling, exploit the pad properties to improve polishing performance. Our previous model has substantially influenced commercial simulation software (e.g., as offered by Cadence among others), and adding the physical connection between the pad parameters and planarization models will help to optimize the CMP process.

An ideal pad must be stiff enough to provide efficient local planarization. The benefit of a stiff pad is that most removal occurs on the raised areas and not in the trenches, reducing the step height faster than if the pad conforms to the trenches and removes material inside them. A stiffer pad is also less likely to be a victim of pattern density

effects when compared to a softer pad. Our model captures this pad property by means of the planarization length, which is the length over which a point in a wafer will affect removal in neighboring points of the wafer.

Lower pattern density areas tend to polish faster than higher pattern density areas because the low pattern density raised areas experience a higher local pressure than do higher density areas – this is because the applied pressure or force is borne by fewer features in the low density region, and thus the applied pressure is distributed across a smaller contact area resulting in faster removal. If the pad is infinitely stiff, the pad will contact all areas at the same height and any area that polishes faster will stop contacting the pad until the heights of features in that area are the same as the rest.

On the other hand, an ideal pad must also be soft enough to conform to the wafer level variations of the wafer, such as wafer bow and nanotopography. If the pad is too stiff, there will be areas of the wafer that will suffer thinning. This means that one area will polish more than another because it is "higher" than the other region, even though initially both were of equal thickness. In other words, assuming a wafer has a constant thickness everywhere, but suffers from nanotopography, the "raised" areas will be polished more than "lower areas."

Pad stiffness is one of the two key parameters to pad performance, the other parameter being the pad surface structure. This latter is determined by the asperity heights and pore size, and it is strongly related to step height reduction performance. The water soluble particles (WSPs) in JSR's pad offer a unique opportunity to tune these two properties to engineer a pad that achieves both excellent feature planarization (low down area removal) and good within-die uniformity through long range stiffness.

1.2.2 Previous pad engineering

There have been several pad designs that include double and even triple layer approaches that combine a stiff pad and a soft pad to provide maximum benefits [20], as illustrated in Figure 2. These composite pads typically consist of a hard pad on the wafer contact surface to prevent pattern dependencies and a soft subpad to allow the pad to conform to wafer level flatness non-uniformities.

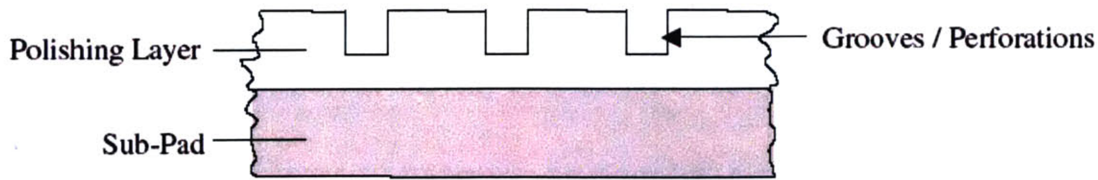


Figure 2: Layered pad design [20].

There have been various approaches by different parties to examine the dynamics of the slurry-pad-wafer interactions in an attempt to improve the process. Some research focused on understanding lubrication mechanisms [17] while others aim at the contact mechanics and hydrodynamics [18], [19].

Of particular interest is the research led by the Pad Engineering Research Group and the Advanced Research Group at Rohm and Hass. With the aid of heavy clustered multi-processing, researchers developed detailed simulations of the fluid dynamics of the slurry between the pad and wafer. By feeding the model different grooving patterns, optimal grooving patterns for different purposes were obtained. One particular paper explores different pad grooving patterns in an attempt to improve slurry distribution and improve material removal [20], [21] while another one investigates the heating performance instead [22]. In the past decade years CMP development has exploded into many different branches, ranging from conventional CMP to electro-chemical mechanical polishing (ECMP) and from orbital tools to rotational tools.

1.2.2 Traditional polishing pads

According to Cook [23], there are four basic pad types. Muldowney [20] summarized them as follows: Type 1 pads are polymer impregnated felts (Suba™, STT 711™), Type 2 are poromerics (synthetic leather, Politex™, Surfin™), Type 3 are filled polymer sheets (IC1000™, FX9™) and Type 4 are unfilled textured polymer sheets (OXP4000™, NCP-1™). These are depicted in the SEM images in Figure 3. It can be seen that Type 1 pads are composed of fibers, Type 2 are vertically oriented pores, Type 3 are closed pores or bubbles, and Type 4 has a very low microtexture (the big lines are from grooves made on the pad).

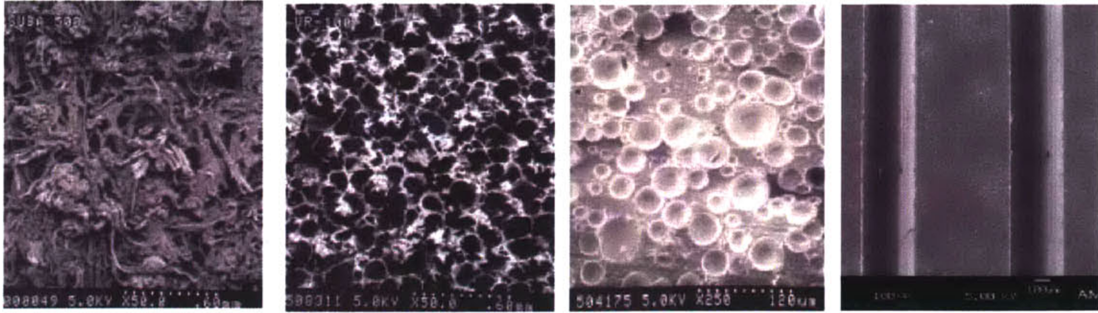


Figure 3: Four main pad types according to L. Cook [23]; Types 1-4 from left to right.

Each pad type has its advantages. Based on its microstructure, their physical characteristics vary. For example Types 1 and 2 are moderately and highly compressible with medium and low stiffness respectively, which make them excellent for tungsten CMP or post CMP buff. On the other hand, Types 3 and 4 have low compressibility but high stiffness, and are therefore good for inter layer dielectric (ILD) CMP and shallow trench isolation (STI) CMP. Type 3 is probably the most versatile of them all, allowing front-end and back-end oxide and metal polishing. The IC1000 is one of the most widely used pads, and for this reason it will initially be used to compare against the experimental pad examined here.

1.2.2 Experimental pad

The experiments considered in this work link the physical properties of the pad to the CMP pattern dependent model parameters. The experiments are performed in collaboration with JSR Micro, a major consumables supplier to the semiconductor industry with an important share of the polishing pad market. JSR has developed a pad that contains water soluble particles (WSPs) embedded and evenly distributed throughout the pad, as illustrated in Figure 4.

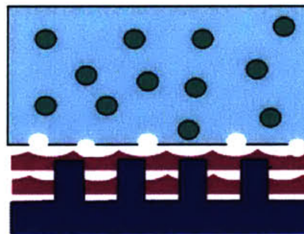


Figure 4: JSR Micro's pad with water soluble particles.

The pad will most likely fall under the Type 3 pad category. Its is made of a proprietary elastomer, in contrast to the usual polyurethane pads by other pad manufacturers. This has the advantage of reducing scratches due to the pad. While in the bulk, the particles remain solid and keep the pad bulk rigid to improve planarization length and efficiency. When the particles reach the surface of the pad and come in contact with the slurry, they are dissolved leaving behind voids. The created large pores improve the pad's ability to transport slurry and to take material out of the removal site. This exposure of new WSPs also helps to keep the pad surface *fresh* as the surface wears down. This last benefit is not sufficient to maintain removal rate alone and therefore in-situ conditioning is also used. Parameters like particle size and density can be tuned to change the physical pad properties. By using our PDSH model, the pad performance can be characterized and related to the tuned physical characteristics of the pad.

1.3 ECMP

Over the past few years, electrochemical-mechanical polishing (ECMP) has become an accepted and robust process to remove bulk copper. As dimensions shrink and low-K dielectrics are introduced, polishing pressures must be decreased to prevent damage to increasingly fragile dielectrics, resulting in lower removal rates. ECMP is an alternative offering high removal rates for pressures below 1 psi, extending pad life and reducing consumables cost by up to 30% [24]. The electrolyte solution that is used instead of the slurry is typically much less expensive than the slurry that would be needed [25] ECMP is believed to be a non-Prestonian process controlled by applied voltage, with an endpoint enabled by measuring the total charge delivered. Using multiple voltage zones enables control of removal rates across the wafer to account for variations in the incoming copper thickness profile. At present, conventional CMP models [4], [8], [5] have been pressed into service to characterize this process, but the nature of ECMP is different from conventional CMP. As the industry moves from hybrid ECMP approaches to full sequence ECMP tools, better models are needed at the wafer level, die level, and feature level.

In this work, we propose a wafer level dynamic ECMP model based on time-evolving current density distributions across the wafer. A single layer of copper on the wafer

surface is assumed, and a discretized element model for this layer, and for multiple electrolyte layers, is generated. The electrical potential and current density distributions are calculated based on the applied voltage zones and the metal film thicknesses across the wafer. These values are then used to calculate the removal rate based on current-controlled copper dissolution.

1.3.1 Electrochemistry of ECMP

In the basic ECMP approach, a voltage is applied between the pad and the wafer. This process does not use slurry, instead it uses electrolyte that provides a suitable medium for copper atoms be oxidized. As the voltage is applied, a current flows from the positively biased pad to the wafer which is grounded via the *bagel*. The bagel is composed of copper beads to provide good contact with the least damage to the contact area. Typical bias voltages range from 2.5V to 3.5V. This current removes electrons from the copper atoms on the wafer, causing them to oxidize and dissolve in the electrolyte. Most of the oxidized copper is in the form of Cu^{+2}, though there are also some Cu^{+1} ions present depending on the pH of the electrolyte. Typical pH levels are around 4.

$$Cu \rightarrow Cu^{+2} + 2e^{-} \text{A} \quad (1)$$

$$Cu \rightarrow Cu^{+1} + e^{-} \text{A} \quad (2)$$

Assuming that most of the Cu oxidizes into Cu^{+2}, the copper removal can be calculated by using the fact that for every two electrons of charge delivered, one copper atom will be removed. The current density J and the time t that the current is applied to the wafer can be used to calculate the amount of copper removed:

$$t \cdot \frac{J \cdot C}{s \cdot cm^2} \frac{1Cu}{2e^-} \cdot \frac{1molCu}{N_A \cdot Cu} \cdot \frac{63.54gCu}{1mols} \cdot \frac{cm^3}{8.96g} \cdot \frac{1e^-}{1.67 \times 10^{-19} C} \cdot \frac{10^7 nm}{1cm} = nm \quad (3)$$

1.3.2 ECMP implementations

The ECMP approach shown in Figure 5 was the first ECMP setup, as developed at IBM and produced by Applied Materials in their Reflexion line of polishing tools. This approach has been adopted by some major companies, including IBM and Phillips. The main drawback of using an approach with a center contact point is that a conductive path on the wafer surface from the removal site to the contact point must exist for current to

flow. When copper is removed so that the current path to the center point is interrupted, removal is cannot continue in the isolated regions, resulting in copper *islands*. This approach is not used as a stand alone single platen tool, but instead is part of a three platen system. It is referred to as hybrid ECMP, as platen 1 provides the ECMP bulk copper removal, followed by conventional CMP clearing of the remaining copper in platen 2, and finally conventional CMP removal of the barrier metal on the third platen.

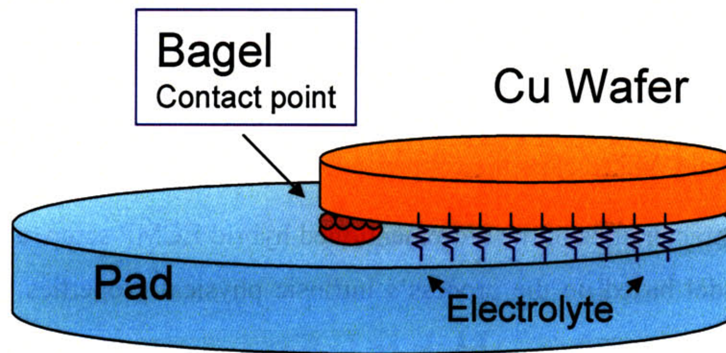


Figure 5: Basic hybrid ECMP setup.

In the case of copper interconnect, the most common barrier used is a thin layer of tantalum on top of another layer of tantalum nitride. This latter layer is needed as a seed to promote the growth of the tantalum barrier layer. A barrier is necessary to prevent copper atoms from migrating through the interconnect and into the silicon, creating deep traps that can hinder the proper behavior of the transistors. The hybrid approach at first seems to be of little benefit, as conventional CMP is still used anyways. The advantage lies in the fact that bulk copper removal is much faster through ECMP, and that electrolyte (mostly citric acid) is used instead of expensive slurry. The cost of the electrolyte is considerably cheaper than the slurry, as the electrolyte is mostly citric acid and a polishing slurry is usually a complex chemical composition and with expensive monodispersed nanoparticles. The increased removal rate might be diminished by the overhead required for an extra platen step, but the reduced pressure also helps to protect the fragile high-K dielectrics used and reduce defectivity.

The contact point with the wafer, the bagel, consists of an intricate contact system that uses small copper beads that provide the contact. Apparently the beads tend to oxidize and loose smoothness, eventually causing scratch-like defects. The defects are attributed to the bagel since they occur only in the outer two inches of the wafer, the exact diameter in which the bagel contacts the wafer. The bagel must be replaced every so often, increasing the maintenance costs.

A different approach is a fully conductive pad [26], consisting of a special pad that contains both the cathode and the anode in a cross grid to eliminate the need for a center contact on the pad (edge contact on the wafer). This approach is called full sequence ECMP, in which ECMP can polish all the way to the barrier metal, eliminating the need to use conventional CMP to clear the copper.

Using the characteristics of the aforementioned hybrid ECMP setup, a dynamic wafer level ECMP model based on the process's intrinsic physical properties is proposed for platen 1.

1.4 Summary

CMP is an essential process in semiconductor processing. The need for better planarity becomes more crucial as dimensions shrink down. The polishing pad plays a center role in CMP and is therefore critical to understand it. Existing pad types, design practices and modeling techniques were described. The experimental pad is a type 3 pad containing water soluble particles that dissolve at the surface. The experiments aim at characterizing the pad and linking its physical properties to its polishing performance.

ECMP presents an efficient and cost effective alternative to conventional copper CMP. The electrochemistry behind the process was discussed followed by descriptions of existing ECMP tool setups. From platen 1 of the hybrid ECMP setup, a dynamic wafer level physical ECMP model is developed in this work.

In Chapter 2, we discuss the plan, experiments and results for our research into pad parameters and performance. In order to measure pad performance, polishing metrics are laid out, together with a description of the oxide CMP model to be used for the data

analysis and simulations. An overview of the data obtained is done followed by the insight gained from its analysis.

Chapter 3 deals with the dynamic wafer level physical ECMP model. It first introduces an empirical model derived from experimental data from the same tool for which the physical model was designed. It then describes the model in detail and follows this explanation with a fitting analysis. This is followed by model optimizations, applications, limitations and future enhancements.

Finally, the conclusion in Chapter 4 summarizes the contributions of this research, and provides suggestion for future work.

Chapter 2 *Experimental Pad Characterization*

The goal of this thesis is to develop models for the chip-scale and wafer-scale performance in two important emerging planarization technologies. The first target is to relate the parameters of a novel WSP pad to its polishing performance during oxide CMP. The second target is to consider the ECMP process, and develop a model to understand the wafer-level uniformity of copper removal, based on the flow of current through the thin surface films on the wafer.

This second chapter is divided into three sections, in which we focus on the experimental pad characterization and results. The first section will establish polishing performance metrics needed to evaluate our results. The second section will explain the model used to evaluate the pad. The third section will go into ample detail about the experimental results and insights obtained.

2.1 Polishing performance metrics

Polishing performance metrics are needed in order to quantify the performance of the polishing processes we study here. There are various ways to measure the polishing performance depending on the process discussed. For oxide CMP, final oxide thickness and thickness range are key parameters to measure since these determine the breakdown voltage of the film as well as its parasitic capacitance and its effect on the frequency response of the circuit. Figure 6 shows how the oxide thickness can vary due to the underlying pattern, in this case aluminum lines in conventional aluminum/oxide interconnect.

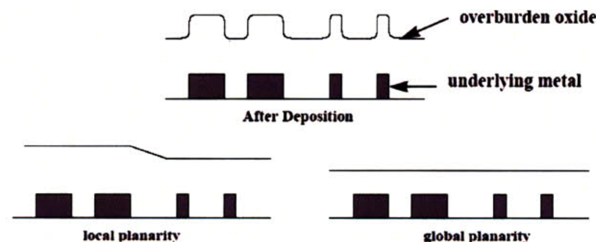


Figure 6: Oxide CMP pre and post polish oxide thickness variations.

In the case of STI CMP, the minimum requirement is that all the oxide be cleared down to the nitride over the active areas or device malfunction will occur. Dishing of the oxide and erosion of the nitride, as pictured in Figure 7, is a key measure of the effectiveness of the process, particularly at the point in time when clearing of the overburden oxide has just been achieved. Copper ECMP is very similar to STI CMP, so that the performance metric to be used in patterned wafers is copper dishing and dielectric erosion. An advantage of copper polishing versus oxide polishing is that copper thickness can be measured by resistivity measurements, which clearly show the effects of copper thinning and are easier and more accessible than ellipsometer film thickness measurements.

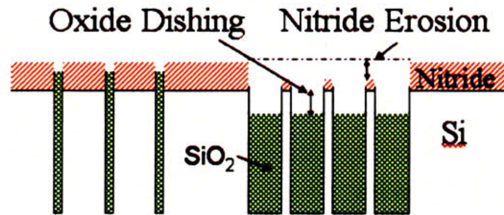


Figure 7: STI CMP dishing and erosion.

An important performance metric for all processes is the final step height, which ideally should be zero. In the case of oxide CMP this is usually the case, but for dual material polish there can often be a finite step height left at the end due to polish rate selectivity, such as between nitride and oxide in STI CMP. As dimensions keep shrinking, this final step height which used to be negligible starts to play a more important role. The time to achieve a target step height is a good measure of planarization efficiency with respect to time. The thickness range at this time is a good measure of wafer non-uniformity. Figure 8 illustrates these points using step height and thickness evolution plots.

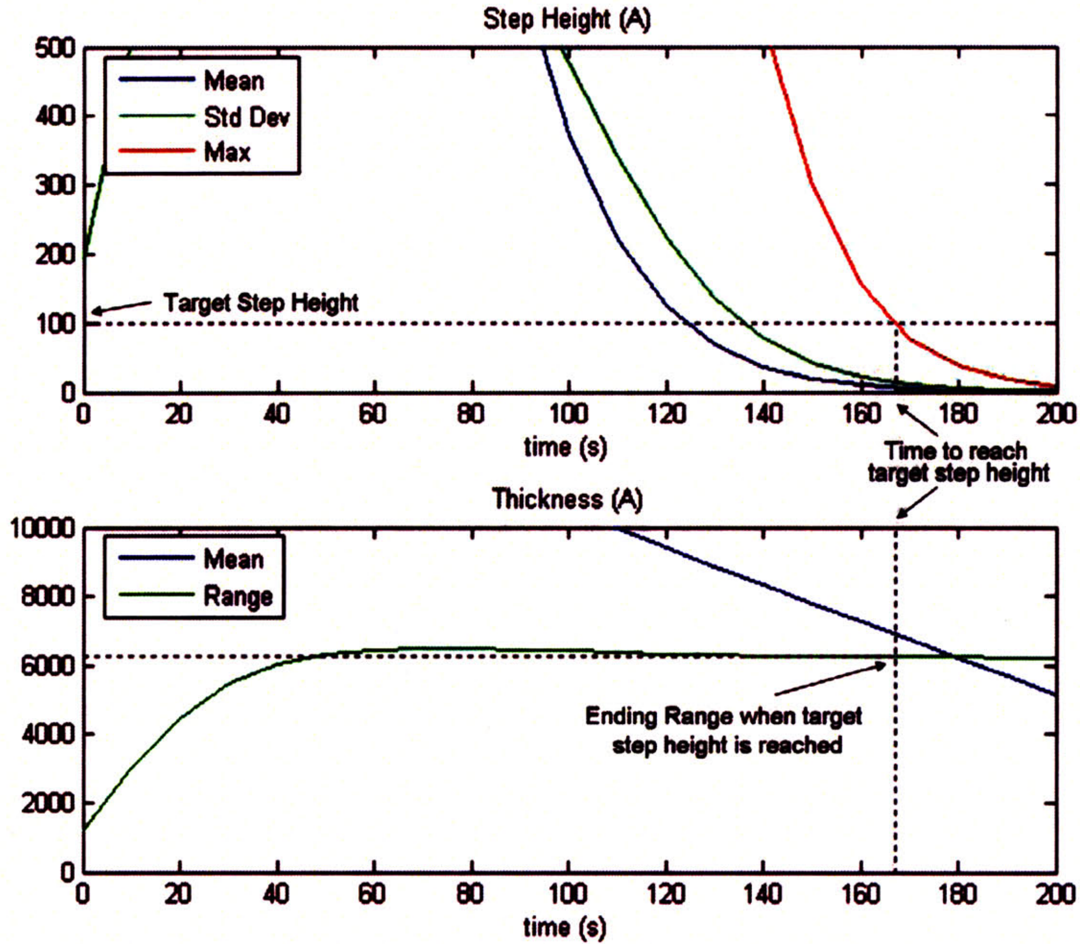


Figure 8: Performance metrics (above – time to target step height; below - ending range).

In terms of efficiency with respect to spatial dimensions, the amount of material removed (thickness) to achieve a target maximum final step height is a useful metric. This metric can be combined with the blanket removal rate to offer a sense of the time efficiency as well. The planarization efficiency [27], [28], or PE, is defined as:

$$PE = \left(1 - \frac{\Delta down}{\Delta up}\right) \times 100\% \tag{4}$$

where Δdown is the amount removed in the trench area and Δupper is the amount removed in the upper areas.

2.2 Conventional oxide CMP PDSH model description

Our research group has previously developed a pattern-density step-height (PDSH) dependent model for CMP. The three main parameters the model uses are planarization length, contact height h_c, and equivalent blanket removal rate. The model takes the raw layout pattern density and creates an effective density map as illustrated in Figure 9.

