Control of Wafer-scale Non-uniformity in Chemical-Mechanical Planarization by Face-up Polishing

by

Catherine Mau

B.S., Mechanical Engineering
University of California, San Diego, 2006

Submitted to the Department of Mechanical Engineering
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE IN MECHANICAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2008

© 2008 Catherine Mau. All Rights Reserved.
The author hereby grants to MIT permission to reproduce
and to distribute publicly paper and electronic
copies of this thesis document in whole or in part
in any medium now known or hereafter created.

Signature of Author: ____________________________
Department of Mechanical Engineering
May 9, 2008

Certified by: ____________________________
Jung-Hoon Chun
Professor of Mechanical Engineering
Thesis Supervisor

Certified by: ____________________________
Nannaji Saka
Research Affiliate, Department of Mechanical Engineering
Thesis Supervisor

Certified by: ____________________________
Lallit Anand
Professor of Mechanical Engineering
Chairman, Department Committee on Graduate Students
Control of Wafer-scale Non-uniformity in Chemical-Mechanical Planarization by Face-up Polishing

by

Catherine Mau

Submitted to the Department of Mechanical Engineering on May 9, 2008 in partial fulfillment of the requirements for the Degree of Master of Science in Mechanical Engineering

ABSTRACT

Chemical-mechanical planarization (CMP) is a key process in the manufacture of ultra-large-scale-integrated (ULSI) semiconductor devices. A major concern in CMP is non-uniform planarization, or polishing, at the wafer-scale – primarily as interconnect metal dishing and dielectric erosion. In conventional face-down CMP, the pad is much larger than the wafer and the wafer is always in contact with the pad. Thus, non-uniform polishing rate at the wafer-scale is due to variations in relative velocity, normal pressure, and especially slurry distribution at the wafer/pad interface. Wafer-scale polishing uniformity requirements are expected to be even more stringent in the future as the ULSI technology advances toward larger wafers (450 mm) and ever shrinking feature sizes (< 20 nm).

This thesis presents the theory and experimental validation of a novel, face-up CMP architecture proposed for achieving a high degree, better than 95 percent of polishing uniformity at the wafer-scale. The novel design utilizes a small, perforated pad that contacts only a portion of the wafer during CMP. Polishing uniformity is achieved by progressively translating the pad away from the polished to the unpolished regions of the wafer. The theory is based on Preston’s Law for material removal rate and an optimal algorithm for pad translation. CMP experiments were conducted on both blanket and patterned wafers to validate the theory. Polishing of blanket wafers by non-translating pads showed that the Preston constant is higher at the center of the pad due to increased slurry flow. Thus, perforations at the pad center were blocked to minimize the variation in Preston constant. Face-up polishing of patterned wafers with the blocked pad showed improved wafer-scale uniformity in material removal rate. Dielectric erosion was below 30 nm, less than 5 percent of the interconnect depth, across a 100-mm circular polished region. However, dishing of the wider interconnects was much greater. Nevertheless, the variation in dishing across the 100-mm region was less than 35 nm for linewidths ranging from 2.5 μm to 100 μm, also less than 5 percent. Based on the theory and experimental results, several suggestions for further improving face-up CMP to minimize Cu dishing and dielectric erosion are offered.

Thesis supervisor: Dr. Jung-Hoon Chun
Title: Professor of Mechanical Engineering

Thesis supervisor: Dr. Nannaji Saka
Title: Research Affiliate, Department of Mechanical Engineering
Acknowledgments

I would like to thank all the people who supported and guided me through my past two years at MIT. This thesis would not have been possible without their contributions.

My utmost gratitude goes to my advisor, Professor Jung-Hoon Chun, who gave me the opportunity to be part of his group and taught me how to be a researcher. Despite his busy schedule, I knew his door was always open whenever I had a problem. I have learned a great deal from him in problem solving and in communication. His suggestions and professional advice were invaluable to my development as an engineer and I will take his wisdom with me as I enter the next stage of my professional career.

I wish to thank Dr. Nannaji Saka, my second advisor, for the countless hours he spent with me to improve my research skills. I cannot begin to enumerate the knowledge I gained from working with him: in research, in writing, and above all, in life. I have the deepest respect for Dr. Saka. His passion towards his work and research philosophy is what I will try to live by.

I would also like to thank my good friend and lab mate, Thor Eusner, who was never too busy to stop what he was doing to help me with my research problems. I appreciate all his comments and advice both inside and outside the lab. My time at MIT would not have been half the experience that it was without Thor and I wish him all the best in his future endeavors.

To all my colleagues in the LMP graduate student office, many thanks for their support and friendship. I will never forget the wonderful time we spent together and I wish them all the very best. I would also like to thank my friends: Eehern Wong, Sai Hei Yeung, Leah Acker, and Jared Thomas, for our many meals together. Our chats about anything and everything kept me sane during the rough times.

I want to express my appreciation for the staff members at MIT who have helped me with my research: Gerald Wentworth, David Dow, and Patrick McAtamney of the LMP machine shop; Tim McClure and Libby Shaw of the Center of Materials Science and Engineering; and Kurt Broderick of the Microsystems Technologies Laboratory.

I wish to acknowledge the financial support of the Semiconductor Research Corporation Education Alliance (SRCEA) and Intel Corporation through their Master’s Scholarship Program and the funding of my research project. I also thank my industrial contacts, Dr. Paul Fischer of Intel and Peter McKeever of Thomas West, Inc., for their invaluable advice and for providing me with materials for my research.

I thank my best friend Marian Lee for encouraging me to always strive for the best.

Finally I thank my family, especially my sister Angela, for their continual love and support. It is they who made me the person I am today, and I will be forever grateful to them for providing me with the opportunities and encouragement to pursue my dreams.
# Table of Contents

Title Page ......................................................................................................................... 1
Abstract ............................................................................................................................... 2
Acknowledgments .................................................................................................................. 3
Table of Contents .................................................................................................................. 4
List of Figures ....................................................................................................................... 6
List of Tables ......................................................................................................................... 10

## CHAPTER 1 INTRODUCTION ......................................................................................... 12
  1.1 Background .................................................................................................................. 12
  1.2 Chemical-Mechanical Planarization ......................................................................... 15
    1.2.1 Current CMP Tools ............................................................................................. 15
    1.2.2 An Integrated, Multi-scale Tribological Model ............................................... 18
    1.2.3 The Face-up CMP Architecture ........................................................................ 22
  1.3 Organization ............................................................................................................... 23
      Nomenclature .............................................................................................................. 24

## CHAPTER 2 FACE-UP CHEMICAL-MECHANICAL POLISHING .................................... 25
  2.1 Introduction ............................................................................................................... 25
  2.2 Geometry .................................................................................................................... 27
  2.3 Kinematic Analysis .................................................................................................... 32
  2.4 Material Removal Rate ............................................................................................. 34
    2.4.1 Non-Translating Pad ......................................................................................... 35
    2.4.2 Translating Pad ............................................................................................... 39
  2.5 Numerical Model for the Pad Translational Velocity .............................................. 41
    2.5.1 Discretization .................................................................................................... 41
    2.5.2 Matrix Formulation .......................................................................................... 42
    2.5.3 An Example ...................................................................................................... 50
    2.5.4 Discretization Error .......................................................................................... 52
  2.6 Summary ................................................................................................................... 57
      Nomenclature .............................................................................................................. 60

## CHAPTER 3 POLISHING EXPERIMENTS WITH A NON-TRANSLATING PAD ........... 61
  3.1 Introduction ............................................................................................................... 61
List of Figures

Figure 1.1: The increase in the number of components in a chip as predicted by Dr. Gordon Moore. [Moore, 1965] ........................................................................................................ 13

Figure 1.2: The number of transistors in recent commercial microprocessors. (Intel Corp.) 13

Figure 1.3: Feature and gate size trends as forecast in the ITRS. [ITRS, 2007] .................... 14

Figure 1.4: Cross-section of a multilayer microprocessor chip built by IBM's 90-nm CMOS technology. (IBM) ................................................................................................. 14

Figure 1.5: Interconnect fabrication steps: (a) dielectric deposition and line etching, (b) via etching if the dual damascene process is used, (c) barrier layer and metal deposition, and (d) planarization by CMP .................................................. 16

Figure 1.6: Schematics of CMP tool architectures: (a) rotary, (b) orbital, (c) web-format with fixed-abrasives, and (d) face-up rotary with an annular, oscillating pad. [Noh, 2005] ................................................................................................. 19

Figure 1.7: Cu dishing and dielectric erosion in a die. [Noh, 2005] ...................................... 19

Figure 1.8: Definition of the feature-scale non-uniformity factor, $\alpha$ ............................... 21

Figure 1.9: Definition of the wafer-scale non-uniformity factor, $\beta$. [Noh, 2005] ............. 21

Figure 2.1: Schematic of the face-up CMP architecture. [Noh, 2005] ................................. 26

Figure 2.2: Pad translation with respect to the polished region during face-up CMP .......... 26

Figure 2.3: Schematic depicting the path of a point $P$ on the wafer during one wafer revolution and the definition of $\theta$. ................................................................. 28

Figure 2.4: Pad locations where (a) the pad edge is covering the center region of the wafer, (b) the pad edge is just touching the center of the wafer, and (c) the pad is away from the center ................................................................. 30

Figure 2.5: Semi-contact angles for three different pad locations for $r^* = 0.7$. .................. 31

Figure 2.6: Definition of the wafer and pad coordinate systems ......................................... 33

Figure 2.7: Schematic of the entrance and exit semi-contact angles ................................. 36

Figure 2.8: $\Delta h^*$ versus $r^*$ for various rotational velocity ratios and pad sizes .......... 38

Figure 2.9: Schematic of the wafer radius and pad translation discretization ..................... 43
Figure 2.10: Flow chart summarizing a numerical method for determining pad translation in face-up CMP. ................................................................. 45
Figure 2.11: A discretization scheme involving clusters of points. ......................... 45
Figure 2.12: Schematic of a translating pad and the subsequent triangular matrix formulation. ........................................................................... 47
Figure 2.13: Normalized pad location versus polishing time for various rotational velocity ratios................................................................. 49
Figure 2.14: Schematic of the discretization of wafer radius and pad translation in the five-step example and the corresponding \( \theta_c'(r) \) ................................................................. 51
Figure 2.15: Dimensionless pad translational motion for the five-step example. .......... 53
Figure 2.16: Pad location versus polishing time for the five-step example. ................. 53
Figure 2.17: Plots of the Cu remaining on the wafer after each pad translation step. .... 54
Figure 2.18: Discretization error for different mesh sizes..................................... 56
Figure 2.19: Error with 2% deliberate overpolishing, \( S_{Cu/e} h_b = 0.02 h_{Cu} \) .......... 56
Figure 2.20: Comparison of errors from two different discretization schemes .......... 58
Figure 3.1: Photograph of the face-up CMP tool. .................................................. 62
Figure 3.2: Schematic of the cross section of a TWI-817 stacked pad. ...................... 62
Figure 3.3: Photograph of the slurry cup............................................................... 64
Figure 3.4: (a) Uniformly spaced and (b) blocked center pad perforation patterns. .... 64
Figure 3.5: Video screenshots from a polishing experiment with an unblocked pad...... 66
Figure 3.6: Material removal by a pad with uniformly spaced perforations compared to theoretical results; \( p = 17 \) kPa, \( \omega_w = \omega_p = 19 \) rad/s (180 rpm). ........................................... 68
Figure 3.7: Video screenshots from the blocked pad polishing experiment. ................. 69
Figure 3.8: Material removal by a blocked pad compared to theoretical results; \( p = 17 \) kPa, \( \omega_w = \omega_p = 19 \) rad/s (180 rpm). ................................................................. 69
Figure 3.9: Variation in \( k_p \) across the wafer when polishing with different pads. ....... 71
Figure 3.10: Comparison of experimental and theoretical \( \Delta h^* \) when \( \omega_w / \omega_p = 0.5 \) ................ 72
Figure 3.11: Comparison of experimental and theoretical $\Delta h_r^*$ when $\omega_w / \omega_p = 1.5$ .......... 73

Figure 3.12: Normalized material removed per wafer rotation for various rotational velocity ratios................................................................. 75

Figure 3.13: $k_p(r)$ for various $r_{cc}'s$ calculated from $MRR$................................................................. 76

Figure 4.1: Photograph of the face-up CMP tool with the 300-mm platen................................. 80

Figure 4.2: A blocked pad with perforations at the intersection of the x-y grooves.............. 82

Figure 4.3: Video screenshots of the wafer and the pad during the course of face-up polishing; $p = 13$ kPa, $\omega_w = \omega_p = 16$ rad/s, and slurry flow $= 150$ ml/min.......... 83

Figure 4.4: Comparison of experimental and calculated pad positions versus polishing time in (a) dimensional and (b) dimensionless form................................................................. 86

Figure 4.5: Die map of the wafers used for dishing and erosion experiments...................... 88

Figure 4.6: Cross-section of an SKW6-2 test wafer. (SKW Associates).......................... 88

Figure 4.7: Measured initial wafer surface geometry: (a) Cu deposition factor, $\alpha$, (b) initial step-height, $h_{si}$, and (c) combined feature-scale geometry, $\alpha \cdot h_{si}$ .................. 90

Figure 4.8: Theoretical and actual dimensionless pad translation path............................... 92

Figure 4.9: Normalized dielectric erosion across the wafer after face-up CMP. ................. 96

Figure 4.10: Normalized dielectric erosion compared with the multi-scale erosion model from Eq. (4.2), where $\alpha$ and $h_{si}$ were approximated using the data shown in Figure 4.7 and $D$ from Eq. (4.5)...................................................................................................................... 98

Figure 4.11: Normalized Cu dishing across the wafer after face-up CMP. ......................... 99

Figure 4.12: Experimental dishing compared with the multi-scale dishing model presented in Eq. (4.5), where $\alpha$ and $h_{si}$ were approximated using the data shown in Figure 4.7. ................................................................................................................................. 101

Figure 4.13: Schematic of a wide trench with a polymer coating........................................ 103

Figure 4.14: Normalized Cu dishing versus linewidth after the polishing of uncoated and spin-coated wafers. ................................................................................................................................. 103

Figure 4.15: Confocal micrograph of a new TWI-817 pad surface.................................... 105

Figure 4.16: Scaled schematic of sinusoidal pad asperities contacting a wide line feature... 105
Figure A.1: Wafer samples after etching with NaOH+H₂O₂................................. 114

Figure B.1: Schematics of (a) an ideal initial feature topography where \( \alpha = 0 \) and \( h_0 = 0 \), (b) an actual feature topography where the feature is replicated on the Cu surface, and (c) a feature coated with polymer to obtain \( \alpha = 0 \) and \( h_0 = 0 \)..................... 117

Figure B.2: Profile of a scratch created on an SU-8 coating to measure initial film thickness. ................................................................. 119

Figure B.3: Comparison of experimental spin coating film thickness with theoretical film thickness............................................................................. 121

Figure B.4: Experimental dimensionless film thickness versus dimensionless time compared with the hydrodynamic model proposed by Emslie et al.......................... 122

Figure B.5: Profile of a subdie (a) before and (b) after spin coating with a 4-μm layer of SU-8 photoresist, \( w = 5 \) μm and \( \lambda = 10 \) μm ...................................................... 123

Figure B.6: Profile of a subdie (a) before and (b) after spin coating with a 4-μm layer of SU-8 photoresist, \( w = 10 \) μm and \( \lambda = 20 \) μm ...................................................... 124

Figure B.7: Profile of a subdie (a) before and (b) after spin coating with a 4-μm layer of SU-8 photoresist, \( w = 20 \) μm and \( \lambda = 40 \) μm ...................................................... 125

Figure B.8: Profile of a subdie (a) before and (b) after spin coating with a 4-μm layer of SU-8 photoresist, \( w = 50 \) μm and \( \lambda = 100 \) μm ...................................................... 126

Figure B.9: Profile of a subdie (a) before and (b) after spin coating with a 4-μm layer of SU-8 photoresist, \( w = 100 \) μm and \( \lambda = 200 \) μm ...................................................... 127
# List of Tables

<table>
<thead>
<tr>
<th>Table 1.1:</th>
<th>Input and output CMP process parameters. .......................................................... 17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1:</td>
<td>Discretized $r$ and $r_{ce}$ values from two different discretization schemes. .......... 58</td>
</tr>
<tr>
<td>Table 3.1:</td>
<td>Wafer, consumables, and process parameters for polishing experiments. .............. 65</td>
</tr>
<tr>
<td>Table 3.2:</td>
<td>Mechanical properties of the materials involved in CMP. ............................................. 65</td>
</tr>
<tr>
<td>Table 3.3:</td>
<td>Uniform pad polishing results with calculated $\theta_c$ and $k_p$. .............................. 66</td>
</tr>
<tr>
<td>Table 3.4:</td>
<td>Blocked pad polishing results with calculated $\theta_c$ and $k_p$. ............................ 68</td>
</tr>
<tr>
<td>Table 3.5:</td>
<td>Wafer, consumables, and process parameters for velocity ratio experiments. ........ 71</td>
</tr>
<tr>
<td>Table 3.6:</td>
<td>Results from the $\omega_w/\omega_p = 0.5$ polishing experiment .................................. 72</td>
</tr>
<tr>
<td>Table 3.7:</td>
<td>Results from the $\omega_w/\omega_p = 1.5$ polishing experiment .................................. 73</td>
</tr>
<tr>
<td>Table 3.8:</td>
<td>Wafer, consumables and process parameters for the $k_p$ versus $r_{ce}$ experiments ... 76</td>
</tr>
<tr>
<td>Table 4.1:</td>
<td>Experimental parameters for the blanket wafer polishing experiments ..................... 82</td>
</tr>
<tr>
<td>Table 4.2:</td>
<td>Pad location measurements from the blanket wafer polishing experiments ................... 84</td>
</tr>
<tr>
<td>Table 4.3:</td>
<td>Measured $\alpha$ and $h_{si}$ on the SKW6-2 test wafers for various linewidths ........... 90</td>
</tr>
<tr>
<td>Table 4.4:</td>
<td>Experimental conditions for patterned wafer polishing ............................................ 92</td>
</tr>
<tr>
<td>Table 4.5:</td>
<td>Measured Cu dishing and dielectric erosion after face-up CMP .................................. 94</td>
</tr>
<tr>
<td>Table 4.6:</td>
<td>Statistical summary of dielectric erosion at various features after face-up CMP .......... 96</td>
</tr>
<tr>
<td>Table 4.7:</td>
<td>Statistical summary of Cu dishing at various features after face-up CMP .................... 99</td>
</tr>
<tr>
<td>Table 4.8:</td>
<td>Comparison of Cu dishing in uncoated wafer and spin-coated wafers ......................... 103</td>
</tr>
<tr>
<td>Table A.1:</td>
<td>Measured SiO$_2$ thickness on the reference sample before and after etching ............. 114</td>
</tr>
<tr>
<td>Table A.2:</td>
<td>Measured SiO$_2$ thickness on the blanket sample ..................................................... 115</td>
</tr>
<tr>
<td>Table A.3:</td>
<td>Measured SiO$_2$ thickness on the patterned sample ................................................... 115</td>
</tr>
</tbody>
</table>
Table B.1: SU-8 spin coating and curing process steps................................................................. 119
Table B.2: Theoretical and experimental results for the viscosity versus thickness spin coating experiments for $h_0 = 5 \mu m$, $\omega = 314 \text{ rad/s}$, and $t = 30 \text{ s}$ ........................................ 121
Table B.3: Comparison of feature step-heights for SU-8 coatings of various thicknesses. . 127
Table B.4: CMP conditions for determining the Preston constant of SU-8 photoresist....... 130
CHAPTER 1

INTRODUCTION

1.1 Background

For the past forty years, the semiconductor industry has been fulfilling Dr. Gordon Moore’s 1965 prediction that the number of transistors in a chip would double roughly every two years, Figure 1.1, [Moore, 1965]. Now, there are 1.7 billion transistors in Intel Corporation’s Itanium server microprocessors. A more recent representation of the increase in the number of transistors per microprocessor is shown in Figure 1.2. The realization of Moore’s Law is made possible by shrinking component sizes and innovations in materials. Decreasing feature size is the logical progression in integrated circuit (IC) technology as it increases the capacity per unit area and functionality, and decreases cost. Reduction in the size of IC components is also driven by incessant consumer demand for smaller electronic devices. The International Technology Roadmap for Semiconductor (ITRS) reports that feature size is expected to be reduced to 20 nm in the year 2017 [ITRS, 2007]. Figure 1.3 shows the decreasing trend of feature and gate sizes as forecast by ITRS. The semiconductor industry’s past and future needs for shrinking devices have thus triggered rapid advancements in manufacturing technology.

Chemical-mechanical planarization (CMP) is one such enabling technology. Developed by International Business Machines in 1990 [Beyer et al., 1990], CMP is used at various stages of IC fabrication for its global surface planarization capabilities. This process is crucial in the manufacture of multilayer devices, like the microprocessor shown in Figure 1.4, since a smooth topography is necessary to meet the depth of focus requirements for the lithography tools used in patterning each additional layer. CMP is used to remove excess metal, or overburden, from the wafer surface during the fabrication of multi-level interconnects. These interconnects provide the electrical connection between two layers of the device and it is crucial that the metal overburden is completely removed to prevent electrical shorts.
Figure 1.1: The increase in the number of components in a chip as predicted by Dr. Gordon Moore. [Moore, 1965]

Figure 1.2: The number of transistors in recent commercial microprocessors. (Intel Corp.)
Figure 1.3: Feature and gate size trends as forecast in the ITRS. [ITRS, 2007]

Figure 1.4: Cross-section of a multilayer microprocessor chip built by IBM's 90-nm CMOS technology. (IBM)
Due to its high electrical conductivity, Cu is the choice metal for interconnects. A schematic of the damascene process used for the fabrication of Cu interconnects is shown in Figure 1.5. The process begins with a layer of dielectric material such as SiO₂. Features are formed on the SiO₂ surface by photolithography followed by an etching process, Figure 1.5(a). Since Cu can be used to form both interconnects and vias, a dual damascene process is commonly employed in industry. This differs from the single damascene method by including an additional via etch process after interconnect etching, Figure 1.5(b). Next, a thin (20 – 50 nm) layer of barrier metal such as Ti/TiN or Ta/TaN is deposited onto the surface to prevent Cu diffusion into SiO₂, which is followed by Cu deposition, Figure 1.5(c). Finally, Cu CMP is used to planarize the wafer surface and remove the Cu overburden, Figure 1.5(d).

While CMP is currently used in multiple processes on various metal and dielectric materials, this thesis focuses only on Cu CMP. It is expected, however, that the ideas presented in this thesis could be transferred to other CMP technologies such as inter-level dielectric (ILD) and shallow trench isolation (STI) CMP.

1.2 Chemical-Mechanical Planarization

Material removal by Cu CMP comprises two major components. Chemicals in the slurry modify the Cu surface to yield a softer, porous layer [Cook, 1990; Du et al., 2004] while a polymeric pad acts on abrasive particles in the slurry to mechanically remove the coating material [Liu et al., 1996; Ahmadi and Xia, 2001; Paul et al., 2007]. Therefore, the chemical content of the slurry, abrasive size and shape, pad material and topography all contribute to material removal rate. Additionally, research has also shown that the process is dependent on pressure, velocity [Preston, 1927], temperature [Mudhivarthi et al., 2005], and feature geometry [Steigerwald et al., 1994]. Due to its complexity, which involves a multitude of inputs at various scales, CMP is a difficult process to control and optimize, and defects are often generated. Table 1.1 lists the primary contributors to the CMP process along with the outputs from the process.

