PARTICLE SCALE MODELING OF MATERIAL REMOVAL AND SURFACE ROUGHNESS IN CHEMICAL MECHANICAL POLISHING

By

SURESH BABU YERUVA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2005
Copyright 2005

by

Suresh Babu Yeruva
This document is dedicated to Sri Paramahansa Yogananda, my parents and sister.
ACKNOWLEDGMENTS

I would like to express my gratitude and appreciation to my advisor, Dr. Brij Moudgil, for his valuable guidance and support throughout this study. His sincere dedication towards his students encouraged me to perform my best and expand my potential. Sincere thanks are also due to Dr. Chang-Won Park, for his valuable guidance and support throughout the study. I would also like to acknowledge the other members of my advisory committee, Dr. Rajiv Singh, Dr. Wolfgang Sigmund, Dr. Hassan El-Shall, Dr. Susan Sinnott and Dr. Dinesh Shah, for their time, helpful discussions and guidance.

I would like to thank Dr. Yakov Rabinovich for help with atomic force microscope (AFM) and Fourier transform infrared spectroscopy (FTIR) analysis, and Gerald Bourne from Major Analytical Instrumentation Center (MAIC) for help with focused ion beam (FIB) and nanoindentation data. Furthermore, I sincerely acknowledge the guidance and support given by Kuide Qin at Praxair Inc.

I gratefully acknowledge the National Science Foundation’s Particle Engineering Research Center (PERC) for financially supporting this research. Special thanks are due to PERC staff members, Gary Schieffele, Gill Brubaker, Mike Beasley, and John Henderson who made this work possible.

My colleagues, Bahar Basim, Jeremiah Abiade, Su-Ho Jung, Vijay Krishna, Madhavan Esayanur, Scott Brown, Parvesh Sharma, Kyoung-Ho Bu, Pankaj Singh, Debamitra Dutta and Ivan Vakarelski, who have contributed to this research by their valuable discussions and friendship, are also gratefully acknowledged.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>xi</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>Multilevel Metallization</td>
<td>2</td>
</tr>
<tr>
<td>Need for Planarization</td>
<td>4</td>
</tr>
<tr>
<td>Planarization Techniques</td>
<td>5</td>
</tr>
<tr>
<td>Overview of the Chemical Mechanical Polishing Process</td>
<td>7</td>
</tr>
<tr>
<td>History</td>
<td>7</td>
</tr>
<tr>
<td>Advantages and Disadvantages of Chemical Mechanical Planarization</td>
<td>8</td>
</tr>
<tr>
<td>General CMP applications</td>
<td>9</td>
</tr>
<tr>
<td>CMP Output Parameters</td>
<td>9</td>
</tr>
<tr>
<td>Removal rate</td>
<td>10</td>
</tr>
<tr>
<td>Global planarization</td>
<td>12</td>
</tr>
<tr>
<td>Selectivity</td>
<td>13</td>
</tr>
<tr>
<td>Measurement Techniques</td>
<td>14</td>
</tr>
<tr>
<td>End Point Detection</td>
<td>15</td>
</tr>
<tr>
<td>CMP Consumables</td>
<td>16</td>
</tr>
<tr>
<td>CMP Slurry</td>
<td>18</td>
</tr>
<tr>
<td>Rate of surface reaction</td>
<td>18</td>
</tr>
<tr>
<td>Slurry agglomeration</td>
<td>19</td>
</tr>
<tr>
<td>Chemical Mechanical Polishing Pads</td>
<td>20</td>
</tr>
<tr>
<td>Effect of pad grooves</td>
<td>20</td>
</tr>
<tr>
<td>Physical and chemical changes in pad during polishing</td>
<td>23</td>
</tr>
<tr>
<td>Effect of temperature change on polishing pad</td>
<td>26</td>
</tr>
<tr>
<td>Non-uniformity of the polishing pad</td>
<td>27</td>
</tr>
<tr>
<td>CMP process in the future</td>
<td>27</td>
</tr>
<tr>
<td>Scope of Dissertation</td>
<td>28</td>
</tr>
<tr>
<td>2 CMP MODELING: OVERVIEW OF PERC I</td>
<td>30</td>
</tr>
<tr>
<td>Literature Review</td>
<td>31</td>
</tr>
<tr>
<td>Table</td>
<td>page</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>1-1. Interconnection delay (RC) in silicon VLSI chip.</td>
<td>3</td>
</tr>
<tr>
<td>1-2. Chemical mechanical polishing applications.</td>
<td>10</td>
</tr>
<tr>
<td>4-1. Parameters included in the revised PERC model.</td>
<td>85</td>
</tr>
<tr>
<td>4-2. Parameter values used for comparing PERC I and II models.</td>
<td>95</td>
</tr>
<tr>
<td>5-1. Process parameter ranges used in sensitivity analysis.</td>
<td>103</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1. Future trends in interconnects</td>
<td>4</td>
</tr>
<tr>
<td>1-2. Schematic of a device fabricated using multilevel metallization</td>
<td>5</td>
</tr>
<tr>
<td>1-3. Different degrees of local and global surface planarity</td>
<td>6</td>
</tr>
<tr>
<td>1-4. Schematic of chemical mechanical polishing process</td>
<td>7</td>
</tr>
<tr>
<td>1-5. Surface layer formation on Cu with various slurry chemistries</td>
<td>11</td>
</tr>
<tr>
<td>1-6. Micro and nanoscale phenomena during CMP</td>
<td>13</td>
</tr>
<tr>
<td>1-7. Parameters involved in CMP process</td>
<td>17</td>
</tr>
<tr>
<td>1-8. Chronoamperometry measurements for tungsten</td>
<td>19</td>
</tr>
<tr>
<td>1-9. Type I CMP pad</td>
<td>22</td>
</tr>
<tr>
<td>1-10. Type II CMP pad</td>
<td>22</td>
</tr>
<tr>
<td>1-11. The absorbance spectra of used and dry IC 1000 pad along with dried silica slurry</td>
<td>25</td>
</tr>
<tr>
<td>2-1. Polished wafer surfaces showing nanoscratch surfaces</td>
<td>33</td>
</tr>
<tr>
<td>2-2 Pad consideration in PERC I</td>
<td>46</td>
</tr>
<tr>
<td>2-3. Schematic of single particle abrasion showing the removal of chemically modified surface layer</td>
<td>51</td>
</tr>
<tr>
<td>2-4. Schematic of PERC I predictions for increasing particle size</td>
<td>57</td>
</tr>
<tr>
<td>2-5. Polishing of silica by nanoslurry showing optimal particle size</td>
<td>57</td>
</tr>
<tr>
<td>2-6. Polishing of silica at various pressure, velocity and particle sizes showing optimal particle size</td>
<td>58</td>
</tr>
<tr>
<td>3-1. Schematic of the fixture used in measuring pad-wafer (dry) contact area</td>
<td>62</td>
</tr>
</tbody>
</table>
3-2. Pad asperity contact spots seen through optical microscope.................................63