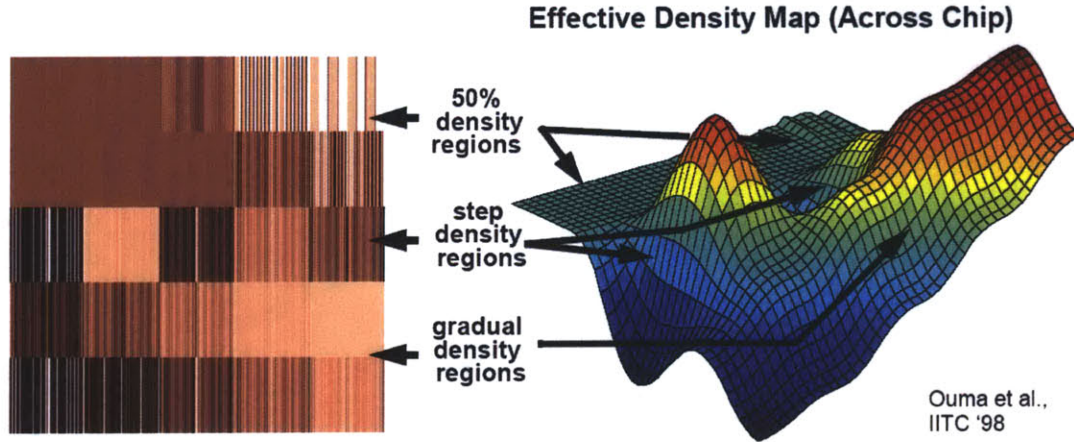


Figure 9: Raw pattern density (left) and effective pattern density (right) across test die.

The effective density of the wafer is intimately tied to the planarization length (*PL*), which measures the distance over which the raw or local pattern density at one point will influence polishing on neighboring points. Common values for planarization length are in the range of millimeters, making it a good measure for wafer non-uniformity as well as local planarization. A higher *PL* means the process is less susceptible to effective pattern density variations.

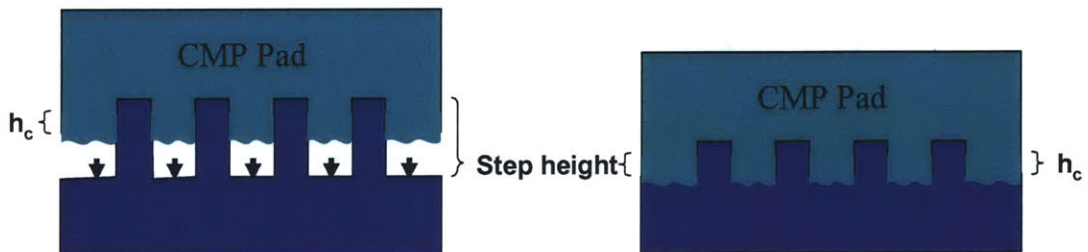


Figure 10: Contact height h_c.

The contact height h_c can be thought of as the step height at which the pad starts to polish inside the trench, as pictured in Figure 10. This contact height is dependent on the pattern density and local deformation of the pad around features. Along with the pattern density, the contact height determines the exponential decay in removal rate with respect to step height as shown in Figure 11.

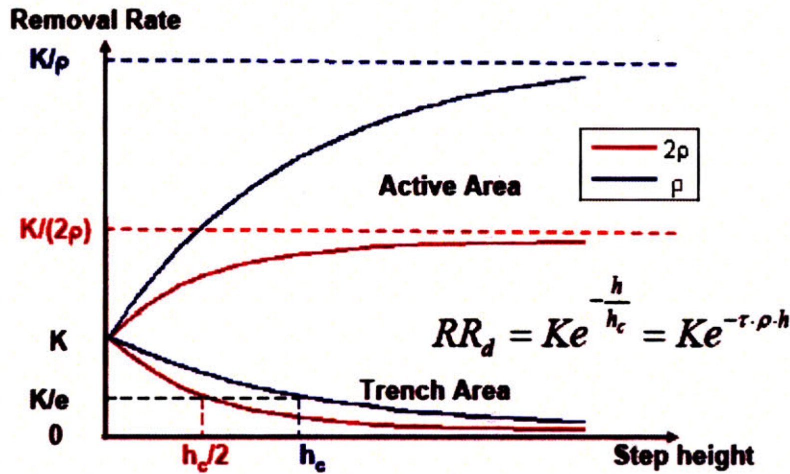


Figure 11: Exponential step height - pattern density model.

A higher h_c will result in a higher planarization efficiency, since there is less material removal in the trench that could otherwise slow down the step height reduction. The model internally uses a more general parameter τ, which is multiplied by the effective pattern density to give the inverse of h_c. That is $h_c = \frac{1}{\tau \cdot \rho}$, since the characteristic step height is different for different pattern densities, and thus step height reduction rates are different for areas with different effective pattern densities. Notice that removal rates are based on *effective* pattern densities and not *raw* pattern densities. The effective pattern density is determined by the physical properties of the pad and is tied to the *PL* parameter previously discussed.

The blanket removal rate *K* is the theoretical blanket removal rate if the wafer had no features on it, in other words, a zero step height all throughout the wafer. The removal rate is an important characteristic of the pad, though it gives into insight on spatial efficiency. A higher blanket removal *K* means the pad will polish the wafer faster, but

offers no information about planarization. A high removal rate is ideal, as long as it not at the expense of planarization, and tradeoffs in these two metrics will be important. Some guidelines for measuring pad performance will be discussed later in the performance metrics section.

The three model parameters, τ (or h_c), PL, and K, together provide a valuable characterization of the polishing behavior. The model is also augmented to include a ramp-up time parameter to account for tool ramp-up time required at the start of a polish step before the process reaches to steady state conditions. In later sections the sensitivity of each parameter to one another and to the polishing performance will be discussed.

All of these empirical model parameters can be correlated to pad properties. The purpose of the experiments in collaboration with JSR is to be able to relate these model parameters to pad properties in an attempt to use the model to engineer a better polishing pad. The planarization length and contact height can be related to the pad stiffness, since the local and longer range pad deformations will determine these. The PL is expected to relate to the pad's stiffness, measured by its lateral Young's modulus E. A higher WSP concentration should make the pad bulk stiffer. The h_c value will depend on the vertical Young's modulus and the WSP size could influence this value. The blanket removal rate may be related to the pad surface roughness, dictated by the pad's porosity and WSP concentration and size. The pad roughness is usually measured in terms of the asperity height distribution [30]. A higher roughness might also impact removal rate, by improving slurry transport to and material removal out of the polishing sites.

2.3 Experimental

2.3.1 Setup and methodology

Removal amounts are measured for the SKW7-2 MIT test mask for different time splits up to 120 seconds. The wafer has blocks of patterned lines to generate different layout pattern densities. Figure 12 shows a clear mapping of the different density areas on the die:

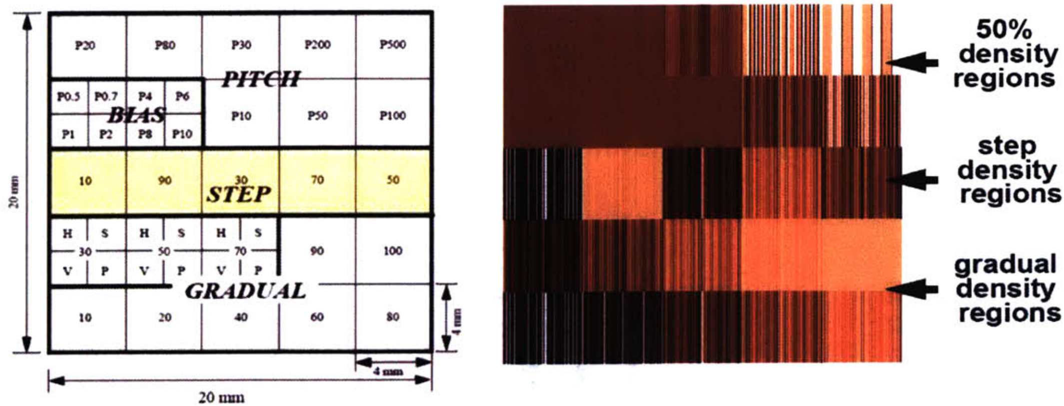


Figure 12: Test die pattern.

For our experiments, only measurements from the horizontal central bar are used; this area includes five blocks of varying density but constant pitch. This was done because pattern density, and not pitch, is of interest for the model. The initial oxide thickness is approximately 20,000 Å with an initial step height around 8000 Å (for the model calibration the separately measured values of thickness and step height are used). Optical film thickness measurements are taken for the active and trench areas for regions with 70%, 50%, and 30% pattern density. These measurements are performed by ellipsometry. Step height profilometry measurements are also obtained for 10%, 30%, 50%, 70% and 90% regions. Polishing was done on a MIRRA/MESA (AMAT) polisher using Cabot SS25 50% diluted slurry at 150 ml/min. The platen speed is 62 rpm and the head speed is 56 rpm. The retainer ring pressure is 6 psi and the membrane and inner tube pressures are both 5 psi. The conditioners used are Mitsubishi 100 and 325 grit conditioners rotating at 56 rpm with a downforce of 4 lbf. A pad break-in time of 600 seconds is used. Figure 13 gives a summary of the polishing conditions, measurement sites and test die.

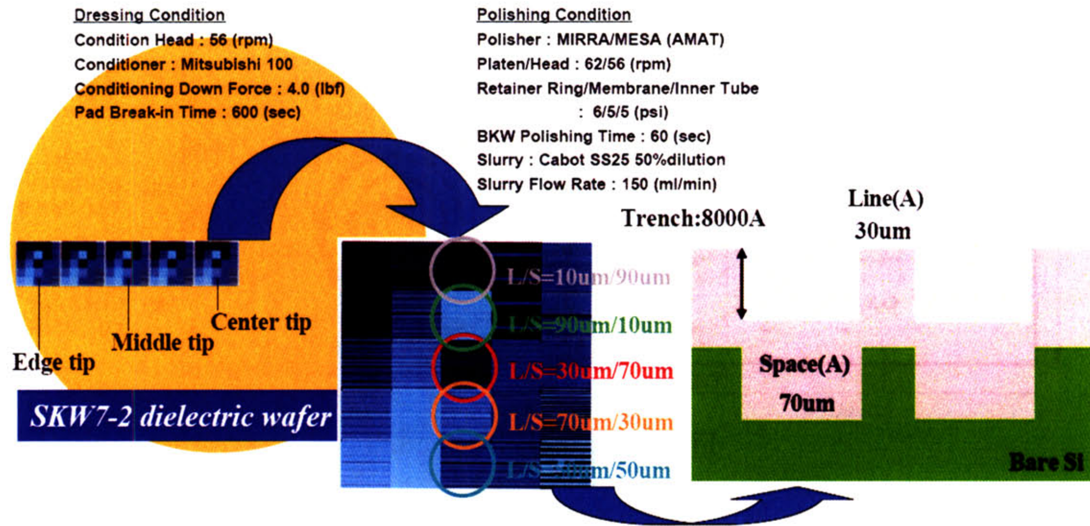


Figure 13: Experimental polishing conditions.

2.3.2 *Ramp-up time*

When a wafer is placed into the tool for polishing, the pad comes in contact with the wafer before the desired speed and pressure are achieved. The time between the initial contact and when the steady state rotational speed and pressure are achieved is referred to as the ramp-up time, and is usually less than ten seconds. During this time the removal rate is ramping up to its full rate, but nonetheless there is a certain amount of material removed. Ideally one would want to have n identical wafers and polish them in incremental time intervals, where the first one polishes for $1/n^{th}$ of the desired time interval, the second one for $2/n^{th}$ of the time interval and so on until the last one polishes for the full time interval. In this case all wafers will share an initial ramp-up time.

Test wafers are sold for at least $500 a piece, therefore minimizing the number of wafers used is imperative. In our case, the data is taken by polishing a wafer for a fraction of the full time interval, taking measurements, and then polishing the same for another fraction of the remaining time interval and so on until the full time interval is completed. This will give the cumulative removal for the same test wafer at a lower experimental cost but at the expense of accumulated ramp-up times. The time reported for the measurements includes ramp-up time, but there is no a priori specification of the ramp up time. Therefore it is extracted so that this transient time of lower polishing rate can be accounted for. In order to mitigate this ramp-up time limitation, an extra fitting parameter

delta is added to the model. This time offset is fitted by minimizing the root mean square of the error of the amount removed data and the model predicted amount.

Two modalities of this fitting time offset are implemented into the model. In the first one, only an initial ramp-up time is added for wafers with different "polish time" experiment split points. This works for the case when a separate test wafer is used for every fraction of the time interval. In the second case, the time offset is added each and every time a new polish time interval measurement is made. This is to account for the case when the same test wafer is polished in time increments to sample the whole time interval. Figure 14 shows qualitatively how the delta parameter affects.

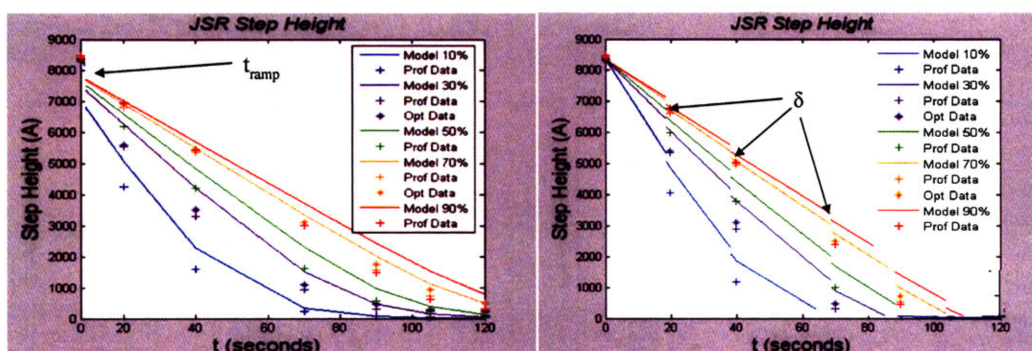


Figure 14: Ramp-up time and multiple ramp-up times effect on measurements.

2.3.3 *Preliminary Results*

Figure 15 shows the first set of data obtained by JSR comparing the WSP pad with one of the most commonly used pads in industry, the IC1000. The data is obtained from optical film measurements of the trench and the active areas by ellipsometry on the 30% and 70% density blocks of the test die at three locations along the wafer radius (center, middle, edge) after polishing with either pad. The trench area is also referred to as the down area, and the active area is known as the up area. The data by itself reveals little information about each pad's polishing performance. It can be said that the IC1000 has a higher removal rate, which is more evident for the 30% than the 70% blocks, for which rates are comparable. Also the JSR pad plots show that the edge removal tends to be higher than the middle and center of the wafer, revealing edge non-uniformity for this pad. The IC1000 pad appears to be fairly uniform throughout the wafer radius. The raw data does not yet reveal much about each pad's physical characteristics and how they

might play a role in efficient planarization. The model will provide a means to squeeze out more insight from the raw data to learn how pad characteristics can be modified to improve polishing performance.

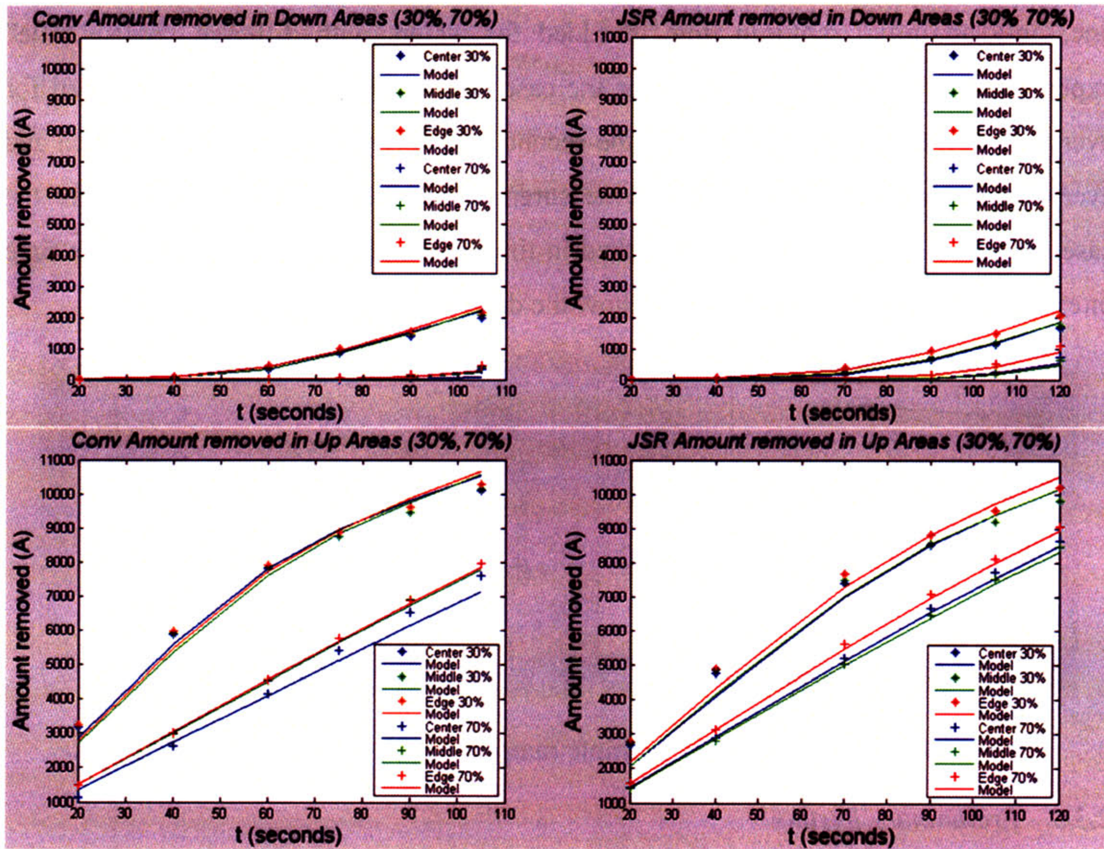


Figure 15: Initial experimental results for JSR WSP vs. IC1000 pad.

Once the model is fitted, the value of this data is exploited; these are the solid lines on the plots. This is done by extracting the model parameters, τ (h_c), *PL*, and *K*, given the data, by running an optimization loop that extracts the best parameters. Their respective values and the fitting error for middle wafer locations are shown in Table 1. The fitting error of the model is calculated by minimizing the square of error between the measurement data and the model prediction given the three model parameters. Given that the full amount removed is 10,000 Å, the error is less than 5% in both cases.

Pad	τ (1/Å)	*PL* (mm)	*K* (Å/min)	Error (Å)
IC1000	0.018	3.6	2780	353
JSR WSP	0.018	4.8	3050	393

Table 1: Model parameters for preliminary runs for the IC1000 and the JSR WSP pad

Recall that $h_c = \frac{1}{\tau \cdot \rho}$, meaning that both pads had the same characteristic heights but not necessarily similar asperity height distributions. The physical reasoning behind this will be explained later in the physical pad properties section. For now it is sufficient to say that most of the removal is done by the tallest asperities as Figure 16 shows, so both pads may have similar a similar distribution of the tallest asperities but a very different short asperity distribution. This explains why two structurally different pads can share the same value for τ.

Figure 16: Pad asperities.

The three parameters above can be fed into the model to extrapolate removal curves for other pattern densities. Figure 17 shows the predicted traces of the amount removed in the up areas and the down areas for the middle wafer locations of the test die. The left side shows the IC1000 and the right side shows the JSR WSP pad. The double sided arrows at t = 100 sec are meant to show that the amount removed spread across all pattern densities is greater for the IC1000 pad than for the JSR pad. The lower spread means that the JSR pad is less pattern dependent, since its removal rate is more even across different pattern densities. This advantage seems to be at the expense of lower removal rate, so there is a tradeoff between speed and planarization which translates into a potential tradeoff between wafer throughput and yield.

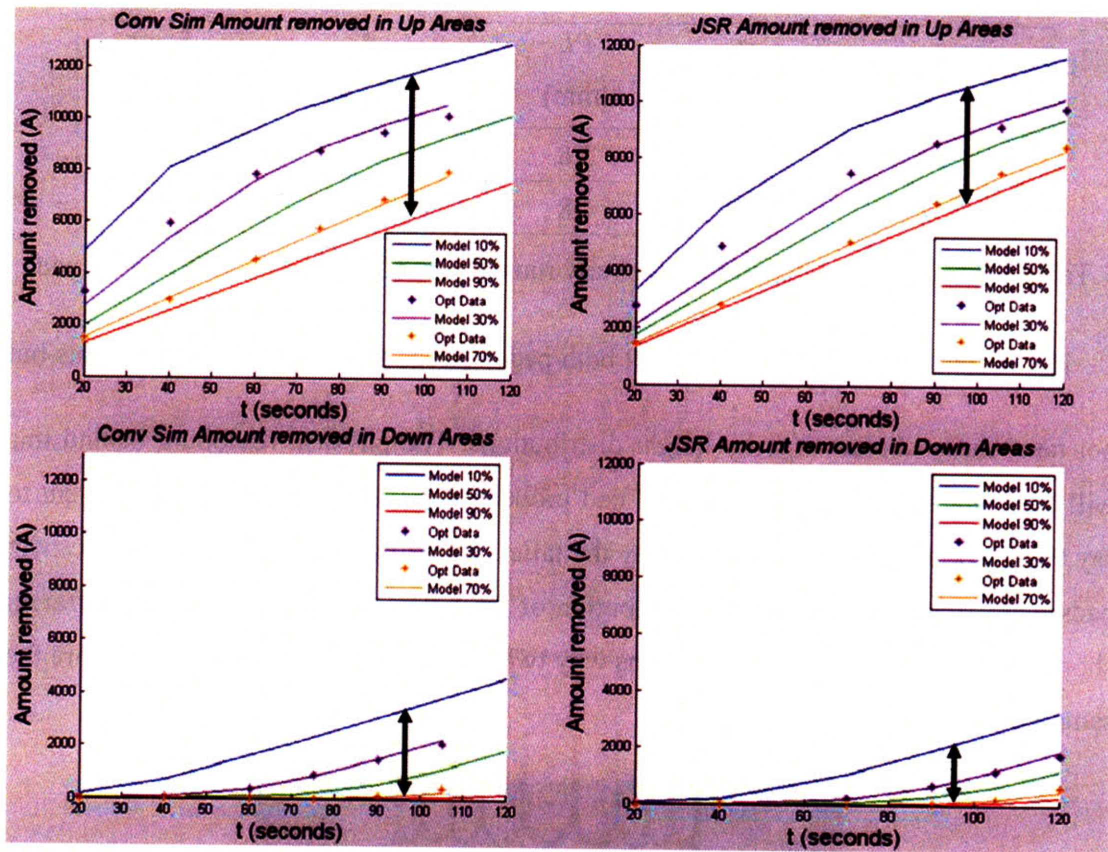


Figure 17: Model predictions for different pattern densities for the JSR pad (right) and the IC1000 pad (left).