1.2.1 Current CMP Tools

The most common type of CMP tool to date is a rotary setup where the wafer is polished faced down on a large pad (Figure 1.6a). Rotation of both the wafer and the pad provides
Interconnect fabrication steps: (a) dielectric deposition and line etching, (b) via etching if the dual damascene process is used, (c) barrier layer and metal deposition, and (d) planarization by CMP.
Table 1.1: Input and output CMP process parameters.

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wafer Parameters:</td>
<td>Material removal</td>
</tr>
<tr>
<td>Curvature</td>
<td>Cu dishing</td>
</tr>
<tr>
<td>Cu thickness</td>
<td>Dielectric erosion</td>
</tr>
<tr>
<td>Feature size</td>
<td>Scratching</td>
</tr>
<tr>
<td>Feature density</td>
<td>Contaminants</td>
</tr>
<tr>
<td>Slurry Parameters:</td>
<td>Waste</td>
</tr>
<tr>
<td>Abrasive size</td>
<td></td>
</tr>
<tr>
<td>Abrasive material properties</td>
<td></td>
</tr>
<tr>
<td>Concentration of abrasives</td>
<td></td>
</tr>
<tr>
<td>Selectivity (chemistry)</td>
<td></td>
</tr>
<tr>
<td>Flow rate</td>
<td></td>
</tr>
<tr>
<td>Pad Parameters:</td>
<td></td>
</tr>
<tr>
<td>Material properties</td>
<td></td>
</tr>
<tr>
<td>Topography</td>
<td></td>
</tr>
<tr>
<td>Conditioning</td>
<td></td>
</tr>
<tr>
<td>Mechanical Parameters:</td>
<td></td>
</tr>
<tr>
<td>Relative velocity</td>
<td></td>
</tr>
<tr>
<td>Pressure</td>
<td></td>
</tr>
</tbody>
</table>
relative motion at the interface for material removal. Other types of CMP tools include orbital (Figure 1.6b) and web formats (Figure 1.6c). The web format, or linear motion, technology is mainly used with slurry-free fixed abrasives for STI applications [Simpson et al., 2001; Kulawski et al., 2003]. Instead of using slurry, a roll of abrasive pad material is passed over the rotating wafer in a linear fashion to remove material. Recently, there has also been some focus on electrochemical-mechanical polishing (ECMP), which employs an electrical potential to oxidize Cu to its ions and a polishing pad for removal of the passivation layer. Researchers claim that pressures much lower than those of conventional CMP can be used with this technique, which is a benefit for planarizing the mechanically weaker low-k materials [Economikos et al., 2004]. Low-pressure material removal is also claimed in a face-up rotary tool that utilizes a high-speed, oscillating pad for polishing a wafer held in the face-up orientation, Figure 1.6d, [Hoshino et al., 2003].

1.2.2 An Integrated, Multi-scale Tribological Model

For the most part, CMP is a highly empirical process. Interest in applying physical models to CMP has increased over the past decade in the hope of gaining a better understanding of the material removal mechanisms and control over non-uniformities and yield.

Two primary non-uniformity issues in interconnect formation by CMP are Cu dishing and dielectric erosion. A schematic of Cu dishing and dielectric erosion is shown in Figure 1.7. Copper dishing is defined as the difference in height between the center of the Cu line and the dielectric at the edge of the feature. Dishing is mainly a problem for features with wide interconnect Cu lines because the pad deforms over the feature and applies non-uniform pressures between the edge and the center of the feature. The reduction of metal due to dishing can cause deterioration in electrical performance. Dielectric erosion is defined as the difference in the dielectric thickness before and after CMP. Erosion is more widespread in features with a high areal density of Cu compared with the dielectric because there is increased pressure on the dielectric, and thus higher dielectric material removal rate.

Most early models of CMP focused on local material removal rates at the feature-scale. Warnock developed a phenomenological model for the effects of feature geometry on the process [Warnock, 1991] and Runnels approached the problem by assuming a hydrodynamic slurry layer [Runnels, 1994] between the pad and the wafer. On the pad side, models were
Figure 1.6: Schematics of CMP tool architectures: (a) rotary, (b) orbital, (c) web-format with fixed-abrasives, and (d) face-up rotary with an annular, oscillating pad. [Noh, 2005]

Figure 1.7: Cu dishing and dielectric erosion in a die. [Noh, 2005]
developed based on the deformation of a smooth, elastic pad [Chekina et al., 1998; Lai et al., 2002] or by assuming that the pad deforms in discrete blocks [Fu and Chandra, 2003; Noh et al., 2004].

Despite the analyses at the local scale, however, CMP remains a multi-scale process that requires uniformity across the subdie (feature), die, and wafer scales. Noh developed a comprehensive, multi-scale tribological model based on pad-wafer contact mechanics at the feature-scale and non-uniform polishing at the wafer-scale [Noh, 2005]. Non-uniformities at the feature- and wafer-scales are defined by the pattern geometry and the material removal rate, respectively. These factors must be controlled in order to obtain uniform planarization.

Feature-scale non-uniformity is described by the Cu deposition factor, $\alpha$, and the initial step-height, $h_{st}$, as shown Figure 1.8. The deposition factor $\alpha$ is defined as:

$$\alpha = \frac{w_s}{w} \quad (0 \leq \alpha \leq 1)$$

where $w_s$ is the surface trench width and $w$ is the underlying interconnect linewidth. If $\alpha = 0$, the initial Cu surface topography is planar regardless of the underlying pattern geometry and if $\alpha = 1$, the trench pattern is exactly reproduced on the Cu surface.

The wafer-scale non-uniformity factor, $\beta$, is defined as the ratio of material removal rates, $MRR$, of the slowest and fastest field regions, shown in Figure 1.9.

$$\beta = \frac{MRR_{\text{slowest field}}}{MRR_{\text{fastest field}}} \quad (0 < \beta \leq 1)$$

If $\beta = 1$, the $MRR$ at the slowest field is equal to that at the fastest field, and the entire wafer is polished uniformly.

Polishing behavior at the feature-scale is characterized by a step-height evolution model, which accounts for the change in pressure distribution on the wafer surface by pad-wafer contact mechanics and $\alpha$. Cu dishing and dielectric erosion are then determined based on the evolution of step-height in the overpolishing stage. In this stage, it is assumed that the fastest die continues to polish until the slowest die reaches the endpoint. The ratio of the times required for these two dies to reach their respective endpoints is described by $\beta$. The maximum dishing and erosion will therefore occur in the fastest die.
Figure 1.8: Definition of the feature-scale non-uniformity factor, $\alpha$.

$w_s = \alpha w$ 

Figure 1.9: Definition of the wafer-scale non-uniformity factor, $\beta$. [Noh, 2005]
Cu dishing, $D$, is defined as the interconnect step-height after CMP. For a rough pad with uniform, fully plastic asperities,

$$D = \frac{S_{\text{Cu/ox}} \left[ \frac{1}{(1-\frac{w}{\lambda})S_{\text{Cu/ox}} + \frac{w}{\lambda}} \right] \left( \frac{1}{p} \right) \left( \frac{\lambda_o^2}{6\pi R_o} \right) \left[ 1 - \exp(-t^*_o) \right]}{(1-\frac{w}{\lambda})S_{\text{Cu/ox}} + \frac{w}{\lambda}}$$

(1.3)

where $S_{\text{Cu/ox}}$ is the selectivity of the slurry, $w$ the linewidth of the interconnect, $\lambda$ the pitch, $p$ the average pressure, $Y_p$ the yield strength of the pad, $R_o$ the radius of curvature of the pad asperities, $\lambda_o$ the spacing between pad asperities, and $t^*_o$ the dimensionless polishing time for the overpolishing stage. And $t^*_o$ is expressed as

$$t^*_o = \left[ \frac{(1-\frac{w}{\lambda})S_{\text{Cu/ox}} + \frac{w}{\lambda}}{S_{\text{Cu/ox}}} \right] \left( \frac{Y_p}{p} \right) \left( \frac{6\pi R_o}{\lambda_o^2} \right) \left[ \frac{1}{\beta} - 1 \right] h_{\text{Cu}} + \frac{\alpha w}{\lambda} h_{\text{si}} + S_{\text{Cu/ox}} \frac{1}{\beta} h_o$$

(1.4)

where $h_{\text{Cu}}$ is the Cu coating thickness, $h_{\text{si}}$ the initial feature step-height, and $h_o$ the thickness of oxide removed at the slowest field region, or the amount of deliberate overpolishing.

Dielectric erosion is defined as the change in height of the high feature (dielectric) during overpolishing:

$$e = \frac{1}{(1-\frac{w}{\lambda})S_{\text{Cu/ox}} + \frac{w}{\lambda}} \left[ \left( \frac{1}{\beta} - 1 \right) h_{\text{Cu}} + \frac{\alpha w}{\lambda} h_{\text{si}} + S_{\text{Cu/ox}} \frac{1}{\beta} h_o - \left( \frac{w}{\lambda} \right) D \right]$$

(1.5)

From his integrated tribological model, Noh concluded that in order to minimize dishing and erosion, $\alpha$ must be close to 0, $\beta$ close to 1, and overpolishing should be minimal. This means that the surface should be initially planar and the whole wafer should have uniform material removal rates.

1.2.3 The Face-up CMP Architecture

The integrated tribological model suggests that wafer-scale polishing should be uniform, $\beta$ close to unity, to minimize dishing and erosion. Attemps to address wafer-scale non-uniformity by the current CMP tools shown in Figure 1.6 include zonal pressure control and pad oscillation, which relies on complicated control systems and empirical data. Often there is a lack of physical understanding for these forms of compensation. Meanwhile, the ITRS anticipates an increase of wafer size to 450 mm in 2012, which indicates that manufacturing technology must be in place in the coming years. Directly applying current technologies to the larger wafer will
result in a reduction in $\beta$, which will lead to an increase in wafer-scale non-uniformity. Therefore, a novel face-up CMP tool architecture was developed to control wafer-scale polishing. This architecture utilizes a smaller diameter pad to enable kinematic control of material removal rate across the wafer. A steep polishing gradient is developed, and a pad translation scheme is used to reduce overpolishing time at all points on the wafer.

1.3 Organization

The object of this thesis is to study wafer-scale polishing uniformity by face-up CMP. A set of process parameters are selected to determine their effects on wafer-scale material removal rate and a numerical model is developed to determine the pad translation scheme for minimizing overpolishing time. In Chapter 1, background information on the CMP process is provided and the integrated tribological Cu CMP model is introduced. Chapter 2 describes the face-up CMP tool architecture by relating geometric and kinematic parameters to material removal rate. Cases for both a non-translating pad and a translating pad are discussed. Based on the requirement that the total material removed across the wafer must be uniform, a numerical method for pad translation is developed. Chapter 3 details an experimental study on the face-up CMP tool with a non-translating pad and blanket Cu wafers. The wafers were polished with pads containing different perforation patterns to investigate the change in Preston constants across the wafer. In Chapter 4, polishing experiments with a translating pad are presented. Experiments were performed on blanket Cu wafers to validate the numerical model by comparing the actual pad position during polishing with the calculated values. Patterned wafers were then polished to measure Cu dishing and dielectric erosion after face-up CMP to demonstrate wafer-scale uniformity. Finally, Chapter 5 summarizes this thesis and suggests future work in reducing Cu dishing and dielectric erosion using the face-up CMP technology.
Nomenclature

\( D = \) Cu dishing (m)  
\( e = \) dielectric erosion (m)  
\( h = \) thickness of Cu coating removed (m)  
\( h_{Cu} = \) initial Cu coating thickness (m)  
\( h_o = \) thickness of oxide removed at the slowest field during overpolishing (m)  
\( h_s = \) step-height (m)  
\( h_{si} = \) initial step-height (m)  
\( MRR = \) material removal rate (m/s)  
\( p = \) pressure (N/m\(^2\))  
\( R_a = \) radius of curvature of the pad asperity (m)  
\( S_{Cu/ox} = \) Cu to oxide slurry selectivity  
\( t_o^* = \) dimensionless overpolishing time  
\( w, w_s = \) interconnect linewidth, surface trench width (m)  
\( Y_p = \) yield strength of the pad asperity (N/m\(^2\))  
\( \alpha = \) feature-scale non-uniformity factor, Cu deposition factor  
\( \beta = \) wafer-scale non-uniformity factor  
\( \lambda = \) pitch of Cu interconnect lines (m)  
\( \lambda_o = \) spacing between pad asperities (m)
CHAPTER 2

FACE-UP CHEMICAL-MECHANICAL POLISHING

2.1 Introduction

It has been widely observed that the material removal rate across the wafer is non-uniform. This non-uniformity is caused by the variation in applied pressure due to wafer curvature [Fu and Chandra, 2001] and non-uniform slurry film thickness due to the single-point slurry delivery system common in conventional tools [Thakurta et al., 2001]. However, little can be done to control wafer-scale polishing with current CMP tools because the entire wafer is in contact with the pad throughout the polishing process.

In conventional face-down CMP, the wafer is always in contact with the pad and it is impossible to independently terminate polishing at different points on the wafer. Thus, overpolishing will occur in areas of high material removal rates. These areas continue to polish after the endpoint has been reached until the regions with low material removal rates also finish polishing. Overpolishing is a primary cause of Cu dishing and dielectric erosion in CMP [Stavreva et al., 1997; Noh, 2005]. The face-up CMP tool architecture, Figure 2.1, is proposed to improve wafer-scale polishing uniformity by allowing the pad to translate away from the region of the wafer that has completed polishing [Noh et al., 2006]. To allow for better control of the pad translation, a polishing gradient is induced by geometry and kinematics in which the highest material removal rate occurs at the center of the wafer. With such variation in material removal rate, the pad is only required to travel uni-directionally away from the center of the wafer, as shown in Figure 2.2.

The face-up CMP tool is also novel in the method by which slurry is delivered to the pad-wafer interface. Non-uniform wafer-scale polishing in CMP is in part due to non-uniform slurry distribution at the pad-wafer interface [Coppeta et al., 2000; Fu et al., 2005]. In face-down CMP, slurry is fed at the periphery of the wafer. Therefore, overpolishing occurs at the edge of the wafer where there is an ample supply of fresh slurry. This issue will continue to gain importance as the industry increases the size of wafers to 450 mm [ITRS, 2007]. The face-up
Figure 2.1: Schematic of the face-up CMP architecture. [Noh, 2005]

Figure 2.2: Pad translation with respect to the polished region during face-up CMP.
CMP tool addresses the issue of uniform slurry distribution by supplying slurry through perforations in the pad, as shown in Figure 2.1.

2.2 Geometry

The primary difference between the face-up and the conventional face-down CMP architecture is that the pad does not cover the entire wafer during the polishing process in face-up CMP. As a result, material removal at a point, \( P(r, \theta) \), on the wafer occurs only during the period when that point is in contact with the pad. Figure 2.3 shows the path of \( P \) as the wafer completes one revolution. The duration of contact between \( P \) and the pad is dependent on the semi-contact angle, \( \theta_c \). From Figure 2.3, \( \theta_c \) can be defined by applying the Law of Cosines with points \( P, O_w, \) and \( O_p \):

\[
\theta_c = \cos^{-1}\left(\frac{r^2 + r_{cc}^2 - r_p^2}{2rr_{cc}}\right)
\]

(2.1)

where \( r_{cc} \) is the distance between the wafer center and the pad center and \( r_p \) is the radius of the pad. The limiting cases for Eq. (2.1) are: \( \theta_c = \pi \) for points on the wafer in continuous contact with the pad, and \( \theta_c = 0 \) for points outside the contact region. The location of the pad in relation to the wafer and pad radius, and its relation to \( \theta_c \), can be categorized into three cases:

\[
r_{cc} < r_p: \quad \theta_c = \begin{cases} 
\pi & \text{if } r < (r_p - r_{cc}) \\
\cos^{-1}\left(\frac{r^2 + r_{cc}^2 - r_p^2}{2rr_{cc}}\right) & \text{if } (r_p - r_{cc}) \leq r \leq (r_{cc} + r_p) \\
0 & \text{if } r > (r_{cc} + r_p)
\end{cases}
\]

(2.2)

\[
r_{cc} = r_p: \quad \theta_c = \begin{cases} 
\cos^{-1}\left(\frac{r}{2r_{cc}}\right) & \text{if } 0 \leq r \leq (r_{cc} + r_p) \\
0 & \text{if } r > (r_{cc} + r_p)
\end{cases}
\]

\[
r_{cc} > r_p: \quad \theta_c = \begin{cases} 
\cos^{-1}\left(\frac{r^2 + r_{cc}^2 - r_p^2}{2rr_{cc}}\right) & \text{if } (r_{cc} - r_p) \leq r \leq (r_{cc} + r_p) \\
0 & \text{otherwise}
\end{cases}
\]

27
Figure 2.3: Schematic depicting the path of a point \( P \) on the wafer during one wafer revolution and the definition of \( \theta_c \).
These pad locations are shown schematically in Figure 2.4 for \( r_p = 0.7r_w \). In Figure 2.4(a), \( r_{cc} < r_p \) and the pad overlaps the center region of the wafer, a practical initial position for face-up CMP. Figure 2.4(b) shows the pad when its edge is just touching the center of the wafer, \( r_{cc} = r_p \), and Figure 2.4(c) shows the pad after it has translated away from the center, \( r_{cc} > r_p \). It is often useful to express Eq. (2.2) in dimensionless form. To do so, the following dimensionless variables are introduced:

\[
\begin{align*}
\theta_c^* &= \theta_c / \pi \\
r^* &= r / r_w \\
r_{cc}^* &= r_{cc} / r_w \\
r_p^* &= r_p / r_w
\end{align*}
\]  

Equation (2.2) can then be expressed as:

\[
\begin{align*}
r_{cc}^* < r_p^* : & \quad \theta_c^* = \left\{ \begin{array}{ll}
1 & r^* < (r_p^* - r_{cc}^*) \\
\left( \frac{1}{\pi} \right) \cos^{-1} \left( \frac{r_{cc}^* + r_p^* - r^*}{2r_{cc}^*} \right) & (r_p^* - r_{cc}^*) \leq r^* \leq (r_{cc}^* + r_p^*) \\
0 & r^* > (r_{cc}^* + r_p^*)
\end{array} \right. \\
r_{cc}^* = r_p^* : & \quad \theta_c^* = \left\{ \begin{array}{ll}
1 & 0 \leq r^* \leq (r_{cc}^* + r_p^*) \\
\left( \frac{1}{\pi} \right) \cos^{-1} \left( \frac{r_p^*}{2r_{cc}^*} \right) & r^* > (r_{cc}^* + r_p^*)
\end{array} \right. \\
r_{cc}^* > r_p^* : & \quad \theta_c^* = \left\{ \begin{array}{ll}
1 & (r_{cc}^* - r_p^*) \leq r^* \leq (r_{cc}^* + r_p^*) \\
0 & \text{otherwise}
\end{array} \right.
\end{align*}
\]  

The change in \( \theta_c^* (r^*) \) with each pad location is shown in Figure 2.5 for \( r_p^* = 0.7 \). When the pad overlaps the center of the wafer, there is continuous contact with the pad in that region and \( \theta_c^* = 1 \). The region of no contact, \( \theta_c^* = 0 \) is outside the wafer due to the pad size relative to the wafer radius. For the case when the edge of the pad just touches the center of the wafer, \( \theta_c^* \) begins at 0.5 which indicates that the center of the wafer is in contact with the pad for exactly
Figure 2.4: Pad locations where (a) the pad edge is covering the center region of the wafer, (b) the pad edge is just touching the center of the wafer, and (c) the pad is away from the center.
Figure 2.5: Semi-contact angles for three different pad locations for $r^*_p = 0.7$. 
half the time of one wafer rotation. Finally, when the pad is moved away from the center of the wafer, the termination of contact in that region is described by $\theta_c^* = 0$ near $r^* = 0$. The magnitude of $\theta_c$ is also decreased in this pad position. Therefore, the period of contact at each point on the wafer has a non-linear dependence on pad size and pad location.

2.3 Kinematic Analysis

The relative velocity between a point on the wafer and a point on the pad is determined by kinematics. A set of wafer and pad coordinate systems are defined in Figure 2.6. A polar coordinate system $(r, \theta)$ is assigned to the wafer with its origin at the center of the wafer, $O_w$. A pad coordinate system, situated at the center of the pad, $O_p$, is defined as $(r', \theta')$. Lastly, a global Cartesian coordinated system, $(x, y)$, is defined at the center of the wafer.

When the wafer rotates at angular velocity, $\omega_w$, the tangential velocity of a point $P(r, \theta)$ on the wafer is expressed as a vector with by $x$ and $y$ components:

$$v_p = -\omega_w r \sin \theta e_x + \omega_w r \cos \theta e_y$$

(2.5)

Similarly, the tangential velocity at a point on the pad $P'(r', \theta')$ with pad angular velocity, $\omega_p$, and pad translation velocity, $v_{cc}$, is expressed in the pad coordinate system as:

$$v_{p'} = (v_{cc} - \omega_p r' \sin \theta') e_x + \omega_p r' \cos \theta' e_y$$

(2.6)

Points from the pad coordinate system can be converted to the wafer coordinate system by equating their global Cartesian positions:

$$x = r \cos \theta = r' \cos \theta' + r_{cc}$$
$$y = r \sin \theta = r' \sin \theta$$

(2.7)

Therefore, Eq. (2.6) can be expressed in wafer coordinates as:

$$v_{p'} = (v_{cc} - \omega_p r \sin \theta) e_x + \omega_p (r \cos \theta - r_{cc}) e_y$$

(2.8)

The velocity of the wafer relative to the pad is:

$$v_R (r, \theta) = v_p - v_{p'}$$

$$v_R (r, \theta) = -\left[ (\omega_w - \omega_p) r \sin \theta + v_{cc} \right] e_x + \left[ (\omega_w - \omega_p) r \cos \theta + \omega_p r_{cc} \right] e_y$$

(2.9)
Figure 2.6: Definition of the wafer and pad coordinate systems.
The magnitude of relative velocity is expressed as:

\[
v_r(r, \theta) = \sqrt{\left[(\omega_w - \omega_p) r \sin \theta + v_{cc}\right]^2 + \left[(\omega_w - \omega_p) r \cos \theta + \omega_p r_{cc}\right]^2}\]

(2.10)

A simplifying case can be drawn from Eq. (2.10). If the wafer and the pad rotates at the same angular velocity, \(\omega_w = \omega_p = \omega\), the relative velocity is constant regardless of the point's position on the wafer:

\[
v_r = \sqrt{v_{cc}^2 + (\omega r_{cc})^2}\]

(2.11)

Furthermore, when the pad is not translating, \(v_{cc} = 0\), or is translating at a negligible velocity, \(v_{cc} \ll \omega r_{cc}\), Eq. (2.11) can be approximated by

\[
v_r = \omega r_{cc}\]

(2.12)

### 2.4 Material Removal Rate

The local material removal rate in CMP is described by the Preston equation [Preston, 1927],

\[
\frac{dh}{dt} = k_p \cdot p \cdot v_r
\]

(2.13)

where \(h\) is the thickness of coating removed, \(t\) the polishing time, \(k_p\) the Preston constant, and \(p\) the nominal pressure. As researchers have previously noted, \(k_p\) is not a fundamental constant. Instead, it depends on polishing conditions such as pad stiffness and surface topography, and slurry concentration and selectivity [Liu et al., 1996; Saka et al., 2001; Luo and Dornfeld, 2003; Noh et al., 2005]. While pressure and relative velocity are mechanical parameters that can be measured and controlled, determining the Preston constant requires exhaustive information gathering during the process. Thus maintaining a uniform Preston constant is an essential, yet challenging, requirement for uniform material removal rate.