3-3. Pad asperity contact area dependence on applied load. Each legend represents a
selected asperity contact..............................................................................................63

3-4. Pad contact area as a function of applied load......................................................64

3-5 Schematic of the FTIR-ATR technique used to measure pad-wafer contact area......65

3-6. Analysis of the FTIR-ATR spectra..........................................................................67

3-7. Absorbance spectra of pad (in pH 4) at different applied pressures.....................67

3-8. Contact area variation with applied load in dry and wet conditions......................68

3-9. Contact area in dry and wet systems in various pH conditions..............................68

3-10. Colloidal probe AFM tip .....................................................................................70

3-11. Force-distance curve indicating softer “modified layer” on the pad surface. ........71

3-12. Force-indentation curves from AFM for different conditions..............................72

3-13. Schematic of pad surface as deduced from AFM data.........................................72

3-14. Schematic of polishing pads................................................................................73

3-15. Particle size distribution of silica slurry..............................................................75

3-16. Schematic of pad, slurry and wafer in PERC I model..........................................76

3-17. Schematic of pad, slurry and wafer in PERC II model........................................77

4-1. Slurries used in oxide polishing............................................................................87

4-2. Silica polishing by slurries in the micron range.....................................................88

4-3. Silica polishing by colloidal nanosilica.................................................................89

4-4. Silica polishing by nanosized slurries.................................................................90

4-5 Copper CMP as a function of particle size............................................................92

4-6. Alumina slurry size distributions..........................................................................93

4-7. Effect of particle size on tungsten CMP...............................................................94

4-8. Comparison between PERC I and PERC II models with constant solids loading of
different particle sizes.............................................................................................96
4-9. Model comparison with varying modified layer thickness. ..........................98
4-10. CMP performance comparison with varying modified layer hardness. ..........100
5-1. Sensitivity analysis of silica CMP system. ...............................................103
PARTICLE SCALE MODELING OF MATERIAL REMOVAL AND SURFACE ROUGHNESS IN CHEMICAL MECHANICAL POLISHING

By

Suresh Babu Yeruva

December 2005

Chemical mechanical polishing (CMP) is widely adopted in producing excellent local and global planarization of microelectronic devices. It has been demonstrated experimentally that the polishing performance is a result of the synergistic effect of both the chemicals and the particles involved in CMP. However, the fundamental mechanisms of material removal and the interactions of the chemical and mechanical effects are not well understood. A comprehensive model for CMP was developed taking into account both the chemical and mechanical effects for slurries with a given particle size distribution. The model developed (PERC II) is based on a previously developed model (PERC I). The chemical aspect is attributed to the chemical modification of the surface layer due to slurry chemistry, whereas the mechanical aspect is incorporated by indentation of particles into the modified layer and the substrate depending on the
operating conditions. In this study, the effects of particle size and pad asperity distributions are included in the model.

The contact area of pad with wafer was measured in dry and wet conditions in different pH solutions using optical microscopy and Fourier transform infrared spectroscopy respectively. Pad surface mechanical properties in dry and wet states were also investigated using atomic force microscopy. The contact area results obtained were utilized in modeling to estimate the pad modulus leading to pad-wafer contact stress distribution.

The predictions of the model show a reasonable agreement with the experimental data. The model is validated for oxide and metal CMP systems. The PERC II model not only predicts the overall removal rate, but also the surface roughness of the polished wafer in selected systems. The developed model can be used to optimize the current CMP systems and provide insights into future CMP endeavors.
CHAPTER 1
INTRODUCTION

The past three decades have witnessed rapid growth of the electronics industry and its impact on human lifestyle. Ultra-large-scale integrated (ULSI) circuits, with $10^8$ or more devices on a chip, can now be fabricated on semiconductor wafers, reducing cost and increasing the performance of electronic products. The number of components on a memory chip has approximately doubled every two years over the past two decades, following the famous Moore’s law [1]. The advancement in fabrication technology shrinks the minimum dimension of the device-feature by about 13% per year. This trend is expected to continue in the future even as 65 and 45 nm processes are set for production in 2005 [2] and near future.