The step height evolution can be extracted from the initial film thickness and the amount removed in the up and down areas using Equation 5:

$$\textit{Step Height} = \textit{Initial Thickness} - \textit{Up Areas Amount Removed} + \textit{Down Areas Amount Removed} \quad (5)$$

The step height reduction can give a more explicit picture of each pad's planarization efficiency. Figure 18 shows the step height evolution for the previous data. The lower pattern density regions are planarized first and are then overpolished until the other regions are planarized. This is reflected in the extra amount removed by the IC1000 in Figure 17. The IC1000 has a much wider overpolish time spread than the JSR pad, by at least 15 seconds. To put this in perspective, consider a typical regular oxide polish rate of 3000 Å/min: fifteen extra seconds add up to an extra 750 Å removed in the lower pattern density areas. An overpolish of that magnitude can also impact wafer non-uniformity. In back-end oxide polishing this can imply large deviations in the designed capacitance of an insulating layer, leading to time constant changes for the metal lines that affect a

circuit's speed performance. In the case of STI, those extra fifteen seconds can lead to serious trench dishing and active area erosion that can reduce yield considerably. Despite the fact that both pads achieve planarization roughly at the same time, the benefit of the reduced overpolish offered by the JSR WSP makes it an attractive option.

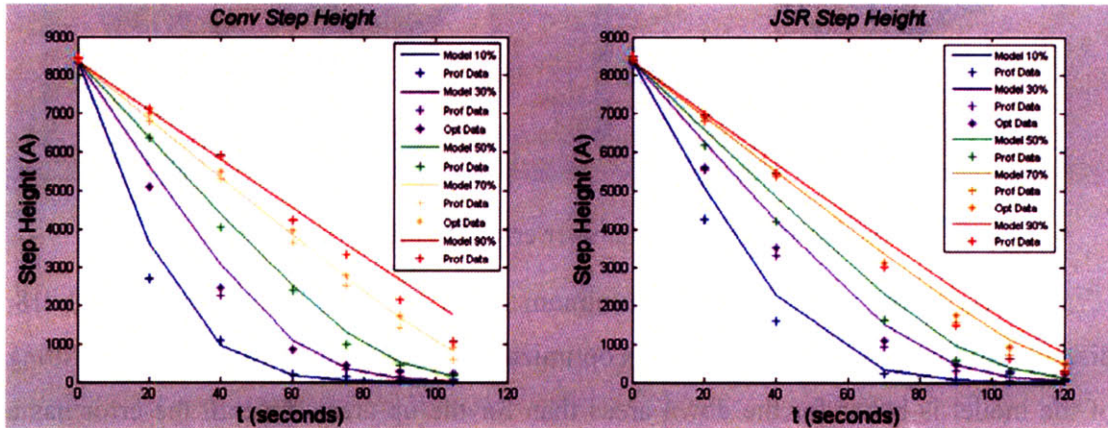


Figure 18: Step height evolution between IC1000 pad (left) and JSR WSP pad (right).

From this data it is clear that the JSR WSP pad is an interesting subject to study. The flexibility offered by the embedded WSPs implies potential for optimization. Their size and concentration are properties that can easily be adjusted. The particle size distribution is homogenous and easily varied; the concentration can be set during the mix with the pad bulk material. The round of experiments discussed in the Sections 2.3.5 and 2.3.6 aims at exploring the effect on polishing performance for different sizes and concentrations of water soluble particles in the JSR pad.

2.3.4 Model parameter sensitivity analysis

Before discussing additional experiments in the following two sections, a sensitivity analysis is instructive to explore which parameters have a greater effect on the fitting error of the model for both the up areas and the down areas. This would allow a better understanding of the contribution of each parameter to the model error and how to minimize it, as well as to prevent optimization errors caused by local minima. Figure 19 shows the model fitting error for varying values of PL and τ.

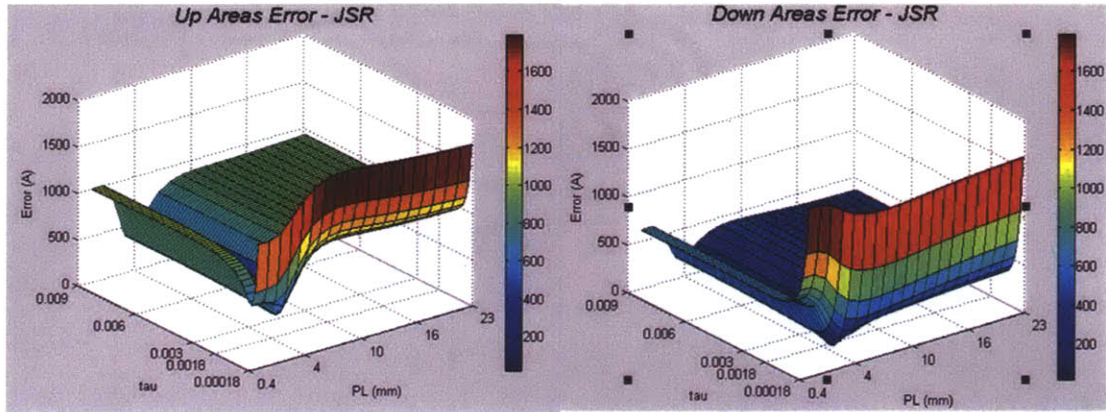


Figure 19: Model parameter error sensitivity analysis.

The plots both show a global minimum along PL = 4.8 mm and τ = 0.018, discounting the possibility of erroneous optimization due to a local minimum. The fitting of the model is better for the down areas than for the up areas. Overall the error has a stronger dependence on the planarization length, except when τ goes below 0.0001, but this would imply contact height h_c on the scale on 2 μm, an order of magnitude higher than usual. These results make sense, since τ is an exponential rate parameter, while PL controls the averaging filter which has a first order effect on the results.

In order to better understand how much the model predictions vary for different parameters, the parameters are varied over a given range and plots of the step height evolution are made. Figure 20 shows the step height evolution for PL values ranging from 2 mm to 5.2 mm, along with the profilometry and optical data available. Keeping in mind that the extracted PL value is 4.8 mm, the data fit is not as good toward the extreme pattern density values (10% and 90%), but gets better towards the middle values. It is impossible for the model to fit a value that has an equally good fit for all densities, because the removal rates have very different trends at low and high densities. For low densities the step height reduction is fastest for lower PL values, showing an exponential decrease. For high pattern densities the trend is opposite, having low densities as the slowest regions to planarize, and showing that the evolution tends to be linear throughout most of the process.

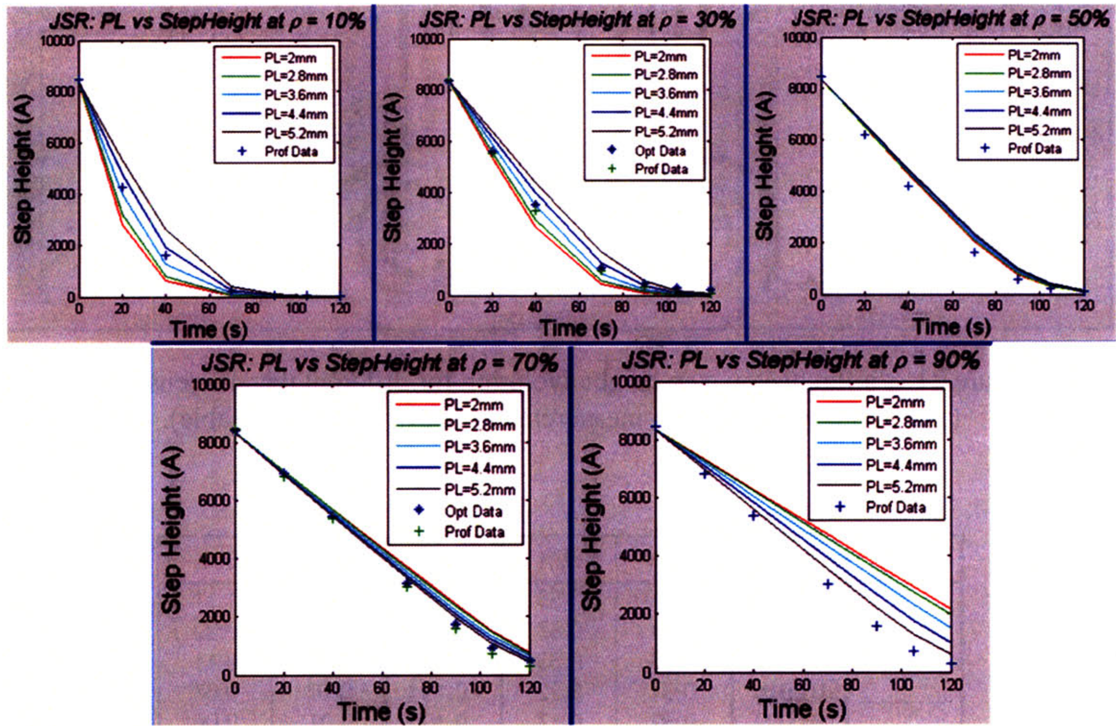


Figure 20: *PL* influence on step height evolution for different pattern densities (profilometry and optical measurements shown where available).

To illustrate the exponential behavior of τ, plots are also generated for values ranging from 0.0001 to 0.0030. The plots of Figure 21 and Table 2 show that as τ gets larger, its effect becomes smaller and eventually negligible. For values below 0.0005, the effect is very large, but the characteristic height is too large and is therefore unphysical. Step height reduction is faster for lower densities and follows a decreasing exponential trend in contrast with the more linear and slower trends for larger densities. These illustrations confirm that the strongest parameter of the model is the planarization length.

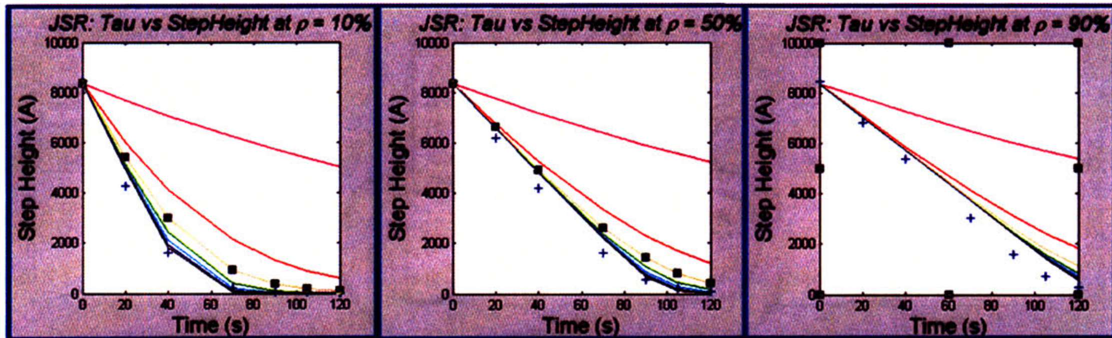


Figure 21: τ influence on step height evolution for different pattern densities. (profilometry and optical measurements shown where available).

Color	τ \ %	10%	30%	50%	70%	90%
	0.0001	10.00	3.33	2.00	1.43	1.11
	0.0005	2.00	0.67	0.40	0.29	0.22
	0.0010	1.00	0.33	0.20	0.14	0.11
	0.0015	0.67	0.22	0.13	0.10	0.07
	0.0020	0.50	0.17	0.10	0.07	0.06
	0.0030	0.33	0.11	0.07	0.05	0.04

Table 2: Equivalent characteristic step height.

2.3.5 Variable WSP concentration experiments

In this section, a second set of experiments are discussed. First the setup conditions used are described with a short summary of the fitted model parameters. The following three subsections describe each of the parameters: *K*, *PL* and τ followed by a section discussing the model fitting error. Finally, the last section summarizes the main trends observed in this second experimental set.

2.3.5.1 Setup conditions and model fitting

After establishing in previous experiments that the JSR WSP pad offers improved planarization efficiency, as dictated by its lower pattern dependence, a second round of experiments is conducted to explore the effect of WSP concentration on polishing performance. In this second round of experiments, three concentrations by weight are explored and will be referred to as low, standard and high concentrations.

Aside from the concentration, the effects of different conditioning are also explored. The grit size is the industrial standards used to classify the size of the diamond particles used in a grinding wheel [31]. Two grit sizes were used to condition the pad: #100 and #325. The first one has an average diamond particle size of 70-80 μm, while the latter one is widely used in industry for CMP pad conditioning with an average diamond particle size of 35-40 μm. Figure 22 shows two SEM images to illustrate the diamond particle density and size (the scale bar is 50 μm):

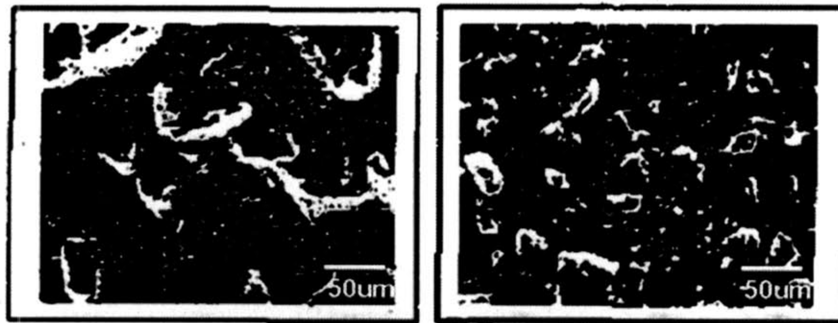


Figure 22: SEM images of pad conditioners: #125 grit (left) and #325 grit (right) [31].

Studying two different grit sizes would allow checking whether or not the more course grit (#100) would be more beneficial than the industry standard (#325), since the WSP pad surface roughness behaves differently due to the water soluble particles.

The wafers polished for this round are also SKW-7 test wafers as used in the first round. A total of eighteen wafers are divided into two lots: half of them are polished using in-situ conditioning with the #325 grit, and the other half with the #100 grit. Out of each lot, one third is polished using a JSR pad with low WSP concentration, another third with a pad with standard WSP concentration and the last third with low WSP concentration. Each wafer has measurements performed on center, middle and edge dies. Within each die, there were profilometry measurements for all five pattern density blocks (10, 30, 50, 70, 90%) and optical measurements for the 30%, 50% and 70% density blocks only. The wafers are polished for two minutes with time splits and measurements at 0, 20, 40, 70, 90, 105, and 120 seconds.

2.3.5.2 *Preliminary results*

The model is able to initially extract the three main parameters (τ, K, PL) with an RMS error lower than 500 Å for center and middle wafer measurements. The edge measurements show a greater error, yet below 800 Å. On average 6,000 Å is removed from the initial 20,000 Å thick oxide, giving an error slightly less than 10% of the amount removed; a relatively good number except when noting that 500 Å can significantly affect the theoretical capacitance of a typical 0.5 μm thick dielectric. The extracted model parameters, as well as the equivalent characteristic height h_c for a 50% pattern density, for each of the 18 runs are shown in Table 3.

Grit	Concentration	Position	K (A/min)	PL (mm)	τ (1/A)	h_c (um) for ρ=50%	Err (A)
100	High	Center	2245.6	4.4856	0.0007	0.2769	441.52
100	High	Middle	2080.1	4.4304	0.0009	0.2167	378.72
100	High	Edge	1904.8	3.9477	0.0007	0.2769	575.94
100	Standard	Center	1999.4	4.1404	0.0008	0.2468	384.43
100	Standard	Middle	1913.1	4.3008	0.0010	0.2077	380.71
100	Standard	Edge	1724.7	3.7557	0.0008	0.2563	559.94
100	Low	Center	2044.6	4.0557	0.0009	0.2152	402.37
100	Low	Middle	1920.2	4.0557	0.0011	0.1810	417.31
100	Low	Edge	1718.9	3.6474	0.0008	0.2408	546.87
325	High	Center	2447.3	4.5220	0.0011	0.1784	418.83
325	High	Middle	2449.6	4.7558	0.0014	0.1479	394.87
325	High	Edge	2371.6	4.0557	0.0010	0.1988	771.34
325	Standard	Center	2515.7	4.4206	0.0013	0.1517	475.2
325	Standard	Middle	2487.8	4.6529	0.0016	0.1219	476.77
325	Standard	Edge	2376.7	4.0236	0.0011	0.1797	757.69
325	Low	Center	2293.2	4.1795	0.0015	0.1372	433.93
325	Low	Middle	2286.4	4.1984	0.0019	0.1052	440.69
325	Low	Edge	2408.0	4.0557	0.0012	0.1612	714.95

Table 3: Extracted model parameters for second experimental set.

The main observation from the above data is that all the parameters are better for the case when the pad is conditioned using the #325 grit instead of the #100 grit. For all three model parameters, K, τ and PL, higher values are indicative to increased polishing performance. This comparison is shown Figure 23.

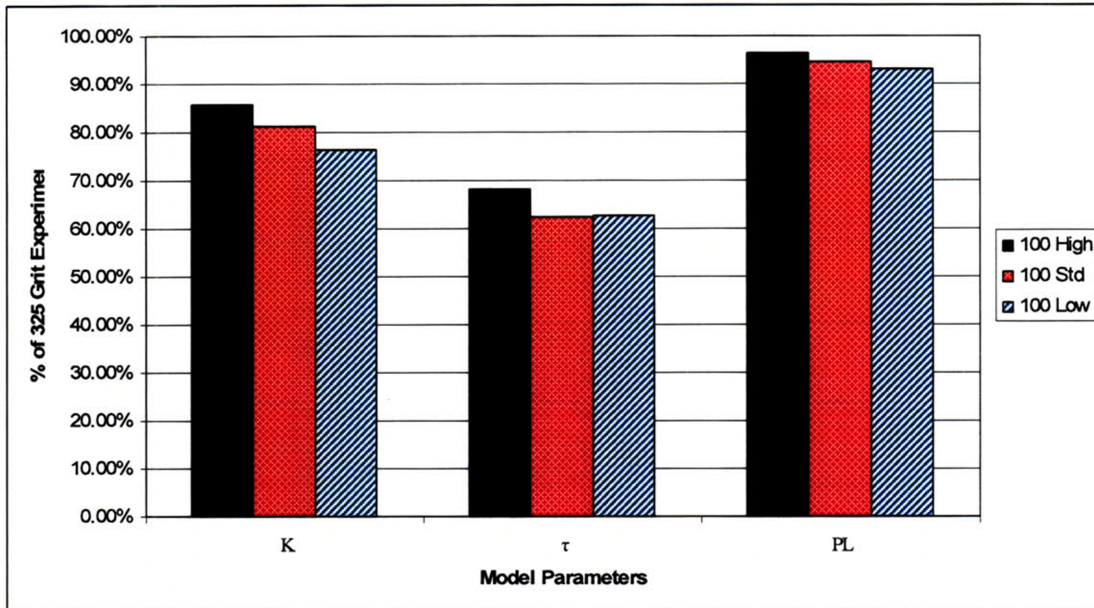


Figure 23: #100 grit parameters normalized to #325 grit parameters.

From this point on, the data presented for analysis will only include the #325 grit data since its performance is superior to that of the #100 grit data. For this same reason all of the following experiments use only the #325 grit.

Aside from the large RMS error, the results raise important concerns: an unexpectedly low value for K and a PL value larger than the density block size. Regarding the first issue, the model seemed to underestimate the blanket removal rate; this is shown in Figure 24. There are three possible explanations for this inconsistency. First, the data provided to compare is from blanket wafers, for which the removal rate might be different than for patterned wafers. Second, these blanket wafer removal rates represent an average over the entire wafer, while the data with which the model is calibrated is for sites in the center, middle and edge of the wafer. Finally, the fact that the data comes from only one wafer in which the time intervals measured are close to the polishing ramp-up time, means that some polishing occurred during the transient polishing time, when the full removal rate has not been reached.

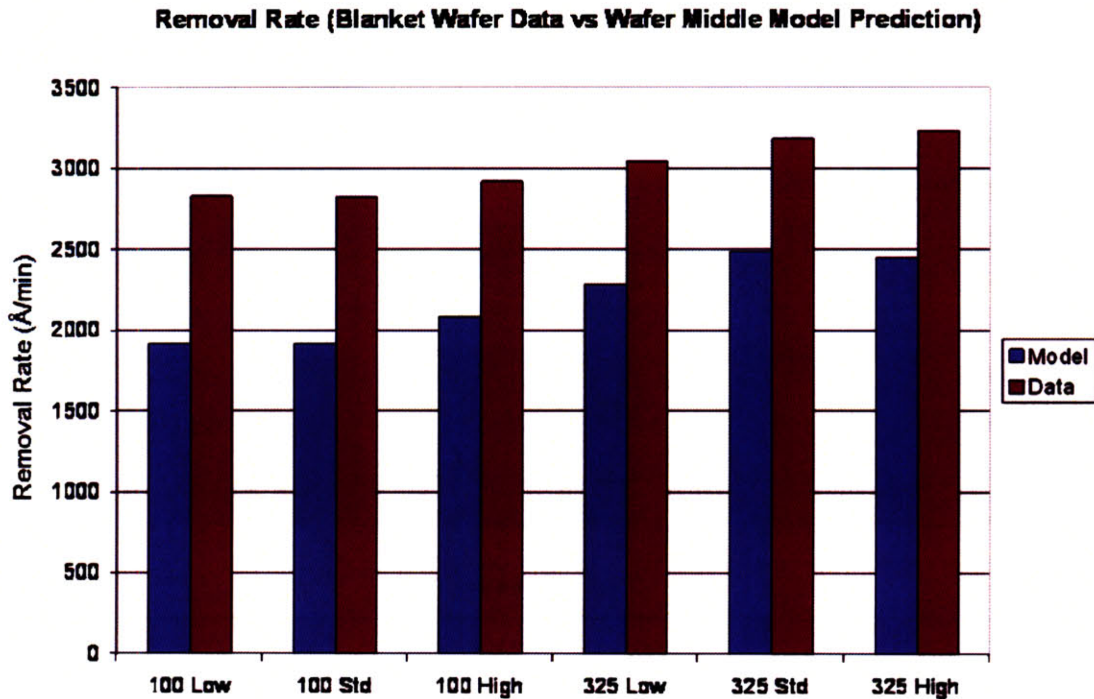


Figure 24: Model predicted blanket removal rate vs. given data on blanket removal rate.

The second issue is that the planarization length magnitude interacts with the experimental pattern density conditions. Specifically, the local layout pattern density is not identical to the effective (a spatially averaged) pattern density, except in the undesirable case of a short *PL*. (A short *PL* is indicative of poor within die uniformity, and so a larger *PL* is desirable to reduce or minimize pattern density dependencies.) As a result, care must be taken interpreting experimental results. The planarization length is around 4.2 mm on average and is as large as 4.7 mm for the case of the #325 grit, high concentration middle wafer location. This is overall a better planarization length than the previous 3.6 mm for the IC1000 pad, showing that the pad produces better die uniformity even with varying conditioning and WSP concentrations. The few places where the *PL* drops below 4 mm are at the edge measurements of the wafers polished with the pads conditioned with the #100 grit, which is uncommon in industry. The challenge lies in the fact that the test die size is 20 mm x 20 mm, and the individual density blocks are 4 mm x 4 mm. This causes a slight difference between the raw and local layout pattern density and the effective pattern density, since the averaging filter includes neighboring squares with different densities; this is illustrated in Figure 25.