In face-down polishing, the pad is always in contact with the wafer. This means that material is being removed from the wafer at all times during one wafer revolution. Consequently, the thickness of coating removed during one revolution, \(\Delta h_r\), is expressed as

\[
\Delta h_r = \int_0^{T} \left| \frac{dh}{dt} \right| dt
\]

(2.14)
where $\Delta t_r$ is the time required for one wafer revolution. In terms of $\theta$, with $dt = d\theta/\omega_w$, Eq. (2.14) can be rewritten as:

$$\Delta h_r = \int_{0}^{2\pi} \frac{1}{\omega_w} \left| \frac{dh}{dt} \right| d\theta \tag{2.15}$$

In face-up CMP, however, only a portion of the wafer is exposed during polishing. Thus, a point on the wafer may not be in contact with the pad for the full rotation. The period of time that the pad contacts the wafer is dependent on the semi-contact angle, $\theta_c$. To account for the non-contact period in the face-up configuration, Eq. (2.15) is expressed as

$$\Delta h_r(r) = \int_{\theta_{c1}(r)}^{\theta_{c2}(r)} \frac{k_p p V_R}{\omega_w} d\theta \tag{2.16}$$

where $\theta_{c1}$ and $\theta_{c2}$ are the entrance and exit semi-contact angles, respectively. Figure 2.7 shows the change in $\theta_c$ for a point on the wafer as it travels one wafer rotation underneath a translating pad. In this case, $\theta_{c1} > \theta_{c2}$ as a result of the change in $r_{cc}$. If the pad is not translating, $\theta_{c1} = \theta_{c2}$.

The average material removal rate for one wafer revolution, $\overline{MRR}$, is thus

$$\overline{MRR}(r, r_{cc}) = \frac{\Delta h_r(r, r_{cc})}{\Delta t_r} = \frac{1}{2\pi} \int_{\theta_{c1}(r_{cc})}^{\theta_{c2}(r_{cc})} k_p p V_R d\theta \tag{2.17}$$

where $\Delta t_r = 2\pi / \omega_w$.

### 2.4.1 Non-Translating Pad

While the face-up tool concept requires that the pad translate away from the center of the wafer, it is useful to first examine the case when the pad does not translate. That is, $v_{cc} = 0$ and $r_{cc}$ is constant. If $k_p, p, \omega_w$ and $\omega_p$ are also assumed constant, the integral in Eq. (2.16) is only dependent on $V_R$

$$\Delta h_r(r) = \frac{k_p p}{\omega_w} \int_{\theta_{c1}(r)}^{\theta_{c2}(r)} \sqrt{\left(\omega_w - \omega_p\right)^2 r \sin \theta + \left(\omega_w - \omega_p\right)^2 \cos \theta + \omega_{pc}^2 r_{cc}^2} d\theta \tag{2.18}$$

Furthermore, suppose that $\omega_w = \omega_p$. From Eq. (2.12), when $\omega_w = \omega_p = \omega$ and $v_{cc} = 0$, $V_R$ is constant over the entire pad-wafer interface: $V_R = \omega r_{cc}$. The material removed per wafer
Figure 2.7: Schematic of the entrance and exit semi-contact angles.
rotation at $r$ is then:

$$\Delta h_r(r) = k_p \rho \int \frac{\theta(r) \rho(r) \theta(r)}{\theta(r)} d\theta$$  \hspace{1cm} (2.19)$$

When the pad is not translating, the entrance semi-contact angle is equal to the exit semi-contact angle, $\theta_1(r) = \theta_2(r) = \theta_c(r)$. Therefore, Eq. (2.19) can be further simplified to

$$\Delta h_r(r) = 2k_p \rho \theta_c(r)$$  \hspace{1cm} (2.20)$$

and the average material removal rate per wafer revolution is given by

$$\frac{MRR(r)}{\pi} = k_p \rho \theta_c(r)$$  \hspace{1cm} (2.21)$$

It is often convenient to determine material removal by the time required to reach the process endpoint, $t_e$. The relation between polishing time and material removal is expressed as

$$t_e(r) = \frac{h_{ce} \pi}{k_p \rho \theta_c(r)}$$  \hspace{1cm} (2.22)$$

where $h_{ce}$ is the initial Cu coating thickness.

A dimensionless variable, $\Delta h_r^*$, is defined as the ratio of the material removed at radius $r$ to the material removed at the center of the wafer ($r = 0$) per wafer rotation:

$$\Delta h_r^*(r) = \frac{\Delta h_r(r)}{\Delta h_r(0)}$$  \hspace{1cm} (2.23)$$

From Eq. (2.20), the only spatially varying parameter in $\Delta h_r^*$ is $\theta_c$. Therefore, $\Delta h_r^*$ for the non-translating pad and $\omega_w = \omega_p$ case is:

$$\Delta h_r^*(r) = \frac{\theta_c(r)}{\theta_c(0)}$$  \hspace{1cm} (2.24)$$

Since $t_e(r)$ is inversely related to $\theta_c(r)$, Eq. (2.24) can also be expressed as

$$\Delta h_r^* = \frac{t_e(0)}{t_e(r)}$$  \hspace{1cm} (2.25)$$

Figure 2.8 shows $\Delta h_r^*(r^*)$ for varying wafer-pad rotational velocity ratios and size ratios. The decreasing polishing gradient with the fastest material removal rate at the center of the wafer is crucial for progressive polishing. This allows the pad to translate uni-directionally toward the edge of the wafer to avoid overpolishing.
Figure 2.8: $\Delta h_r^*$ versus $r^*$ for various rotational velocity ratios and pad sizes.
In all the cases shown, the pad slightly overlaps the center of the wafer, \( r_{cc} < r_p \), which is evident in Figure 2.8 by the simultaneous polishing of the central region. This sort of pad initial position is necessary in practice because it guarantees that the center of the wafer is completely polished. An additional benefit of the overlap is that it increases the material removal rate across the wafer, thereby decreasing polishing time. Most importantly, the overlap produces a steeper polishing gradient for better control over progressive polishing.

2.4.2 Translating Pad

The novel concept in face-up CMP is that the pad translates laterally across the wafer to minimize overpolishing time. In the previous section, it was possible to obtain a closed-form solution for \( \overline{MRR} \) because the period of contact between a point on the wafer and the pad for every wafer revolution is constant when the pad is not translating. For a translating pad the semi-contact angle may change over one wafer rotation, \( \theta_c_\neq \theta_c_\neq \), and \( \overline{MRR} \) is a function of the displacement of the pad with respect to time, \( r_{cc}(t) \).

The total thickness of material removed at a point on the wafer, \( \Delta h(r) \), can be obtained by integrating Eq. (2.17), from \( t = 0 \) to the completion of the polishing process, \( t = t_e (r) \):

\[
\Delta h(r) = \frac{1}{2\pi} \int_{\theta_{c1}(r_{cc})}^{\theta_{c2}(r_{cc})} \int_0^{r_{cc}(r)} k_p p \left[ \left( \omega_w - \omega_p \right) r \sin \theta + v_{cc} \right] \frac{1}{2} d\theta dt (2.26)
\]

For uniform material removal across the wafer, \( \Delta h(r) \) should be independent of \( r \). Furthermore, the total material removed should be equal to the initial Cu thickness with some additional overpolishing thickness. In the previous chapter, \( h_o \) was defined as the amount of deliberate oxide overpolishing. Therefore, the equivalent Cu overpolishing is equal to \( S_{Cu/ox} \cdot h_o \) and

\[
\Delta h(r) = h_{Cu} + S_{Cu/ox} h_o \tag{2.27}
\]

where \( h_{Cu} \) is the initial Cu coating thickness and \( S_{Cu/ox} \) is the Cu to oxide slurry selectivity. The average material removal rate at \( r \) is the total material removed divided by the total polishing time

\[
\overline{MRR}(r) = \frac{h_{Cu} + S_{Cu/ox} h_o}{t_e (r)} \tag{2.28}
\]

39
When the pad translates continuously, $r_{cc}$ is a function of time, $t$, and $v_{cc}$ is the time derivative of $r_{cc}(t)$. Equation (2.26) should then be evaluated to obtain the displacement of the pad, $r_{cc}(t)$. It is important to note that while $\Delta h(r)$ is equated to a constant $h_{Cu} + S_{Cu/ox}h_{0}$, the right hand side of Eq. (2.26) still has $r$ dependencies.

The case when $k_p$ and $p$ are constants and $\omega_w = \omega_p$ is again considered. Equation (2.26) can then be expressed as

$$\frac{k_p P}{2\pi} \int_{0}^{t(r)} \int_{-\theta_2(r,r_0)}^{\theta_2(r,r_0)} \sqrt{\left[ \frac{dr_{cc}(t)}{dt} \right]^2 + \left[ -\omega_p r_{cc}(t) \right]^2} \, d\theta \, dt = h_{Cu} + S_{Cu/ox}h_{0}$$  (2.29)

If the pad translates slowly, the entrance and exit semi-contact angles will be nearly constant per wafer rotation and the time derivative of $r_{cc}(t)$ will be small. Assuming $\theta_1 \approx \theta_{e2}$ and neglecting $v_{cc}$, the inner integral can be evaluated:

$$\frac{k_p P}{\pi} \int_{0}^{t(r)} \theta_0(r, r_{cc}(t)) \omega_p r_{cc}(t) \, dt = h_{Cu} + S_{Cu/ox}h_0$$  (2.30)

From Eq. (2.2), $\theta_c$ is a piecewise function with a constant $\pi$ term, a $\cos^{-1}$ term, and a $0$ term. Therefore, Eq. (2.30) can be rewritten as a sum of three integrals with one term being zero:

$$\pi \int_{0}^{t_{e1}(r)} r_{cc}(t) \, dt + \int_{t_{e1}(r)}^{t_{e2}(r)} \cos^{-1} \left( \frac{r^2 + r_{cc}(t)^2 - r_p^2}{2rr_{cc}(t)} \right) r_{cc}(t) \, dt = \frac{(h_{Cu} + S_{Cu/ox}h_0)\pi}{k_p p\omega_p}$$  (2.31)

where $t_{e1}$ and $t_{e2}$ are $\theta_c$ transition times at $r$:

$$r_{cc}(t) \leq r_p - r, \quad 0 \leq t \leq t_{e1}$$
$$r_p - r < r_{cc}(t) \leq r_p + r, \quad t_{e1} < t \leq t_{e2}$$
$$r_{cc}(t) > r_p + r, \quad t > t_{e2}$$  (2.32)

Equation (2.31) assumes that the pad is large enough to initially contact the edge of the wafer, $r_p > r_w - r_{cc}$, and translates unidirectionally away from the center of the wafer. In the case of a smaller pad that does not initially reach the edge of the wafer, there will be additional regions of $\theta_c = 0$ and the transition times should be redefined to reflect that.

Equation (2.31) cannot be simply evaluated because the nature of the function $r_{cc}(t)$ is unknown. The task is even more difficult due to $t_{e1}(r)$ and $t_{e2}(r)$ also being dependent on $r_{cc}(t)$. Thus, numerical methods are appropriate for the resolution of pad translational motion for uniform material removal.
2.5 Numerical Model for the Pad Translational Velocity

Pad translation can be described by the change in wafer center to pad center distance over time: \( r_{cc}(t) \). The numerical approach to obtaining \( r_{cc}(t) \) discretizes the wafer radius into a set of points where the total material removed will be computed. The pad displacement is also discretized into steps. The model then computes the duration of time the pad needs to stay at each translation step in order to completely polish the wafer at all the discretized radial locations. By reducing the size of the translation steps, it is possible to obtain an estimate of the continuous pad translation function, \( r_{cc}(t) \) [Mau et al., 2008].

2.5.1 Discretization

The first step in the numerical procedure involves the discretization of radial location, \( r \), and \( r_{cc} \). First, the initial pad position is defined as \( r_{cc_0} \). To ensure that the center of the wafer is completely polished, \( r_{cc_0} \) should be so chosen that the pad overlaps the center of the wafer, as in Figure 2.4(a). From Figure 2.5, when \( r_{cc_0} < r_p \), the overlap region on the wafer is always in contact with that pad. Hence \( \theta_e \), or material removal rate, is constant. The radius of the central region that will polish concurrently is given by:

\[
   r_{overlap} = r_p - r_{cc_0}
\]

(2.33)

To avoid redundancy, \( r \) should be discretized so that only one point, \( r_0 \), is within the overlap region. Therefore, let

\[
   r_0 \equiv r_{overlap}
\]

(2.34)

Now, \( r_1, \ldots, r_i, \ldots, r_n \) can be arbitrarily chosen as long as \( r_i > r_0 \) and the points span the entire wafer. Because that \( r \) and \( r_{cc} \) are coupled through \( \theta_e \), the discretization of \( r_{cc} \) should be dependent on \( r \). From Figure 2.5, when the pad is away from the center of the wafer, \( \theta_e \) is zero at the edge of the pad, and hence, \( MRR \) is also zero. Therefore, to avoid underpolishing between \( r_i \) and \( r_{i+1} \), \( r_{cc_i} \) should be so chosen that the edge of the pad is at \( r_i \) for every translation step after the pre-defined \( r_{cc_0} \):

\[
   r_{cc_{i+1}} = r_i + r_p, \quad i = 0, 1, \ldots, n - 1
\]

(2.35)
Figure 2.9 shows a schematic of the wafer radius and the pad translation discretized into six points each as defined by Eqs. (2.33)-(2.35). Note that \( r_1, \ldots, r_n \) need not be uniformly spaced as long as the edge of the pad lies on each point on the discretized wafer radius.

### 2.5.2 Matrix Formulation

The objective of the matrix formulation is to obtain the duration, \( \Delta t_j \), the pad should reside at each position, \( r_{ce,j} \), to guarantee that the radial location, \( r_b \), is completely polished. Moreover, since translation occurs in steps, \( v_{cc} = 0 \) in each time interval \( \Delta t_j \).

For a particular pad position, \( r_{ce,j} \), the material removal rate at \( r_i \) can be calculated by Eq. (2.17). Thus, the total thickness of material removed at \( P(r_i) \) by a discretely translating pad is the sum of the material removed by the pad at each position

\[
\Delta h_i = \sum_{j=0}^{n} \overline{MRR}(r_i, r_{ce,j}) \cdot \Delta t_j, \quad \text{for } i = 0, 1, \ldots, n
\]  

where \( n \) is the number of pad translation steps. To completely remove the Cu layer, the total thickness of material removed should be equal to the initial Cu thickness, \( h_{Cu} \). In practice, there is some deliberate overpolishing to deal with the uncertainties of the process and ensure that the Cu is completely removed from the wafer. This is incorporated into the model by \( S_{Cu/ox} h_o \), the equivalent thickness of Cu removed during overpolishing. Finally, a set of equations can be written as:

\[
\sum_{j=0}^{n} \overline{MRR}(r_i, r_{ce,j}) \cdot \Delta t_j = h_{Cu} + S_{Cu/ox} h_o
\]  

for \( i = 0, 1, \ldots, n \).

Equation (2.37) is expressed in matrix form as:

\[ \begin{bmatrix} \overline{MRR}(r_0, r_{ce_0}) & \cdots & \overline{MRR}(r_0, r_{ce_n}) \\ \vdots & \ddots & \vdots \\ \overline{MRR}(r_n, r_{ce_0}) & \cdots & \overline{MRR}(r_n, r_{ce_n}) \end{bmatrix} \begin{bmatrix} \Delta t_0 \\ \vdots \\ \Delta t_n \end{bmatrix} = \begin{bmatrix} h_{Cu} + S_{Cu/ox} h_o \end{bmatrix} \]  

(2.38)
Figure 2.9: Schematic of the wafer radius and pad translation discretization.
where $M$ is an $(n+1) \times (n+1)$ square matrix of material removal rates, $t$ a set of time intervals, and $h$ a set of initial Cu thickness.

It may be noted that the rows of $M$ correspond to the $r_i$ parameter while the columns are related to $r_{ccj}$. The unknown in Eq. (2.38) is $\Delta t_j$, where each entry $\Delta t_j$ represents the time the pad stays at a particular $r_{ccj}$. Figure 2.10 is a flow chart summarizing the steps in the pad translation algorithm.

Each individual term of the matrix, $M_{ij}$, is the average MRR at $r_i$ when the pad is at $r_{ccj}$, which can be found by Eq. (2.17). Since there is only one solution to Eq. (2.38), it is possible to have a matrix $M$ such that negative $\Delta t$'s are required to satisfy the equality. Physically, this means that overpolishing will occur at some $r_i$'s and uni-directional pad translation is not feasible for the chosen $r_i$'s and $r_{ccj}$'s. Negative solutions are more likely to occur when random discretization schemes are used. For example, if $r_w$ is discretized in clusters of points and the pad translation steps are large, as in Figure 2.11, one translation step will be required to satisfy several $\Delta h$ equations. The appropriate discretization of $r$ and $r_{cc}$ is therefore crucial in obtaining $M$ such that the resulting $\Delta t$ has only positive terms.

The matrix $M$ can be further simplified if $k_p$ and $p$ are constants and $\omega_u = \omega_p$:

\begin{equation}
\frac{k_p p \omega}{\pi} \begin{bmatrix}
  r_{cc_0} \theta_c(r_0, r_{cc_0}) & \cdots & r_{cc_a} \theta_c(r_0, r_{cc_a}) \\
  \vdots & \ddots & \vdots \\
  r_{cc_0} \theta_c(r_n, r_{cc_0}) & \cdots & r_{cc_a} \theta_c(r_n, r_{cc_a})
\end{bmatrix} \begin{bmatrix}
  \Delta t_0 \\
  \vdots \\
  \Delta t_n
\end{bmatrix} = \begin{bmatrix}
  h_{Cu} + S_{Cu/ox} \omega_i h_{ox}
\end{bmatrix} \begin{bmatrix}
  1 \\
  \vdots \\
  1
\end{bmatrix}
\end{equation}

Equation (2.39) shows that MRR is dependent on two geometric parameters, $\theta_c$ and $r_{cc}$. It is important to also note that $\theta_{cy}$ is a non-linear function of $r_i$, $r_{ccj}$, and $r_p$ as defined by Eq. (2.2).
Choose \( r_{cc0} \)  

Define \( r_0 \) according to \( r_{cc0} \) and \( r_p \)  

Discretize \( r \) into \( n \) increments  

Discretize \( r_{cc} \) according to \( r_i \)  

Calculate \( MRR \) for each set of \( r_i \) and \( r_{ccj} \)  

Build matrix \( M \) where entry \( M_{ij} \) is the \( MRR \) at \( r_i \) when pad is at \( r_{ccj} \)  

Build vector \( h \) such that each entry \( h_i \) equals the desired \( Cu \) removal at \( r_i \)  

Solve vector equation \( M*\Delta t = h \) for column vector \( \Delta t \)  

\( \Delta t \) represents the time interval the pad should stay at \( r_{ccj} \)  

Figure 2.10: Flow chart summarizing a numerical method for determining pad translation in face-up CMP.

\[
M = \begin{bmatrix}
M_{00} & M_{00} & 0 & 0 \\
M_{10} & M_{10} & 0 & 0 \\
M_{20} & M_{20} & 0 & 0 \\
M_{30} & M_{30} & M_{30} & M_{30}
\end{bmatrix}
\]

Figure 2.11: A discretization scheme involving clusters of points.
To express Eq. (2.39) in dimensionless form, the following variables are introduced:

\[
\begin{align*}
    r_i^* &\equiv r_i / r_w \\
    r_{cc_i}^* &\equiv r_{cc_i} / r_w \\
    \theta_{cc_i}^* &\equiv \theta_{cc_i} / \pi \\
    \Delta t_j^* &\equiv \Delta t_j / \Delta t_0
\end{align*}
\]  (2.41)

where \( \Delta t_0 \) is the time interval the pad stays at the initial position \( r_{cc_i} \), or the time required for the center of the wafer to be fully polished. When there is an initial overlap of the pad at the wafer center, \( \theta_c(r = 0) = \pi \). From Eq. (2.17),

\[
\Delta t_0 = \frac{h_c + S_{ctc} h_o}{k_p \cdot p \cdot \omega \cdot r_{cc_i}}
\]  (2.42)

Thus the dimensionless form of Eq. (2.40) is:

\[
\begin{bmatrix}
    \theta_{cc_0}^* & \cdots & \theta_{cc_a}^* \\
    \vdots & \ddots & \vdots \\
    \theta_{cc_a}^* & \cdots & \theta_{cc_a}^*
\end{bmatrix}
\begin{bmatrix}
    r_{cc_0}^* \\
    \vdots \\
    r_{cc_a}^*
\end{bmatrix}
= \begin{bmatrix}
    \Delta t_0^* \\
    \vdots \\
    \Delta t_a^*
\end{bmatrix}
\begin{bmatrix}
    1 \\
    \vdots \\
    1
\end{bmatrix}
\]  (2.43)

where \( \mathbf{M}^* = \theta_c^* \cdot r_{cc}^* \). In this form, the dependence of \( \text{MRR} \) on process parameters, such as \( k_p, p, \) and \( \omega \), only appears within the normalizing constant, \( \Delta t_0 \). The pad radius, however, is integrated into the matrix through the computation of \( \theta_c^* \) and the discretization of \( r_{cc} \).

If \( \theta_c \), and therefore \( \theta_c^* \), is triangular, the resulting \( \mathbf{M} \) matrix will also be triangular. By geometry, it is possible to obtain a lower triangular \( \theta_c \) matrix. As the pad moves away from the wafer center, it ceases contact with that region, resulting in \( \theta_c = 0 \) for those points. Since the rows of \( \theta_c \) represent increasing \( r \) and the columns represent increasing \( r_{cc} \), if the pad leaves one discretization point at every translation step, every column will have one additional zero at the top row(s) than the column to its left – creating an upper-right corner of zeros. The pad translation necessary for the formulation of a lower triangular matrix is shown in Figure 2.12.