The relentless competitor and customer driven demand for increased circuit density, functionality and versatility has led to revolutionary advances in the “front end” of the chip manufacturing line where the circuit elements are fabricated, and the “back end” where these elements are appropriately wired within the integrated circuit (IC) [3]. Chip interconnections, or “interconnects,” serve as local and global wiring, connecting circuit elements and distributing power [4]. To incorporate and accommodate the improvements such as decreased feature size, increased device speed and more intricate designs, research in the back end of line (BEOL) processes has become equally important as the development of the front end of line (FEOL) processes to reduce gate oxide thickness and channel length.
While recent path breaking innovations [5-7] in the field of lithography and patternning have brought progressive device scaling, the development of a planar BEOL approach, which incorporates the use of chemical mechanical polishing (CMP) to planarize inter-level dielectrics and metal stud levels, represents a significant advancement in BEOL processes. Innovation in BEOL technology is required in each technology generation, since only part of the density increase could be achieved with improvements in lithography.

**Multilevel Metallization**

In earlier stages, the complementary metal oxide semiconductor (CMOS) device structure had multiple isolated devices connected by a single level of interconnect wiring. Scaling down of the devices was very effective in achieving the goals of increased device density, functional complexity and performance. However, device scaling became less profitable, and speed and complexity were dependant on the characteristics of interconnects that wired them [8]. With the single level metallization scheme, the total area occupied by the wiring on the chip significantly increased with the increase in the active density on the chip.

The interconnect delay is usually measured as RC time delay, which is the total time taken by the voltage at one end of the metal line to reach 63% of the total value of the step input applied at the other end. This reduction is due to resistance of the interconnect wiring metal (R) and the interlayer dielectric capacitance (C) [9]. The resistance of the line is given by

\[
R = \rho \frac{l}{wd} \quad (1.1)
\]

where \(\rho\) is the resistivity of the wiring material, \(l\), \(w\), and \(d\) are the length, width, and thickness of the wiring material respectively. The capacitance of the line is given by
\[ C = \varepsilon \frac{w}{t} \]  
(1.2)

where \( \varepsilon \) is the relative permittivity and \( t \) is the thickness of the insulator. The total RC delay can be given by

\[ RC = \frac{\rho}{\varepsilon} \frac{l}{wd} \frac{wl}{t} = \rho \varepsilon \frac{l^2}{td} \]  
(1.3)

Thus, it can be seen from Eq. (1.3) that RC delay is independent of the line width and further scaling of line width translates into reduction of IC line thickness which in turn increases the RC delay. Other factors such as parasitic capacitances and cross-talk become dominant for sub 0.5 \( \mu \)m integrated circuits. One approach to decrease the RC delay is by incorporating metals of low resistivity, and interlayer dielectrics of lower dielectric constant. Table 1-1 gives the simple RC time constants calculated for a few metals of given \( R_s \) (sheet resistance) and 1 mm length on 1\( \mu \)m thick SiO\(_2\) [9]. The other approach implemented widely in the industry is forming multilevel metallization schemes where different levels of metal interconnections are isolated by dielectrics and are connected by vertical vias.

Table 1-1. Interconnection delay (RC) in silicon VLSI chip.

<table>
<thead>
<tr>
<th>Metal</th>
<th>Bulk resistivity (( \mu \Omega ) cm)</th>
<th>Polycrystalline film resistivity (( \mu \Omega ) cm)</th>
<th>Film thickness (Å)</th>
<th>( R_s ) (( \Omega ) /square)</th>
<th>Delay(^i) (ps/mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polysilicon</td>
<td>-</td>
<td>~1000</td>
<td>5000</td>
<td>20</td>
<td>690</td>
</tr>
<tr>
<td>CoSi(_2)</td>
<td>10</td>
<td>15</td>
<td>2500</td>
<td>0.6</td>
<td>21</td>
</tr>
<tr>
<td>MoSi(_2)</td>
<td>~35</td>
<td>~100</td>
<td>2500</td>
<td>4</td>
<td>138</td>
</tr>
<tr>
<td>TaSi(_2)</td>
<td>45</td>
<td>55</td>
<td>2500</td>
<td>2.2</td>
<td>76</td>
</tr>
<tr>
<td>TiSi(_2)</td>
<td>13</td>
<td>15</td>
<td>2500</td>
<td>0.6</td>
<td>21</td>
</tr>
<tr>
<td>WSi(_2)</td>
<td>~25</td>
<td>70</td>
<td>2500</td>
<td>2.8</td>
<td>97</td>
</tr>
<tr>
<td>W</td>
<td>5.65</td>
<td>8-10</td>
<td>2500</td>
<td>0.32-0.4</td>
<td>11-14</td>
</tr>
<tr>
<td>Mo</td>
<td>5.2</td>
<td>8-10</td>
<td>2500</td>
<td>0.32-0.4</td>
<td>11-14</td>
</tr>
<tr>
<td>Al</td>
<td>2.65</td>
<td>2.7</td>
<td>2500</td>
<td>0.11</td>
<td>4</td>
</tr>
<tr>
<td>Cu</td>
<td>1.67</td>
<td>2.0</td>
<td>2500</td>
<td>0.08</td>
<td>3</td>
</tr>
</tbody>
</table>

\(^i\)Delay=RC=34.5 \( R_s \) (ps/mm) for 1 mm length conductor on 1\( \mu \)m thick SiO\(_2\) [10].
The increase in the levels of the metallization lines implies that packing density need not keep pace with the device density and the minimum metal line feature does not have to scale at the same pace as the gate width. The foremost reason behind the implementation of multilevel metallization schemes is the reduction in the length of the metal lines, thereby reducing the RC delay sizably (Eq. 1.3), and allowing direct routing of the active devices. In places where metal wiring length cannot be reduced, routing can be done at the upper levels without reducing the metal line width, thus reducing the RC delay due to the higher surface area. Starting with two levels of metallization, the levels of metallization have increased up to 8 by 2001 [11]. The future trends in the levels of metallization can be seen in Fig. 1-1 [12].