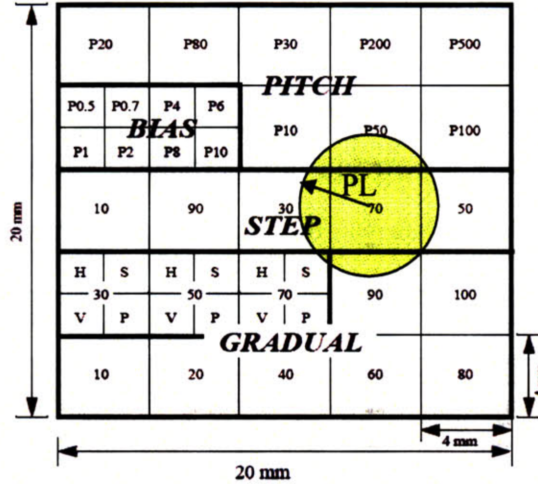


Figure 25: Planarization length larger than the test density region.

Based on averaging due to the larger planarization length, the 50% designed region does not show much change, but the 30% and 70% regions have considerable shifts. The center of the 30% region has an effective density over 40%, and the center of the 70% region is shifted by a lesser degree, down to around 65%. The main concern is that the model fitting may be better if a wider and more representative set of possible pattern densities were available. One could argue that in industry most design patterns fall between 40% and 60% effective pattern density (through the use of dummy fills and other layout optimization tools and rules), so that our available data helps to make the model more accurate for the most meaningful range for real world applications. The overall trends of the model for the different densities follow the expected inverse relation between pattern density and removal rate, but to ensure a better model fit, more measurements near the edges are obtained for the #325 grit wafers to attempt a better fit. Figure 26 shows the additional off center measurement sites. The new measurements reduce the fitting error by 50%. All of the model parameters increased, suggesting even better pad performance than previously recognized. These are shown in Table 4. Each of the parameters will be discussed in the following sections.

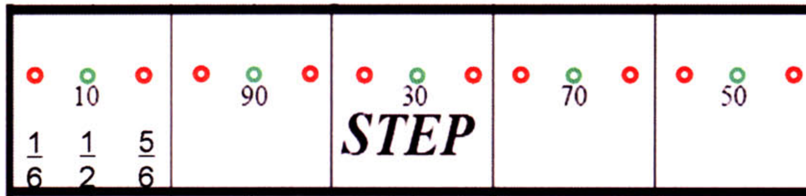


Figure 26: New sites near the edge to improve model fitting.

Grit	Concentration	Position	WSP Size	K (A/min)	PL (mm)	τ (1/A)	h_c (um) for p=50%	Err (A)
325	High	Center	Medium	3341.3	6.383	0.001521	0.132	225.0
325	High	Middle	Medium	3478.1	6.249	0.001594	0.125	236.5
325	High	Edge	Medium	3557.8	6.420	0.001182	0.169	332.4
325	Standard	Center	Medium	3437.0	6.100	0.001668	0.120	229.0
325	Standard	Middle	Medium	3621.3	6.186	0.001870	0.107	231.6
325	Standard	Edge	Medium	3841.0	6.411	0.001352	0.148	335.2
325	Low	Center	Medium	3337.8	5.875	0.001766	0.113	280.9
325	Low	Middle	Medium	3470.1	5.962	0.002015	0.099	254.5
325	Low	Edge	Medium	3713.2	5.944	0.001525	0.131	362.5

Table 4: Improved fit results for second experimental set.

2.3.5.3 Blanket Removal Rate

The blanket removal rate has a relatively large radial dependence. Figure 27 plots this variation across the wafer, showing center to edge variations of around 12%. This variation also follows a clear linear trend as it moves towards the edge and can cause wafer non-uniformities. A 400 Å/min difference would imply a 400 Å across-wafer range for a one minute polish, which depending on the process might or might not be important.

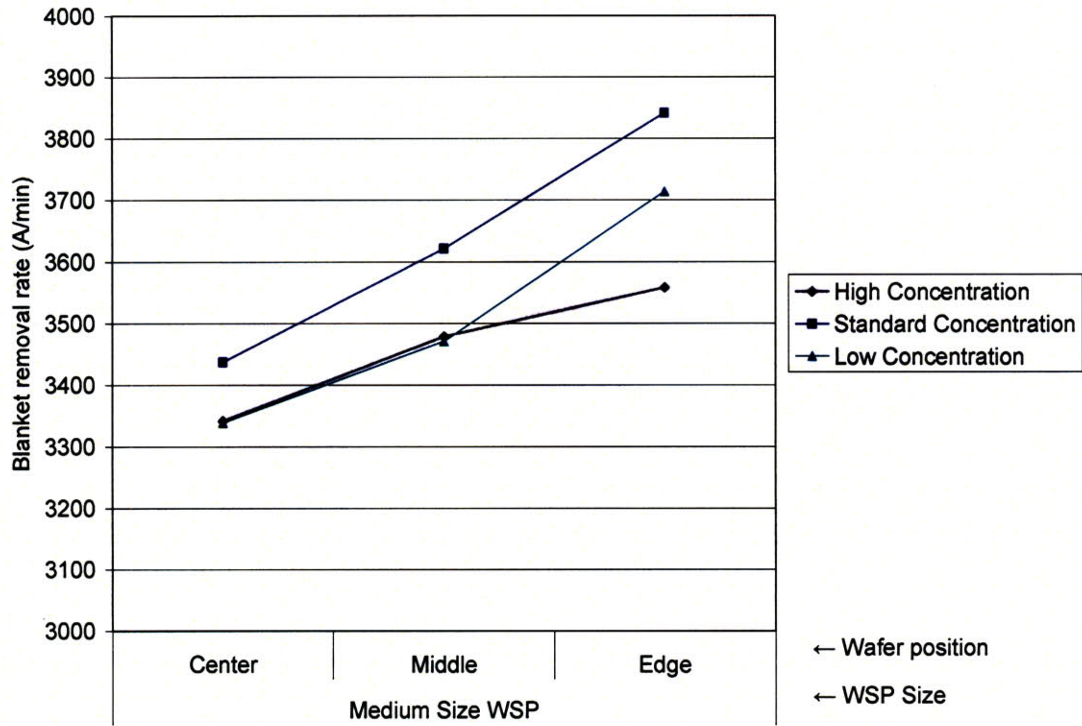


Figure 27: Blanket removal rate across wafer locations for the second experimental set.

Figure 28 shows the removal rates for the three WSP concentrations used. The high and low concentrations behaved similarly, except that the low concentration has a greater *edge bite* (higher relative removal in wafer-edge region). The standard concentration seems to be the best choice to maximize the removal rate, but the difference is less than 5%. We next consider the other two parameters to understand how concentration affects overall performance.

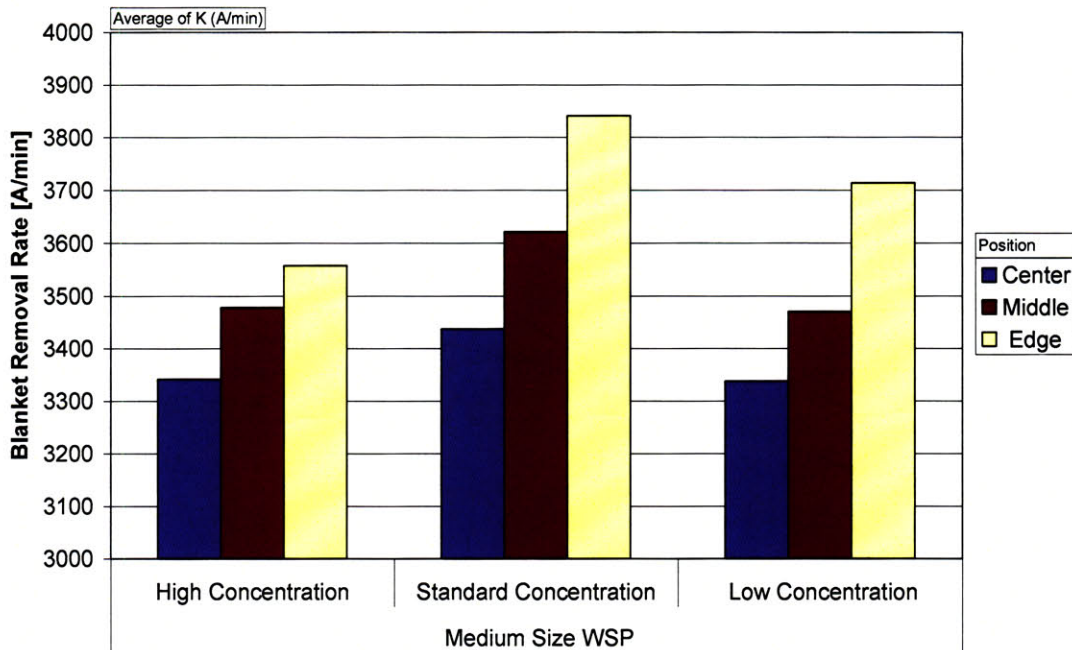


Figure 28: Blanket removal rates for all grit and WSP concentrations tested.

2.3.5.4 Planarization Length

The extracted planarization length variation across the wafer is depicted in Figure 29. There is no general trend; *PL* could be higher at the edge but the wafer non-uniformity is below 5% for all WSP concentrations, which is comparable to the range of wafer to wafer variations.

On the other hand, the WSP concentration does show a clear effect on the *PL*. Figure 30 shows a direct relationship between *PL* and WSP concentration: the high concentration yields the largest *PL*, which is the most desirable case.

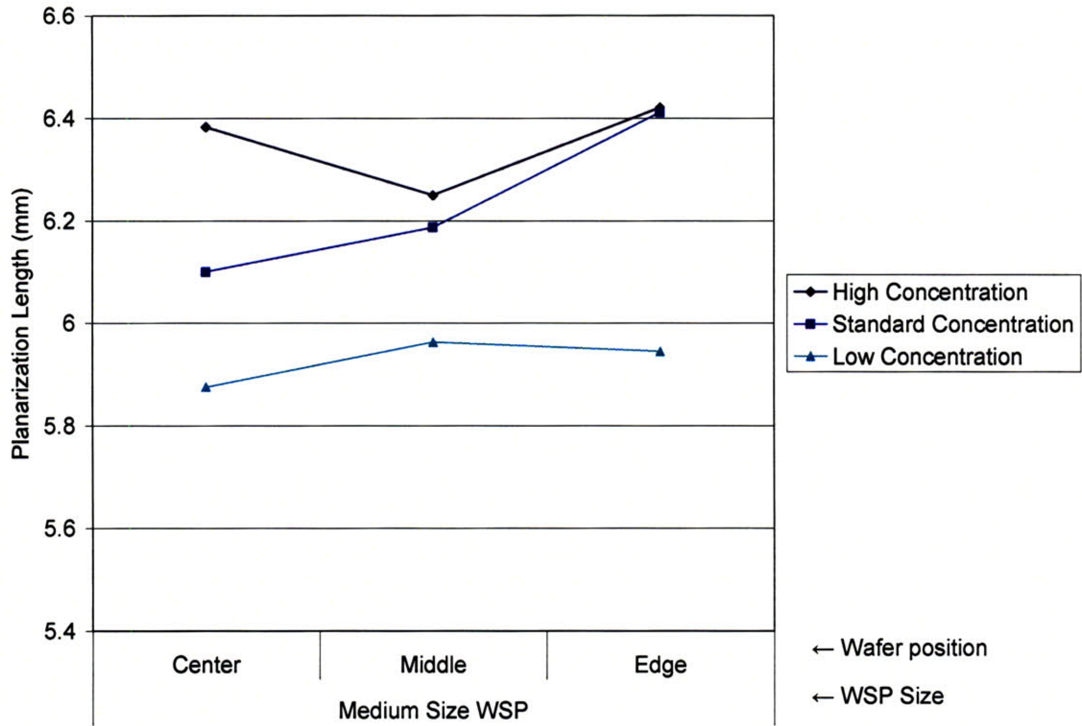


Figure 29: Planarization length across wafer locations for the second experimental set.

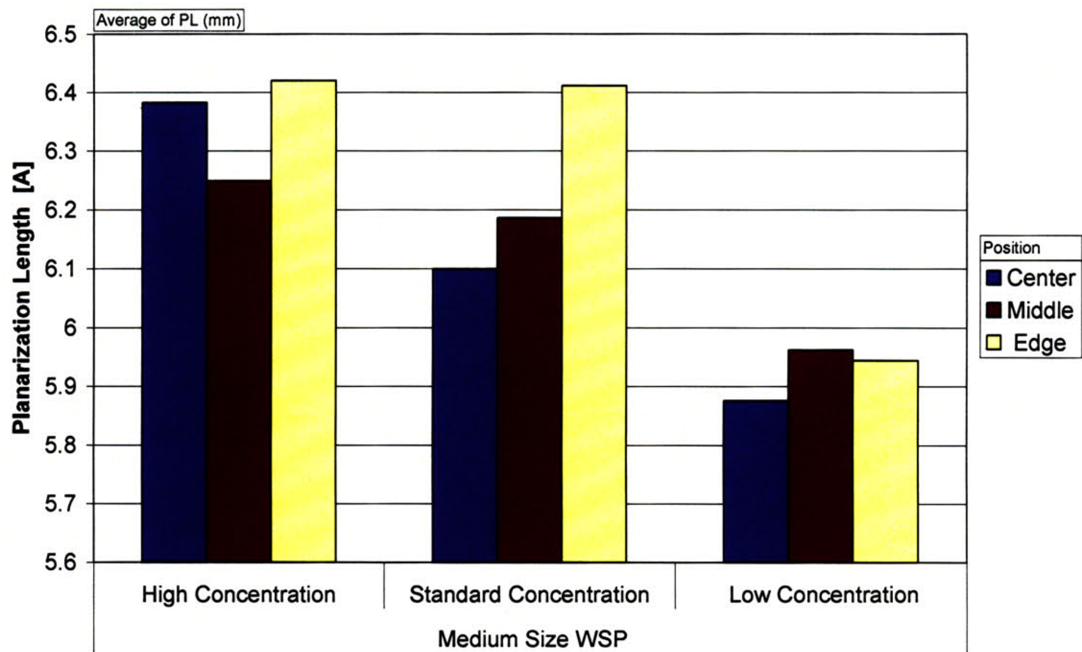


Figure 30: Planarization length for all grit and WSP concentrations tested.

2.3.5.5 *τ and characteristic height*

Figure 31 and Figure 32 show the behavior for τ. The first plot shows that the values are mostly constant across the wafer except for a drop at the edge. The second plot shows that low WSP concentration is beneficial to increase τ. Changing the concentration from high to low can increase τ by about 15%. Recall that τ is an exponential rate parameter, so a small change can potentially translate into a large difference.

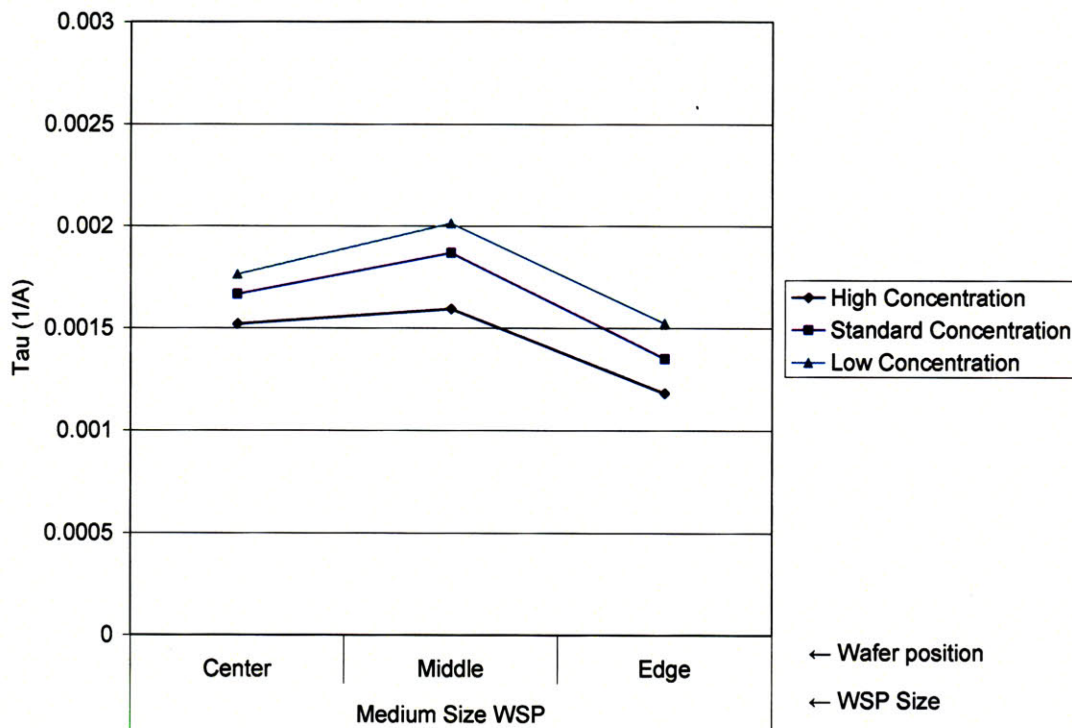


Figure 31: τ values across wafer locations for the second experimental set.

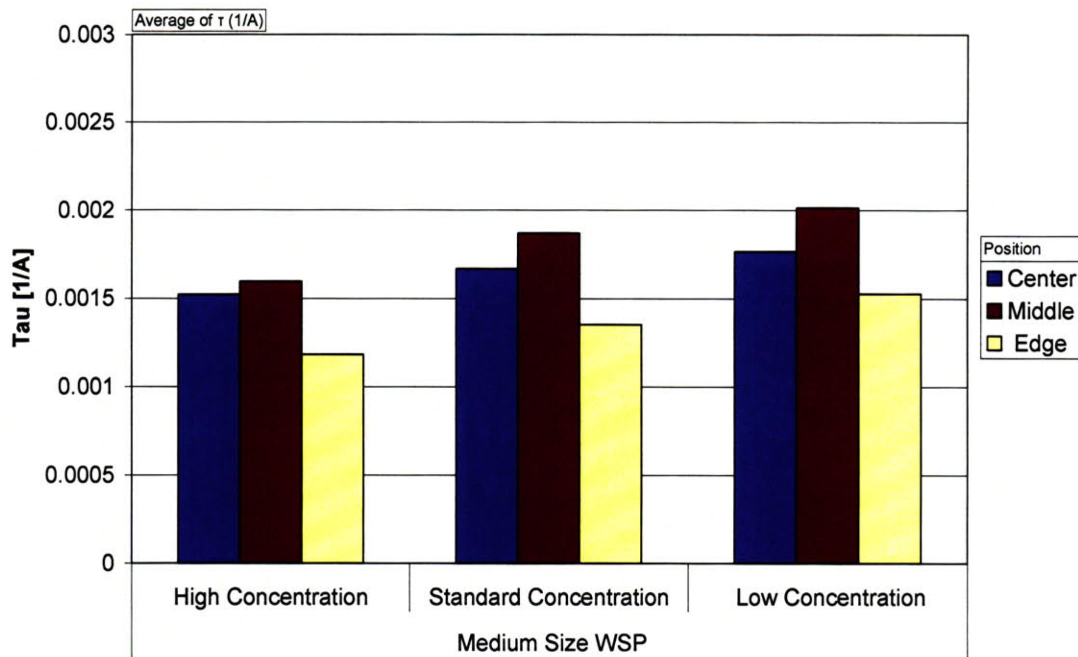


Figure 32: τ values for all grit and WSP concentrations tested.

2.3.5.6 Fitting Error

The fitting error for the second set of experiments is low overall; lower than 300 Å for center and middle measurements, and lower than 400 Å for edge measurements, as shown in Figure 33. The edge measurements usually have a higher error due to the model's inability to capture the edge effects present in the process. Figure 34 shows that the low WSP concentration pad had the highest error, probably because it is a softer pad and therefore the surface topography will be more vulnerable to feature level effects that the model does not account for. Even so, the error seems relatively low.

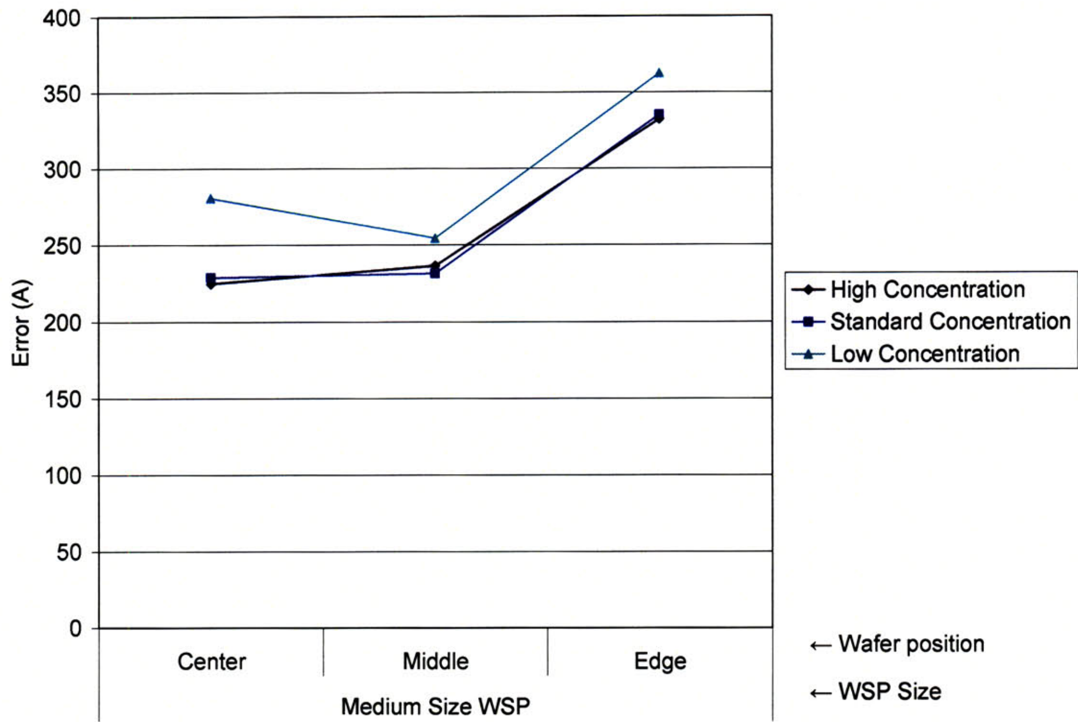


Figure 33: Model error across wafer locations for the second experimental set.