The ability to obtain a triangular \( \mathbf{M} \) matrix is already in place under the discretization scheme given by Eqs. (2.33)-(2.35). Figure 2.9 shows how the pad terminates contact with one radial point at every translation step.
Figure 2.12: Schematic of a translating pad and the subsequent triangular matrix formulation.
Equation (2.43) can then be rewritten as:

\[
\begin{bmatrix}
\theta_{c_0}
& 0 \\
\vdots & \ddots & \ddots \\
\theta_{c_m}
& \cdots 
& \theta_{c_m}
\end{bmatrix}
\begin{bmatrix}
r_{c_0}^* \\
\vdots \\
r_{c_m}^*
\end{bmatrix}
= \begin{bmatrix}
\Delta t_0^* \\
\vdots \\
\Delta t_n^*
\end{bmatrix}
\begin{bmatrix}
1 \\
\vdots \\
1
\end{bmatrix}
\]

(2.44)

A lower triangular \( \theta_c^* \) matrix simplifies the solution of Eq. (2.43) by facilitating forward substitution. By definition, \( \Delta t_0^* = 1 \). Solving the equation formed by the first row of the matrix will result in the same solution:

\[
\theta_{c_0}^* \cdot r_{c_0}^* \cdot \Delta t_0^* = r_{c_0}^*
\]

\[
\Delta t_0^* = \frac{\pi}{\pi} = 1
\]

The equation obtained from multiplying the second row of \( \theta_c^*(r) \) will then only contain the unknown \( \Delta t_i^* \). Solving for \( \Delta t_i^* \) will leave \( \Delta t_{i+1}^* \) as the only unknown in the third equation, and so on. Thus

\[
\begin{align*}
\Delta t_0^* &= 1 \\
\theta_{c_0}^* \cdot r_{c_0}^* \cdot \Delta t_0^* + \theta_{c_1}^* \cdot r_{c_1}^* \cdot \Delta t_1^* &= r_{c_0}^* \\
\theta_{c_2}^* \cdot r_{c_2}^* \cdot \Delta t_0^* + \theta_{c_1}^* \cdot r_{c_1}^* \cdot \Delta t_1^* + \theta_{c_2}^* \cdot r_{c_2}^* \cdot \Delta t_2^* &= r_{c_0}^* \\
&\vdots \nonumber \\
\theta_{c_n}^* \cdot r_{c_n}^* \cdot \Delta t_0^* + \theta_{c_{n-1}}^* \cdot r_{c_{n-1}}^* \cdot \Delta t_{n-1}^* + \cdots + \theta_{c_2}^* \cdot r_{c_2}^* \cdot \Delta t_2^* + \theta_{c_1}^* \cdot r_{c_1}^* \cdot \Delta t_1^* + \theta_{c_0}^* \cdot r_{c_0}^* \cdot \Delta t_0^* &= r_{c_0}^*
\end{align*}
\]

(2.46)

Figure 2.13 compares the computed dimensionless pad location as a function of polishing time for various wafer-pad rotational velocity ratios. For all cases \( k_p \) and \( p \) are constants and \( r_p = 0.7r_w, \ r_{cc_n} = r_p - 0.1r_w, \) and \( n = 50. \) It was shown in the non-translating pad case that the polishing gradient is steeper if \( \omega_w / \omega_p = 0.5 \). Therefore, the pad translation motion is more gradual. The steeper gradient offers easier control over pad velocity since the pad stays at each \( r_{cc} \) for a longer period of time. However, this control is gained at the expense of longer polishing times. The non-translating pad case also shows that a large portion of the wafer will polish at nearly the same time when the larger ratio is used, so for quicker pad motion and shorter polishing times, \( \omega_w / \omega_p = 1.5 \) is appropriate. Because of the higher velocities involved, tighter control of pad displacement is required for the desired results.
Figure 2.13: Normalized pad location versus polishing time for various rotational velocity ratios.
2.5.3 An Example

Let \( r_w = 50 \text{ mm} \), \( r_p = 35 \text{ mm} \) and \( r_{co} = 30 \text{ mm} \). By Eq. (2.34), \( r_o = 5 \text{ mm} \). The rest of the wafer from \( r_0 \) to \( r_w \) is then evenly divided into five increments: \( r = \{5, 14, 23, 32, 41, 50\} \text{ mm} \).

Using Eq. (2.35), \( r_{co} = \{30, 40, 49, 58, 67, 76\} \text{ mm} \). The dimensionless values will be

\[
\begin{bmatrix}
0.10 \\
0.28 \\
0.46 \\
0.64 \\
0.82 \\
1.00 \\
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
0.60 \\
0.80 \\
0.98 \\
1.16 \\
1.34 \\
1.52 \\
\end{bmatrix}
\]

(2.47)

A schematic of the discretized wafer radius and pad translation is shown in Figure 2.14 along with \( \theta_{c}^*(r^*) \) for each \( r_{co} \). From Eq. (2.43),

\[
\begin{bmatrix}
0.60 \\
0.33 & 0.26 \\
0.27 & 0.27 & 0.22 \\
0.23 & 0.25 & 0.25 & 0.20 \\
0.19 & 0.23 & 0.24 & 0.24 & 0.20 \\
0.15 & 0.20 & 0.23 & 0.24 & 0.23 & 0.19 \\
\end{bmatrix}
\]

(2.48)

As noted previously, the zeros in the upper-right corner of \( \mathbf{M}^* \) represent the absence of contact between the pad and the wafer.

Solving for \( \Delta t^* \) by Eq. (2.46) will result in

\[
\begin{bmatrix}
1.00 \\
1.03 \\
0.23 \\
0.26 \\
0.30 \\
0.37 \\
\end{bmatrix}
\]

(2.49)

which represents the dimensionless time that the pad remains at each translation step. Equation (2.49) can be used to control the pad motion.
Figure 2.14: Schematic of the discretization of wafer radius and pad translation in the five-step example and the corresponding $\theta_c(r)$. 

\[ r' = r / r_w \]
Figure 2.15 plots $r_e^*$ versus $\Delta t^*$ for this five-step example. Points A and B represent the time the pad remains at its initial position before the center of the wafer is clear. The pad then translates quickly from B to C as it moves away from the center overlap region that is completely polished. From C to D is another long timestep due to the steep decrease in $\theta_c$ when the pad moves away from the center of the wafer to $r_{c_{1_{st}}}$, as shown in Figure 2.14. Lastly, the pad translates more or less uniformly from D to E to finish polishing the rest of the wafer.

Assuming some common face-up polishing parameters: $h_{Cu} = 1 \mu m$, $h_o = 0 \mu m$, $k_p = 5.0 \times 10^{-13} 1/\text{Pa}$, $p = 13 \text{kPa}$, and $\omega = 16 \text{rad/s}$, the time for the center to polish is $\Delta t_0 = 320 \text{s}$ from Eq. (2.42). This value can be applied to the dimensionless result in Eq. (2.49) to obtain the polishing time for each translation step. Figure 2.16 plots the computed pad translational motion for this five-step example. From the obtained time durations, the thickness of Cu remaining on the wafer after every pad translation step is computed and shown in Figure 2.17.

2.5.4 Discretization Error

The set of $\Delta t^*$'s obtained from Eq. (2.38) guarantees that exactly $h_{Cu} + S_{Cu/r_c} h_o$ coating thickness is removed at each $r_i$. The model, however, does not consider what happens between $r_i$ and $r_{i+1}$ and errors may occur in that region. After solving for $\Delta t$, the thickness of material removed at any point $P(r)$ along the wafer radius can be obtained by summing the material removed at each translation step, as in Eq. (2.36), up until $r_{c_{i_{st}}}$. From the Eq. (2.35), $r_{c_{i_{st}}}$ is positioned so that the edge of the pad is at $r_i$. Beyond this step, the pad no longer contacts the annulus between $r_i$ and $r_{i+1}$ and no material removal takes place (Figure 2.9). In the case where $k_p$ and $p$ are constants and $\omega_o = \omega_p$,

$$\Delta h(r) = \frac{k_p p \omega_{o}}{\pi} \sum_{j=0}^{i_{st}} r_{c_j} \cdot \Delta t_j \cdot \theta_c(r, r_{c_j})$$ (2.50)

for $r_i \leq r \leq r_{i+1}$.

Since $\Delta h(r_i) = \Delta h(r_{i_{st}}) = h_{Cu} + S_{Cu/r_c} h_o$, if $\Delta h$ is a non-constant, continuous function, a maximum or minimum must exist between $r_i$ and $r_{i+1}$. The extremum represents the largest
Figure 2.15: Dimensionless pad translational motion for the five-step example.

Figure 2.16: Pad location versus polishing time for the five-step example.
Figure 2.17: Plots of the Cu remaining on the wafer after each pad translation step.
deviation from the endpoints, and therefore the largest polishing error. Because the discretization scheme places the minimum $\theta_c$ at $r_i$, $\Delta h$ must increase from that point, indicating that the extremum is a maximum and the error describes overpolishing. To find the point where the maximum occurs, Eq. (2.50) is differentiated with respect to $r$ and set equal to zero:

$$\frac{\partial \Delta h}{\partial r} = \frac{k_p \rho \omega}{\pi} \sum_{j=0}^{i+1} r_{cc_j} \cdot \Delta t_j \cdot \frac{\partial \theta_c(r, r_{cc_j})}{\partial r} = 0 \quad (2.51)$$

Solving Eq. (2.51) for $r$ will result in the point of maximum polishing, $r_{\text{max}}$, and subtracting $Ah(r_i)$ from $Ah(r_{\text{max}})$ will yield the maximum overpolishing error.

By increasing the size of the matrix or the number of discretizations, $n$, the region between $r_i$ and $r_{i+1}$ is decreased. As a result, $\theta_c(r_{i+1})$ is close to $\theta_c(r_i)$ and the variation in material removal rate between the two points will be small. The error is therefore expected to decrease as $n$ increases. Figure 2.18 compares the discretization error for different mesh sizes: $n = 5, 10, \text{and} 50$. The errors between $r_i$ and $r_{i+1}$ are obtained by finding $\Delta h$ by Eq. (2.50) for a fine mesh of points at 0.1 mm increments along the radius and comparing the result with $h_0$. When the discretized intervals become smaller, the magnitude of error between the wafer mesh points also decrease. Figure 2.18 shows, as expected, that there is no error in material removal at the points along the discretized wafer radius.

From previous discussion, the algorithm guarantees exact material removal at $r_i$. However, the computed polishing error does not always have to be 0%, since there can be deliberate overpolishing, $h_o > 0$. In those cases, the whole error curve will be shifted up by the percentage of specified overpolishing, as shown in Figure 2.19. Such computation is useful for deciding how much overpolishing time is allowable for the entire wafer to still meet specifications. The other parameters used in the simulations are: $r_w = 50 \text{ mm}$, $r_p = 35 \text{ mm}$, and $r_{cc_o} = 30 \text{ mm}$. Additionally, $k_p$ and $p$ are assumed constant, and $\omega_w = \omega_p$.

Computing the discretization error is also valuable when comparing different discretization schemes. The scheme described by Eqs. (2.33)-(2.35) is designed to avoid underpolishing between $r_i$'s. However, the matrix $\mathbf{M}$ in Eq. (2.38) can be formulated with any set of $r_i$'s and $r_{cc_j}$'s. For example, consider the following two discretization schemes: uniform intervals and no underpolishing as outlined by Eqs. (2.33)-(2.35). In the uniform intervals scheme, $r_w$ was divided into 5 increments of 10 mm, and $r_{cc}$ was defined as 10 mm steps starting.
Figure 2.18: Discretization error for different mesh sizes.

Figure 2.19: Error with 2% deliberate overpolishing, $S_{Cu/ox}h_o = 0.02h_{Cu}$.
from $r_{cc}$. In the no underpolishing scheme, $r_0 = 5$ mm by Eq. (2.34). The rest of the wafer, from $r_0$ to $r_w$ was evenly divided into 5 sections. The values for $r_{cc_1}$ to $r_{cc_5}$ was obtained using Eq. (2.35). Table 2.1 lists the discretized values for both methods. Figure 2.20 compares the errors from the two discretization schemes. The length parameters for the simulations are $r_w = 50$ mm, $r_p = 35$ mm and $r_{cc_0} = 30$ mm.

Although the magnitude of error is similar for both cases, in practice it is required to avoid underpolishing. The excess metal on an underpolished wafer can lead to electrical shorts, resulting in unusable dies and reduced yield. While overpolishing may still cause defects, small amounts of overpolishing is necessary to counteract the uncertainties of the process and ensure that all the Cu overburden is completely removed. Moreover, Figure 2.18 has shown that overpolishing can be controlled by using finer discretizations in the model. A discretization scheme like the one outlined by Eqs. (2.33)-(2.35) is therefore the more practical choice in pad translation computation.

2.6 Summary

This chapter introduces the face-up CMP tool architecture and presents a numerical method for determining a pad translation scheme that would minimize overpolishing by face-up polishing. First, the geometric relations between the pad and the wafer are described to define the time of contact between a point on the wafer and the pad. Second, a kinematic analysis is used to obtain the relative velocity of a point on the wafer with respect to the pad. Third, the geometrical and kinematical parametes are incorporated into the Preston equation for local material removal rate to express the material removed at a point on the wafer per wafer revolution, which can then be used to obtain the average material removal rate at that point.

Two cases of pad motion are discussed: non-translating pad and translating pad. In the former case, $k_p, p, \omega_w$, and $\omega_p$ are assumed to be constants. Since the pad is not translating, $r_{cc}$ is also constant and hence, $v_{cc} = 0$. If the pad rotates at the same speed as the wafer, the resulting material removal rate is only dependent on the semi-contact angle, $\theta_c$. The polishing gradient is controlled by the change in contact angle across the wafer.
Table 2.1: Discretized $r$ and $r_{cc}$ values from two different discretization schemes.

<table>
<thead>
<tr>
<th>Index, $i$</th>
<th>Uniform Intervals</th>
<th>No Underpolishing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r_i$ (mm)</td>
<td>$r_{cc}$ (mm)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>70</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>80</td>
</tr>
</tbody>
</table>

Figure 2.20: Comparison of errors from two different discretization schemes.
For the pad translation case, the objective is to obtain the pad displacement with time, $r_{cc}(t)$, for uniform material removal across the wafer. Even when $k_p$, $p$, $\omega_w$, and $\omega_p$ are constant, the material removal rate cannot be solved analytically due to the dependence of the time and geometric parameters on the unknown, $r_{cc}(t)$. Therefore, a numerical approach is developed. In this method, the wafer radius is discretized into a set of points. The pad displacement is treated as a series of steps so that the non-translating pad material removal rates can be used to compute the material removed at each step. Finally, the discretization error is evaluated and its effects on overpolishing are obtained. The polishing error is used to show the effects of overpolishing and to compare different discretization schemes.
Nomenclature

\( h \) = thickness of Cu coating removed (m)
\( h_{Cu} \) = initial Cu coating thickness (m)
\( h_o \) = thickness of oxide removed during deliberate overpolishing (m)
\( \Delta h_r, \Delta h_r^* \) = thickness of coating removed in one wafer rotation (m), normalized value
\( \Delta h \) = total thickness of coating removed (m)
\( k_p \) = Preston constant (m²/N)
\( MRR, MRR^* \) = material removal rate, average per wafer rotation (m/s)
\( p \) = pressure (N/m²)
\( O_w, O_p \) = center of the wafer and the pad
\( P, P' \) = point in the wafer coordinate system, pad coordinate system
\( r, \theta \) = coordinates in the wafer polar coordinate system
\( r', \theta' \) = coordinates in the pad polar coordinate system
\( r_i \) = discretized radial wafer coordinate (m)
\( r_{cc}, r_{cc}^* \) = wafer center to pad center distance (m) and normalized value
\( r_{cc_i} \) = discretized wafer center to pad center distance (m)
\( r_{max} \) = point of maximum material removal between \( r_i \) and \( r_{i+1} \) (m)
\( r_{overlap} \) = radius of central area on wafer that will polish concurrently due to the pad/wafer center overlap (m)
\( r_w, r_p, r_p^* \) = wafer radius, pad radius (m), and normalized pad radius
\( S_{Cu/ox} \) = Cu to oxide slurry selectivity
\( t, t^* \) = polishing time (s) and normalized value
\( t_e \) = time required to reach the process endpoint (s)
\( t_{e1}, t_{e2} \) = semi-contact angle transition times (s)
\( \Delta t_0 \) = time for center of the wafer to be completely polished (s)
\( \Delta t_j, \Delta t_j^* \) = time duration of pad translation step \( j \) (s) and normalized value
\( v_{cc} \) = translational velocity of the pad; rate of change of \( r_{cc} \) (m/s)
\( v_R \) = relative velocity of the wafer with respect to the pad (m/s)
\( \theta_c, \theta_c^* \) = semi-contact angle (rad) and normalized value
\( \theta_{c1}, \theta_{c2} \) = entrance and exit semi-contact angle (rad)
\( \omega_w, \omega_p \) = angular velocities of the wafer and the pad (rad/s)
CHAPTER 3

POLISHING EXPERIMENTS WITH A NON-TRANSLATING PAD

3.1 Introduction

A steep polishing gradient is fundamental for progressive polishing [Saka and Chun, 2007], and therefore the material removal rate for a non-translating pad must first be validated. Accordingly, polishing experiments were conducted on blanket Cu wafers to measure the material removal rate and Preston constant across the wafer. Two pad perforation patterns were used to optimize slurry distribution at the pad-wafer interface, and hence the Preston constant. Experiments were conducted at various wafer-pad rotational velocity ratios to compare their effect on the polishing gradient across the wafer. Finally, experiments were performed with the pad positioned at different distances from the center of the wafer to compare the change in Preston constants.

3.2 Equipment and Consumables

All polishing experiments were performed on a rotary face-up CMP tool, shown in Figure 3.1. The 100-mm wafers were held with the coating facing upward in a vacuum chuck. A perforated pad was attached to a slurry cup and the normal load was applied by compressed air. Slurry was supplied to the cup by a peristaltic pump. Both the wafer and pad were rotated in the same direction during polishing. To prevent interruption during the course of polishing, a video camcorder captured the experiments and the video clips were used to analyze the areas of Cu removal afterwards.

Commercial stacked pads manufactured by Thomas West, Inc. were used to polish the wafers. The proprietary pad face material, TWI-817, was made of polymer-impregnated fibers, and the SP-7 subpad was of felt-type material. X-Y grooves on the pad face aid in distributing the slurry across the surface during polishing. A schematic of the pad cross section is shown in Figure 3.2. The face-up CMP architecture requires the delivery of slurry through perforations in the pad; however, this feature is not yet available commercially. Therefore, holes were punched
Figure 3.1: Photograph of the face-up CMP tool.

Figure 3.2: Schematic of the cross section of a TWI-817 stacked pad.
into the pad at the center of the squares formed by the X-Y grooves using a stainless steel punch and a drill press. Due to the lack of in-situ conditioning which has been shown to extend pad life [Muldowney and James, 2004], new pads were used for each experiment. This ensured that the material removal rates were comparable between experiments. Prior to polishing, the pads were conditioned with deionized water and a stiff nylon brush. This procedure revitalized the pad surface and removed the debris left on the surface from hole-punching.

Cabot Microelectronics iCue 5001 Cu CMP slurry, was used for the polishing. The slurry contained fumed alumina particles with an average particle diameter of 2.8 μm that make up 3% of the total slurry volume. Hydrogen peroxide making up 3% the total volume was added to the slurry and the pH of the mixture was measured prior to usage.

3.3 Kinematic Effects of Slurry Cup Rotation

The face-up CMP tool architecture addresses the issue of uniform slurry flow to the pad-wafer interface by supplying slurry through perforations in the pad. In order to do so, the pad must be attached to a slurry cup as shown in Figure 3.3. To investigate the kinematic effects of the cup rotation on slurry distribution and material removal rate, polishing experiments were performed on blanket Cu wafers using pads with different perforation patterns: uniformly spaced and sized perforations as shown in Figure 3.4(a) and central region blocked, Figure 3.4(b) [Mau et al., 2007]. The process parameters were kept constant for the two tests to isolate the effects of slurry flow. These parameters are shown in Table 3.1 and the material properties of the coatings, abrasive particles, and pad are listed in Table 3.2. Due to the surface topography of the pad, the measured local material properties can span a wide range of values [Eusner, 2008].

Table 3.3 lists the polishing time, \( t \), the radial location of the cleared region, \( r \), and the calculated values of semi-contact angle, \( \theta_c \), and Preston constant, \( k_p \), from Eqs. (2.2) and (2.22), respectively. When polishing with a uniformly perforated pad, the Cu was first removed in an annular pattern at \( r = 27 \) mm and the polished area extended in both directions as time progressed. One minute after the initial substrate opening, the Cu at the center of the wafer was cleared, creating discontinuous exposed SiO\(_2\) regions. Finally, the two regions merged to form the final circular polished region. The polishing pattern can be seen in the video screenshots in Figure 3.5.
Figure 3.3. Photograph of the slurry cup.

Figure 3.4: (a) Uniformly spaced and (b) blocked center pad perforation patterns.
Table 3.1: Wafer, consumables, and process parameters for polishing experiments.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{Cu}$ (μm)</td>
<td>1.0</td>
</tr>
<tr>
<td>$r_w$ (mm)</td>
<td>50</td>
</tr>
<tr>
<td>Pad type</td>
<td>TWI-817</td>
</tr>
<tr>
<td>$r_p$ (mm)</td>
<td>35</td>
</tr>
<tr>
<td>$r_{cc}$ (mm)</td>
<td>30</td>
</tr>
<tr>
<td>Slurry</td>
<td>Cabot iCue 5001</td>
</tr>
<tr>
<td>Slurry additive</td>
<td>H$_2$O$_2$ - 3% vol</td>
</tr>
<tr>
<td>pH</td>
<td>8</td>
</tr>
<tr>
<td>$\omega_w$ (rad/s) (rpm)</td>
<td>19 (180)</td>
</tr>
<tr>
<td>$\omega_p$ (rad/s) (rpm)</td>
<td>19 (180)</td>
</tr>
<tr>
<td>$v_{cc}$ (m/s)</td>
<td>0</td>
</tr>
<tr>
<td>$p$ (kPa) (psi)</td>
<td>17 (2.4)</td>
</tr>
<tr>
<td>Slurry flow rate (ml/min)</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 3.2: Mechanical properties of the materials involved in CMP.

<table>
<thead>
<tr>
<th>Material</th>
<th>$E$ (GPa)</th>
<th>$H$ (GPa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu</td>
<td>128.0</td>
<td>1.22</td>
</tr>
<tr>
<td>Ta</td>
<td>186.0</td>
<td>0.80</td>
</tr>
<tr>
<td>SiO$_2$</td>
<td>92.0</td>
<td>15.0</td>
</tr>
<tr>
<td>Al$_2$O$_3$</td>
<td>350.0</td>
<td>20.0</td>
</tr>
<tr>
<td>TWI-817 Pad (wet)$^+$</td>
<td>0.09 – 0.68</td>
<td>0.02 – 0.17</td>
</tr>
</tbody>
</table>

$^+$[Eusner, 2008]
Table 3.3: Uniform pad polishing results with calculated $\theta_c$ and $k_p$.