![Fig. 1-1. Future trends in interconnects [12]](image)

**Need for Planarization**

Deposition of multiple layers leads to conformation of deposited layer on the substrate features or already deposited pattern. When multilevel metallization is carried out without any planarization step, non-planarized surface topography (Fig. 1-2a) results
in several processing difficulties. The irregular surface causes a hindrance in conformal coating of the photoresist and efficient pattern transfer with contact lithography. The anomalies in the surface cause the variation of the thickness in fine line widths (sub 0.5 μm) depending on photoresist thickness. Fig. 1-2 shows a comparison between planarized and non-planarized surface topography. Effectively planarized surfaces have a number of benefits such as (a) higher photolithography and dry etch yields, (b) elimination of step coverage concerns, (c) minimization of prior level defects, (d) elimination of contact interruption, undesired contacts and electro-migration effects, (e) reduction of high contact resistance and inhomogeneous metallization layer thickness, and (f) limitation in the stacking height of metallization layers. Thus the demand for developing devices with enhanced performance necessitates planarization.

Fig. 1-2. Schematic of a device fabricated using multilevel metallization. (a) Non-planarized structure. (b) Planarized structure [8, 13]

**Planarization Techniques**

Techniques such as spin on deposition (SOD), reflow of boron phosphorous silicate glass (BPSG), spin etch planarization (SEP), reactive ion etching and etch back (RIE EB), SOD + EB and chemical mechanical polishing (or chemical mechanical planarization) (CMP) are the most prominent of several competing technologies available
to achieve local and global planarization (Fig. 1-3). The plasma enhanced chemical vapor deposited oxides have limited capability of gap filling and are restricted in their gap filling ability below patterns having 0.3-μm feature size. High-density plasma deposited oxides have acceptable gap filling capabilities; however, they produce variation in surface topography at local as well as global level. Even though spin-on deposited (SOD) doped and undoped oxides and polymeric materials have acceptable ability for gap filling, CMP is the only technique which produces excellent local and global planarization of microelectronic materials.

Fig. 1-3. Different degrees of local and global surface planarity [14].
Overview of the Chemical Mechanical Polishing Process

Current semiconductor fabrication technology for logic and memory devices requires CMP to achieve the required multilevel interconnection densities. Indeed, each silicon wafer can be exposed to 15 or more CMP steps before final device assembly. A schematic diagram of the CMP process is shown in Fig. 1-4. In CMP, the wafer is pressed face down against a rotating polishing pad, while a chemical and physically abrasive slurry fed between pad and wafer planarizes the wafer. The material removal in simplest terms takes place by synergistic mechanism between chemical reaction on the wafer surface due to the slurry, and physical removal of the reacted material by the slurry abrasives. The various mechanisms between pad, slurry and wafer are tailored to enhance material removal, planarization, surface finish and higher selectivity compared to the underneath layer.

Fig. 1-4. Schematic of chemical mechanical polishing process.

History

The modern day application of the CMP process in the semiconductor industry was for polishing the raw silicon wafers to achieve a global flatness. One of the most widespread applications of CMP outside the semiconductor industry is in optical lens
polishing. In fact, the first semiconductor CMP machine was an innovation of the lens polishing industry. The first wafers polished using CMP were commercially available in the early 1960s [15]. There was wide spread industrial implementation of the different variants of the CMP by companies such as NEC, National Semiconductor, Hitachi, etc. [10]. This led to the introduction of the first commercial polisher designed specifically for CMP by Cybeg, Japan in 1988. Later, International Sematech identified CMP as a technology critical for the future of IC manufacturing and launched a project to develop competitive and advanced CMP tools in the US [16]. Once the technology and required equipment were available, CMP was used beyond polishing of inter-level dielectric layers to polishing of tungsten, which was deposited in via openings between metal layers. Copper was not used for interconnects until the availability of CMP even though it has lower resistivity than aluminum [17], as it could not easily be etched by reactive ion etching. This enabling polishing technology has spread to several other applications, the significant one being in shallow trench isolation (STI) [18].

**Advantages and Disadvantages of Chemical Mechanical Planarization**

The advantages of CMP are (a) it achieves global planarization, (b) a wide range of materials can be planarized, sometimes in the same polishing step, (c) it reduces severe topography to allow fabrication with tighter design rules, (d) it provides an alternate means of patterning metal, eliminating the need to plasma etch, (e) it improves metal step coverage due to reduction in topography and (f) it does not use hazardous gasses common in dry etch processes.

Conversely, the disadvantages of the CMP process are (a) poor control over the process with narrow latitude in process variables, (b) new types of defects introduced by the CMP process itself (scratches, particle residue, etc.), (c) difficult end-point detection
for desired thickness and (d) expensive to operate because of costly equipment and consumables.

General CMP applications

The process of CMP was initially developed and implemented for planarization of SiO$_2$ which is used as an interlayer dielectric in a multilevel metallization scheme. Tungsten is used as an interconnect plug to the source, drain, and gates of transistors in Si microprocessor chips. Initially Ti and TiN barrier layers are deposited, followed by chemical vapor deposition of W to fill the contact vias. Proceeding from achieving local and global planarization of SiO$_2$, removal of excessive tungsten from the horizontal surfaces on the wafer pattern proved to be an asset for subsequent Al metallization. Hence CMP was developed with a two-fold approach of (a) planarizing oxide and (b) removing the via fill metal from the horizontal surfaces. Along with its successful implementation for the achievement of the above-mentioned objectives, CMP is now extended to (a) polishing of different metals like Al, Cu, Pt, Au, Ti, Ta, etc., (b) polishing of different insulators like SiO$_2$, Si$_3$N$_4$, various low-k dielectrics, doped and undoped oxides of silicon, (c) polysilicon, (d) ceramics like SiC, TiN, TaN, etc., (e) multichip modules, (f) packaging; (g) optoelectronic components, (h) flat panel displays, (i) microelectromechanical systems (MEMS), and (j) magnetic recording heads and CD read write drives [19]. The major applications of CMP are summarized in Table 1-2.