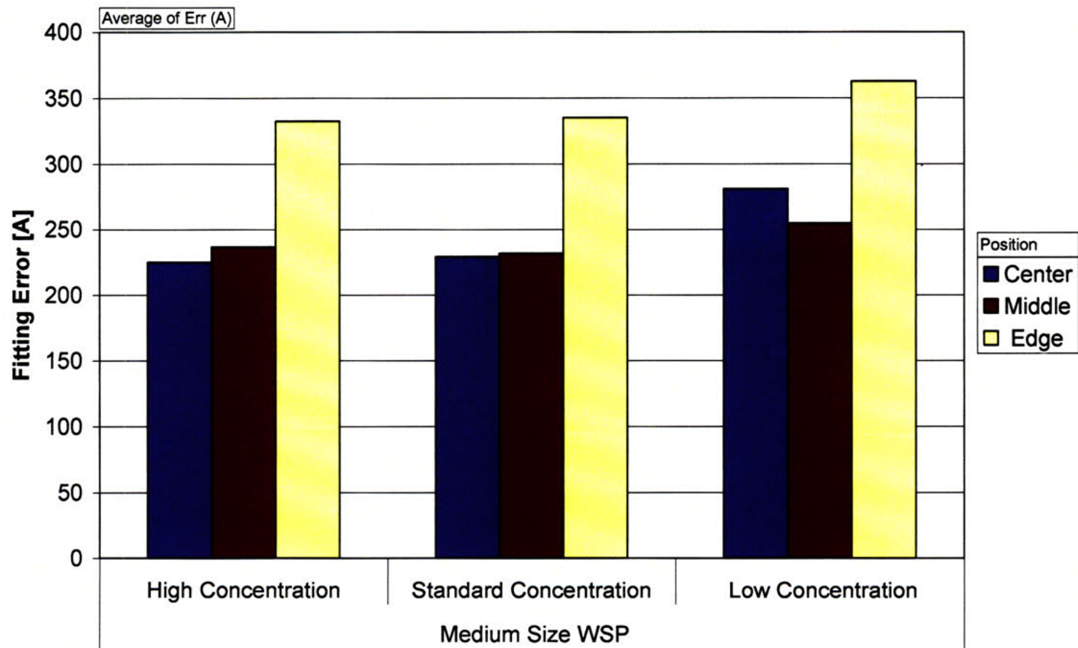


Figure 34: Model error for all grit and WSP concentrations tested.

2.3.5.7 Trend summary for the second experimental set

Overall, the second experimental set shows some valuable trends. First we review the parameters and their effects in a general sense. A greater τ implies a lower characteristic contact height $h_c = \frac{1}{\tau \cdot \rho}$, which gives a more effective or rapid planarization since the pad mostly polishes on the up areas. Ideally, the least amount of material to be removed to achieve planarization is only the up areas. Removing material from the down areas actually counteracts and slows down planarization. Recall that planarization efficiency is defined as follows:

$$PE = \left(1 - \frac{\Delta down}{\Delta up}\right) \times 100\% \tag{4}$$

As the contact height becomes smaller, so does the amount removed in the down areas (inside the trench), and therefore the planarization efficiency increases. A better step height removal efficiency helps achieve a greater throughput by achieving a better time efficiency, which is dependent on the contact height as well as the blanket removal rate *K* and the planarization length *PL*. A larger planarization length helps to avoid within wafer non-uniformities. It helps avoid strong pattern density dependence, so that all densities polish roughly at the same rate, and overpolishing is avoided for less dense areas while more dense areas finish planarizing. So in summary we desire a high *PL*, a high τ and a high *K*.

Examining the data from our second set of experiments, the first and most important trend is the evident tradeoff between *PL* and τ. The results show that *PL* is positively correlated with WSP concentration, while τ is negatively correlated; this was attributed to the WSP concentration effect on pad stiffness. The second trend is a weak direct relation between *K* and the WSP concentration, which can be attributed to the larger amount of pores created on the pad surface by the greater amount of water soluble particles when dissolving in the slurry. The last trend is that using the #325 grit for the in-situ pad conditioning gives better results across all model parameters than using the #100 grit. Due to this, for the subsequent round of experiments only the #325 grit is used.

Choosing which parameter to prioritize is tricky. A higher *PL* is desired up to the point where it does not affect the pad's compliance to the wafer's underlying nanotopography; this is why pads are not made out of steel. Prioritizing τ by means of a lower concentration can slightly decrease *K*, which decreases the time efficiency. This is the part where the performance metrics previously established are helpful. The extracted model parameters for the middle wafer measurement sites for the #325 grit conditioned wafers are used to simulate the surface evolution of the whole test die. The plots from Figure 35 show the model output for the step height and film thickness evolutions of the whole test die area. The top two plots show the final surface topography at the end of the simulation after 200 seconds. The bottom two plots keep track of the surface evolution with time.

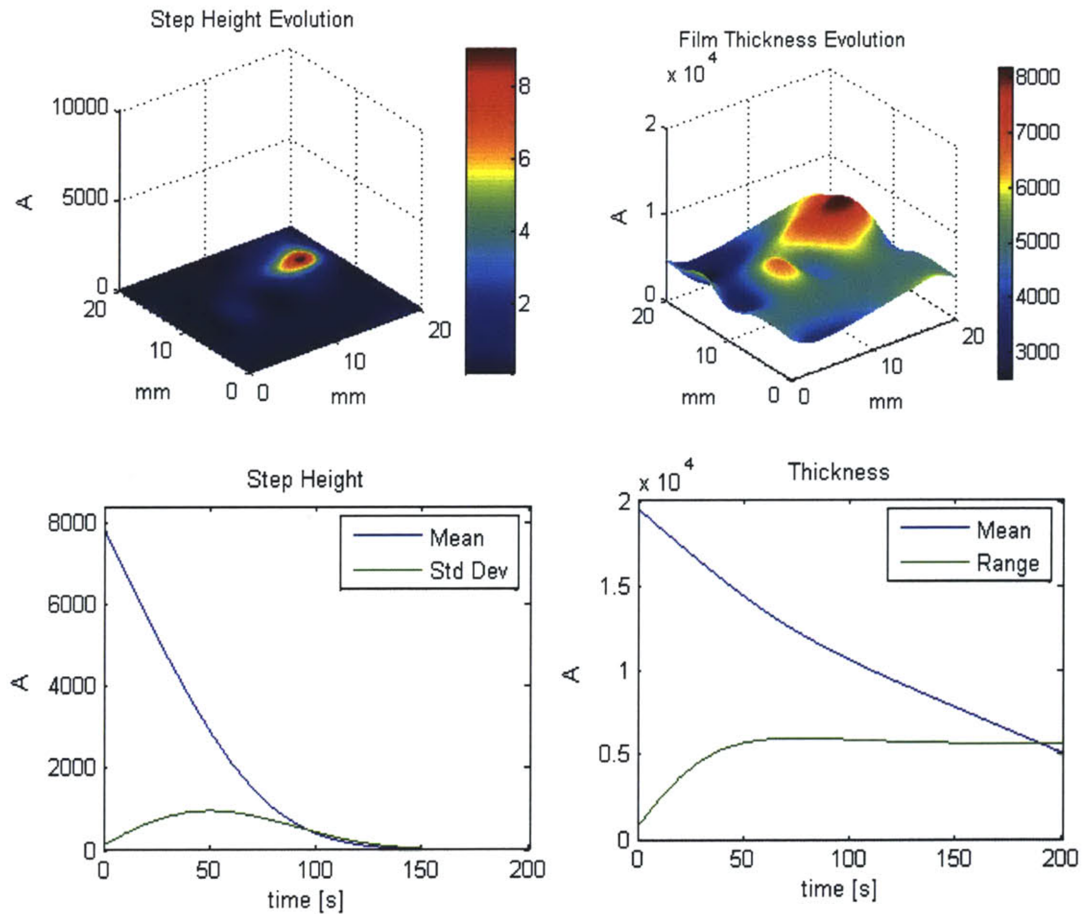


Figure 35: Model step height and thickness evolution plots. The chip-scale plots at top correspond to $t = 200$ sec.

An arbitrary target step height of 100 Å is chosen and the time to reach this value as well as the thickness range at this point is simulated. This is only done for the model based on the #325 grit conditioned pads, since their performance is clearly superior to those conditioned with the #100 grit. The results for these simulations are shown in Figure 36 and Figure 37. The first plot shows that the low concentration of WSP gives a faster planarization, in agreement with the increased τ value. The latter plot shows that the high WSP concentration results in a lower ending die thickness range (less die non-uniformity). Notice that the wafer level range follows the same trend but is larger due to the thickness variations across the wafer. This also agrees with the increased *PL* value at low WSP concentrations.

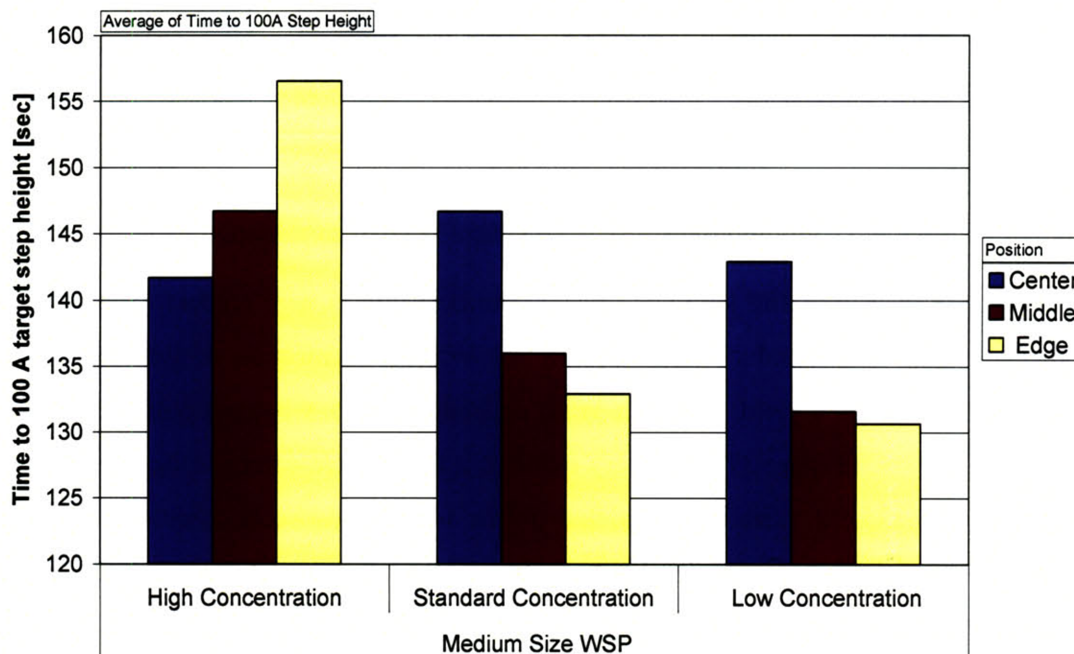


Figure 36: Time to reach step height target for #325 grit conditioned pads.

Despite the fact that different areas of the wafer planarize faster, the range does not change significantly in the areas that are *waiting* for the unplanarized areas to finish. This was previously shown in Figure 35 where the thickness evolution plot showed asymptotic behavior of the range after about a minute. The range variations for Figure 37 if plotted after all wafer locations reach the target step height are less than 25 Å.

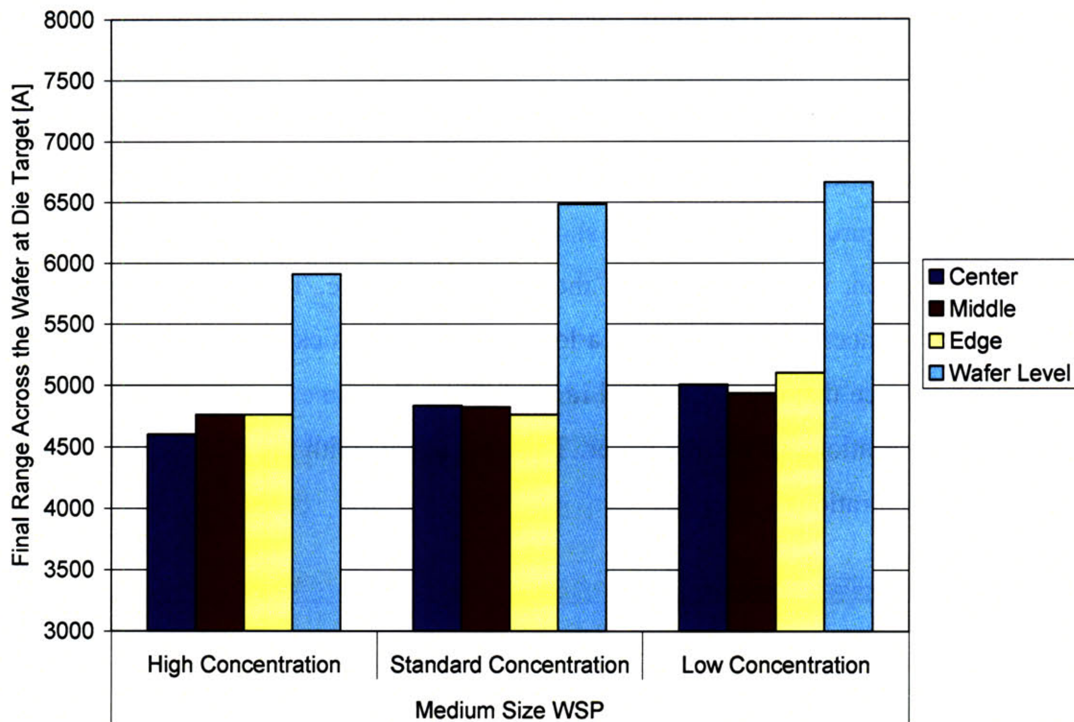


Figure 37: Ending thickness ranges for #325 grit conditioned pads.

From the measured and simulated results the best option with respect to throughput and lower cost of ownership would be the low WSP concentration. From the point of view of uniformity or yield, the high WSP concentration is a better option, though its benefit as shown in Figure 37 is not very evident. If the process can bear the die level and wafer level thickness variations depicted above for the low concentration, then this concentration would be the optimal option. If not, then the compromise solution is the standard concentration pad. The high concentration WSP pad has a serious drawback in that its time to reach the target step height is 10 more seconds than the standard concentration, and 15 more seconds than the low concentration. More experiments, or at least process tolerance margins, are required to find the optimal tradeoff concentration that would optimize the process, but these experiments and models help provide insight into the goals and tradeoffs.

2.3.6 Variable WSP size experiments

In this section, a third set of experiments is discussed. First, the setup conditions are described, with a short summary of the fitted model parameters. The next three subsections describe each of the parameters: K, PL and τ, followed by a section discussing the model fitting error. Finally, the last section summarizes the main trends observed in this third round of experiments.

2.3.6.1 Setup conditions and model fitting

The purpose of the third round of experiments is to understand the effect of the water soluble particle size on polishing performance. Three WSP sizes are used and will be referred to as small, medium and large. The medium particle size has a diameter three times larger than the small particle size, and one fourth of the large particle size. The particle size used for the previous sets of experiments was the medium size. For this round, only the #325 grit is used to condition the pad, since in the second round, the effects on polishing performance of the #100 grit were clearly inferior.

The wafers polished for this round are also SKW-7 test wafers as used in the first two rounds. A total of nine wafers are polished: one third is polished using a JSR pad with small WSPs, another third with the standard WSPs and the last third with the large WSPs. Each wafer has measurements performed on center, middle and edge dies. Within each die, there are profilometry measurements for all five pattern density blocks (10, 30, 50, 70, 90%) and optical measurements for the 30%, 50% and 70% density blocks only. The wafers are polished for two minutes with time splits and measurements at 0, 20, 40, 70, 90, 105, and 120 seconds.

The model is able to extract the three parameters (τ, K, PL) with an RMS error lower than 265 Å for center and middle wafer measurements. The edge measurements show a greater error, still below 350 Å. Taking into consideration that the initial wafer has a deposited oxide thickness of 2 μm, the model error is less than 2% of the amount removed. The extracted model parameters summary is shown in Table 5:

Grit	Concentration	Position	WSP Size	K (A/min)	PL (mm)	τ (1/A)	h_c (um) for p=50%	Err (A)
325	Standard	Center	Small	3182.3	5.944	0.002459	0.081	260.6
325	Standard	Middle	Small	3452.7	5.944	0.002587	0.077	262.7
325	Standard	Edge	Small	4005.7	6.012	0.001713	0.117	315.8
325	Standard	Center	Medium	3437.0	6.100	0.001668	0.120	229.0
325	Standard	Middle	Medium	3621.3	6.186	0.001870	0.107	231.6
325	Standard	Edge	Medium	3841.0	6.411	0.001352	0.148	335.2
325	Standard	Center	Large	3372.9	6.180	0.001438	0.139	243.9
325	Standard	Middle	Large	3480.6	6.199	0.001572	0.127	237.9
325	Standard	Edge	Large	3681.6	6.549	0.001184	0.169	367.2

Table 5: Extracted model parameters for the third experimental set.

2.3.6.2 Blanket Removal Rate

The following plots summarize the trends observed for the blanket removal rates. Figure 38 shows that, just as in the variable concentration experiments, there is once again a clear radial dependence shown by *K*. The small WSP size shows the largest variation, and the large particle size has the smallest wafer level variation.

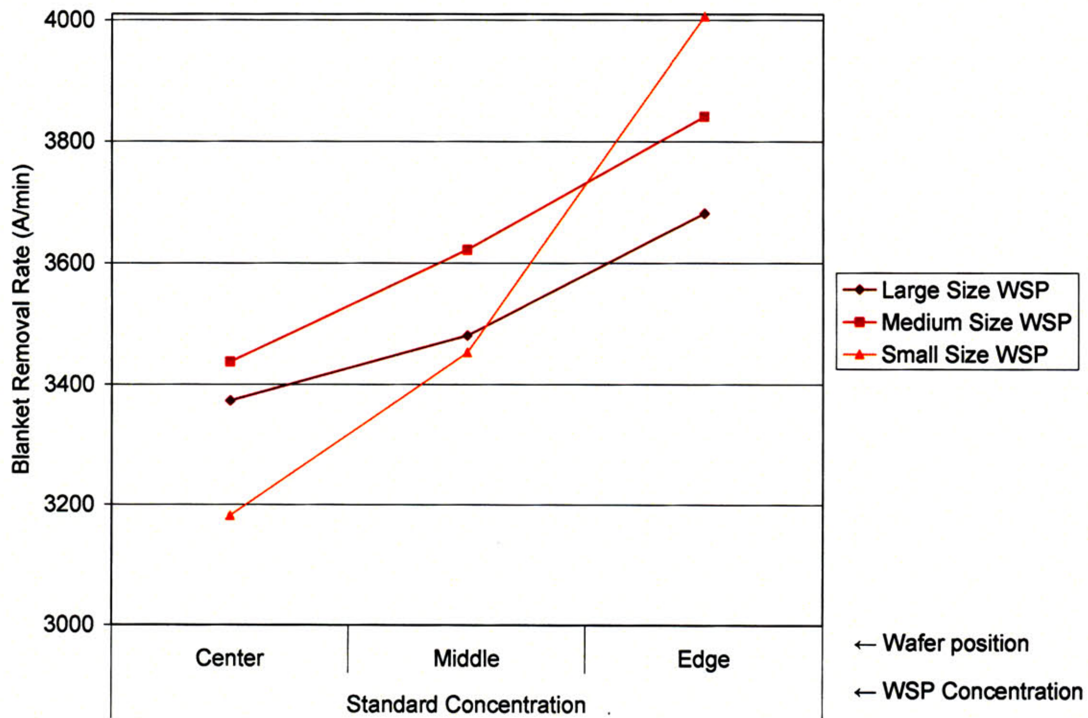


Figure 38: Blanket removal rate across wafer locations for the third experimental set.

Figure 39 shows that the medium size particles have the greatest removal rate, followed by the large particle size. Interestingly enough, the trend is not linear, but instead peaks somewhere near the medium size particle. Despite the fact that the small WSP has higher *K*, the large WSP has a more even *K* which will help to reduce die and wafer level non-uniformities. The blanket removal rate *K* is a parameter that is preferably high, but only as long as it is uniform or else it can be counterproductive. For this reason, in this case the large particle is a better choice, from the perspective of blanket removal rate and uniformity.

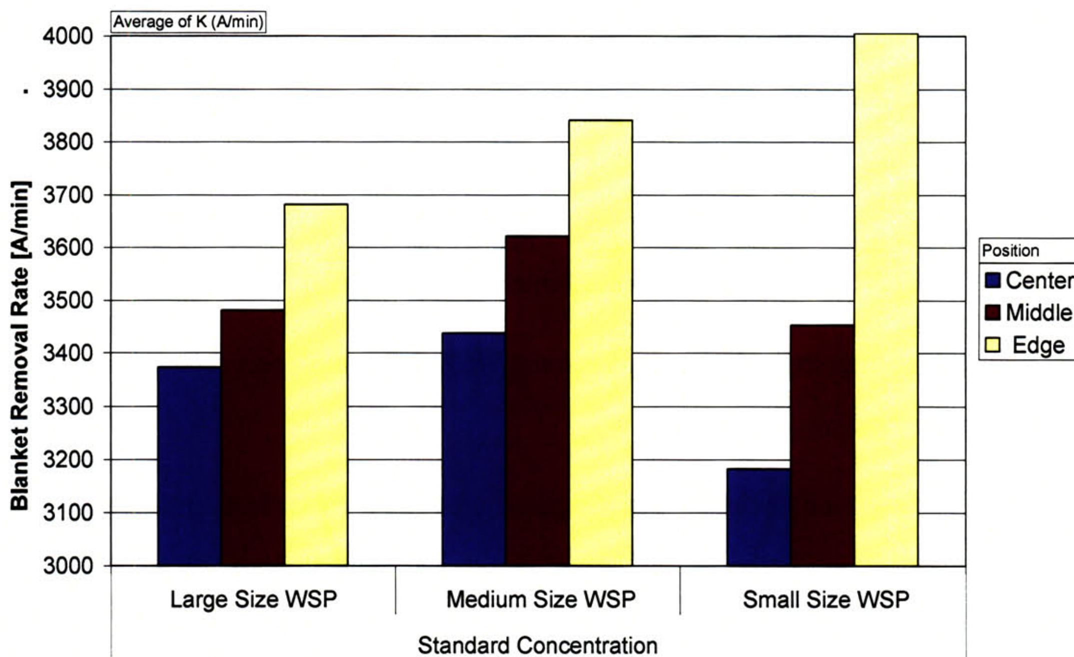


Figure 39: Blanket removal rate for different WSP sizes.

Recalling that the large particle has a radius four times as large as the medium size particle, the results would imply that if the particle becomes too large the performance benefit diminishes. Also the fact that the small particle is three times the medium one shows that once a critical size is reached, the removal rate goes up considerably (10% or so) and then slowly diminishes as size increases. Figure 40 shows the same data as before but normalizing the particle size on a continuous axis to give a better perspective of the suggested possible underlying trend. The number of points is limited, but it suggests that the maximum overall *K* can be achieved with medium size particles, unless there are more in between sizes available can provide a higher value.