<table>
<thead>
<tr>
<th>$t$ (s)</th>
<th>$\Delta t_0 / t$</th>
<th>$r$ (mm)</th>
<th>$r / r_w$</th>
<th>$\theta_c$ (rad)</th>
<th>$\Theta / \pi$</th>
<th>$k_p \times 10^{-13}$ Pa$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>420</td>
<td>1.14</td>
<td>25</td>
<td>0.50</td>
<td>1.37</td>
<td>0.44</td>
<td>5.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>29</td>
<td>0.57</td>
<td>1.28</td>
<td>0.41</td>
<td>6.19</td>
</tr>
<tr>
<td>480</td>
<td>1.00</td>
<td>4</td>
<td>0.08</td>
<td>3.14</td>
<td>0.94</td>
<td>2.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>0.30</td>
<td>1.67</td>
<td>0.53</td>
<td>4.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>36</td>
<td>0.72</td>
<td>1.11</td>
<td>0.35</td>
<td>6.26</td>
</tr>
<tr>
<td>540</td>
<td>0.89</td>
<td>5</td>
<td>0.10</td>
<td>3.14</td>
<td>1.00</td>
<td>1.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13</td>
<td>0.26</td>
<td>1.77</td>
<td>1.77</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>40</td>
<td>0.80</td>
<td>1.01</td>
<td>1.01</td>
<td>6.10</td>
</tr>
<tr>
<td>600</td>
<td>0.80</td>
<td>7</td>
<td>0.14</td>
<td>2.28</td>
<td>0.73</td>
<td>2.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>0.24</td>
<td>1.82</td>
<td>0.58</td>
<td>3.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>42</td>
<td>0.84</td>
<td>0.96</td>
<td>0.31</td>
<td>5.78</td>
</tr>
<tr>
<td>660</td>
<td>0.73</td>
<td>9</td>
<td>0.18</td>
<td>2.06</td>
<td>0.65</td>
<td>2.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>0.20</td>
<td>1.97</td>
<td>0.63</td>
<td>2.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>44</td>
<td>0.88</td>
<td>0.91</td>
<td>0.29</td>
<td>5.53</td>
</tr>
<tr>
<td>720</td>
<td>0.67</td>
<td>45</td>
<td>0.90</td>
<td>0.89</td>
<td>0.28</td>
<td>5.20</td>
</tr>
<tr>
<td>780</td>
<td>0.62</td>
<td>46</td>
<td>0.93</td>
<td>0.85</td>
<td>0.27</td>
<td>5.01</td>
</tr>
</tbody>
</table>

Figure 3.5: Video screenshots from a polishing experiment with an unblocked pad.
Figure 3.6 shows the normalized thickness of material removed per wafer rotation, $\Delta h^*_r$, along the wafer radius and compares the experimental results with the calculated values from Eq. (2.24). It should be noted that Eq. (2.24) assumes that $k_p$ and $p$ are constant across the pad-wafer interface. The deviation between the experimental and predicted polishing gradients is a result of a variation in $k_p$ induced by non-uniform slurry distribution in the contact area. The initial ring of Cu removal indicates that the material removal rate was greatest near the center of the pad where the tangential velocity of the slurry in the cup was zero. Since the slurry at this region was essentially motionless, its downward gravitational flow was not hindered by the inertial forces from the rotating cup. As a result, more abrasive particles were delivered to this area and the material removal rate was enhanced.

The blocked pad in Figure 3.4(b) has no perforations at the center of the pad. Thus, slurry flow from the center of the pad, or the region of zero tangential velocity, is suppressed. Polishing with the blocked pad produced a uni-directional polishing gradient, which is necessary for face-up polishing. Table 3.4 summarizes the time and location of complete Cu removal for the blocked pad polishing experiment. The Cu at the center of the wafer first cleared at $t = 7$ min and extended outward as shown in the video screenshots, Figure 3.7. The initial radius of Cu clearance was 2 mm, which is less than the 5 mm pad-wafer center overlap. This indicates that the effective pad radius is smaller than the actual pad radius. Some causes for the pad edge effect are reduced applied pressure at the edge of the pad or reduced slurry flow due to the lack of perforations.

Figure 3.8 compares the experimental $\Delta h^*_r$ along the radius of the wafer with theoretical values computed from Eq. (2.24). The plot shows that polishing progressed uni-directionally as predicted by the theory. The disparity between the experimental and theoretical values can again be attributed to non-constant $k_p$. Table 3.4 also lists the calculated $k_p$ at various radii. While the modified pad produced desirable polishing results, the calculated $k_p$ still varied within the wafer, supporting the general notion that the Preston constant is difficult to manipulate. Nonetheless, the modified pad mitigated kinematical effects to some extent and provided better control of $k_p$ than the pad with uniformly spaced holes.
Material removal by a pad with uniformly spaced perforations compared to theoretical results; \( p = 17 \) kPa, \( \omega_w = \omega_r = 19 \) rad/s (180 rpm).

Table 3.4: Blocked pad polishing results with calculated \( \theta_c \) and \( k_p \).

<table>
<thead>
<tr>
<th>( t ) (s)</th>
<th>( \Delta t_0/t )</th>
<th>( r ) (mm)</th>
<th>( r/r_w )</th>
<th>( \theta_c ) (rad)</th>
<th>( \theta_c/\pi )</th>
<th>( k_p \times 10^{-13} ) Pa(^{-1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>420</td>
<td>1.00</td>
<td>2</td>
<td>0.04</td>
<td>3.14</td>
<td>1.00</td>
<td>2.52</td>
</tr>
<tr>
<td>480</td>
<td>0.88</td>
<td>4</td>
<td>0.09</td>
<td>3.14</td>
<td>1.00</td>
<td>2.21</td>
</tr>
<tr>
<td>540</td>
<td>0.78</td>
<td>8</td>
<td>0.16</td>
<td>2.15</td>
<td>0.68</td>
<td>2.87</td>
</tr>
<tr>
<td>600</td>
<td>0.70</td>
<td>27</td>
<td>0.53</td>
<td>1.32</td>
<td>0.42</td>
<td>4.21</td>
</tr>
<tr>
<td>660</td>
<td>0.64</td>
<td>37</td>
<td>0.73</td>
<td>1.08</td>
<td>0.34</td>
<td>4.67</td>
</tr>
<tr>
<td>720</td>
<td>0.58</td>
<td>39</td>
<td>0.78</td>
<td>1.03</td>
<td>0.33</td>
<td>4.47</td>
</tr>
<tr>
<td>780</td>
<td>0.54</td>
<td>43</td>
<td>0.87</td>
<td>0.94</td>
<td>0.30</td>
<td>4.55</td>
</tr>
<tr>
<td>840</td>
<td>0.50</td>
<td>44</td>
<td>0.89</td>
<td>0.91</td>
<td>0.29</td>
<td>4.33</td>
</tr>
</tbody>
</table>
Figure 3.7: Video screenshots from the blocked pad polishing experiment.

Figure 3.8: Material removal by a blocked pad compared to theoretical results; $p = 17$ kPa, $\omega_w = \omega_p = 19$ rad/s (180 rpm).
Figure 3.9 compares the variation in $k_p$ along the radius of the wafer for the two pads. Both cases showed an increase of $k_p$ with $r$, which also indicates that $MRR$ is greater at the outer edge of the wafer. Accordingly, the data points are above the predicted material removal curve in Figures 3.6 and 3.8. The Preston constant ranged from $2.0 \times 10^{-13}$ to $6.3 \times 10^{-13} \text{ 1/Pa}$ when polishing with the unblocked pad with a mean of $4.3 \times 10^{-13} \text{ 1/Pa}$, compared with $2.2 \times 10^{-13}$ to $4.7 \times 10^{-13} \text{ 1/Pa}$ and a mean of $3.7 \times 10^{-13} \text{ 1/Pa}$ when the blocked pad was used. Therefore, $k_p$ must be somewhat controlled to produce the uni-directional polishing gradient necessary for progressive polishing. Through further optimization of the perforation geometry, e.g., add or remove holes at the center, it may be possible to gain greater control over $k_p$.

### 3.4 Variation of $k_p$ with Angular Velocity Ratios

The polishing gradient can be further controlled by manipulating the relative angular velocity between the pad and the wafer as shown in Figure 2.8. Using pads with the blocked center perforation pattern, blanket Cu wafers were polished at different wafer to pad angular velocity ratios, $\omega_w/\omega_p$, to study its effect on material removal rate. The process parameters for these experiments are in shown Table 3.5.

Decreasing $\omega_w/\omega_p$ should increase the control of material removal by creating a steeper polishing gradient, thus allowing more time to translate the pad. Table 3.6 lists the time of polish of points along the radius of the wafer and the $k_p$ calculated from $MRR$. The steep gradient was attained partially in the central region of the wafer when $\omega_w/\omega_p = 0.5$, as shown in Figure 3.10. However, the issue of increasing $k_p$ towards the edge of the wafer continues to be present, and the outer region of the wafer polished much faster than the equations predicted.

According to Figure 2.8, increasing $\omega_w/\omega_p$ will result in a more gradual polishing gradient. This may be useful for one-step polishing where the pad does not need to translate in as many steps. Table 3.7 and Figure 3.11 show the polishing results when $\omega_w/\omega_p = 1.5$. The initial Cu removal occurred over approximately one-third of the wafer, then extended outward relatively slowly. The polishing gradient was therefore flatter near the center of the wafer. In this case, the initial overlap did not have the predicted effect on material removal rate at the center. It was expected that since $\theta_c$ was much larger at the overlap region, the Cu in the region...
Figure 3.9: Variation in $k_p$ across the wafer when polishing with different pads.

Table 3.5: Wafer, consumables, and process parameters for velocity ratio experiments.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\omega_w / \omega_p = 0.5$</th>
<th>$\omega_w / \omega_p = 1.0$</th>
<th>$\omega_w / \omega_p = 1.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{cw}$ (µm)</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>$r_w$ (mm)</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Pad type</td>
<td>TWI-817</td>
<td>TWI-817</td>
<td>TWI-817</td>
</tr>
<tr>
<td>$r_p$ (mm)</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>$r_{cc}$ (mm)</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Slurry</td>
<td>Cabot iCue 5001</td>
<td>Cabot iCue 5001</td>
<td>Cabot iCue 5001</td>
</tr>
<tr>
<td>Slurry additive</td>
<td>$\text{H}_2\text{O}_2$ - 3% vol</td>
<td>$\text{H}_2\text{O}_2$ - 3% vol</td>
<td>$\text{H}_2\text{O}_2$ - 3% vol</td>
</tr>
<tr>
<td>pH</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>$\omega_w$ (rad/s) (rpm)</td>
<td>10 (100)</td>
<td>19 (180)</td>
<td>24 (225)</td>
</tr>
<tr>
<td>$\omega_p$ (rad/s) (rpm)</td>
<td>21 (200)</td>
<td>19 (180)</td>
<td>16 (150)</td>
</tr>
<tr>
<td>$v_{cc}$ (m/s)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$p$ (kPa) (psi)</td>
<td>17 (2.4)</td>
<td>17 (2.4)</td>
<td>17 (2.4)</td>
</tr>
<tr>
<td>Slurry flow rate (ml/min)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 3.6: Results from the $\omega_w / \omega_p = 0.5$ polishing experiment.

<table>
<thead>
<tr>
<th>$t$ (s)</th>
<th>$\Delta t_0 / t$</th>
<th>$r$ (mm)</th>
<th>$r / r_w$</th>
<th>$\theta_c$ (rad)</th>
<th>$\theta_c / \pi$</th>
<th>$k_p \times 10^{-13}$ Pa$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>480</td>
<td>1.00</td>
<td>6</td>
<td>0.13</td>
<td>2.50</td>
<td>0.80</td>
<td>2.55</td>
</tr>
<tr>
<td>540</td>
<td>0.89</td>
<td>8</td>
<td>0.15</td>
<td>2.15</td>
<td>0.68</td>
<td>2.71</td>
</tr>
<tr>
<td>600</td>
<td>0.80</td>
<td>8</td>
<td>0.15</td>
<td>2.15</td>
<td>0.68</td>
<td>2.44</td>
</tr>
<tr>
<td>660</td>
<td>0.73</td>
<td>11</td>
<td>0.23</td>
<td>1.89</td>
<td>0.60</td>
<td>2.62</td>
</tr>
<tr>
<td>720</td>
<td>0.67</td>
<td>31</td>
<td>0.63</td>
<td>1.22</td>
<td>0.39</td>
<td>5.08</td>
</tr>
<tr>
<td>780</td>
<td>0.62</td>
<td>35</td>
<td>0.70</td>
<td>1.13</td>
<td>0.36</td>
<td>5.52</td>
</tr>
<tr>
<td>840</td>
<td>0.57</td>
<td>40</td>
<td>0.79</td>
<td>1.01</td>
<td>0.32</td>
<td>6.44</td>
</tr>
<tr>
<td>900</td>
<td>0.53</td>
<td>41</td>
<td>0.83</td>
<td>0.99</td>
<td>0.31</td>
<td>6.33</td>
</tr>
<tr>
<td>960</td>
<td>0.50</td>
<td>43</td>
<td>0.87</td>
<td>0.94</td>
<td>0.30</td>
<td>6.60</td>
</tr>
<tr>
<td>1020</td>
<td>0.47</td>
<td>44</td>
<td>0.88</td>
<td>0.91</td>
<td>0.29</td>
<td>6.57</td>
</tr>
<tr>
<td>1080</td>
<td>0.44</td>
<td>45</td>
<td>0.90</td>
<td>0.89</td>
<td>0.28</td>
<td>6.58</td>
</tr>
</tbody>
</table>

Figure 3.10: Comparison of experimental and theoretical $\Delta h^*_r$ when $\omega_w / \omega_p = 0.5$. 

72
Table 3.7: Results from the $\omega_w / \omega_p = 1.5$ polishing experiment.

<table>
<thead>
<tr>
<th>$t$ (s)</th>
<th>$\Delta t_0 / t$</th>
<th>$r$ (mm)</th>
<th>$r / r_w$</th>
<th>$\theta_c$ (rad)</th>
<th>$\theta_c / \pi$</th>
<th>$k_p$ (x $10^{-13}$ Pa$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>461</td>
<td>1.00</td>
<td>17</td>
<td>0.33</td>
<td>1.97</td>
<td>0.63</td>
<td>4.50</td>
</tr>
<tr>
<td>480</td>
<td>0.96</td>
<td>32</td>
<td>0.63</td>
<td>1.31</td>
<td>0.42</td>
<td>4.77</td>
</tr>
<tr>
<td>540</td>
<td>0.85</td>
<td>36</td>
<td>0.72</td>
<td>1.17</td>
<td>0.37</td>
<td>4.40</td>
</tr>
<tr>
<td>600</td>
<td>0.77</td>
<td>40</td>
<td>0.79</td>
<td>1.06</td>
<td>0.34</td>
<td>4.10</td>
</tr>
<tr>
<td>660</td>
<td>0.70</td>
<td>41</td>
<td>0.82</td>
<td>1.01</td>
<td>0.32</td>
<td>3.80</td>
</tr>
<tr>
<td>720</td>
<td>0.64</td>
<td>43</td>
<td>0.86</td>
<td>0.94</td>
<td>0.30</td>
<td>3.58</td>
</tr>
<tr>
<td>780</td>
<td>0.59</td>
<td>44</td>
<td>0.88</td>
<td>0.92</td>
<td>0.29</td>
<td>3.34</td>
</tr>
</tbody>
</table>

Figure 3.11: Comparison of experimental and theoretical $\Delta h^*_r$ when $\omega_w / \omega_p = 1.5$. 
would polish much quicker than the rest of the wafer. The gradual polishing gradient caused by
the increase in rotational velocity ratio will then result in the areas outside the overlap region to
be polished at nearly the same time. The experiment, however, showed that the whole wafer
polished in a short span of time, with the edge taking longer than expected.

In Figure 3.12, results from the velocity ratio experiments are compared with those from
the blocked pad experiment in the previous section where \( \omega_w / \omega_p = 1.0 \). The plot shows that
there is no significant change in \( \Delta h^* \) in the outer region of the wafer between the cases when
\( \omega_w / \omega_p = 0.5 \) and \( \omega_w / \omega_p = 1.0 \). It is likely that rotational effects on slurry output were still
influencing material removal rates, since the pad rotated at different speeds for each of the three
tests. Therefore, one may not be able to control Preston constant by velocity ratio alone without
considering the absolute rotational velocity of the pad. Finally, it should be noted that due to the
limitations of visual endpoint detection, there may have been some discrepancies in the time of
initial substrate exposure, thus causing some disagreement between the experimental and
theoretical curves.

3.5 Pad Position Effects

It has been shown thus far that it is difficult to maintain a uniform Preston constant
during polishing. In progressive polishing, the pad is required to translate uni-directionally
across the wafer. Therefore, it is important to study the variation in \( k_p \) with pad location.
Experiments were performed by polishing three blanket Cu wafers at various pad locations on
the wafer. The consumables and process parameters are summarized in Table 3.8. These
process variables were kept constant while \( r_{cc} \) was varied. To calculate \( \overline{MRR} \) and \( k_p \), the time
and location of complete Cu removal required across the wafer was measured.

Figure 3.13 compares \( k_p(r) \) when polishing with the pad center at different distances from
the wafer center. There was a general trend that \( k_p \) increases with \( r \), independent of \( r_{cc} \). However,
translating the pad also had an effect on \( k_p(r) \), particularly when \( r_{cc} > r_p \). In all cases, \( k_p \) was
lowest at the inner edge of the pad, which suggested that edge effects were influencing material
removal rate. Possible causes for the decrease in material removal rate include the pressure not
being uniformly distributed to the edge of the pad or a decrease in slurry flow to the edge area.
Figure 3.12: Normalized material removed per wafer rotation for various rotational velocity ratios.
Table 3.8: Wafer, consumables and process parameters for the $k_p$ versus $r_{cc}$ experiments.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{cw}$ (µm)</td>
<td>1.0</td>
</tr>
<tr>
<td>$r_w$ (mm)</td>
<td>50</td>
</tr>
<tr>
<td>Pad type</td>
<td>TWI-817</td>
</tr>
<tr>
<td>$r_p$ (mm)</td>
<td>35</td>
</tr>
<tr>
<td>Slurry</td>
<td>5% vol. 300 nm Al$_2$O$_3$ particles</td>
</tr>
<tr>
<td>pH</td>
<td>4</td>
</tr>
<tr>
<td>$\omega_w$ (rad/s) (rpm)</td>
<td>18 (170)</td>
</tr>
<tr>
<td>$\omega_p$ (rad/s) (rpm)</td>
<td>18 (170)</td>
</tr>
<tr>
<td>$v_{cc}$ (m/s)</td>
<td>0</td>
</tr>
<tr>
<td>$p$ (kPa) (psi)</td>
<td>12 (1.7)</td>
</tr>
<tr>
<td>Slurry flow rate (ml/min)</td>
<td>50</td>
</tr>
</tbody>
</table>

Figure 3.13: $k_p(r)$ for various $r_{cc}$'s calculated from $MRR$. 
due to the lack pad perforations. The experiments also showed that there was less variation in $k_p$ toward the periphery of the wafer, even when polishing with different $r_{cc}$'s. Due to the small $\theta_c$ in that region, the residence time of the pad was limited and the change in material removal rate caused by variation in $k_p$ and $r_{cc}$ may not be as apparent.

3.6 Summary

This chapter presented a series of polishing experiments performed on a face-up CMP tool with no pad translation. First, pads with different perforations patterns were used to polish blanket Cu wafers to determine the effects of the pattern on slurry flow distribution. It was determined that uniformly spaced perforations in the pad led to increased slurry flow from the central area of the slurry cup, presumably due to a lower tangential velocity of the slurry in the cup. The higher slurry flow resulted in an increase in Preston constant and material removal rate. A pad with no perforations in the central area was used to block slurry flow from that region. Polishing with the blocked pad resulted in a uni-directional polishing gradient as predicted by the theory. Thus, the experimental results showed that the Preston constant was better controlled with the blocked pad.

The blocked pad was then used to polish wafers at different wafer-pad angular velocity ratios. It was shown that decreasing the ratio resulted in a steeper polishing gradient at the central region of the wafer. However, there was minimal difference at the outer region of the wafer when compared with the case where the angular velocities were equal. The experiments also showed that increasing the velocity ratio created a flatter polishing gradient, which may be useful for cases where fewer pad translation steps are desired.

Finally, blanket wafers were polished with the pad at different distances from the center of the wafer. These experiments were used to determine the effect of pad location on Preston constant. The results showed some variation in Preston constant, but the disparity was less at the periphery of the wafer. This effect was likely due to the small contact angles in that region. The variation in Preston constant was not significant, possibly due to the low material removal rates.
Nomenclature

\[ h = \text{thickness of Cu coating removed (m)} \]
\[ h_{\text{Cu}} = \text{initial Cu coating thickness (m)} \]
\[ \Delta h_{\text{e}}, \Delta h'_{\text{e}} = \text{thickness of coating removed in one wafer rotation (m), normalized value} \]
\[ k_p = \text{Preston constant (m}^2/\text{N)} \]
\[ MRR, \overline{MRR} = \text{material removal rate, average per wafer rotation (m/s)} \]
\[ p = \text{pressure (N/m}^2\text{)} \]
\[ r_{cc} = \text{wafer center to pad center distance (m)} \]
\[ r_w, r_p = \text{wafer radius, pad radius (m)} \]
\[ \Delta t_0 = \text{time for center of the wafer to be completely polished (s)} \]
\[ t = \text{polishing time (s)} \]
\[ v_{cc} = \text{translational velocity of the pad; rate of change of } r_{cc} \text{ (m/s)} \]
\[ v_R = \text{relative velocity of the wafer with respect to the pad (m/s)} \]
\[ \theta_c = \text{semi-contact angle (rad)} \]
\[ \omega_w, \omega_p = \text{angular velocities of the wafer and the pad (rad/s)} \]
4.1 Introduction

An experimental study to validate the numerical pad translation model is outlined in this chapter. First, the pad was manually translated at one-minute intervals to uniformly polish a 100-mm region on 300-mm blanket Cu wafers. The pad location was measured at each step and compared with those obtained by the model. Multiple wafers were polished to test for repeatability. Patterned 200-mm wafers were then polished using the pad translation velocities determined by the model to study wafer-scale polishing uniformity. Cu dishing and dielectric erosion were measured at subdies across the wafer to validate uniform material removal at the wafer-scale.

4.2 Equipment and Consumables

Experiments were performed by polishing wafers on a rotary face-up CMP tool, shown in Figure 4.1. A perforated pad, blocked in the center region, was attached to the slurry cup and slurry was delivered to the cup by a peristaltic pump. Normal load was applied to the pad by compressed air. To enable lateral movement, the pad assembly was attached to a linear stage that consisted of a lead screw driven by a stepper motor. The stage could either be controlled manually with a jogger, as in the blanket wafer polishing experiments, or by a computer, as in the patterned wafer experiments.

A 300-mm wafer carrier was used for the translating pad experiments to facilitate the polishing of 200- and 300-mm wafers. The carrier assembly comprised of an aluminum platen and a glass plate stacked between two rubber sheets as shown in the inset of Figure 4.1. The glass plate provided a planar reference surface and the rubber sheets prevented slippage during the process. The wafer was situated on the top rubber sheet and was held in place by friction during polishing. While the objective of this thesis is to obtain uniform wafer-scale polishing for
Figure 4.1: Photograph of the face-up CMP tool with the 300-mm platen.
100-mm wafers, polishing larger wafers to 100 mm ensured that the pad traveled across a planar surface. Concerns regarding the pad sliding over the interface between the wafer edge and the carrier were avoided.

The wafers were polished using a mixture of Cabot Microelectronics iCue 5001 Cu polishing slurry and H₂O₂, which made up 3% the total volume. The pH of the slurry was measured prior to each experiment. Thomas West TWI-817 polishing pads with SP-7 sub-pads were the pad materials used. These consumables were the same as those used for the non-translating pad experiments described in Chapter 3. Due to the non-uniform Preston constant measured in the non-translating pad experiments, blocked pads were used. To further enhance slurry flow rate, the perforations were placed at the intersection of the grooves as shown in Figure 4.2. Each pad was conditioned with deionized water and a stiff nylon brush before usage.

4.3 Blanket Wafer Polishing

To validate the numerical pad translation model, 300-mm blanket Cu wafers were partially polished to a circular area of 100 mm in diameter at the center of the wafer. Three wafers were polished using constant process parameters to test for repeatability. The process parameters are listed in Table 4.1. In these experiments, the pad was initially positioned to cover the center of the wafer. The pad remained at the initial position until the Cu coating at the center of the wafer was completely removed. Next, the pad was translated toward the edge of the wafer in one-minute intervals so that the edge of the pad touched the edge of the polished region at the beginning of each time step. The overlap between the pad and polished region was about 2 mm. This overlap was kept at a minimum to reduce both over- and underpolishing, yet still allow the Cu at the boundary of the polished region to be completely removed. The process ended when the radius of the central SiO₂ region reached 50 mm. Figure 4.3 is a series of video screenshots showing the wafer and the pad at various times during a polishing experiment.