Cmp Output Parameters

The critical output parameters of CMP are described below. Depending on the application, different output parameters are optimized.
Table 1-2. Chemical mechanical polishing applications.

<table>
<thead>
<tr>
<th>Materials</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metal</strong></td>
<td></td>
</tr>
<tr>
<td>Al</td>
<td>Interconnection</td>
</tr>
<tr>
<td>Cu</td>
<td>Interconnection</td>
</tr>
<tr>
<td>Ta</td>
<td>Diffusion barrier/adhesion promoter</td>
</tr>
<tr>
<td>Ti</td>
<td>Diffusion barrier/adhesion promoter</td>
</tr>
<tr>
<td>TiN, TiNxCy</td>
<td>Diffusion barrier/adhesion promoter</td>
</tr>
<tr>
<td>W</td>
<td>Interconnection e-emitter</td>
</tr>
<tr>
<td><strong>Dielectric</strong></td>
<td></td>
</tr>
<tr>
<td>Cu-alloys</td>
<td>Interconnection</td>
</tr>
<tr>
<td>Al-alloys</td>
<td>Interconnection</td>
</tr>
<tr>
<td>Polysilicon</td>
<td>Gate/interconnect</td>
</tr>
<tr>
<td>SiO$_2$</td>
<td>ILD</td>
</tr>
<tr>
<td>BPSG</td>
<td>ILD</td>
</tr>
<tr>
<td>PSG</td>
<td>ILD</td>
</tr>
<tr>
<td>Polymers</td>
<td>ILD</td>
</tr>
<tr>
<td>Si$_3$N$_4$ or SiO$_x$N$_y$</td>
<td>Passivation layer, hard mask</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
</tr>
<tr>
<td>Aerogels</td>
<td>ILD</td>
</tr>
<tr>
<td>ITO</td>
<td>Flat panel</td>
</tr>
<tr>
<td>High k dielectrics</td>
<td>Packaging</td>
</tr>
<tr>
<td>High T$_2$ superconductors</td>
<td>Interconnection/packaging</td>
</tr>
<tr>
<td>Optoelectronic materials</td>
<td>Optoelectronics</td>
</tr>
<tr>
<td>Plastics, ceramics</td>
<td>Packaging</td>
</tr>
<tr>
<td>Silicon on Insulator (SOI)</td>
<td>Advanced device/circuits</td>
</tr>
<tr>
<td>Ref. [19]</td>
<td></td>
</tr>
</tbody>
</table>

**Removal rate**

The polishing steps involving CMP of STI and metal mainly emphasize removing an overburden of one material and stop on another material leaving a planar surface. A high material removal rate is one of the key outputs desired from such CMP processes. The removal rate of the material being polished is strongly influenced by slurry chemistry and pressure distribution, i.e., chemical action of the slurry chemicals on the material, interplay of the different complexing agents, oxidizers and corrosion inhibitors and the mechanical abrasion of the particles on the wafer. Chemicals such as oxidizers and corrosion inhibitors vastly affect the reaction rate of the slurry with similar particle nature, size and distributions.
Fig. 1-5. Surface layer formation on Cu with various slurry chemistries. [20]

Fig. 1-5 shows the variation of the reaction rate of the different slurry components on Cu when the reaction kinetics were studied using electrochemical chronoamperometry [20]. It can be seen from the figure that the surface rate kinetics reach approximately 60 Å/s when Cu is immersed in distilled water. The reaction rate increases to around 120 Å/s when 5% H₂O₂ is added. However upon addition of 10 mM BTA, the reaction rate decreased considerably. Therefore, the chemical contribution is very important in the CMP process. The other important factor is the down pressure used. However, for effective CMP of materials, the enhancement of removal rate by increasing the down force should not be at the cost of surface quality (for example, increase in wafer to wafer and within wafer non-uniformity, delamination, dishing, erosion, etc.) [21, 22] of the wafer. For this reason, the process throughput is compromised to polish the wafers at lower applied load, thereby achieving global planarization and better surface quality.
**Global planarization**

In contrast to glass or silicon wafer polishing, the objective of the CMP process in interlayer dielectric (ILD) polishing is to planarize the topography created by previous semiconductor processing steps. The pad and slurry play key roles in achieving global planarization, concurrently maintaining the removal rate and producing a defect free wafer surface.

If mechanical removal predominates, high frictional forces are produced which result in surface defectivity. Also variation in local polishing pressure leads to variable removal rates within the wafer, which seriously compromises global planarization [20], whereas, if chemical removal dominates, etching and corrosion [23] can occur. The key to a good polishing process is dependent on synergy between chemical etching and mechanical planarization. For this purpose, a passivation layer is created on the wafer surface by reaction with slurry and removal of this passivation (or modified) layer occurs due to mechanical abrasion of slurry particles (Fig. 1-6). For example, in Cu polishing, the formation of the passivation layer is achieved by oxidizers such as hydrogen peroxide, potassium ferricyanate, ferric chloride, and ferric iodate. Corrosion damage to the surface is prevented by adding corrosion inhibitors such as benzotriazole (BTA). In tungsten polishing, a rapid formation of surface passivation layer is due to the use of peroxxygen compounds and stabilizing agents [24]. The purpose of the passivation layer in the case of silica polishing is to soften the surface, which is inherently hard resulting in minimal material removal. For this purpose, maintenance of alkaline pH is sufficient in most cases [25, 26]. During Ta polishing, formation of stable Ta$_2$O$_5$ helps in uniform removal of material from the surface [27]. To achieve optimal CMP performance, the time to form the passivation layer of desired thickness should be minimized.
Selectivity