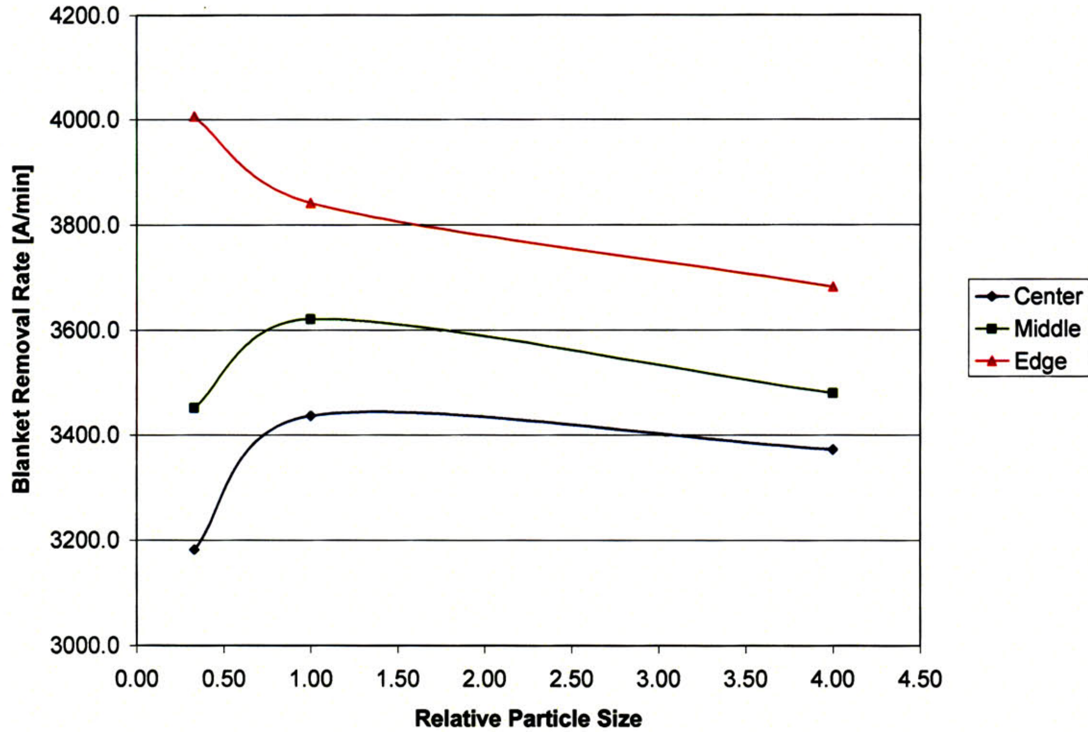


Figure 40: Suggested relationship between WSP size and blanket removal rate *K*.

2.3.6.3 Planarization Length

In this third set of experiments, once again there is an increase in *PL* as measurements move towards the edge of the wafer; except for the small particle which seems very uniform despite it having the largest variation in the case of *K*. This is shown in Figure 41. Regarding the different WSP sizes, Figure 42 shows that the small WSP offers the best *PL*, followed very closely by the large particle. The large particle's performance follows closely, showing *PL* values about 3% lower, except for at the edge. Figure 43 could suggest that gains in *PL* diminish as the particle gets larger. The best option from the perspective of planarization length would be the large size particle, unless the large *PL* variation towards the edge causes large wafer non-uniformities. This will be analyzed later with the performance metrics. Other constraints, such as how narrow and controlled can the particle size distribution be, might also play a role.

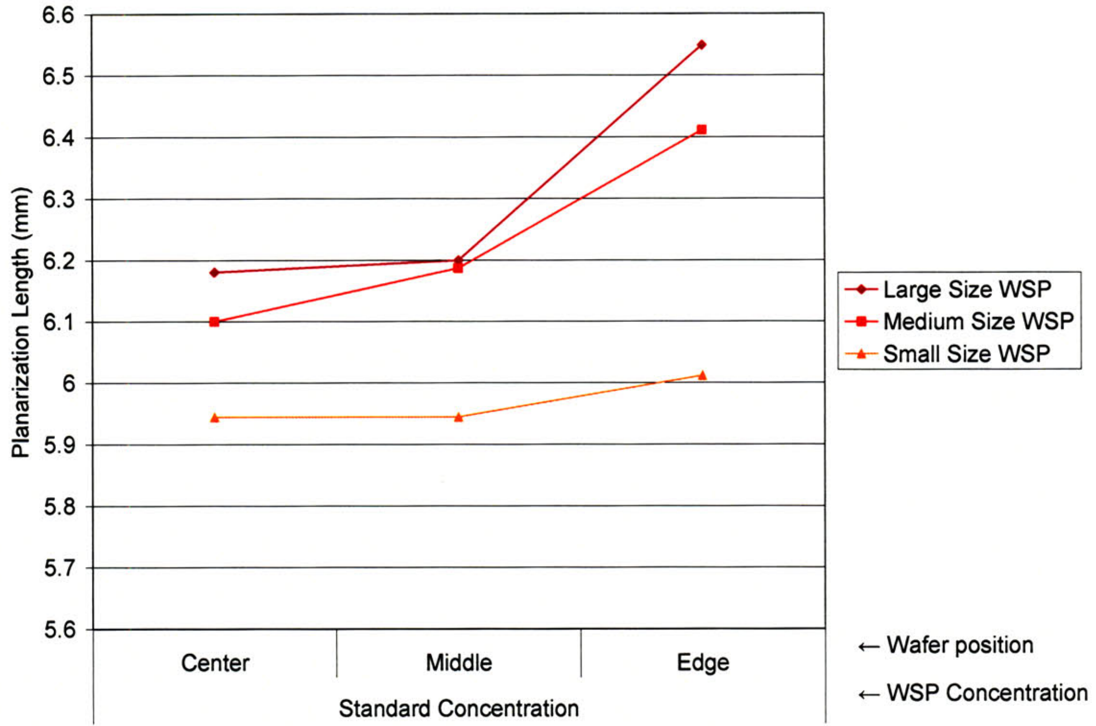


Figure 41: Planarization length across the wafer for the third experimental set.

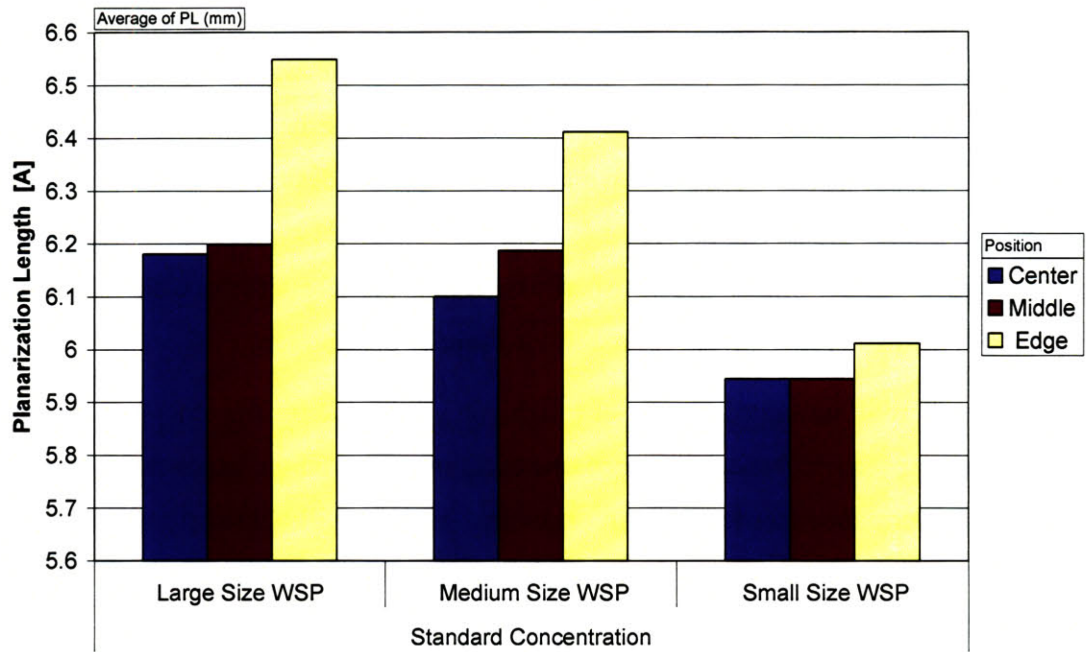


Figure 42: Planarization length for different WSP sizes.

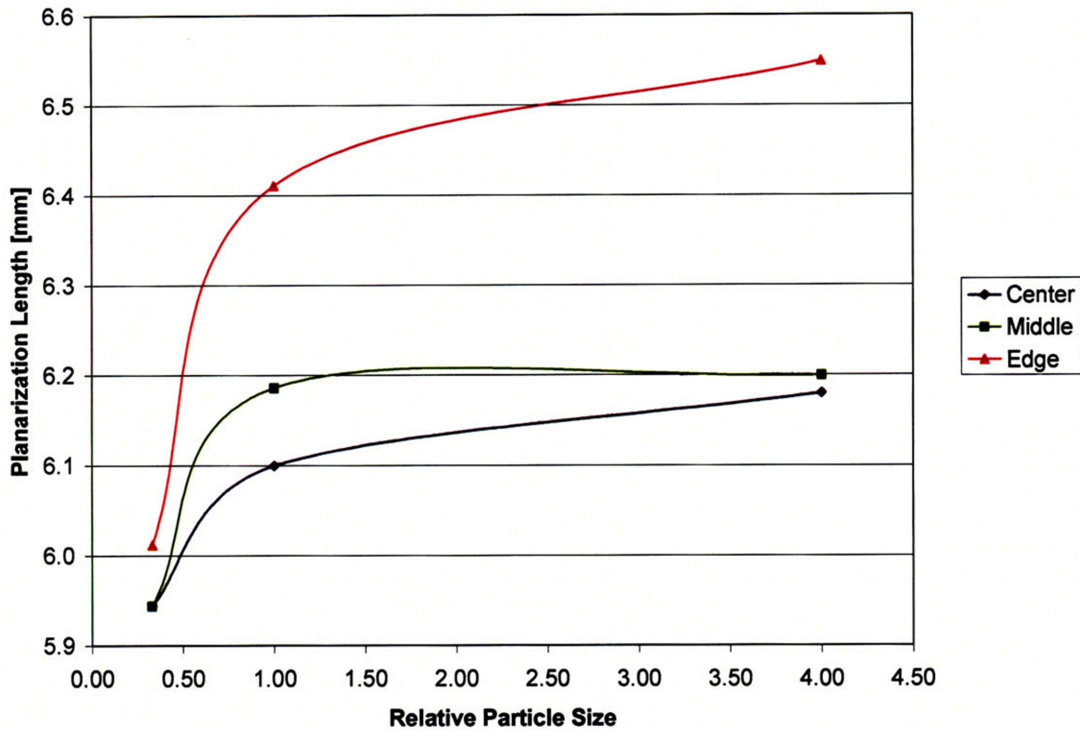


Figure 43: Suggested relationship between WSP size and *PL*.

2.3.6.4 τ and characteristic step height

Just as with *PL*, the τ value is almost constant throughout the wafer and then shows a substantial difference, in this case a drop, for dies near the wafer edge, as shown in Figure 44. It can also be seen that the small particle has the largest variability, though Figure 45 shows that smaller WSP give a larger and more desirable τ value. Figure 46 suggests that once the particles get large enough, their effect on τ is minimal; perhaps there exists a relationship between asperity size and particle size in which the pores created by the dissolved WSP skew or shift the asperity distribution.

The small WSP has an advantage of roughly 2/3 the value for the other sizes. This is an important difference, especially when one takes into account the exponential behavior of τ. The performance metric charts in upcoming sections will help sort out whether or not the *PL* and *K* advantages of the large WSP offset the advantageous τ value for the small WSP.

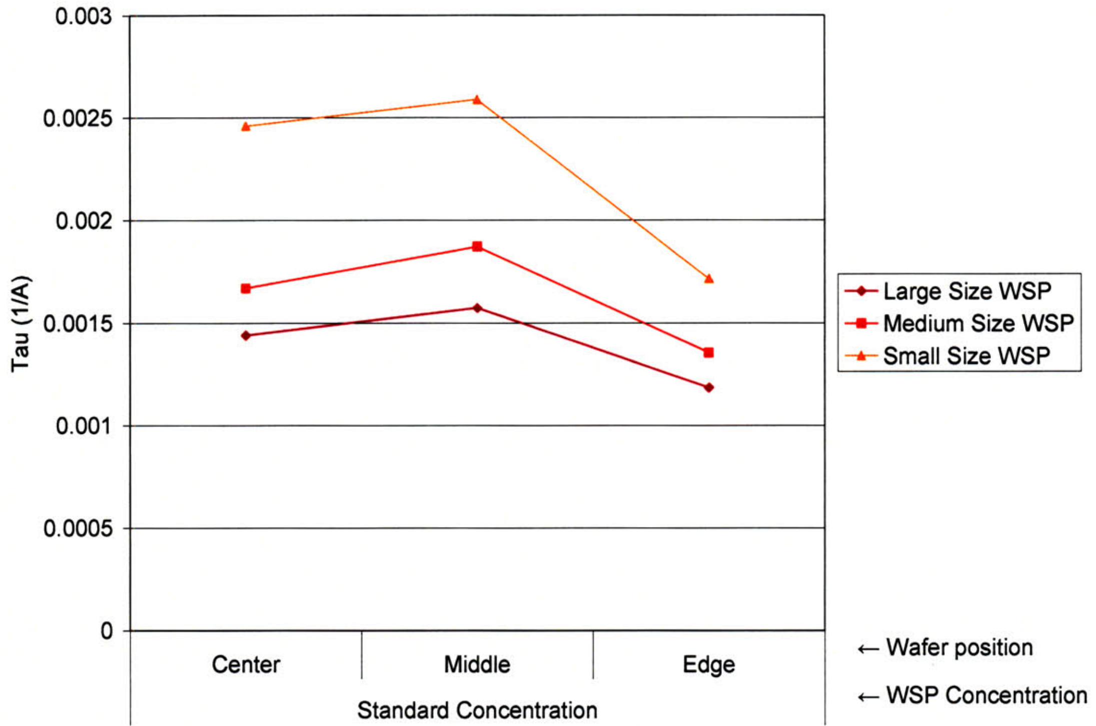


Figure 44: τ across wafer for the third experimental set.

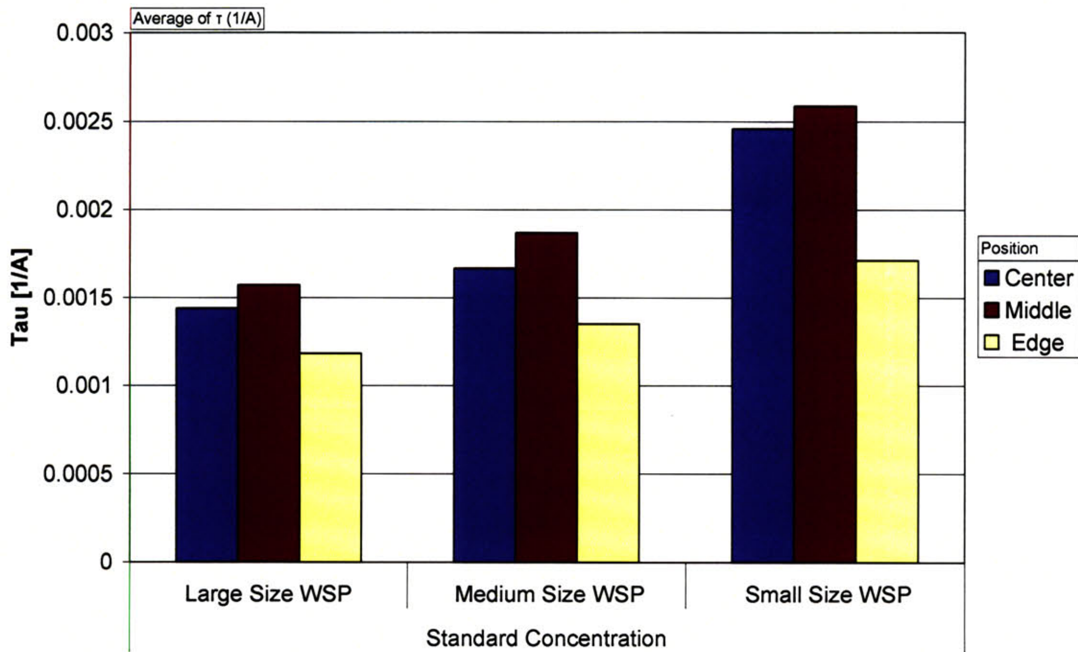


Figure 45: τ values for different WSP sizes.

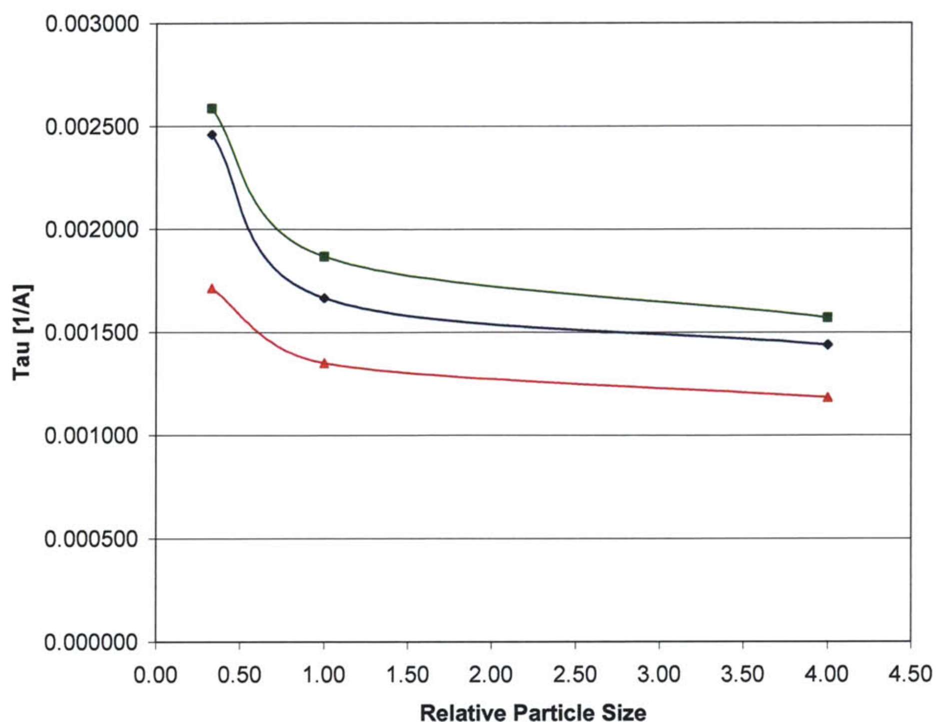


Figure 46: Suggested relationship between WSP size and τ.

2.3.6.5 Fitting Error

Once again, just as in the variable concentration experiments, the fitting error is around or less than 250 Å for the center and middle measurements, and around 350 Å for the edge measurements. These are shown in Figure 47 and Figure 48.

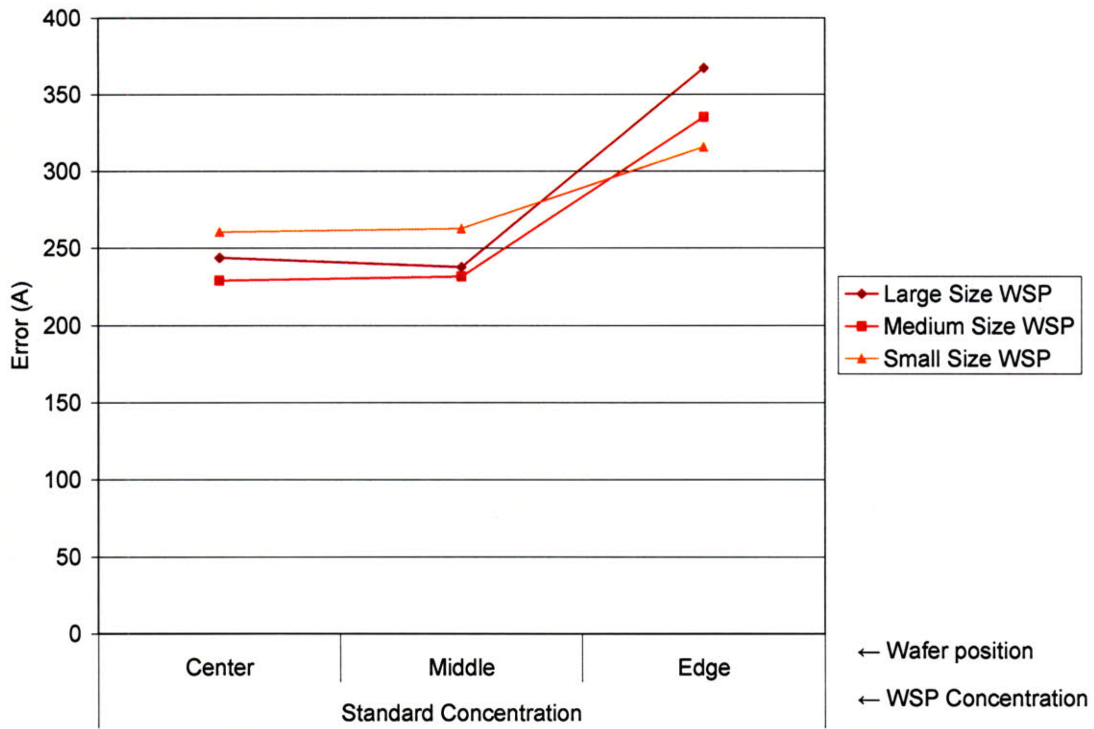


Figure 47: Fitting error across the wafer for the third experimental set.