The distance between the wafer center and the pad center, \( r_{cc} \), was measured after each translation step and the results are shown in Table 4.2. The data show that the progression of polishing was similar for the three runs with the center clearing between 5.0 and 5.5 minutes and the total polishing time ranging from 14 to 16 minutes. This suggests that the Preston constant, \( k_p \), was consistent among the experiments.
Figure 4.2: A blocked pad with perforations at the intersection of the x-y grooves.

Table 4.1: Experimental parameters for the blanket wafer polishing experiments.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{Cu}$ (µm)</td>
<td>1.0</td>
</tr>
<tr>
<td>$r_w$ (mm)</td>
<td>50</td>
</tr>
<tr>
<td>Pad type</td>
<td>TWI-817</td>
</tr>
<tr>
<td>$r_p$ (mm)</td>
<td>35</td>
</tr>
<tr>
<td>$r_c$ (mm)</td>
<td>30</td>
</tr>
<tr>
<td>Slurry</td>
<td>Cabot iCue 5001</td>
</tr>
<tr>
<td>Slurry additive</td>
<td>H$_2$O$_2$ - 3% vol</td>
</tr>
<tr>
<td>pH</td>
<td>8</td>
</tr>
<tr>
<td>$\omega_w$ (rad/s) (rpm)</td>
<td>16 (155)</td>
</tr>
<tr>
<td>$\omega_p$ (rad/s) (rpm)</td>
<td>16 (155)</td>
</tr>
<tr>
<td>$p$ (kPa) (psi)</td>
<td>13 (1.9)</td>
</tr>
<tr>
<td>Slurry flow rate (ml/min)</td>
<td>150</td>
</tr>
</tbody>
</table>
Figure 4.3: Video screenshots of the wafer and the pad during the course of face-up polishing; $p = 13$ kPa, $\omega_w = \omega_p = 16$ rad/s, and slurry flow = 150 ml/min.
<table>
<thead>
<tr>
<th>Run</th>
<th>$t$ (s)</th>
<th>$t/\Delta t_o$</th>
<th>$r_{cc}$ (mm)</th>
<th>$r_{cc}/r_w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run 1</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>330</td>
<td>1.00</td>
<td>31</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>420</td>
<td>1.27</td>
<td>34</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>1.45</td>
<td>38</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>540</td>
<td>1.64</td>
<td>40</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>1.82</td>
<td>43</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>660</td>
<td>2.00</td>
<td>50</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>690</td>
<td>2.09</td>
<td>54</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>720</td>
<td>2.18</td>
<td>58</td>
<td>1.16</td>
</tr>
<tr>
<td></td>
<td>780</td>
<td>2.36</td>
<td>60</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>840</td>
<td>2.55</td>
<td>67</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>2.73</td>
<td>72</td>
<td>1.44</td>
</tr>
<tr>
<td>Run 2</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>330</td>
<td>1.00</td>
<td>35</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>420</td>
<td>1.27</td>
<td>36</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>1.45</td>
<td>37</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>540</td>
<td>1.64</td>
<td>38</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>1.82</td>
<td>45</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>660</td>
<td>2.00</td>
<td>55</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>720</td>
<td>2.18</td>
<td>62</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td>780</td>
<td>2.36</td>
<td>68</td>
<td>1.36</td>
</tr>
<tr>
<td></td>
<td>840</td>
<td>2.55</td>
<td>72</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>2.73</td>
<td>76</td>
<td>1.52</td>
</tr>
<tr>
<td></td>
<td>960</td>
<td>2.91</td>
<td>80</td>
<td>1.60</td>
</tr>
<tr>
<td>Run 3</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>1.00</td>
<td>36</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>360</td>
<td>1.20</td>
<td>36</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>420</td>
<td>1.40</td>
<td>37</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>1.60</td>
<td>38</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>540</td>
<td>1.80</td>
<td>40</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>2.00</td>
<td>43</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>660</td>
<td>2.20</td>
<td>55</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>720</td>
<td>2.40</td>
<td>60</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>780</td>
<td>2.60</td>
<td>70</td>
<td>1.40</td>
</tr>
<tr>
<td></td>
<td>840</td>
<td>2.80</td>
<td>73</td>
<td>1.46</td>
</tr>
</tbody>
</table>
Figures 4.4(a) and (b) compare the experimental and calculated pad positions versus polishing time in dimensional and dimensionless form, respectively. The $M$ matrix in Eq. (2.38) was computed by discretizing the wafer radius, $r_w$, into 50 intervals and assuming that $k_p$ and pressure, $p$, are constant, i.e. $k_p$ is independent of $r$ and $r_{cc}$. Using the time of initial Cu clearance $\Delta t_0$ with Eq. (2.42), the $k_p$ at the center of the wafer was computed for each polishing experiment and the mean value, $k_p = 5.0 \times 10^{-13}$ 1/Pa, was used as the constant value in the model. The amount of oxide overpolishing, $h_o$, was assumed to be zero because the pad was moved away from the polished region shortly after the endpoint was reached.

In practice, the face-up CMP process must start with the pad slightly overlapping the center of the wafer to guarantee that the center is completely polished. AB and CD in Figure 4.4 show two long time-steps at the beginning of the process. AB describes the time the pad must stay at its initial position, $r_{cc}$, before the center of the wafer is polished. Next, CD is due to the steep decrease in the semi-contact angle, $\theta_c$, when the pad leaves the center of the wafer. From Figure 2.5, when the pad terminates contact with the wafer center ($r_{cc} > r_p$), there is a significant decrease in $\theta_c(r)$. Therefore, after the center is cleared, the pad must remain at the next step for a long period of time in order to finish polishing the region immediately outside the overlap.

Due to the 5-mm overlap of the pad at the center of the wafer, the model assumes that a circular region 10-mm in diameter will be first polished concurrently. Then the pad translates quickly so that its edge is directly on the boundary of the polished region. This is shown in Figure 4.4 by the sharp increase in $r_{cc}$ from B to C and constant $r_{cc}$ from C to D. Experimentally, the constant pad position during time-step CD is reflected by small, discrete displacements for the first three minutes after the center has been cleared. The pad translational velocity was not zero because the model assumes the edge of the pad moves directly to the boundary of the polished region, while experimentally there was an overlap between the pad and polished region. Furthermore, edge effects reduced the effective pad radius and resulted in an initial polished area that is smaller than the assumed 10 mm (from the pad/wafer center overlap). Therefore, a number of small steps were required for the pad to travel from B to D.

The discrepancy between the experimental and calculated values between points D and E in Figure 4.4(a) is primarily due to the spatial variation in Preston constant. The non-translating pad experiments in the previous chapter, Figure 3.9, had shown an increase of $k_p$ with radial
Figure 4.4: Comparison of experimental and calculated pad positions versus polishing time in (a) dimensional and (b) dimensionless form.
location, \( r \), which increased material removal rate. This increase in \( k_p \) was likely due to chemical effects from the slurry that remained on the surface of the wafer after exiting the pad-wafer interface. Screenshots from the polishing video in Figure 4.3 show that there was a layer of slurry on the surface of the wafer throughout the experiment. Since the pad contacts the points on the periphery of the wafer for a shorter period of time (due to smaller \( \theta_c \)'s), the chemicals from residual slurry layer had a longer reaction time with the coating before the pad mechanically removes the material. This may have altered the mechanical properties of the coating and created a surface that was easier to remove. Therefore, as the pad moved further away from the center of the wafer, it was not required to stay at each position for as long as the calculated time. The experiments show that beyond point D the pad generally moved to a particular \( r_{cc} \) before the predicted time, i.e. the data points lie above the model curve.

Finally, the spatial variation of \( k_p \) also affected the total pad displacement. Point E shows that the model requires a final pad position, \( r_{cc} \), of 84 mm in order to polish a 100-mm region, while experimentally, \( r_{cc} \) ranged from 72 to 80 mm. Since \( MRR \) was higher toward the edge of the wafer, the pad was not required to travel as far as the model predicts to remove Cu up to the edge region. However, the overprediction of total polishing time should not have a negative effect on polishing uniformity. If the pad continued to \( r_{cc} = 84 \) mm as prescribed by the model, the only result would have be a larger area polished. Due to the continuous translation of the pad, overpolishing would still be avoided.

### 4.4 Patterned Wafer Polishing

Patterned Cu/SiO\(_2\) test wafers, type SKW6-2, from SKW Associates were polished on the face-up CMP tool to determine Cu dishing and dielectric erosion after face-up CMP. The 200-mm wafer consisted of 20-mm by 20-mm dies with pitch and density structures as shown in Figure 4.5. Pitch structures with linewidths ranging from 2.5 \( \mu \)m to 100 \( \mu \)m were used for dishing measurements and linewidths of 5 \( \mu \)m to 100 \( \mu \)m were used for erosion measurements. A highly schematic cross-section of the wafer is shown in Figure 4.6. The SiO\(_2\) thickness was found to be 850 nm by chemical etching experiments described in Appendix A. This value was used in lieu of the listed value (800 nm) as the interconnect depth, \( h_l \).
Figure 4.5: Die map of the wafers used for dishing and erosion experiments.

Figure 4.6: Cross-section of an SKW6-2 test wafer. (SKW Associates)
Prior to polishing, the initial surface topography – surface trench width, \( w_s \), and initial step-height, \( h_{si} \) – was obtained with a Tencor P10 surface profilometer with a 2-\( \mu \)m stylus tip. The Cu deposition factor, \( \alpha \), was determined from \( w_s \) for each subdie by Eq. (1.1). The \( \alpha \) and \( h_{si} \) values are listed in Table 4.3. Figures 4.7(a) and (b) plot \( \alpha \) and \( h_{si} \) with linewidth, \( w \), respectively. As expected, both \( \alpha \) and \( h_{si} \) increased with \( w \). For \( w \geq 10\ \mu m \), \( \alpha \) is close to 1 and \( h_{si} \) is over 1 \( \mu m \). Therefore, the underlying structure is closely reproduced on the Cu surface. For \( w < 5\ \mu m \), both \( \alpha \) and \( h_{si} \) are low and the initial wafer surface is fairly planar. Figure 4.7(c) shows the combined effects of initial surface geometry, \( \alpha \cdot h_{si} \), with \( w \).

The CMP process parameters for patterned wafer polishing are listed in Table 4.4. A 200-mm wafer was partially polished to a circular area of 100 mm in diameter for Cu dishing and dielectric erosion measurements. A dimensionless pad translation model was used to determine pad motion. This is the appropriate choice for face-up CMP experiments because the model is only dependent on \( \Delta t_0 \), which is a function of the polishing conditions. Therefore, any change in Cu thickness from wafer to wafer as well as any process variation will be accounted for by \( \Delta t_0 \). Because the dies in the test wafers are large and feature-scale non-uniformity is not controlled, fully polishing the entire die would result in excessive dishing in the wide lines. Therefore, \( \Delta t_0 \) was chosen to completely polish the eight subdies in consideration, which are highlighted in Figure 4.5.

A dimensionless velocity parameter, \( v_{cc}^* \) is defined as

\[
\frac{v_{cc}^*}{r_c} = \frac{\Delta t_0}{r_w} \cdot v_{cc}
\]

(4.1)

To simplify the pad translation control, the theoretical motion path was approximated with a linear path where \( v_{cc}^* = 0.6 \). Additionally, the length of time of the second translation step was shortened to reduce overpolishing at the center of the wafer. This adjustment is attributed to the non-constant \( k_p \) and pad edge effects observed from the blanket wafer polishing experiments. The revised path is shown in Figure 4.8. During the experiment, a Labview program recorded the time of center polishing, \( \Delta t_0 \), and used the model to compute the pad translation velocity for the remainder of the process. The velocity and displacement were sent to the stepper motor controller which translated the pad assembly.
Table 4.3: Measured $\alpha$ and $h_{si}$ on the SKW6-2 test wafers for various linewidths.

<table>
<thead>
<tr>
<th>$w$ (µm)</th>
<th>$\alpha$</th>
<th>$h_{si}$ (µm)</th>
<th>$\alpha h_{si}$ (µm)</th>
<th>$\alpha h_{si} / h_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>0.17</td>
<td>0.084</td>
<td>0.014</td>
<td>0.017</td>
</tr>
<tr>
<td>3.5</td>
<td>0.25</td>
<td>0.192</td>
<td>0.048</td>
<td>0.056</td>
</tr>
<tr>
<td>4.5</td>
<td>0.33</td>
<td>0.150</td>
<td>0.050</td>
<td>0.058</td>
</tr>
<tr>
<td>5</td>
<td>0.95</td>
<td>0.732</td>
<td>0.695</td>
<td>0.818</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1.135</td>
<td>1.135</td>
<td>1.335</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1.132</td>
<td>1.132</td>
<td>1.332</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>1.133</td>
<td>1.133</td>
<td>1.333</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>1.173</td>
<td>1.173</td>
<td>1.380</td>
</tr>
</tbody>
</table>

Figure 4.7: Measured initial wafer surface geometry: (a) Cu deposition factor, $\alpha$, (b) initial step-height, $h_{si}$, and (c) combined feature-scale geometry, $\alpha h_{si}$.
Figure 4.7 (cont.): Measured initial wafer surface geometry: (a) Cu deposition factor, $\alpha$, (b) initial step-height, $h_{si}$, and (c) combined feature-scale geometry, $\alpha \cdot h_{si}$. 
Table 4.4: Experimental conditions for patterned wafer polishing.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{Cv}$ (μm)</td>
<td>1.5</td>
</tr>
<tr>
<td>$r_w$ (mm)</td>
<td>100</td>
</tr>
<tr>
<td>$r_{w,\text{effective}}$ (mm)</td>
<td>50</td>
</tr>
<tr>
<td>Pad type</td>
<td>TWI-817</td>
</tr>
<tr>
<td>$r_p$ (mm)</td>
<td>35</td>
</tr>
<tr>
<td>$r_{co}$ (mm)</td>
<td>30</td>
</tr>
<tr>
<td>Slurry</td>
<td>Cabot iTue 5001</td>
</tr>
<tr>
<td>Slurry additive</td>
<td>$\text{H}_2\text{O}_2$ - 3% vol</td>
</tr>
<tr>
<td>pH</td>
<td>8</td>
</tr>
<tr>
<td>$\omega_w$ (rad/s) (rpm)</td>
<td>16 (155)</td>
</tr>
<tr>
<td>$\omega_p$ (rad/s) (rpm)</td>
<td>16 (155)</td>
</tr>
<tr>
<td>$p$ (kPa) (psi)</td>
<td>13 (1.9)</td>
</tr>
<tr>
<td>Slurry flow rate (ml/min)</td>
<td>150</td>
</tr>
<tr>
<td>$\Delta t_0$ (min)</td>
<td>7</td>
</tr>
</tbody>
</table>

Figure 4.8: Theoretical and actual dimensionless pad translation path.
After polishing, a Tencor P10 surface profilometer was used to measure the step-height of the features at various radial locations across the wafer within the 100-mm central region. Since the Cu overburden has been removed outside the line features, by definition the remaining step-height is Cu dishing. Dielectric erosion was determined by comparing the difference in height between the SiO₂ at the center of the subdie and the edge of the subdie. Because Δh₀ was chosen so that the pad translated when the subdies were just polished, and slurry selectivity was high, there was negligible overpolishing at the edge of the subdie, or the reference field region. Therefore, wafer-scale erosion was neglected and the measured feature-scale erosion was assumed to be the total erosion. The patterned wafer polishing results are shown in Table 4.5. The experimental data show that the material removal rate across the wafer was controlled reasonably well by face-up polishing.

4.4.1 Dielectric Erosion

Figure 4.9 plots the normalized dielectric erosion data for each subdie as a function of its radial location on the wafer. Erosion was at most 5% of the total interconnect depth for all subdies, which is within industry standards. The low magnitude of erosion is to be expected, since the face-up CMP pad translation scheme is based on minimizing erosion. The algorithm is designed so that the pad only remains at each position long enough to just expose the SiO₂ layer. Therefore, there was no SiO₂ overpolishing. Additionally, the high selectivity slurry had low material removal rates on Ta and SiO₂, which further reduced the amount of SiO₂ removed. The low erosion values, however, also resulted in a large degree of scatter in the measured data. The statistical summary listed in Table 4.6 shows that the average erosion for each feature ranged from 1 nm to 29 nm and the standard deviation from 1 nm to 12 nm.

The multi-scale, tribological erosion model from Chapter 1 states:

\[
e = \left[ \frac{1}{(1-w/\lambda)S_{Cu/ox} + w/\lambda} \right] \left[ \left( \frac{1}{\beta} - 1 \right) h_{Cu} + \frac{\alpha w}{\lambda} h_{Ta} + S_{Cu/ox} \frac{1}{\beta} h_{o} - (w/\lambda)D \right] \tag{1.5}
\]

However, the purpose of face-up CMP is to maintain uniform material removal at the wafer-scale, so β is approximately unity – the material removal rate at the “slowest” field region of the wafer is equal to that at the “fastest”. Furthermore, deliberate oxide overpolishing is not necessary in face-up CMP because of wafer-scale polishing uniformity and the ease of endpoint
Table 4.5: Measured Cu dishing and dielectric erosion after face-up CMP.

<table>
<thead>
<tr>
<th>Feature</th>
<th>$r$ (mm)</th>
<th>$r/r_w$</th>
<th>$D$ (nm)</th>
<th>$D/h_I$</th>
<th>$e$ (nm)</th>
<th>$e/h_I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w=5\ \mu$m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda=10\ \mu$m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_f=0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.5</td>
<td>0.19</td>
<td>73.22</td>
<td>0.09</td>
<td>22.11</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>14.0</td>
<td>0.28</td>
<td>66.18</td>
<td>0.08</td>
<td>31.09</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>16.5</td>
<td>0.33</td>
<td>71.69</td>
<td>0.08</td>
<td>19.58</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>20.0</td>
<td>0.40</td>
<td>60.70</td>
<td>0.07</td>
<td>16.64</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>26.5</td>
<td>0.53</td>
<td>86.05</td>
<td>0.10</td>
<td>28.82</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>28.5</td>
<td>0.57</td>
<td>74.32</td>
<td>0.09</td>
<td>42.50</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>31.0</td>
<td>0.62</td>
<td>98.24</td>
<td>0.12</td>
<td>46.03</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>36.0</td>
<td>0.72</td>
<td>69.81</td>
<td>0.08</td>
<td>44.62</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>37.5</td>
<td>0.75</td>
<td>68.87</td>
<td>0.08</td>
<td>11.79</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>46.0</td>
<td>0.92</td>
<td>55.73</td>
<td>0.07</td>
<td>23.38</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>$w=10\ \mu$m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda=20\ \mu$m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_f=0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.5</td>
<td>0.13</td>
<td>194.09</td>
<td>0.23</td>
<td>19.23</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>15.5</td>
<td>0.31</td>
<td>153.44</td>
<td>0.18</td>
<td>24.46</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>17.5</td>
<td>0.35</td>
<td>166.79</td>
<td>0.20</td>
<td>19.30</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>22.5</td>
<td>0.45</td>
<td>148.01</td>
<td>0.17</td>
<td>16.83</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>25.5</td>
<td>0.51</td>
<td>167.86</td>
<td>0.20</td>
<td>44.04</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>28.0</td>
<td>0.56</td>
<td>159.45</td>
<td>0.19</td>
<td>17.14</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>30.5</td>
<td>0.61</td>
<td>148.89</td>
<td>0.18</td>
<td>28.01</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>34.5</td>
<td>0.69</td>
<td>150.57</td>
<td>0.18</td>
<td>14.23</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>39.0</td>
<td>0.78</td>
<td>138.37</td>
<td>0.16</td>
<td>22.02</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>45.0</td>
<td>0.90</td>
<td>130.92</td>
<td>0.15</td>
<td>26.06</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>$w=20\ \mu$m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda=40\ \mu$m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_f=0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>0.10</td>
<td>269.27</td>
<td>0.32</td>
<td>25.56</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>15.0</td>
<td>0.30</td>
<td>205.12</td>
<td>0.24</td>
<td>4.37</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>21.0</td>
<td>0.42</td>
<td>232.39</td>
<td>0.27</td>
<td>13.88</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>25.0</td>
<td>0.50</td>
<td>221.73</td>
<td>0.26</td>
<td>5.40</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>26.0</td>
<td>0.52</td>
<td>220.45</td>
<td>0.26</td>
<td>6.28</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>32.0</td>
<td>0.64</td>
<td>227.78</td>
<td>0.27</td>
<td>12.71</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>33.0</td>
<td>0.66</td>
<td>226.08</td>
<td>0.27</td>
<td>9.15</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>35.0</td>
<td>0.70</td>
<td>214.16</td>
<td>0.25</td>
<td>15.04</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>41.0</td>
<td>0.82</td>
<td>196.19</td>
<td>0.23</td>
<td>16.40</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>45.5</td>
<td>0.91</td>
<td>213.22</td>
<td>0.25</td>
<td>19.15</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Feature</td>
<td>$r$ (mm)</td>
<td>$r/r_w$</td>
<td>$D$ (nm)</td>
<td>$D/h_l$</td>
<td>$e$ (nm)</td>
<td>$e/h_l$</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>$w = 50 \mu m$</td>
<td>7.0</td>
<td>0.14</td>
<td>501.19</td>
<td>0.59</td>
<td>10.32</td>
<td>0.01</td>
</tr>
<tr>
<td>$\lambda = 100 \mu m$</td>
<td>17.0</td>
<td>0.34</td>
<td>423.10</td>
<td>0.50</td>
<td>1.73</td>
<td>0.01</td>
</tr>
<tr>
<td>$A_f = 0.5$</td>
<td>22.0</td>
<td>0.44</td>
<td>428.39</td>
<td>0.50</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>25.0</td>
<td>0.50</td>
<td>487.95</td>
<td>0.57</td>
<td>12.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>29.0</td>
<td>0.58</td>
<td>402.11</td>
<td>0.47</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>30.0</td>
<td>0.60</td>
<td>448.14</td>
<td>0.53</td>
<td>9.62</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>35.0</td>
<td>0.70</td>
<td>434.14</td>
<td>0.51</td>
<td>4.87</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>35.5</td>
<td>0.71</td>
<td>415.24</td>
<td>0.49</td>
<td>12.69</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>38.5</td>
<td>0.77</td>
<td>404.13</td>
<td>0.48</td>
<td>9.93</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>45.5</td>
<td>0.91</td>
<td>405.07</td>
<td>0.48</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>$w = 100 \mu m$</td>
<td>10.0</td>
<td>0.20</td>
<td>592.95</td>
<td>0.70</td>
<td>2.07</td>
<td>0.00</td>
</tr>
<tr>
<td>$\lambda = 200 \mu m$</td>
<td>13.5</td>
<td>0.27</td>
<td>588.90</td>
<td>0.69</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>$A_f = 0.5$</td>
<td>17.5</td>
<td>0.35</td>
<td>563.53</td>
<td>0.66</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>19.0</td>
<td>0.38</td>
<td>574.95</td>
<td>0.68</td>
<td>1.31</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>26.5</td>
<td>0.53</td>
<td>637.48</td>
<td>0.75</td>
<td>1.53</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>29.0</td>
<td>0.58</td>
<td>606.41</td>
<td>0.71</td>
<td>1.92</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>32.5</td>
<td>0.65</td>
<td>551.89</td>
<td>0.65</td>
<td>0.16</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>36.0</td>
<td>0.72</td>
<td>561.25</td>
<td>0.66</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>38.0</td>
<td>0.76</td>
<td>553.06</td>
<td>0.65</td>
<td>1.18</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>45.0</td>
<td>0.90</td>
<td>533.80</td>
<td>0.63</td>
<td>0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Figure 4.9: Normalized dielectric erosion across the wafer after face-up CMP.