The difference in the removal rate of one material versus another in a given system represents the measure of the slurry selectivity for these two materials. The higher the slurry selectivity, the more effective the end point detection is for the given CMP process. This is due to a marked change between the tribological properties of the material being polished, and the underlayer. The difference in tribological properties can be monitored in-situ using techniques such as motor current, force or acoustic emission for better end point detection. Selectivity is a very important criterion in designing any slurry for multilevel metallization. For STI, the slurry must affect the oxide that is being planarized, and not the underlying nitride layer. For polishing Cu, the slurry needs to react selectively on Cu and spare the barrier layer Ta and underlying layers of silica or low-k dielectric material. Typically slurry selectivity of 10–25 has been reported for Cu polishing [20].
The selectivity could be considerably improved up to 1000 by introducing a particle free slurry [28, 29]. The reduction in mechanical component due to lack of abrasive implies that the majority of material removal occurs due to the solution chemistry which can be made highly selective.

Addition of specialized chemicals that can act as catalysts in the chemical interaction between slurry and the polished material thereby increasing the rate of the chemical reaction considerably is a common approach for improving the slurry selectivity. Tetramethyl ammonium hydrate can be added to Cu slurries to considerably decrease the silica polishing rate [30]. An increase in the polishing rate of TaN with addition of phosphoric acid in alumina and colloidal silica slurry is also observed. Mixed abrasive slurries (MAS) containing alumina/silica and alumina/ceria particles have been developed to improve the selectivity of the CMP slurry. The details of the experiments with these novel mixed abrasive slurries have been published by Jindal et al. [31, 32]. Though highly selective slurries might help in avoiding defects such as erosion of sub layers, planar defects such as dishing still exist causing wafer-to-wafer nonuniformity (WTWNU) due to the differential polishing pressures in the process.

**Measurement Techniques**

The outputs of the CMP process are quantified by a host of metrics. Material removal rate (e.g., nm/min) is commonly measured in terms of thickness loss per time or weight loss per time. In terms of global planarization at the wafer scale, film thickness is measured at points across the wafer surface and used to calculate within-wafer-non-uniformity (WIWNU). Most commonly, a 49-point measurement grid is used to determine the WIWNU, which is the variation of those measurements with respect to the average thickness. Similarly, a within-die non-uniformity (WIDNU) calculates a similar
variation but just across the area of one die. At the feature-scale of the process, metrics such as step height reduction are used to indicate efficiency of the CMP process. Selectivity of the slurry is more often characterized by estimating the erosion or polishing of the underneath layer.

**End Point Detection**

Theoretically, the end of a polishing step can be realized by setting a time limit to the process. However, this is only a rough estimate because variations of polishing pad surface property, fluctuations in slurry chemistry, particle size distribution, feeding rate, pad conditioning issues, and other potential variables can cause over- or under-polishing. Therefore, in-line monitoring or automatic end-point detection of CMP is needed. Initial oxide CMP relied on process times and control of removal rate stability to achieve target thicknesses. As the CMP environment is hostile for in-situ metrology, the earlier endpoint techniques relied on indirect methods. One of the commercialized indirect methods monitors the motor current change [33] while driving the carrier or platen rotation. This technique measures changes in friction forces between the pad and the wafer surface. It is mostly successful for tungsten CMP than for copper or STI CMP processes, because the frictional forces between the polishing pad and exposed films do not differ significantly in the case of copper or STI scenarios. The method sufficiently sensitive for STI, dielectric and metal CMP processes is the vibration monitoring method [34]. Vibrations during the CMP process are detected using an accelerometer on the carrier behind the wafer. This method is used to detect delamination of low-k dielectric films in copper dual damascene process [35].

Optical reflectometry is the preferred method for endpoint detection. But, the method is highly intrusive, requiring extensive polisher modifications. Various other
endpoint techniques have been developed in recent years, many of which are currently adapted to the high volume production stage [35].

**CMP Consumables**

The CMP market in the year 2003 was $1.7B with slurry (24%) and pad (15%) amounting to 40%. The CMP equipment was estimated to comprise 55% with the remaining 6% accounting for conditioner and other consumables. The market is expected to grow to $3.3B in the year 2008. The rapid growth of CMP and its application for polishing various materials, necessitated development of a variety of slurries, pads, complex process recipes, more complex slurry mixing and distribution systems. Polishing of different materials and customized needs of the various semiconductor industries resulted in more complex CMP equipment with different process dynamics such as linear, orbital and fixed head machines [35]. The process of CMP has gone through a lot of evolution from first being used just for silicon dioxide planarization to the present day planarization applications in pre-metal dielectric (PMD), ILD, STI, metal and gate oxide, etc.

While the CMP process seems fundamentally easy to understand, there are a staggering number of process variables (Fig. 1-7) that complicate the underlying interaction between the pad, slurry and wafer, preventing full comprehension of the process. The most widely accepted mechanism for the material removal is that the chemicals in the slurry react with the surface of the wafer (e.g., oxide, copper, etc.) and create a “modified” layer that can be removed easily. The mechanical action of the abrasives trapped in pad asperities then physically abrade and remove the modified material [24, 36, 37].
In terms of the CMP input parameters, the process variables are generally categorized into chemical and mechanical parameters. The chemical nature of the CMP process typically relates to the constituent makeup of the slurry and how those chemicals react with the wafer surface to create a passivated or modified layer, keeping the particles well suspended in the slurry. The chemicals in the slurry are highly dependent upon the type of CMP being used. For example, in oxide polishing, an alkaline solution is used along with various buffering and oxidizing agents, whereas for metal CMP as in Cu CMP, acidic slurry can be used. The mechanical aspects of CMP largely depend on the polishing pad, wafer pattern, and process conditions like pressure and relative pad-wafer velocity.