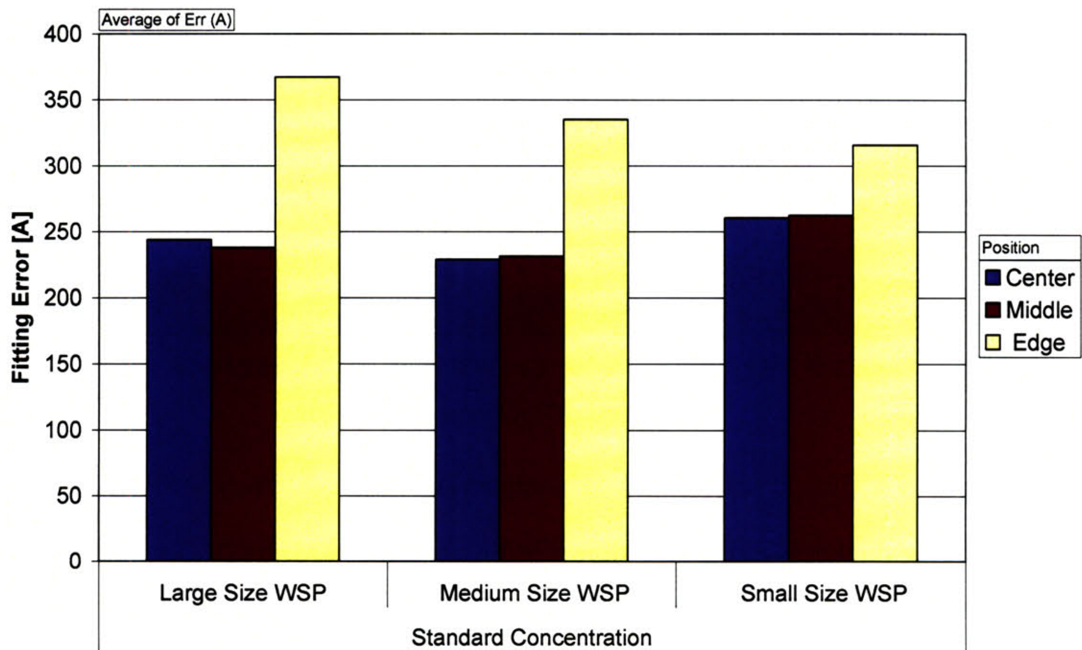


Figure 48: Fitting error for the third experimental set.

2.3.6.6 *Performance metrics*

A careful study of the previously established performance metrics reveals that the large particle size appears to offer a good tradeoff overall. Figure 49 shows that the small particle pad planarizes the fastest, but at the expense of large wafer level variability. At first this might seem challenging but recall from Figure 35 that the range stabilizes after about a minute of polishing and then stays fairly constant. Unless wafer level thickness control is an issue for the process, then this behavior won't be problematic. The only issue with this time differences is that it is useless that the small particles can planarize the edge in less than 125 seconds, if it needs to wait 20 more seconds for the center to be planarized, so that its speed benefits are neglected by its large within wafer nonuniformity. In the end, all three particles size pads take roughly the same 145 seconds to completely planarize the entire wafer. From this plot it can only be concluded that the small size has a minimal 5 second advantage over the other sizes.

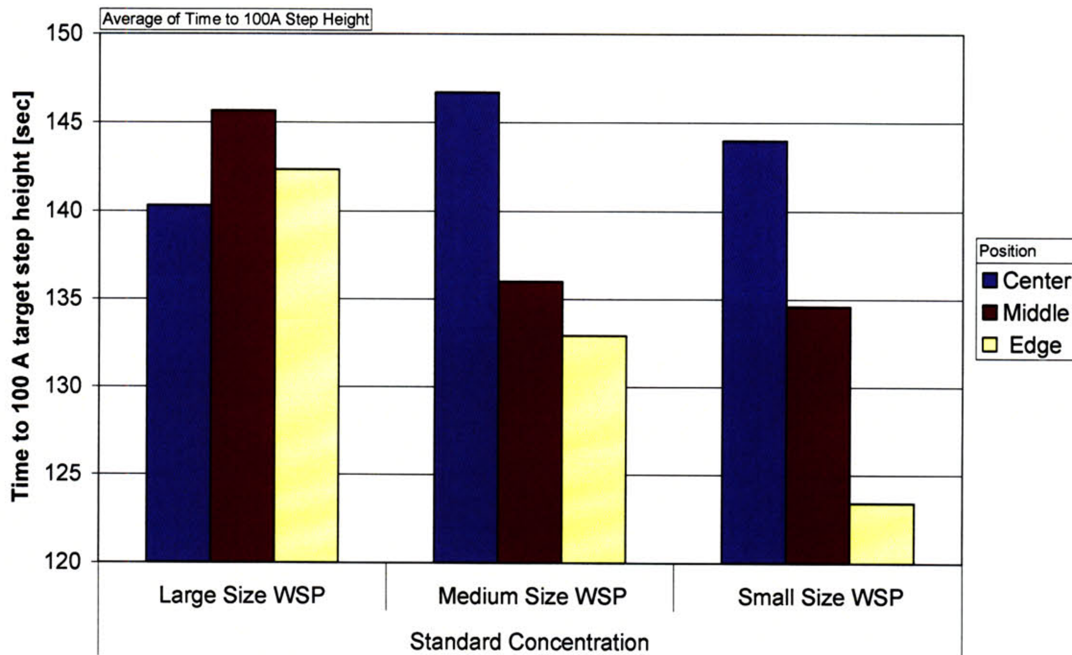


Figure 49: Time to reach step height target (200 Å) for different WSP sizes.

The ending thickness ranges are shown in Figure 50. The within wafer range is very similar for all particle sizes, slightly lower for the large particle. The wafer level range is about 1500 Å lower for the large particle. It could be the case that this thickness variation is not important for the process, in which case any particle size would work.

The best option overall for the third experimental set is the large particle size pad. Other factors not included in this analysis, such as cost, reproducibility or complexity of making the pads with the different WSP sizes might change this decision.

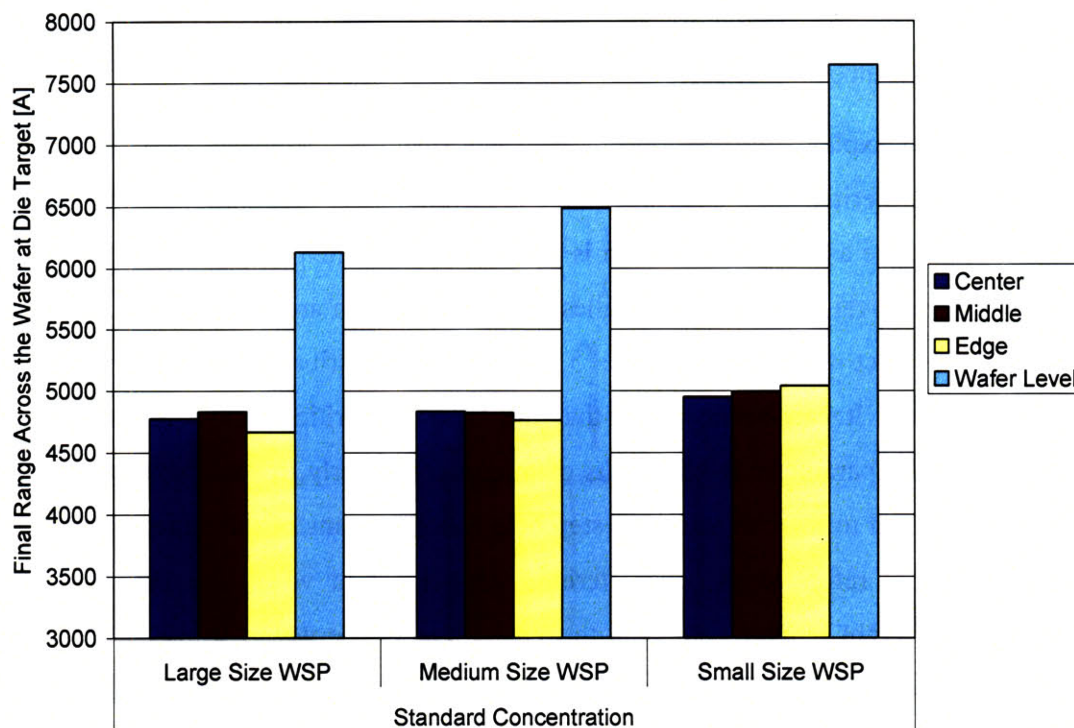


Figure 50: Ending thickness ranges for different WSP sizes.

2.4 Physical modeling

Many physical models have been presented to model CMP. Of particular interest is the physical model presented by Xie [29]. In his model, Xie reconciles the previous Step Height / Pattern Dependency model with physical parameters for the polishing pad. While the SHPD model uses *PL*, *K*, and τ, the physical model uses *E* and λ, the pad's Young's modulus and tall asperity height distribution coefficient, respectively. These parameters are closely related with one another. The *PL* goes hand in hand with *E*, and τ behaves similarly to λ. A stiffer pad will have a longer planarization length resulting from its increased Young's modulus. Experimental Young modulus are made for the JSR pads and they follow a linear increasing trend with the WSP concentration. The particles are harder than the elastomer the pad is made of, so it is natural for the pad to get stiffer as

the WSP concentration increases. The pads with the different particle sizes had almost constant experimental Young modulus values, which makes sense since all these pads had the same WSP concentration by weight, not by volume. Since all particle sizes are made of the same material, the density is the same and therefore the total mass of WSP material in the pad is the same for all pads with different WSP sizes. This could explain why the performance metrics were so close for the third set of experiments.

In order to explain the relationship between λ and τ, the meaning of λ must be clarified. The asperities on the surface of a pad have a height distribution. When the pad is conditioned, the asperity distribution is narrowed, which helps have a more even pad. Most of the removal is made by the tallest asperities, which are the ones tall enough to reach and contact the wafer. Figure 51 shows the distribution of asperities of a conditioned pad. It can be observed that the tallest asperities follow an exponential distribution with parameter λ, with units of microns. Similarly, it was mentioned before how $1/\tau$ is related to the characteristic step height h_c by means of the pattern density. In this fashion, $1/\tau$ refers to the characteristic step height at which the asperities with characteristic height λ make contact with trench areas of the wafer. In this sense, a pad with a wider exponential tall asperity height distribution (hence a larger λ) will have taller asperities that will contact the pad and start removing material in the trench. In the opposite scenario, a lower λ means fewer tall asperities that will cause that unwanted behavior. Ideally all asperities would be the same size so that the pad is perfectly flat, with a very small λ value. Borucki et al. [30] showed that the asperity distribution is dependent on the conditioning used. It was previously mentioned in our experiments how different grit sizes provide different levels of roughness to polish the pad; therefore by changing the conditioning, the pad polishing performance can be enhanced without the need of engineering a different pad.

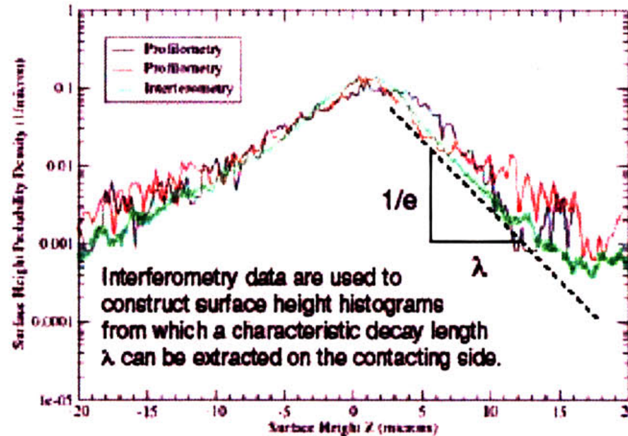


Figure 51: Exponential distribution of conditioned pad asperities [30].

Further experiments can be performed to see how the JSR pad responds to a wider set of conditioning grits, and asperity height distribution measurements can be performed to confirm the effect on the pad. A tradeoff can be made between enhancing the pad's physical composition and enhancing the way in which is conditioned. An alternate approach is to enhance the pad's composition to respond favorably to a given conditioning.

2.5 Summary

A first set of experiments shows that the JSR pad containing water soluble particle is superior to the most commonly used pad, the IC1000. A second set of experiments observes the effect of two different pad conditioner grit sizes and three WSP concentrations on the pad's polishing performance. The third experimental set compared three WSP sizes and their effect on polishing performance.

The #325 grit, widely used in industry, is the best choice for pad conditioning. Overall, the best WSP concentration and size combination from the perspective of planarization efficiency, pattern density response and wafer uniformity, is the standard concentration of large-sized particle. A more optimal size would probably lie between the large and medium sizes, but more experiments would be needed to conclude the exact size. The advantage of a larger particle lies in the reduced die level and wafer level thickness non-uniformities. The results match the intuition provided by the physical modeling and measurements.

Chapter 3 *ECMP Model and Simulations*

Currently in industry, either regular CMP models are used for ECMP or empirical models that vary from tool to tool are used. A physical model will fill in this void and help process engineers as well as tool designers to understand the process dynamics better and along with existing empirical models learn how to optimize it. The proposed model is a dynamic wafer level physical ECMP model based on the electrochemical properties of the ECMP.

This chapter is divided into five sections. The first section shows an empirical model fitted to data from the same ECMP tool in which the physical model will be based on. The second section describes the physical model and how it is implemented. The third section will compare the physical model's performance against an empirical model provided by Applied Materials. The fourth section discusses the applications of the model. Finally the fifth section outlines future possible model enhancements.

3.1 Empirical modeling

In order to gain more insight into the process and have another reference point aside from AMAT's empirical model and to further validate the proposed model, a full factorial design of experiment with four replicates per run is performed. The runs include the combination of voltage setting in each of the three zones as shown in Table 6.

V1	V2	V3
3	0	0
0	3	0
3	3	0
0	0	3
3	0	3
0	3	3
3	3	3

Table 6: Full factorial ECMP experiment.

At each experimental point, two wafers are polished under the same conditions, and 240 radial measurements of copper thickness are taken. The 240 measurements are

composed of two measurements per radial distance from the center, giving two replicates per wafer for a total of four replicates at each spatial radial position. This allows for a simple within wafer and wafer to wafer variation analysis of variance. From the data, 120 multivariate regressions are run to give 120 site models, one for each radial point measurement. These models are plotted against the Applied Materials model and the measured amount removed data in Figure 52.

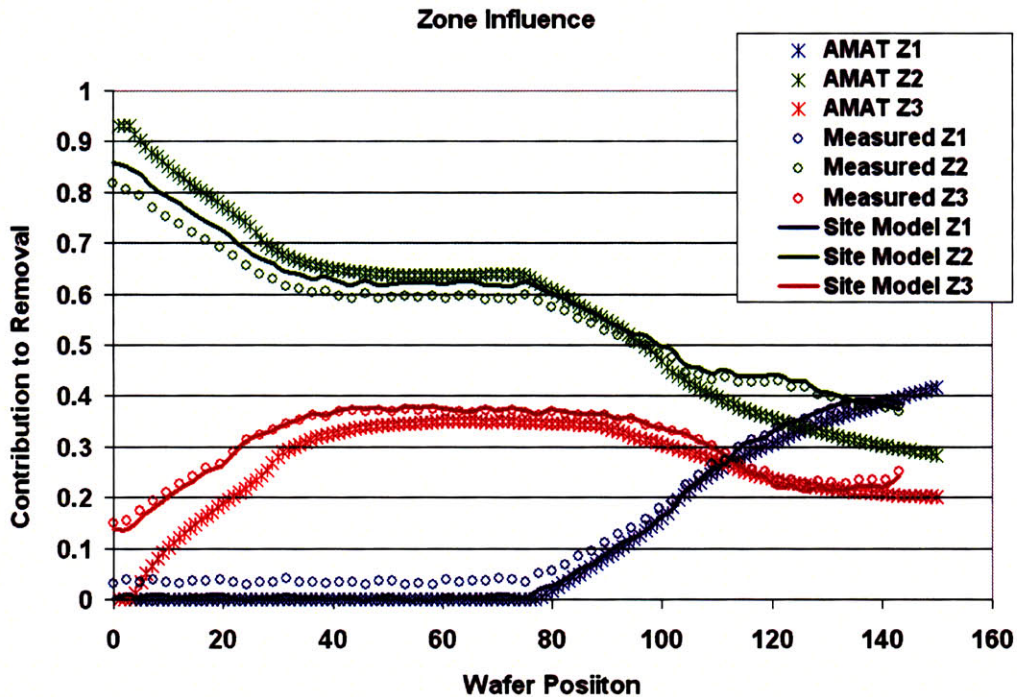


Figure 52: Empirical model vs. Applied Materials model.

The empirical model fit closely follows the removal data from which it was derived, indicating that the fit was good. The data overall follows the Applied Materials (AMAT) model. Despite the fact that the data came from the exact same tool model, an AMAT Reflexion LK ECMP, the subtle differences might be due to tool to tool variation. The data was acquired at IBM's Albany Nanotech Facility, while AMAT's model was developed at IBM's Fishkill facility. This emphasizes the need for both physical and empirical models to have appropriate model parameters that can be fine-tuned, depending on the particular tool state, to achieve optimal control.

3.2 Physical Model Description

A physical model has been implemented for wafer level removal in an ECMP tool. The model is a time-stepped finite element approach to calculate copper removal during ECMP. The model can be modified to fit different tool setups, but the present model has been constructed for the Applied Materials Reflexion LK ECMP tool. In this setup, the conducting polishing pad has three concentric voltage zones that affect the removal rate in different regions of the wafer. These zones can be tuned to achieve a more even removal rate across the wafer. A central contact point on the pad, known as the bagel, provides a current path to ground; the bagel contacts the wafer at the edge. An electrolyte, composed of a diluted citric acid solution mixed with hydrogen peroxide and proprietary additives, flows between the wafer and the pad. The electrolyte solution provides conductivity, lubrication, additional oxidants (H_2O_2) and ligands to complex and dissolve the copper ions. The edge contact point will close the circuit and allow for current to flow and copper to be oxidized. In addition to the ECMP platen used to remove the bulk copper, a standard CMP platen available on the tool is used to clear the metal; a third platen is typically used to remove the barrier metal, usually tantalum lined with tantalum nitride. The proposed model only covers the first platen; the other two platens use conventional CMP processes that can be modeled by existing approaches [5]. The current delivered can be used to determine the endpoint, though usually a set time is used, and copper clearing is left to the next platen.

An empirical approach can be used and is available to control the process. The time and voltages used to polish to an even, planar surface can be calculated using a radial measurement table lookup model from Applied Materials, which takes the radial eddy current measurement of initial copper film thicknesses as input. The tabular model predictions provide a first order approximation, and fine tuning is mostly done through further empirical characterization. Our goal in this work is to develop a more detailed, physically based model for ECMP that can assist in process design, control, and tool optimization.

The key relationship of the model proposed here is between removal rate and the applied charge. This relationship is derived as shown below:

$$I[\frac{C}{sec}]\cdot\frac{1}{\text{area}[cm^2]}\cdot\frac{1atom_{Cu}}{2e^-[C]}\cdot\frac{1mole_{Cu}}{N_A\,atoms_{Cu}}\cdot\frac{63.546g_{Cu}}{1mole_{Cu}}\cdot\frac{1cm^3}{8.941g_{Cu}}\cdot\frac{10^7nm}{1cm}\cdot\frac{60sec}{1min}=RR\left[\frac{nm}{min}\right]$$

$$area_{wafer}=706.85cm^2\rightarrow\frac{RR}{I}=29.93\frac{nm}{A\cdot min} \qquad (6)$$

The applied current is divided by the area to obtain the current density. Two electrons are needed to remove one atom of copper, so by using the molar weight of copper along with its density, the removal rate can be obtained. In the equation above scaling factors are used to obtain a result in nm/min. If the removal rate is divided by the current, then the removal per unit charge can be obtained. For a 300 mm wafer this constant is about 30 nm/(A·min), which means that for every minute that 1 amp of current is applied to the wafer polish process, 30 nm of average copper thickness will be removed. If we include a 2.5 mm wafer edge exclusion, the number increases to 31 nm/(A·min). As a check, a typical ECMP process applies around 15 A on average, giving a removal rate in the range of 4500 Å/min as typically observed in ECMP. An equivalent derivation to that presented here was shown by Smekalin [25] using the copper crystal lattice constant, giving similar results of 30.8 nm/(A·min) for 2.5 mm edge exclusion. The actual removal rate is expected to be larger due to grain boundaries in a multi-crystal structure. There are also etching effects that increase the rate, mainly due to the oxidizing action of the H_2O_2 added to the electrolyte. The model can be modified by adding a constant removal rate component to account for mechanical and chemical etching rates. Liu [35] reports that this constant rate is less than 200 Å/min; other data shows values as high as 450 Å/min.

In the proposed model, the current density distribution is calculated by means of a 3D discretization of the wafer and electrolyte, using layered resistance meshes as shown in Figure 51. The first mesh level represents the copper layer, composed of the deposited copper on the wafer. The nodes in this copper mesh are connected to respective nodes in similar meshes that sit on top, representing the electrolyte film. The electrolyte level is made up of *n* mesh levels. The additional electrolyte meshes are of possible interest, to better model lateral as well as vertical current paths from the pad through the electrolyte and onto the wafer. This also allows for modeling of the smooth transitions between pad voltage zones; if no levels were employed, the pad voltage zone transitions would be abrupt and non-physical. The topmost electrolyte level connects to the voltage zones of

the pad, which are treated as ideal voltage sources with no resistive losses across their surface.

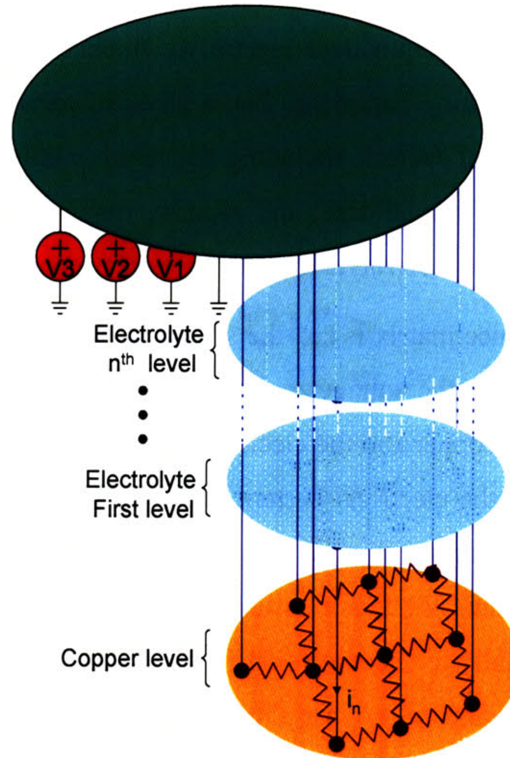


Figure 53: Model mesh levels.

The model algorithm computes a single conductance matrix that represents all layers. This matrix is systematically assembled by using patterns observed in the Kirchoff's Current Law equations for all nodes. Separate matrices keep track of the each node's position, copper thickness, which and how many neighbors the node has, and in which layer the node is located. Based on this information, connecting conductances between neighboring nodes are assigned. All horizontal conductivities in the copper level are the same, but the conductances vary according to the average thickness between two nodes such that $g_{cu} = \frac{\sigma_{cu}}{t}$, where σ_{cu} is the conductivity of the annealed copper after electroplating and t is the copper thickness. For the other electrolyte layers, conductances are derived according to the cross sectional areas and thickness given by their position relative to the current path. This means that horizontal conductances vary as $g_h = \frac{\sigma \cdot h}{n}$

and vertical conductances bridging two layers vary as $g_h = n \cdot \frac{\sigma_{elect} \cdot d^2}{h}$, where n is the number of mesh levels used, d is the discretization size used, σ_{elect} is the electrolyte conductivity, and h is the equivalent electrolyte thickness. The equivalent electrolyte thickness is more of a fitting parameter than a physical parameter. Its physical meaning encompasses a number of factors, including the relative hole area of the polishing pad with respect to the pad's total area, the asperity height distribution of the pad, and assumed infiltration of the electrolyte into the pores between the pad asperities.