Table 4.6: Statistical summary of dielectric erosion at various features after face-up CMP.

<table>
<thead>
<tr>
<th>( w ) (( \mu \text{m} ))</th>
<th>Mean (nm)</th>
<th>Std. Dev (nm)</th>
<th>Mean / ( h_f )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>28.66</td>
<td>12.19</td>
<td>0.03</td>
</tr>
<tr>
<td>10</td>
<td>23.13</td>
<td>8.55</td>
<td>0.03</td>
</tr>
<tr>
<td>20</td>
<td>12.79</td>
<td>6.70</td>
<td>0.02</td>
</tr>
<tr>
<td>50</td>
<td>6.12</td>
<td>5.32</td>
<td>0.01</td>
</tr>
<tr>
<td>100</td>
<td>0.82</td>
<td>0.86</td>
<td>0.001</td>
</tr>
</tbody>
</table>
detection. The overpolishing of SiO\textsubscript{2} is therefore minimal and \( h_o \) can be neglected. Hence, dielectric erosion is only dependent on the feature-scale non-uniformity factors, \( \alpha \) and \( h_{si} \), and Cu dishing. 

Equation 1.5 can be reduced to

\[
\epsilon = \frac{w/\lambda}{(1-w/\lambda)S_{Cu/ax}+w/\lambda}(\alpha \cdot h_{si} - D) \tag{4.2}
\]

Erosion is low when slurry selectivity is high. For the current experimental conditions, \( w/\lambda = 0.5 \) and \( S_{Cu/ax} = 50 \), \( \epsilon = 0.02(\alpha \cdot h_{si} - D) \). Therefore, even if \( \alpha = 1 \) and dishing is neglected, erosion is still only 2% of the \( h_{si} \). Moreover, if \( D \approx \alpha \cdot h_{si} \), erosion is close to zero.

The erosion results are compared with the multi-scale, tribological erosion model in Figure 4.10. The model and the data are in agreement in the magnitude of erosion being less than 3% the interconnect depth. The model, however, suggests that erosion should increase with linewidth while the experimental data shows erosion decreasing with linewidth. However, the experiments were conducted with a compliant pad which conformed to the wafer surface and dishing was large for the large linewidths. The experimental erosion results were therefore dominated by the dishing term and decreased with linewidth.

The low magnitude of erosion in features with large lines is to be expected when polishing with a compliant pad because the pad asperities contact both the high and low features simultaneously. Thus, the applied pressure is evenly distributed between the Cu and SiO\textsubscript{2}, and erosion is reduced. For the smaller lines, the pad contacts only the high feature outside the interconnect, which increases the applied pressure on SiO\textsubscript{2}, thereby increasing erosion. Nevertheless, both the data and the model show that erosion is not a major defect issue when overpolishing is minimized by face-up CMP.

### 4.4.2 Cu Dishing

Figure 4.11 shows the measured Cu dishing across the wafer normalized with interconnect depth. For each subdie, dishing was fairly constant across the wafer, demonstrating that wafer-scale polishing can be controlled by the face-up CMP architecture. Table 4.7 summarizes the statistical parameters from the dishing measurements. The standard deviation for all features is less than 35 nm, or 4% of the interconnect depth. While the magnitude of dishing is higher than the industry specification of 5% interconnect depth, the low variation
Experimental -- Theoretical

Model Parameters:

\[ w/ \lambda = 0.5, \quad Y_p = 20 \text{ MPa}, \]
\[ R_s = 5 \text{ µm}, \quad \lambda_s = 250 \text{ µm}, \]
\[ S_{\text{Cu/ox}} = 50, \quad p = 13 \text{ kPa} \]

Figure 4.10: Normalized dielectric erosion compared with the multi-scale erosion model from Eq. (4.2), where \( \alpha \) and \( h_{st} \) were approximated using the data shown in Figure 4.7 and \( D \) from Eq. (4.5).
Figure 4.11: Normalized Cu dishing across the wafer after face-up CMP.

Table 4.7: Statistical summary of Cu dishing at various features after face-up CMP.

<table>
<thead>
<tr>
<th>( w ) (( \mu \text{m} ))</th>
<th>Mean (nm)</th>
<th>Std. Dev (nm)</th>
<th>Mean / ( h_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>27.31</td>
<td>3.04</td>
<td>0.03</td>
</tr>
<tr>
<td>3.5</td>
<td>40.72</td>
<td>2.72</td>
<td>0.05</td>
</tr>
<tr>
<td>4.5</td>
<td>22.96</td>
<td>2.07</td>
<td>0.03</td>
</tr>
<tr>
<td>5</td>
<td>72.48</td>
<td>12.16</td>
<td>0.09</td>
</tr>
<tr>
<td>10</td>
<td>155.84</td>
<td>17.67</td>
<td>0.18</td>
</tr>
<tr>
<td>20</td>
<td>222.64</td>
<td>19.67</td>
<td>0.26</td>
</tr>
<tr>
<td>50</td>
<td>434.95</td>
<td>34.74</td>
<td>0.51</td>
</tr>
<tr>
<td>100</td>
<td>576.42</td>
<td>30.59</td>
<td>0.68</td>
</tr>
</tbody>
</table>
suggests that once the optimal process parameters and consumables are determined to minimize dishing, face-up CMP can control dishing to approximately the same value across the wafer.

The multi-scale tribological dishing model introduced in Chapter 1 states:

\[
D = \left[ \frac{S_{Cu/ox} - 1}{(1 - w/\lambda) S_{Cu/ox} + w/\lambda} \right] \left( \frac{p}{Y_p} \right) \left( \frac{\lambda_o^2}{6\pi R_a} \right) \left[ 1 - \exp(-t_o^*) \right] 
\]

where

\[
t_o^* = \left[ \frac{(1 - w/\lambda) S_{Cu/ox} + w/\lambda}{S_{Cu/ox}} \right] \left( \frac{Y_p}{p} \right) \left( \frac{6\pi R_a}{\lambda_o^2} \right) \left[ \frac{1}{\beta} - 1 \right] h_{Cu} + \frac{\alpha w}{\lambda} h_{si} + S_{Cu/ox} \frac{1}{\beta} h_o 
\]

Assuming again that \( \beta = 1 \) and \( h_o = 0 \),

\[
t_o^* = \left[ \frac{(1 - w/\lambda) S_{Cu/ox} + w/\lambda}{S_{Cu/ox}} \right] \left( \frac{Y_p}{p} \right) \left( \frac{6\pi R_a}{\lambda_o^2} \right) \left[ \frac{\alpha w}{\lambda} h_{si} \right] 
\]

For the current experimental conditions: \( w/\lambda = 0.5 \), \( \alpha = 1 \), \( h_{si} = 1 \) \( \mu m \), \( Y_p = 20 \) MPa, \( R_a = 5 \) \( \mu m \), \( \lambda_o = 250 \) \( \mu m \), and \( S_{Cu/ox} = 50 \), \( t_o^* = 0.6 \). Then, \( 1 - \exp(-t_o^*) = 0.45 \) or \( 0.75 t_o^* \). Equation (1.3) can be simplified to

\[
D = \left[ \frac{S_{Cu/ox} - 1}{(1 - w/\lambda) S_{Cu/ox} + w/\lambda} \right] \left( \frac{p}{Y_p} \right) \left( \frac{\lambda_o^2}{6\pi R_a} \right) \cdot 0.75 t_o^* = 0.75 \left( \frac{S_{Cu/ox} - 1}{S_{Cu/ox}} \right) \frac{\alpha w}{\lambda} h_{si} 
\]

which for high selectivities can be approximated as

\[
D \approx 0.75 \frac{\alpha w}{\lambda} h_{si} 
\]

Furthermore, when \( w/\lambda \) is constant, \( D \) is only a function of \( \alpha \cdot h_{si} \) which in turn varies with linewidth. More importantly, this approximation is independent of pad properties \( (Y_p, \lambda_o, R_a) \), selectivity, \( S_{Cu/ox} \), and pressure, \( p \). Therefore even when wafer-scale polishing uniformity is maintained by face-up CMP, Cu dishing is not zero due to feature-scale non-uniformity.

Figure 4.12 shows that the experimental data agree fairly well with the dishing model. It should be noted that the 3.5 and 4.5-\( \mu m \) features do not have an area fraction of 0.5 as assumed by the model. However, since \( \alpha \) and \( h_{si} \) are low for small features, dishing is low and the area fraction effects are less significant. It is also important to note that the model presented here assumes plastic pad asperities. Because the Thomas West TWI-817 pad has higher compliance than the more common Rohm and Haas IC1000 pad, it may be more appropriate to consider an
Model Parameters:
\[ w/\lambda = 0.5, \gamma_p = 20 \text{ MPa}, \]
\[ R_a = 5 \mu m, \lambda_a = 250 \mu m, \]
\[ S_{Cu/air} = 50, p = 13 \text{ kPa} \]

Figure 4.12: Experimental dishing compared with the multi-scale dishing model presented in Eq. (4.5), where \( \alpha \) and \( h_{si} \) were approximated using the data shown in Figure 4.7.
elastic rough pad model [Noh, 2005]. In the elastic model, dishing is dependent on the Young’s modulus of the pad instead of hardness. It is expected that this model will predict higher dishing values. However, the elastic model is not considered in this work due to the need for iterative computation.

Reducing Cu Dishing

The increase in dishing with linewidth shown in Figures 4.11 and 4.12 is due to feature-scale effects, which is described in the model through $\alpha$ and $h_{si}$. From Figure 4.7, $\alpha$ and $h_{si}$ are large when the features are wide. Large $\alpha$ and $h_{si}$ values increase the material removal rate so that those features reach their endpoint earlier than the smaller features. Therefore, there will be more overpolishing at the wide features.

To reduce Cu dishing, feature-scale uniformity must be controlled, meaning that the initial wafer surface should be fairly planar. Appendix B presents a novel technique for controlling initial feature topography by spin coating a layer of polymer onto the wafer prior to polishing. The polymer material partially fills the trenches on the Cu coating so that the initial wafer surface is close to planar, as shown in Figure 4.13. To examine the effects of spin coating, 200-mm patterned wafers were coated with SU-8 photoresist that measured 2 and 4 $\mu$m thick, and polished on the face-up CMP tool. The polishing conditions are the same as those listed in Table 4.4 except that polishing time was increased due to the extra coating.

Results from the experiments show that dishing was reduced for all features when the photoresist coating was present. Table 4.8 compares dishing for the uncoated and spin-coated wafers. The magnitude of dishing decreased as the SU-8 coating thickness increased. Dishing on the wafer coated with 4 $\mu$m SU-8 was reduced by 58% for the 100-$\mu$m features and 71% for the 50-$\mu$m features. These results show that it is possible to reduce dishing by controlling the initial wafer topography with spin coating. Furthermore, there was also less variation in dishing among different feature sizes, as shown in Figure 4.14. This would suggest that the material removal rates at each feature were similar, which is important for minimizing overpolishing because all the subdies will finish polishing around the same time.

Another concern in planarizing wide lines is the increased contact from pad asperities at the low feature which can excessively remove Cu from inside the interconnect lines.
Table 4.8: Comparison of Cu dishing in uncoated wafer and spin-coated wafers.

<table>
<thead>
<tr>
<th>$w$ (μm)</th>
<th>Without Coating</th>
<th>With 2-μm Coating</th>
<th>With 4-μm Coating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$D$ (nm)</td>
<td>$D/h_t$</td>
<td>$D$ (nm)</td>
</tr>
<tr>
<td>5</td>
<td>72.48</td>
<td>0.09</td>
<td>69.68</td>
</tr>
<tr>
<td>10</td>
<td>155.84</td>
<td>0.18</td>
<td>122.26</td>
</tr>
<tr>
<td>20</td>
<td>222.64</td>
<td>0.26</td>
<td>151.53</td>
</tr>
<tr>
<td>50</td>
<td>434.95</td>
<td>0.51</td>
<td>192.55</td>
</tr>
<tr>
<td>100</td>
<td>576.42</td>
<td>0.68</td>
<td>323.36</td>
</tr>
</tbody>
</table>

Polishing Conditions: $r_p = 35$ mm, $r_{cv} = 30$ mm, $\omega_w = \omega_p = 16$ rad/s (155 rpm), $p = 13$ kPa

Figure 4.14: Normalized Cu dishing versus linewidth after the polishing of uncoated and spin-coated wafers.
Figure 4.15 is a micrograph of a new TWI-817 pad obtained from an Olympus LEXT OLS3000 confocal microscope. The asperity heights range from 20 μm to 80 μm and the radius of curvature range from 2 μm to 20 μm. According to these values, the asperities can easily fit within the Cu lines of larger linewidths. Figure 4.16 is a scaled schematic of pad asperities contacting a wide feature, which shows that the deformation of the pad asperity outside the line is similar to that inside the line. Additionally, the high global compliance of the TWI-817 further enables the pad to conform to the feature topography. Therefore, the pressures applied to the high and low features are comparable, and the Cu from these regions is removed concurrently from the start of polishing. The step-height does not change significantly with polishing and the final step-height is close to the initial value. When the linewidths are small, however, the larger pad asperities glide over the high features without contacting the Cu lines. Thus, material is only removed at the high feature and step-height is reduced.

To facilitate step-height reduction, the pad should be smooth and stiff. A 200-mm wafer was polished with a Rohm and Haas IC1000 pad on the face-up tool using the conditions listed in Table 4.4. By using a smoother, stiffer pad such as the IC1000, dishing in the 100-μm lines was reduced to 80 nm, or 14% of what was obtained when polishing with the TWI-817 pad. However, it was also more difficult to achieve global planarity and further work is necessary to utilize the IC1000 in the face-up CMP tool.

Finally, the high magnitude of dishing is also a result of the high selectivity slurry used. The iCue 5001 slurry, with a Cu to Ta selectivity of 50 to 60, is used as a first-step Cu polishing slurry in industrial processes due to its high Cu material removal rate. Therefore, while the barrier layer was being removed during the last stage of polishing, proportionally more Cu was also being removed from the interconnects. The increase in dishing is significant even when the overpolishing time is low. It is expected that the degree of dishing can be reduced using a barrier layer polishing slurry with low Cu selectivity.

### 4.5 Summary

This chapter described the face-up CMP experiments regarding pad translation. The face-up CMP architecture requires the pad to translate away from the center of the wafer during the polishing process to achieve uniform material removal at the wafer-scale. The numerical
Figure 4.15: Confocal micrograph of a new TWI-817 pad surface.

Figure 4.16: Scaled schematic of sinusoidal pad asperities contacting a wide line feature.
model for pad translation was validated in this chapter by blanket and patterned wafer polishing experiments.

In the blanket wafer polishing experiments, the pad was translated at one-minute intervals to the edge of the polished region. The pad position was measured and compared with the computed values for those times. The results showed that the model was able to predict the general motion of the pad. Due to the variation of Preston constant across the pad-wafer interface, however, the model over-predicts the time durations for the translation steps towards edge of the wafer. For more accurate results, it is possible to incorporate the change in Preston constant into the model. However, not only does this term vary with radial location and pad location, but it is also affected by a multitude of parameters such as slurry properties, pad properties, slurry flow rate, and rotational velocity. An enormous effort to empirically measure Preston constants at various locations on the wafer under different polishing conditions is required. Therefore, it is more useful to use the constant Preston constant model as a guideline for process development and then either make fine adjustments through experimentation or utilize an in-situ endpoint sensor.

A linearized dimensionless translation model was used to control pad motion during the polishing of a patterned wafer. Dielectric erosion and Cu dishing at subdies with line features ranging from 2.5 µm to 100 µm were measured using a surface profilometer. The results showed fairly uniform erosion and dishing across the wafer for each feature, demonstrating that wafer-scale polishing was controlled. Due to the minimized overpolishing time and high slurry selectivity, dielectric erosion was below 5% of the interconnect depth for all the measured features. When compared with the multi-scale tribological erosion model, both the model and the experimental data showed that erosion will be low when overpolishing is minimized by face-up CMP.

Cu dishing ranged from 9% to 68% the interconnect depth. The higher than expected values were attributed to feature-scale non-uniformity, and the use of a compliant pad and high selectivity slurry. However, the standard deviation of dishing across the wafer for each feature was below 35 nm, or 4% of the interconnect depth. Therefore, it is expected that once the consumables and process parameters are optimized, low dishing values can be obtained uniformly across the wafer.
Comparison of the dishing results with the multi-scale tribological dishing model also showed good agreement. Both the model and data showed that while wafer-scale polishing was controlled, pattern dependence was still a dominant factor in material removal rate. A spin coating method for improving initial wafer surface topography was used to control feature-scale effects. Polishing of a coated wafer resulted in as much as 71% reduction in dishing. Large linewidths were also more susceptible to dishing due to increased material removal inside the lines by pad asperity contact. Experiments showed that polishing with a smoother, stiffer Rohm and Haas IC1000 pad reduced dishing by 86% when compared to the TWI-817 pad. Lastly, it is suggested that a barrier layer slurry with low Cu selectivity be used to further reduce dishing.
Nomenclature

\[ A_f = \text{Area fraction of Cu to SiO}_2 \]
\[ D = \text{Cu dishing (m)} \]
\[ e = \text{dielectric erosion (m)} \]
\[ h_{Cu} = \text{initial Cu coating thickness (m)} \]
\[ h_I = \text{interconnect depth (m)} \]
\[ h_o = \text{thickness of oxide removed during overpolishing (m)} \]
\[ h_{si} = \text{initial feature step-height (m)} \]
\[ k_p = \text{Preston constant (m}^2/\text{N)} \]
\[ MRR = \text{material removal rate (m/s)} \]
\[ p = \text{pressure (N/m}^2) \]
\[ R_a = \text{radius of curvature of a pad asperity (m)} \]
\[ r_{cc}, r_{cc}^* = \text{wafer center to pad center distance (m), normalized value} \]
\[ r_{wo}, r_p = \text{wafer radius, pad radius (m)} \]
\[ t, t^* = \text{polishing time (s), normalized value} \]
\[ \Delta t_0 = \text{time for center of the wafer to be completely polished (s)} \]
\[ v_{cc}, v_{cc}^* = \text{translational velocity of the pad; rate of change of } r_{cc} \text{ (m/s), normalized value} \]
\[ v_R = \text{relative velocity of the wafer with respect to the pad (m/s)} \]
\[ w, w_s = \text{interconnect linewidth, surface trench width (m)} \]
\[ \alpha = \text{feature-scale non-uniformity factor, Cu deposition factor} \]
\[ \beta = \text{wafer-scale non-uniformity factor} \]
\[ \lambda = \text{pitch of Cu interconnect lines (m)} \]
\[ \lambda_a = \text{pad asperity spacing (m)} \]
\[ \omega_w, \omega_p = \text{angular velocities of the wafer and the pad (rad/s)} \]
CHAPTER 5

CONCLUSION

5.1 Summary

In this thesis, a face-up CMP tool architecture was introduced to control material removal rate non-uniformity at the wafer-scale, and to minimize such defects as Cu dishing and dielectric erosion. Models of material removal have been developed for both non-translating and translating pads. A numerical approach for determining pad translation for uniform wafer-scale polishing was also introduced. Polishing experiments on blanket and patterned Cu wafers were then conducted to validate these models.

Chapter 2 introduced the face-up CMP architecture by relating geometrical and kinematic parameters to the Preston Law for material removal rate. Equations for material removal rates for both non-translating and translating pads were derived based on the period of contact of a point on the wafer with the pad. A numerical method for determining that pad translation velocity was developed based on a system of equations that equate the total material removed to the initial Cu thickness. To practically implement the model, a discretization scheme that eliminates under-polishing and a method of determining the discretization error between the computational nodes were developed.

Face-up polishing experiments with a non-translating pad were presented in Chapter 3. Blanket Cu wafers were polished with pads containing different perforation patterns to investigate the kinematic effect of a rotating slurry cup on material removal rate and the Preston constant. It was found that slurry flow from the central region of the pad must be blocked in order to control the variation in Preston constant and obtain uni-directional polishing, a necessary condition for face-up CMP. Experiments were also performed with various rotational velocity ratios to examine kinematical methods of controlling the polishing gradient across the wafer. Increasing the wafer-to-pad velocity ratio resulted in a more gradual polishing gradient. However, decreasing the ratio did not show much difference when compared with the baseline case of equal velocities. Thus, absolute velocities also affect the polishing gradient and must be
controlled. The variation of Preston constants with pad location was also obtained by polishing blanket wafers. While the pad location can affect the Preston constant, the dominating trend is that Preston constant increases with radial position. The Preston constant was found to be more uniform at the periphery of the wafer.

Chapter 4 introduced face-up CMP experiments with a translating pad. Experiments were performed on blanket Cu wafers to validate the numerical pad translation model. The results showed that the model predicts the general pad motion fairly well. To implement face-up CMP for industrial use, the model can be used in conjunction with empirical data to compensate for pad edge effects and non-uniform Preston constant. Patterned wafers were then polished to determine Cu dishing and dielectric erosion across the wafer. The post-CMP topography was characterized for five different subdies at ten die locations on the wafer. The results showed that face-up polishing achieved fairly constant material removal rate at the wafer scale. Dielectric erosion was maintained below 5% of the interconnect depth for all features. Due to feature-scale non-uniformities, however, Cu dishing was significant and increased with feature linewidth. Nevertheless, the variation of dishing across the wafer for each feature was less than 4% the interconnect depth which also indicates good wafer-scale uniformity. Several suggestions for reducing dishing were presented.

5.2 Suggestions for Future Work

Based on the models and experiments presented in this thesis, further work is recommended to better utilize the face-up CMP architecture for minimizing Cu dishing and dielectric erosion.

Feature-scale Non-uniformity: Although this thesis demonstrates the ability of face-up CMP to control wafer-scale polishing uniformity, to fully eliminate dishing and erosion, the initial surface topography too must be controlled. Dishing is a pressing problem for large linewidths, \( w > 10 \mu m \), such as the interconnects used at the global wiring level. At these features, the underlying trench geometry is more or less reproduced on the Cu surface. Thus, alternative plating methods should be explored to fill the trenches so that polishing may be performed on a sufficiently planar surface.

Appendix B examines a method for controlling feature-scale uniformity by spin coating the wafer with a layer of polymer prior to CMP. Polishing a coated wafer showed promising
results for reducing dishing. Further work on optimizing this technique should be conducted to address issues such as: material compatibility, delamination, and ease of removal. The spin coating and curing parameters should also be examined for improved control over the process.