Fig. 1-7. Parameters involved in CMP process.
Polishing in its simplest sense is controlled chemo-mechanical material removal to produce a globally flat, defect free surface. The main components of CMP are the CMP slurry and the pad. The complexity of these CMP components and their interactions are discussed further in the following sections.

**CMP Slurry**

CMP slurry consists of two major components: abrasives and chemicals. Depending on the material of the wafer, the chemistry of the slurry, and the synergy between them, each kind of slurry behaves differently [38]. CMP is influenced to a great extent by numerous slurry parameters such as pH, solution chemistry, particle charge type, concentration and particle size distribution of abrasives, complexing agents, oxidizers, buffering agents, surfactants, corrosion inhibitors, etc.[19, 39]. An ideal CMP slurry should be well suspended with good slurry stability, and be able to achieve high removal rate, excellent global planarization, prevent corrosion (in case of metal, especially Cu), and result in good surface finish, low defectivity and high selectivity.

**Rate of surface reaction**

The reaction between the wafer and slurry chemicals to form a modified layer of desired thickness should be fast enough to avoid material removal by mechanical indentation of the bare wafer. The time scale in which the passivation layer is formed should be less than the average time of successive particle indentations into the wafer for abrasion to achieve optimal CMP performance. Fig. 1-8 shows the electrochemical chronoamperometry (potentiostatic) analysis of Tungsten by first keeping the samples at cathodic potential to avoid surface oxidation and then “anodizing” them. The generation of current, which corresponds to surface reaction rate, is monitored on a millisecond scale.
as shown. It can be seen from the figure that Tungsten surface quickly passivates which is conducive for mechanical removal of the material.

![Chronoamperometry measurements for tungsten](image)

Fig. 1-8. Chronoamperometry measurements for tungsten [20]

For selectivity purposes, the reaction of slurry chemicals with the underneath layer should be significantly low compared to the surface layer. In terms of operational hazard, the chemicals used in the slurry should not make it hostile for the operators performing CMP and the post-CMP chemicals should be treated to avoid environmental damage.

**Slurry agglomeration**

The ideal slurry will have abrasives behaving as discrete single particles during the CMP process. However, particles in a real CMP slurry apart from being discrete can also exist in the form of agglomerates [38, 40]. An agglomerate is a collection of particles that come together into packed clumps due to physical or chemical interactions and act as a single entity. Although CMP slurries are composed of very fine particles up to 200 nm (0.2 μm), ‘‘large’’ particles of 1–3 μm and greater are often observed during the process. Such particles can be due to agglomeration or contamination from foreign material. These are not desirable for CMP as they can cause micro scratches and pits due to deep
indentation or non-uniformity due to varying polishing pressure. Even minimal and undetectable concentration of larger sized particles are shown to produce surface defects in CMP [41]. The agglomeration of slurry can be prevented by proper slurry design, which includes appropriate use of chemicals, and/or surfactants and the suitable order of their addition to design the slurry [42]. Other than adding chemicals, the particle surfaces can also be altered to make the suspension stable during the CMP process [42, 43].

In order to prevent the particle agglomeration during the distribution stage, dispense filters are used which filter out the agglomerated particles before slurry delivery. Care is taken so that the filtration of the agglomerated particles does not change the particle distribution and slurry composition. The drop in slurry pressure and flow due to filtration is minimized by using a series of filters instead of a single filter.

**Chemical Mechanical Polishing Pads**

The pad traps the slurry particles on its asperities, and executes the polishing action by transmitting the normal and shear forces for polishing, thereby playing a very crucial role in process optimization [19]. The polymeric polishing pad must have sufficient mechanical integrity and chemical resistance to be effective for numerous polishing runs through its life. The pad must be able to withstand the aggressive slurry chemistries used in CMP without degrading, delaminating, warping or blistering. Slurry chemistries range from highly alkaline near pH 11 used for polishing of inter-layer dielectric to highly acidic near pH 2 used for metal polishing. Slurries also contain oxidizing agents such as hydrogen peroxide and potassium iodate. The pad material should be resistant to such environments. As the polishing slurries are aqueous based, pad should be hydrophilic to allow circulation of slurry to whole wafer during polishing.
The polishing pad is made up of a matrix of cast polyurethane foam with filler material to tailor the hardness of the polyurethane impregnated felts for the specific CMP process. Polyurethanes have a unique property of combining high strength to resist tearing during polishing and, an acceptable level of hardness and good abrasion resistance to prevent excessive pad wear during polishing. The CMP pads can be categorized into four types differentiated by their microstructure [44]. They are: Polymer impregnated felts (type 1), Poromerics (type 2), Filled polymer sheets (type 3), and Unfilled textured polymer sheets (type 4). Pads of type 1 and 2 are shown in Figs. 1-9 and 1-10. The figures indicate the variety of pad surface textures or asperities involved in CMP. The polyurethane polymer consists of two major components/blocks of alternating hard organic isocyanates and soft polyols. Each of these segments has different properties, the characteristic properties of a given pad depend on the relative ratios of these hard and soft segments. Isocyanates, the more rigid blocks, influence the hardness, modulus, tear strength and stability, specifically at elevated temperatures. Polyols on the other hand tend to affect the elasticity, toughness, abrasion resistance and the low temperature stability of polymer.

For uniform polishing, the pad needs to be in contact with the whole wafer surface with the same contact pressure and relative velocity between pad and wafer in all contact regions. But wafers are not flat and typically have some degree of curvature due to stresses induced during manufacture process or differing coefficients of thermal expansion of the various deposited oxide and metal layers on the substrate. This necessitates the polishing pad to be flexible enough to conform to the wafer-scale flatness. This is achieved by laminating a stiff polishing pad to a flexible underlying base.
pad, which is typically a more compressible material. This improves polishing uniformity across the wafer without unduly compromising the stiffness of the top polishing pad. A stacked pad consisting of an IC 1000 top layer on a Suba IV sub-pad (Rodel Inc., Newark, DE) is currently the pad system of choice to obtain both good global wafer uniformity and high removal rate for various CMP systems.