Once the conductance matrix is set, the system is solved for the voltage drops across the mesh system based on the voltage zones applied, the bagel contact point (ground) and the conductance of the copper and the electrolyte layers. Since calculations of the amount removed are based on the electrical current, the modeling of the dynamic current is the interesting part. This is done by using the calculated voltage drops and conductances to find the current going into the wafer assuming the wafer and pad are static. The values are then averaged radially to account for the rotation of the pad and wafer in the ECMP tool; the rotation is assumed to be fast enough to allow for an even radial average in the given time frame of roughly a minute. These currents are then divided by the discretized area to obtain the current density per node and then using Equation

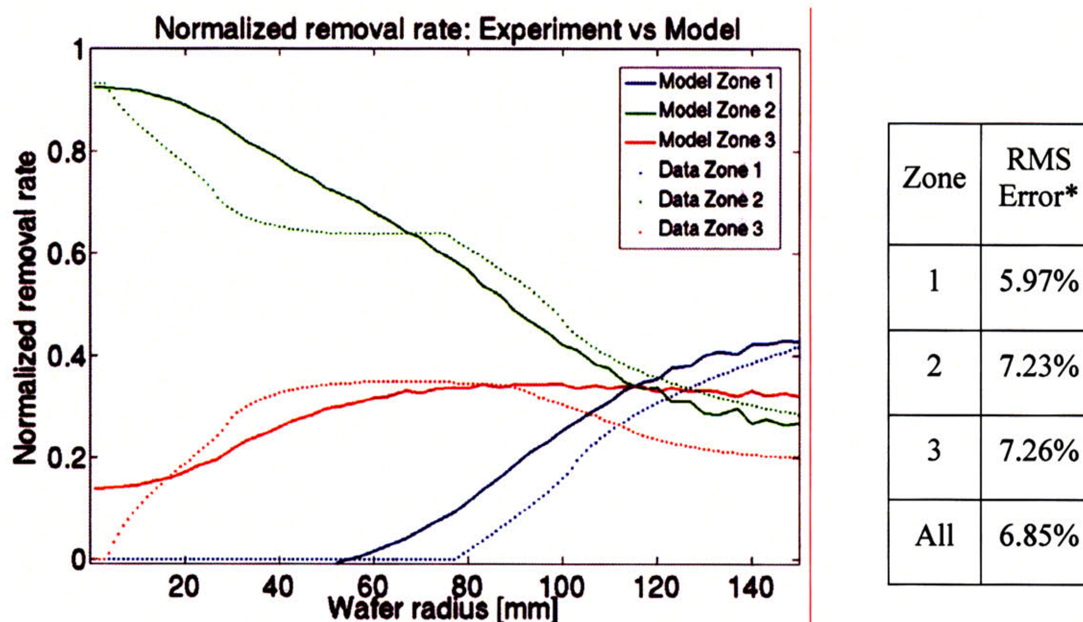
$$I[\frac{C}{sec}] \cdot \frac{1}{\text{area}[cm^2]} \cdot \frac{1 atom_{Cu}}{2e^-[C]} \cdot \frac{1 mole_{Cu}}{N_A atoms_{Cu}} \cdot \frac{63.546 g_{Cu}}{1 mole_{Cu}} \cdot \frac{1 cm^3}{8.941 g_{Cu}} \cdot \frac{10^7 nm}{1 cm} \cdot \frac{60 sec}{1 min} = RR\left[\frac{nm}{min}\right]$$

$$area_{wafer} = 706.85 cm^2 \rightarrow \frac{RR}{I} = 29.93 \frac{nm}{A \cdot min} \qquad (6),$$

the removal rate is calculated. This value is then multiplied by the time step to obtain the removal amount. The copper thickness map is updated and new conductances are calculated for every node. These are used to obtain the new voltage drops and current density distributions, to give the amount removed for the next time step. The model keeps iterating and updating until the full specified time interval is completed, giving the final copper thickness as a function of radial position on the wafer.

3.3 Experimental Data Fitting

The model is tuned to fit experimental data from the AMAT tool, by minimizing the sum of squared error between relative removal rate measurements and model predictions for varying values of the parameters for the electrolyte thickness and resistivity. The tuned model is able to capture the amount removed per zone on the AMAT tool; Figure 54 shows the model predicted normalized amount removed against the experimental data. The pad voltage zone 1 fit is closest to the data. Zone 2 is slightly overestimated, by about as much as zone 3 is underestimated. The fit for zone 3 deteriorates near the outer edge of the wafer, where the model is unable to capture the edge effects. The present



Zone	RMS Error*
1	5.97%
2	7.23%
3	7.26%
All	6.85%

* 91 x 91 mesh with 2 levels and 10 sec iterations

Figure 54: Normalized amount removed fit and RMS zone error table.

model assumes that the current always flows towards the wafer. In the real equipment setup, current from the electrolyte near the wafer edges may flow laterally into the wafer or to other sink points and account for the overestimate of the model near the edge. This can be explored in future model enhancements.

Figure 55 shows the output of the model for an arbitrary initial wafer topography with 3 V applied to zone 1, 1 V applied to zone 2, and 2 V applied to zone 3. The top left shows the initial copper thickness as deposited on a blanket wafer. The next plot on the first line shows the static voltage drop across the wafer, as induced by the voltage zones on the pad. This plot is followed by the effective voltage drop from the electrolyte to the copper, responsible for the copper dissolution. The first plot on the second line shows the radially averaged current density, which is based on the conductance of the electrolyte and the copper as well as the radially averaged voltage drop. The next plot shows the instantaneous removal rate, and finally the last plot is the copper thickness at the final time of 60 seconds.

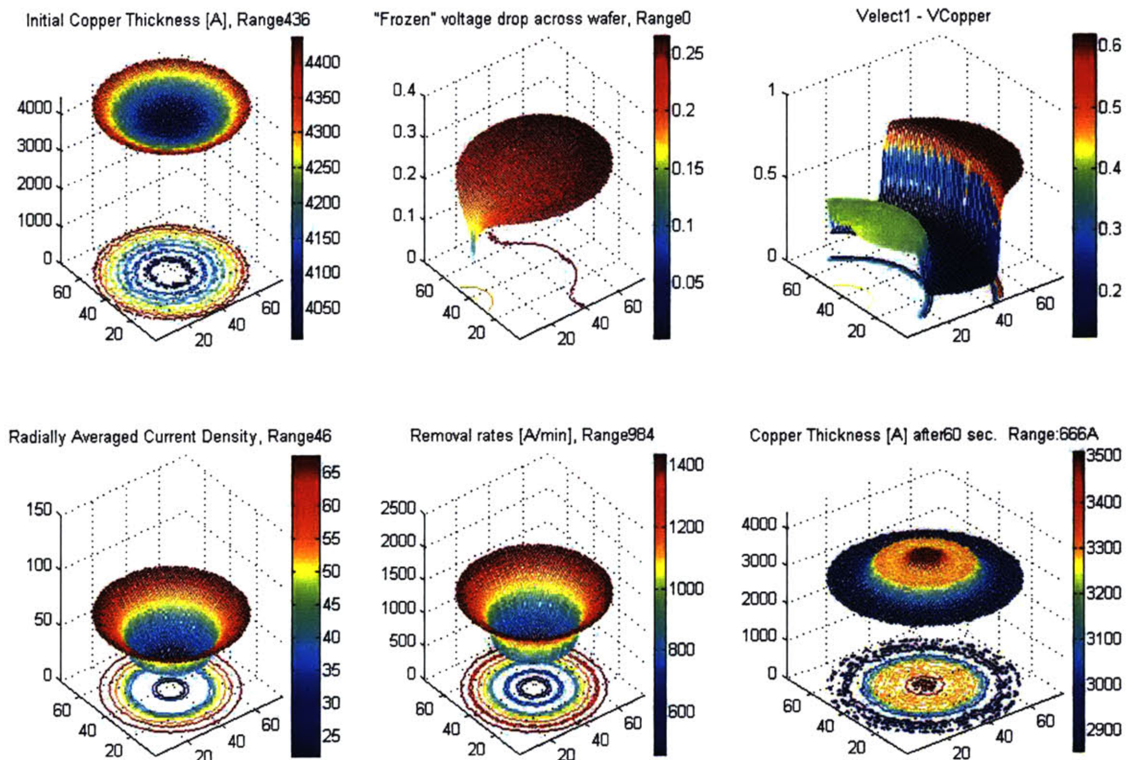


Figure 55: Model output: voltage drop, current distribution, removal rates, and final thickness.

3.4 Physical Model Optimization and Application

The ECMP model can be applied to a number of problems. It can accurately and effectively predict removal rates based on the voltage zones used. Alternatively, the model can be used to find the optimal voltage zone values for a given time and amount to be removed, in order to achieve the best wafer level uniformity. This approach could avoid the need to have separate time endpoints for each zone based on the charge delivered, and the model can be used to calculate the best applied voltage values to synchronize all zones, possibly increasing throughput. On the other hand, the model can also be used to estimate the time endpoints for each voltage zone based on the theoretical charge delivered given a set of voltage values, and eliminate or be used in conjunction with in situ endpointing measurements.

The simulations of the voltage drop and the current density distribution evolution offer insight that helps understand the fundamental mechanisms that drive the process. This allows for optimization of other parameters that can influence the output, such as the electrolyte chemical properties. In particular, the electrolyte conductivity can be expected to play a role in both the overall current density for a given applied voltage, and the spatial distribution of current across the wafer, including lateral current flow between voltage zones.

Finally, the model can be used to aid in tool design or optimization. In order to model rotational tools, like the AMAT Reflexion LK ECMP tool, the voltage zone locations can be changed and the possible benefit of more zones can be considered. In the case of orbital tools, such as the Novellus Xceda, the voltage zones can be varied as well, but also more ground contact points, or "bagels," can be added. The model can also be adapted to consider fully-conducting pad setups. The finite element model developed here provides a simple but flexible model structure that captures the key physical effects involved in wafer level uniformity and control in ECMP.

3.5 Model Possible Enhancements

The present model ignores a number of details that might be of value to include in the future. First, the details of the polishing pad structure used in ECMP could be considered. Macroscopically, the pad has a fraction of its area opened up into ~1 cm cells in which no polishing pad material is present, exposing the cathode. In these regions, the thickness of the electrolyte is approximately equal to the thickness of the pad. In the other regions on the pad, in contrast, the electrolyte is only present in the pad-wafer contact zone, where the electrolyte thickness of about 50 μm is set by the pad asperity and pore structure. These asperities also act as additional obstructions in the electrolyte film and increase the electrolyte effective resistance. The model can be extended to accommodate for the different open cell and pad contact regions, with different vertical and lateral resistances to account for the details of the pad pore structure. Modeling of any conductivity gradients caused by copper ion concentration changes in the electrolyte can also aid in this effort.

The model can also be coupled to pattern density and step height CMP models [8] to calculate the feature and die scale planarization behavior of ECMP, in which raised areas experience higher pressure and thus have preferential removal of the passivation layer that forms on copper during ECMP. Both the present wafer scale and the chip scale models can be iterated one after another to update the conditions as polishing evolves, to better reflect the coupling between chip scale and wafer scale effects.

Chapter 4 *Conclusion*

The studies presented in this thesis are able to link the physical pad properties to polishing performance by means of dynamic wafer level modeling. The large WSP size pad with a standard concentration appears to be the most effective combination with respect to the extracted model parameters and the polishing performance metrics studied. The model parameters are related to physical pad characteristics, which can help in designing better polishing pads and pad conditioning.

The proposed ECMP model helps understand better the underlying processes in this newly emerging process. The ECMP model fit to the Applied Materials model is relatively accurate. Furthermore, the experimental analysis done for the ECMP runs validates both the proposed physical ECMP model and the Applied Materials empirical model used as a benchmark.

Future work can include more experimentation with different variations of the JSR pad. A pad with a high concentration of WSP and large particle size, for example, can help to develop an empirical model if used with the current data to form a full factorial set of experiments. One of the benefits of such a model would be to identify any interactions between the particle size and concentration that are not evident when using one factor at a time experiments.

Another interesting variation is testing the pad performance for STI CMP. The parameters from the oxide modeling were used with the STI CMP model, and preliminary simulations show improved STI polishing performance of the JSR pad over the IC1000 pad. It would be valuable to conduct STI CMP experiments to empirically measure the pad performance.

The dynamic wafer level physical ECMP can be enhanced by adding lateral currents flowing into the wafer, to attempt modeling the edge effect and the inherent radial dependency in ECMP.

The ECMP model can be coupled to the SHPD model to provide die level modeling. The action of the passivation layer can be incorporated into the dual material STI CMP model by adjusting the removal rate selectivities for the passivation layer and the copper.

The conductances would also be modified accordingly; the upper areas where the passivation layer has been removed, would fully conduct electrical current, and trench areas where passivation layer remains, would conduct a fraction of the amount of current in the upper areas as set by a model parameter.

Furthermore, the ECMP model can be extended to model other tool setups, existing and non-existing. In the case of existing tools, modeling can ease a direct comparison between two opposing methodologies, for example rotational tools versus orbital tools. The model will help to identify performance differences and how to address them. For the case of prototype tools, the amount of voltage zones, their respective diameters, the electrolyte conductivity, and many other parameters can be varied to get a better idea of how to improve the process.

References

[1] D. Boning and S. Nassif, "Models of Process Variations in Device and Interconnect," in Design of High Performance Microprocessor Circuits, Eds. A. Chandrakasan, W. Bowhill and F. Fox, IEEE Press, 2000.

[2] Diagrid® Product Brochure, Rohm and Haas Electronic Materials website, http://electronicmaterials.rohmhaas.com/products/brochures/Diagrid.pdf

[3] Tucker, Tom. Laredo Technologies Inc.

[4] D. O. Ouma, D. S. Boning, J. E. Chung, W. G. Easter, V. Saxena, S. Misra, and A. Crevasse, "Characterization and Modeling of Oxide Chemical Mechanical Polishing Using Planarization Length and Pattern Density Concepts," *IEEE Transactions on Semiconductor Manufacturing*, vol. 15, no. 2, pp. 232-244, May 2002.

[5] B. Lee, "Modeling of Chemical Mechanical Polishing for Shallow Trench Isolation," Ph.D. Thesis, MIT Dept. of Electrical Engineering and Computer Science, May 2002.

[6] J. Sorooshian, A. Philipossian, L. Borucki, R. Timon, D. Stein, D. Hetherington, and D. Boning, "Impact of Pattern Density on the Effective Pressure During STI CMP," *Chemical-Mechanical Planarization for ULSI Multilevel Interconnect Conference*, pp. 358-361, Marina Beach, CA, Feb. 2004.

[7] X. Xie, T. Park, Brian Lee, T. Tugbawa, H. Cai and D.Boning, "Re-examining the Physical Basis of Planarization Length in Pattern Density CMP Models," *MRS Spring Meeting, Symposium F: Chemical Mechanical Planarization*, San Francisco, CA, April 2003.

[8] D. Boning, B. Lee, T. Tubawa, and T. Park, "Models for Pattern Dependencies: Capturing Effects in Oxide, STI, and Copper CMP," *Semicon/West Technical Symposium: CMP Technology for ULSI Manufacturing,*, San Francisco, CA, July 2001.

[9] T. Tugbawa, T. Park, B. Lee, D. Boning, P. Lefevre, and L. Camilletti, "Modeling of Pattern Dependencies for Multi-Level Copper Chemical-Mechanical Polishing Processes," *Materials Research Society (MRS) Spring Meeting,* San Francisco, CA, April 2001.

[10] X. Xie, D. Boning, F. Meyer, R. Rzehak, and P. Wagner, "Analysis and Modeling of Nanotopography Impact in Blanket and Patterned Silicon Wafer Polishing". *Chemical-Mechanical Planarization for ULSI Multilevel Interconnect Conference 2006 Proceedings*, pp. 243-253, Freemont, CA 2006.

[11] X. Xie and D. Boning, "Integrated Modeling of Nanotopography Impact in Patterned STI CMP," *Chemical-Mechanical Planarization for ULSI Multilevel Interconnect Conference*, pp. 159-164, Marina Beach, CA, Feb. 2003.

[12] D. Boning, B. Lee, N. Poduje, J. Valley, and W. Baylies, "Impact of Nanotopography on STI CMP in Future Technologies," *Workshop on Metrology for*

Silicon Wafers for 100 nm Technology Generations and Beyond, SEMI/Europa, Munich, April 2002.

[13] B. Lee, D. Boning, W. Baylies, N. Poduje, and J. Valley, "Modeling and Mapping of Nanotopography Interactions with CMP," *Materials Research Society (MRS) Spring Meeting*, San Francisco, CA, April 2002.

[14] D. Boning and B. Lee, "Nanotopography Issues in Shallow Trench Isolation CMP," *MRS Bulletin*, pp. 761-765, Oct. 2002.

[15] B. Lee, D. Boning, W. Baylies, N. Poduje, P. Hester, Y. Xia, J. Valley, C. Koliopoulus, D. Hetherington, H. Sun, and M. Lacy, "Wafer Nanotopography Effects on CMP: Experimental Validation of Modeling Methods," *Materials Research Society (MRS) Spring Meeting*, San Francisco, CA, April 2001.

[16] D. Boning, B. Lee, W. Baylies, N. Poduje, P. Hester, J. Valley, C. Koliopoulos and D. Hetherington, "Characterization and Modeling of Nanotopography Effects on CMP," *International CMP Symposium 2000*, Tokyo, Japan, December 2000.

[17] D. P. Thakurta, C. Borst, D. Schwendeman, R. Gutman, and W. Gill. "Three Dimensional Chemical Mechanical Planarization Slurry Flow Model Based on Lubrication Theory," *J. Electrochem. Soc.*, vol. 148, no. 4, pp. G207-214, 2001.

[18] A. Kim, J. Seok, J. Tichy, and T. Cale, "A Multiscale Contact Mechanics and Hydrodynamics Model of Chemical Mechanical Polishing," *VMIC*, pp. 129-134, Santa Clara, CA, 2001.

[19] W. Jeng, C. Liang, and J. Yeuan, "Numerical Investigation of Particulate Flow of Slurry in Chemical Mechanical Planarization Process," *CMPMIC*, pp. 49-52, Santa Clara, CA, 2001.

[20] G. P. Muldowney, "Characterization of CMP Pad Surface Texture and Pad-Wafer Contact," *MRS Symposium Proceeding*, Vol. 816, pp. 147-157, Spring 2004.

[21] G. P. Muldowney, "Modeling CMP Transport and Kinetics at the Pad Groove Scale," *MRS Symposium Proceedings*, Vol. 816, pp. 159-164, Spring 2004.

[22] G. P. Muldowney, "An investigation of transient CMP wafer surface heating at the pad groove scale," *Chemical-Mechanical Planarization for ULSI Multilevel Interconnect Conference*, pp 455-462, Fremont, CA, 2006.

[23] L. M. Cook, "Semiconductors and Semimetals," Vol. 63, Chapter 6, Academic Press 2000.

[24] L. Chen, "Breakthrough technology for CMP," Semiconductor Fabtech, 24th Edition.

[25] K. Smekalin, "ECMP: Novel Planarization Solution for 65nm and Below," *International Semiconductor Technology Conference Proceedings on Semiconductor Technology*, Santa Clara, CA, 2005.

[26] S. Kondo, S. Tominaga, A. Namiki, K. Yamada, D. Abe, K. Fukaya, M. Shimada, and N. Kobayashi, "Novel Electro-Chemical Planarization using Carbon Polishing

Pad to Achieve Robust Ultra low-k/Cu Integration," *Proceedings of the International Interconnect Technology*, pp 203-205, San Francisco, CA, June 2005.

[27] Y. Hong, S. Pandija, V. K. Devarapalli, S. Jha, and S.V. Babu, *Proceedings of the 2nd PAC-RIM International Conference on Planarization CMP and its Applications*, Seoul, South Korea, Nov. 2005.

[28] Y. Hong, V. K. Devarapalli, D. Roy, and S.V. Babu, "Synergistic roles of dodecyl sulfate and benzotriazole in enhancing the efficiency of chemical-mechanical planarization of copper," to be submitted.

[29] X. Xie, D. Boning, "Physically-Based Die-Level CMP Model," *MRS Symposium Proceedings,* Vol. 991, San Francisco, CA, Spring 2007.

[30] L. Borucki, R. Zhuang, T. Sun, A. Philliposian, D. Slutz, "Mechanical and optical analysis of pad surface micro-texture differences caused by conditioning," *International Conference on Planarization Technologies*, 2006.

[31] C. Yew Wee, "Study of Interaction Between the Function of Grit Size and Residual Damage on an Ultra Thin Wafer," *ICSE Proc.*, Penang, Malaysia, 2002.

[32] F. Q. Liu, L. Chen, A. Duboust, S. Tsai, A. Manens, S. Neo, Y. Wang, and W. Hsu, "Cu Planarization in Electrochemical Mechanical Planarization," *J. Electrochem. Soc.*, vol. 153, no. 6, pp. C377-381, 2006.

[33] Economikos, X. Wang, A. Sakamoto, P. Ong, M. Naujok, R. Knarr, L. Chen, Y. Moon, S. Neo, J. Salfelder, A. Duboust, A. Manens, W. Lu, S. Shrauti, F. Liu, S. Tsai, and W. Swart, "Integrated Electro-Chemical Mechanical Planarization (Ecmp) for Future Generation Device Technology," *International Interconnect Technology Conference*, June 2004.

[34] B. Lee, "Modeling of Chemical Mechanical Polishing for Shallow Trench Isolation," Ph.D. Thesis, MIT Dept. of Electrical Engineering and Computer Science, May 2002.

[35] F. Q. Liu, L. Chen, A. Duboust, S. Tsai, A. Manens, S. Neo, Y. Wang, W. Hsu, "High Planarization Efficiency and Wide Process Window Using Electro Chemical Mechanical Planarization (ECMP^{TM})," *Mater. Res. Soc. Symp. Proc.*, Vol. 867, 2005.

[36] A. Brown, "Flat, Cheap, and Under Control: Applied Materials' New Polishing Technology Could be the Key to the Coming Generation of Microchips," *IEEE Spectrum*, January 2005.

[37] J. Lu, C. Rogers, V. P. Manno, A. Philiposian, S. Anjur, and M. Moinpur, "Measurements of slurry film thickness and wafer drag during CMP," *J. Electrochem. Soc.*, vol. 151, no. 4, pp. G241-G247, 2004.