**Endpoint Detection:** Perhaps the greatest advantage of the face-up CMP architecture is the improved accessibility for in-situ endpoint detection. Since some portion of the wafer is exposed at all times during face-up polishing, merely two “snapshots” of the surface is adequate to obtain the planarization status of the entire wafer. While the pad translation model provides a functional basis for face-up CMP, better control may be attained through in-situ endpoint detection. Furthermore, endpoint sensing can be adapted for the variation in Preston constant without the need for the empirical data required by the model. Therefore, methods for obtaining the process endpoint during face-up polishing should be explored so that the pad can be precisely translated away from a feature soon after its endpoint is reached.
For ellipsometric erosion measurements after patterned wafer polishing, the initial SiO\textsubscript{2} layer thickness must be precisely known. A method for chemically etching the Cu coating and Ta barrier layer is therefore developed so that the thickness of an unpolished layer of SiO\textsubscript{2} can be measured by ellipsometry.

Copper is considered a “noble metal” because of its resistance to corrosion and oxidation. Atomically, this property is due to a filled d-band electronic structure – Cu has ten 3-d electrons and one 4-s electron. Nitric acid (HNO\textsubscript{3}) was therefore chosen to dissolve the Cu due to its oxidizing properties. The reaction of Cu with concentrated HNO\textsubscript{3} produces Cu\textsuperscript{2+} ions, nitrogen oxide, and water, and can be described by [Pauling, 1970]

\[
\text{Cu} + 4\text{HNO}_3 \rightarrow \text{Cu(NO}_3)_2 + 2\text{NO}_2 + 2\text{H}_2\text{O}
\]  

(A.1)

For the experiments, 10 to 20 mm square samples were cleaved from a blanket Cu wafer and a SKW6-2 patterned wafer. One milliliter HNO\textsubscript{3} was applied to the middle of the sample. After a minute of reaction time, the acid was cleaned off the sample with deionized water and the process was repeated to ensure that all the Cu was removed.

While effective in dissolving the Cu, nitric acid was unable to dissolve the Ta barrier layer. Tantalum, due to its resistance to acids, usually requires etchants that include hydrofluoric acid (HF). However, HF is also a known etchant for SiO\textsubscript{2}, and therefore must be avoided to ensure the entire initial layer of SiO\textsubscript{2} remains. For this reason, Grossman and Herman proposed a basic solution for dissolving Ta[Grossman and Herman, 1969].

After Cu etching, the samples were immersed in a heated solution of 30\% sodium hydroxide (NaOH) and 15\% hydrogen peroxide (H\textsubscript{2}O\textsubscript{2}) at 60°C for 30 seconds. The solution had a pH of 12. A blanket SiO\textsubscript{2} sample that was previously measured on the ellipsometer was also subjected to the same solution as a reference. The samples were then removed from the solution and rinsed with deionized water. A photograph of the samples after etching is shown in Figure A.1.

The SiO\textsubscript{2} film thickness on the etched samples was measured with a Gaertner ellipsometer using the quoted thickness as the initial guess. The quoted thickness is 1000 nm for
the blanket samples and 800 nm for the patterned sample. Measurements on the reference SiO$_2$ blanket sample are listed in Table A.1. There was no discernable difference in the average film thickness before and after etching, and a slight 1.7 Å increase in standard deviation. Thus the reference sample showed that the basic solution of NaOH and H$_2$O$_2$ does not remove SiO$_2$.

Tables A.2 and A.3 list the film thickness on the blanket Cu and patterned sample. Due to the lack of a field subdie in the patterned sample, measurements were taken at the oxide region in-between two subdies. An average thickness of 988.0 nm was obtained from the blanket sample with a standard deviation of 1.68 nm, yielding a 1.2% variation from the quoted thickness. For the patterned sample, the average measured thickness was 827.1 nm with a standard deviation of 20.2 nm for the patterned sample, which is only 3.4% off from the quoted value. The average SiO$_2$ thickness for the patterned sample was used as the initial SiO$_2$ thickness for quantifying erosion.
Figure A.1: Wafer samples after etching with NaOH+H$_2$O$_2$.

Table A.1: Measured SiO$_2$ thickness on the reference sample before and after etching.

<table>
<thead>
<tr>
<th>Measurement Number</th>
<th>Thickness (nm)</th>
<th>Before Etching</th>
<th>After Etching</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1023</td>
<td>1022</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1023</td>
<td>1024</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1024</td>
<td>1024</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1024</td>
<td>1023</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1023</td>
<td>1024</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1024</td>
<td>1023</td>
<td></td>
</tr>
<tr>
<td>Mean (nm)</td>
<td>1023</td>
<td>1023</td>
<td></td>
</tr>
<tr>
<td>Std. Dev. (nm)</td>
<td>0.47</td>
<td>0.64</td>
<td></td>
</tr>
</tbody>
</table>
Table A.2: Measured SiO$_2$ thickness on the blanket sample.

<table>
<thead>
<tr>
<th>Measurement Number</th>
<th>Thickness (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>991.0</td>
</tr>
<tr>
<td>2</td>
<td>986.3</td>
</tr>
<tr>
<td>3</td>
<td>987.2</td>
</tr>
<tr>
<td>4</td>
<td>986.8</td>
</tr>
<tr>
<td>5</td>
<td>988.4</td>
</tr>
<tr>
<td>6</td>
<td>988.2</td>
</tr>
<tr>
<td>Mean (nm)</td>
<td>988.0</td>
</tr>
<tr>
<td>Std. Dev. (nm)</td>
<td>1.68</td>
</tr>
</tbody>
</table>

Table A.3: Measured SiO$_2$ thickness on the patterned sample.

<table>
<thead>
<tr>
<th>Measurement Number</th>
<th>Thickness (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>853.0</td>
</tr>
<tr>
<td>2</td>
<td>827.4</td>
</tr>
<tr>
<td>3</td>
<td>832.1</td>
</tr>
<tr>
<td>4</td>
<td>851.0</td>
</tr>
<tr>
<td>5</td>
<td>818.3</td>
</tr>
<tr>
<td>6</td>
<td>823.6</td>
</tr>
<tr>
<td>7</td>
<td>836.8</td>
</tr>
<tr>
<td>8</td>
<td>832.1</td>
</tr>
<tr>
<td>9</td>
<td>827.7</td>
</tr>
<tr>
<td>10</td>
<td>780.9</td>
</tr>
<tr>
<td>11</td>
<td>801.2</td>
</tr>
<tr>
<td>12</td>
<td>840.7</td>
</tr>
<tr>
<td>Mean (nm)</td>
<td>827.1</td>
</tr>
<tr>
<td>Std. Dev. (nm)</td>
<td>20.2</td>
</tr>
</tbody>
</table>
APPENDIX B

CONTROL OF FEATURE-SCALE NON-UNIFORMITY BY SPIN COATING

B.1 Introduction

Feature-scale non-uniformity affects the polishing rates of individual subdies. Therefore, even when wafer-scale polishing is uniform, subdies with different geometry will have different polishing rates. According to the multi-scale tribological CMP model, the ideal condition is when $\alpha = 0$ and $h_i = 0$, i.e. the initial wafer surface is planar and independent of the underlying feature geometry, as shown in Figure B.1(a). Current Cu deposition methods, however, reproduce the feature geometry on the Cu surface for linewidths greater than 10 $\mu$m. A schematic of an actual Cu feature is shown in Figure B.1(b). Previous attempts to fill the metal features using methods such as electroplating have been unsuccessful, especially when the features are wide: $w \geq 10\ \mu$m. Hence, a novel method is proposed for creating a planar initial polishing surface by spin coating a layer of polymer on the wafer prior to polishing, as shown in Figure B.1(c). An epoxy-based photoresist, SU-8, was chosen as the polymer for spin coating due to its compatible mechanical properties and availability. Researchers have reported a Young’s modulus of 4.02 – 4.95 GPa based on screw tensile and beam deflection testing [Dellmann et al., 1997; Lorenz et al., 1997].

B.2 Film Thickness

Spin coating is a commonly employed technique for photoresist deposition. In this process, the photoresist solution is poured at the center of the wafer and the wafer is rotated at high speed to uniformly distribute the liquid on the surface. Depending on the fluid properties of the resist, and the spin speed and time, a uniform thin film can be obtained.

B.2.1 Hydrodynamic Model

Emslie et al. modeled the film thickness of a viscous Newtonian fluid on a rotating disk
Figure B.1: Schematics of (a) an ideal initial feature topography where $\alpha = 0$ and $h_{si} = 0$, (b) an actual feature topography where the feature is replicated on the Cu surface, and (c) a feature coated with polymer to obtain $\alpha = 0$ and $h_{si} = 0$. 
by assuming: (a) the disk is planar and radially infinite, (b) the fluid flow is axisymmetric, and (c) the fluid layer is thin, so that the viscous effects are much greater than the inertial effects [Emslie et al., 1958]. If the initial film thickness is uniform, the thickness of the fluid layer with time is given by:

\[ h = \frac{h_0}{1 + \left( \frac{4 \omega^2 h_0^2}{3 \nu} \right) \frac{t}{4 \omega^2 h_0^2}} \]  

(B.1)

where \( h \) is the film thickness, \( h_0 \) the initial film thickness, \( \omega \) the rotational speed of the disk, \( \nu \) the kinematic viscosity of the fluid, and \( t \) the time. The time constant, \( \tau \), represents the time required for the film thickness to reduce by \( 1/\sqrt{2} \), and is defined as

\[ \tau = \frac{3\nu}{4\omega^2 h_0^2} \]  

(B.2)

If \( t/\tau \gg 1 \), Eq. (B.1) can be reduced to

\[ \frac{h}{h_0} = \left( \frac{t}{\tau} \right)^{1/2} \]  

(B.3)

From Eq. (B.2), \( \tau \) is dependent on the initial film thickness. To measure the initial film thickness of the film, SU-8 photoresist was spin coated on a Cu wafer using the spreading speed of \( \omega = 105 \) rad/s (1000 rpm) for \( t = 10 \) s. The film was then cured according to the baking steps listed in Table B.1. After curing, a scratch was formed on the resist surface using a carbon steel blade and measured with Tencor P10 surface profilometer. A profile of the scratch is shown in Figure B.2. The Cu coating thickness was subtracted from the measured step-height to arrive at an initial film thickness of 5 \( \mu m \).

### B.2.2 Experiments with Different Viscosities

To validate the model, SU-8 photoresist of different viscosities were spin coated on Cu wafers at \( \omega = 314 \) rad/s (3000 rpm) for \( t = 30 \) s. The solids concentration of SU-8 was varied by diluting MicroChem SU-8 2002 photoresist with cyclopentanone and mixing the solution in a magnetic stirrer for 10 minutes. The viscosities of the solutions were estimated by linearly interpolating the available solids content versus viscosity data for SU-8 2000 series resists [MicroChem]. Table B.1 lists the steps for photoresist spin coating and curing. After curing, a
Table B.1: SU-8 spin coating and curing process steps.

<table>
<thead>
<tr>
<th>Step</th>
<th>Process Parameter</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wafer pre-bake</td>
<td>$T = 95 , ^\circ\text{C}$</td>
<td>3 min</td>
</tr>
<tr>
<td>SU-8 dispense</td>
<td>$V = 20 , \text{ml (200mm wafer)}$</td>
<td></td>
</tr>
<tr>
<td>Spin</td>
<td>$\omega = 105 , \text{rad/s (1000 rpm)}$</td>
<td>10 s</td>
</tr>
<tr>
<td></td>
<td>$\omega = 314 , \text{rad/s (3000 rpm)}$</td>
<td>30 s</td>
</tr>
<tr>
<td>Soft bake</td>
<td>$T = 95 , ^\circ\text{C}$</td>
<td>3 min</td>
</tr>
<tr>
<td>Exposure</td>
<td>Hg arc bulb</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\lambda_{\text{light}} = 405 , \text{nm}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$J = 3 , \text{mW/cm}^2$</td>
<td></td>
</tr>
<tr>
<td>Post-expose bake</td>
<td>$T = 95 , ^\circ\text{C}$</td>
<td>5 min</td>
</tr>
<tr>
<td>Hard bake</td>
<td>$T = 135 , ^\circ\text{C}$</td>
<td>10 min</td>
</tr>
</tbody>
</table>

Figure B.2: Profile of a scratch created on an SU-8 coating to measure initial film thickness.
scratch was made on the coated wafer surface and the scratch depth was measured with a profilometer to determine the thickness of the resist. The Cu coating thickness is subtracted from the total scratch depth to obtain the final thickness of the coating.

Table B.2 lists the results for film thickness when spin coating SU-8 of different viscosities. The computed time constants ranged from 0.82 to 2.28 s. Considering the spin time of 30 s, the resulting dimensionless time, $t/\tau$, ranges from 13 to 36. Thus, the approximation made in Eq. (B.3) is valid for the experimental conditions. The measured thickness was within 9% the theoretical thickness for the thinnest resist tested. However, the deviation between theoretical and experimental values increased with viscosity, with the thickest resist resulting in 44% deviation. The larger discrepancy when coating viscous fluids may be due to the variation in initial film thickness from test to test. Increasing viscosity results in larger time constants, which reduces the accuracy of the approximation made in Eq. (B.3). For this case, the model has a stronger dependence on the initial film thickness, which is not measured in-situ. Figure B.3 plots theoretical thickness versus experimental thickness. If the experimental data agreed with the theory completely, the points will fall on the 45-degree line. The data points being above the line indicates that the experimental results are higher than predicted. As thickness increased, a result of increased viscosity, the difference between the data points and the line also grew. The experimental dimensionless film thickness, $h/h_0$, as a function of $t/\tau$ is compared with the hydrodynamic model in Figure B.4.

B.3 Step Coverage

SKW6-2 test wafers were coated with SU-8 photoresist of various thicknesses to determine the planarity of the wafer surface after spin coating. The coating process parameters are as described in Table B.1. Figures B.5–B.9 show the profiles of features with lines ranging from 5 to 100 μm (a) before and (b) after spin coating a 4-μm thick layer of resist. For all features, the photoresist was able to fill the trenches reasonably well. However, the profiles also show a reduction in film thickness directly over the features. This is likely due to the extra material required to fill the trenches, which decreases the coating thickness when compared to a blanket region. Table B.3 compares the feature step-heights on an uncoated wafer with those on wafers coated with a 1 to 4 μm layer of SU-8 photoresist. The step-heights were reduced as the...
Table B.2: Theoretical and experimental results for the viscosity versus thickness spin coating experiments for $h_0 = 5 \text{ \mu m}$, $\omega = 314 \text{ rad/s}$, and $t = 30 \text{ s}$.

<table>
<thead>
<tr>
<th>% Solids</th>
<th>$\nu$ (m$^2$/s)</th>
<th>$\tau$ (s)</th>
<th>$\log(t/\tau)$</th>
<th>$h_{\text{theory}}$ (\mu m)</th>
<th>$\log(h/h_0)_{\text{theory}}$</th>
<th>$h_{\text{exp}}$ (\mu m)</th>
<th>$\log(h/h_0)_{\text{exp}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>$2.7 \times 10^{-6}$</td>
<td>0.82</td>
<td>1.56</td>
<td>0.82</td>
<td>-0.780</td>
<td>0.896</td>
<td>-0.747</td>
</tr>
<tr>
<td>22</td>
<td>$5.1 \times 10^{-6}$</td>
<td>1.55</td>
<td>1.29</td>
<td>1.14</td>
<td>-0.645</td>
<td>1.333</td>
<td>-0.574</td>
</tr>
<tr>
<td>29</td>
<td>$7.5 \times 10^{-6}$</td>
<td>2.28</td>
<td>1.12</td>
<td>1.38</td>
<td>-0.560</td>
<td>1.988</td>
<td>-0.401</td>
</tr>
</tbody>
</table>

Figure B.3: Comparison of experimental spin coating film thickness with theoretical film thickness.
Experimental dimensionless film thickness versus dimensionless time compared with the hydrodynamic model proposed by Emslie et al.

Figure B.4: Experimental dimensionless film thickness versus dimensionless time compared with the hydrodynamic model proposed by Emslie et al.
Figure B.5: Profile of a subdie (a) before and (b) after spin coating with a 4-μm layer of SU-8 photoresist, $w = 5 \mu m$ and $\lambda = 10 \mu m$. 
Figure B.6: Profile of a subdie (a) before and (b) after spin coating with a 4-μm layer of SU-8 photoresist, $w=10$ μm and $\lambda=20$ μm.
Figure B.7: Profile of a subdie (a) before and (b) after spin coating with a 4-µm layer of SU-8 photoresist, \( w = 20 \) µm and \( \lambda = 40 \) µm.
Figure B.8: Profile of a subdie (a) before and (b) after spin coating with a 4-μm layer of SU-8 photoresist, \( w = 50 \, \mu m \) and \( \lambda = 100 \, \mu m \).
Figure B.9: Profile of a subdie (a) before and (b) after spin coating with a 4-μm layer of SU-8 photoresist, \( w = 100 \, \mu \text{m} \) and \( \lambda = 200 \, \mu \text{m} \).

Table B.3: Comparison of feature step-heights for SU-8 coatings of various thicknesses.

<table>
<thead>
<tr>
<th>Linewidth (μm)</th>
<th>No Coating</th>
<th>( h_{SU-A} = 1 , \mu \text{m} )</th>
<th>( h_{SU-A} = 2 , \mu \text{m} )</th>
<th>( h_{SU-A} = 4 , \mu \text{m} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>732</td>
<td>28</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>1135</td>
<td>58</td>
<td>33</td>
<td>30</td>
</tr>
<tr>
<td>20</td>
<td>1132</td>
<td>77</td>
<td>45</td>
<td>36</td>
</tr>
<tr>
<td>50</td>
<td>1133</td>
<td>455</td>
<td>63</td>
<td>61</td>
</tr>
<tr>
<td>100</td>
<td>1173</td>
<td>1096</td>
<td>396</td>
<td>194</td>
</tr>
</tbody>
</table>
resist thickness increased, which is expected since more material was available to fill the trenches. The step-heights on the 4-μm coating ranged from 12 to 194 nm, a significant reduction from 732 to 1173 nm before coating.

B.4 Preston Constant

The Preston constant, \( k_p \), of the fill material should be comparable to Cu so that the structure will remain relatively planar throughout polishing. Therefore, a polishing experiment was performed to determine \( k_{p,\text{SU-8}} \). A 200-mm Cu wafer coated with SU-8 photoresist was polished on the face-up CMP tool with a blocked perforated TWI-817 CMP pad and iCue 5001 slurry. The process parameters are listed in Table B.4. The time required to remove the SU-8 from the central region was recorded and \( k_{p,\text{SU-8}} \) was found to be \( 1.9 \times 10^{-13} \) 1/Pa using Eq. (2.42). This value is about half of \( k_{p,\text{Cu}} \) from the Cu polishing experiments, \( 5.0 \times 10^{-13} \) 1/Pa. Considering the variation in \( k_p \), these values are within the same order of magnitude and therefore SU-8 photoresist is a suitable fill material for Cu CMP. It may be beneficial to find a material with a \( k_p \) even closer to that of \( k_{p,\text{Cu}} \) during process optimization to reduce polishing time and non-uniformity.

It was also observed that there was delamination at various points on the wafer during the polishing of resists thicker than 2 μm. This result is likely due to the thickness of the film which can affect exposure and curing. Therefore, the optimization of resist thickness for planarity and removability should be further studied.

B.5 Summary

A method of spin coating a polymer thin film on the wafer surface to improve feature-scale uniformity has been proposed. SU-8, an epoxy-based photoresist, was used as the fill-material for creating a planar initial polishing surface. First, experiments were performed to control the coating thickness by changing the resist viscosity. The results compared well with a hydrodynamic model for fluid film thickness on a rotating disk. To determine step coverage, SU-8 was spin coated on 200-mm patterned wafers to various thicknesses. Profiles of the features with linewidths ranging from 5 μm to 100 μm were obtained using a surface profilometer. Comparison of the feature topographies before and after spin coating shows an
improvement in planarity. The feature step-heights were reduced by 83 to 98%. Finally, the coated wafer was polished on a face-up CMP tool. The Preston constant of the SU-8 was found to be $1.9 \times 10^{-13}$ 1/Pa, which is on the same order of magnitude as that of Cu, $5.0 \times 10^{-13}$ 1/Pa. SU-8 is therefore an appropriate fill material for Cu CMP. However, delamination was observed at various regions across the wafer during the polishing of resists thicker than 2 μm. Further study is recommended to optimize the coating thickness for planarity and removability.
Table B.4: CMP conditions for determining the Preston constant of SU-8 photoresist.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{SU.8} \text{ (µm)}$</td>
<td>2.2</td>
</tr>
<tr>
<td>$r_w \text{ (mm)}$</td>
<td>100</td>
</tr>
<tr>
<td>$r_{w,\text{effective}} \text{ (mm)}$</td>
<td>50</td>
</tr>
<tr>
<td>Pad type</td>
<td>TWI-817</td>
</tr>
<tr>
<td>$r_p \text{ (mm)}$</td>
<td>35</td>
</tr>
<tr>
<td>$r_{cc} \text{ (mm)}$</td>
<td>30</td>
</tr>
<tr>
<td>Slurry</td>
<td>Cabot iCue 5001</td>
</tr>
<tr>
<td>Slurry additive</td>
<td>H$_2$O$_2$ - 3% vol</td>
</tr>
<tr>
<td>pH</td>
<td>8</td>
</tr>
<tr>
<td>$\omega_w \text{ (rad/s) (rpm)}$</td>
<td>16 (155)</td>
</tr>
<tr>
<td>$\omega_p \text{ (rad/s) (rpm)}$</td>
<td>16 (155)</td>
</tr>
<tr>
<td>$v_{cc} \text{ (m/s)}$</td>
<td>0</td>
</tr>
<tr>
<td>$p \text{ (kPa) (psi)}$</td>
<td>13 (1.9)</td>
</tr>
<tr>
<td>Slurry flow rate (ml/min)</td>
<td>150</td>
</tr>
<tr>
<td>$\Delta t_0 \text{ (min)}$</td>
<td>31</td>
</tr>
</tbody>
</table>
Nomenclature

\( h = \) spin coating film thickness (m)
\( h_0 = \) initial film thickness prior to spin coating (m)
\( h_{Cu}, h_{SU-8} = \) initial Cu coating thickness, SU-8 coating thickness (m)
\( h_s = \) step-height (m)
\( h_{si} = \) initial feature step-height (m)
\( J = \) photoresist exposure source flux (J/m\(^2\)/s)
\( k_{p,Cu}, k_{p,SU-8} = \) Preston constant for Cu, SU-8 (m\(^2\)/N)
\( p = \) pressure (N/m\(^2\))
\( r_{cc} = \) wafer center to pad center distance (m), normalized value
\( r_w, r_p = \) wafer radius, pad radius (m)
\( T = \) temperature (°C)
\( t = \) spin coating time (s)
\( V = \) volume of photoresist solution dispensed for spin coating (m\(^3\))
\( \Delta t_0 = \) time for center of the wafer to be completely polished (s)
\( v_c = \) translational velocity of the pad; rate of change of \( r_{cc} \) (m/s)
\( v_R = \) relative velocity of the wafer with respect to the pad (m/s)
\( w = \) feature linewidth (m)
\( \alpha = \) feature-scale non-uniformity factor, Cu deposition factor
\( \lambda = \) pitch of the Cu interconnect lines (m)
\( \lambda_{light} = \) wavelength of photoresist exposure source (m)
\( \nu = \) kinematic viscosity of the spin coating material (m\(^2\)/s)
\( \tau = \) time constant (s)
\( \omega = \) angular speed of wafer during spin coating (rad/s)
\( \omega_w, \omega_p = \) angular velocities of the wafer and the pad during CMP (rad/s)
References