Fig. 1-9. Type I CMP pad. (a) Surface. (b) Cross-section [44].

Fig. 1-10. Type II CMP pad. (a) Surface. (b) Cross-section [44].
Effect of pad grooves

Polishing pads have macrostructure, which can be either perforations through the pad thickness or surface groove designs. The motivation behind designing pad grooves is to prevent hydroplaning of the wafer, ensure uniform distribution of slurry over the wafer, and to remove the polishing debris from the pad surface through these chemicals [44]. The groove designs can be circular with either concentric or spiral grooves, cross-hatched patterns arranged as an X-Y grid across the pad surface, other regular design such as hexagons, triangles and tire-tread type patterns, or irregular designs such as fractal patterns, or their combinations. The effect of pad texture on tribological and kinetic properties of polishing pad has been studied by Philipossian and Olsen [45]. Real time monitoring of coefficient of friction (COF) was done to estimate the normal shear forces originating during a particular CMP processes for a given pad. Frictional and removal rate data were taken on Rodel IC 1000 flat, perforated, XY- and K-grooved pads. The Somerfeld number, COF, a new parameter called tribological mechanism indicator (TMI) and materials removal rate were some of the parameters monitored during the experiments. Philipossian and Olsen [45] reported that flat and XY-grooved pads begin in boundary lubrication and migrate to partial lubrication as Sommerfeld numbers increase. Thus different pad surface textures showed different material removal mechanisms. However, the variation of the mechanical properties need to be studied further on a microscale to gage its impact on the CMP process and material removal mechanism.

Physical and chemical changes in pad during polishing

Since polishing is a wet process, the mechanical properties of the pad must not vary considerably when the pad is wet during CMP. However, the hydrophilic pad material
looses modulus and hardness and gains ductility because water acts as a plasticizer. These changes are reversible [44] and pad properties revert to the dry values, when the pad is in dry state, indicating that the chemical attack is minimal [46]. The surface of the polyurethane pad has been studied and reported in literature [47]. The quantitative analysis of the physical and chemical changes that occur in the polishing pad during polishing was performed by Lu et al. [48]. In their research, Lu et al. studied the effect of polishing cycles on the pore size and shape of the commercially available IC 1000/Suba IV polishing pad. It was reported that the pore size and shape distribution were modified by polishing cycles. Polishing induced a permanent distortion in the radial direction due to the oscillatory motion of the wafer, while no change was seen in pad pore size in the transverse direction (direction of pad motion). The change was attributed to pore closure by surface reflow of the polymer during polishing or conditioning and not due to debris filling. The surface profiles generated using white light interferometry showed that the overall micro roughness of the pad decreases with polishing and there is a smoothening effect over a period of time due to polishing cycles. Along with physical changes, the pad undergoes surface chemical modification and degradation due to polishing slurry chemicals. No pad degradation of the bulk material of the polishing pad was observed. The change in spectrum (Fig. 1-11) and shifting of the peaks of the used pad as compared to the new one has been attributed to the realignment of the molecules of the damaged pad surface layer.

The porous structure of the polishing pad encourages water seepage. The presence of water in the polymer pore significantly affects the mechanical properties of the polishing pad. Li et al. [47] studied the effect of pad soaking time on the mechanical
properties of the polishing pad using dynamic mechanical analysis (DMA). The dynamic shear modulus decreased to two thirds of its original value when the pad is soaked for 5 h. Further, the rate of decrease in dynamic shear modulus decreases at around 14 h but does not reach a steady state. No significant change in the removal rate was observed when pads with different soaking time are used in oxide polishing. This shows that removal rate is more of a surface characteristic of the pad and is not affected by the bulk material.

Fig. 1-11 The absorbance spectra of used and dry IC 1000 pad along with dried silica slurry. [48]

Modification of the pad-wafer contact mechanics can directly impact the CMP process performance. The modified surface can even change the mechanism of material removal during a polishing process [49]. During polishing, the pad surface layer degrades decreasing the pad surface roughness. The pad surface which is flattened or “glazed” needs to be rejuvenated to bring the polishing performance back to the original value. The roughening of the pad by a diamond conditioner is known to result in pad revival. Along with wafer polishing, the conditioner made of diamond grit considerably alters the surface of the polishing pad. The process of conditioning essentially removes some amount of top pad surface and is performed either in-situ (during the polishing run) or ex-
situ in between two polishing runs. The pad damage due to insufficient conditioning is termed as “glazing”. Experiments were performed on the CMP pad by Lawing [49] at different conditioning levels of aggressiveness (expressed as low, medium and high). The change in polishing rate correlated better with the predicted surface contact area rather than the average surface roughness.

**Effect of temperature change on polishing pad**

The polishing pad might be subjected to elevated temperatures during the course of the CMP process [19]. The heating is due to the frictional forces generated by solid–solid contact between pad and wafer during polishing. There is some alleviation of heating when the polishing is done in hydrodynamic contact mode [48]. However, there is an overall increase in the pad temperature irrespective of the mechanism of polishing. The effects of pad heating are compounded if the chemical reaction between slurry and pad is exothermic [38]. The mechanical, physical and chemical properties of the polyurethane material can be permanently or temporarily altered if the pad is heated beyond a particular temperature [50]. DMA, thermomechanical analysis (TMA) and thermogravimetric analysis have been used for material characterization of the polishing pad [47, 48].

The applicability for DMA for polishing pads has been elaborately studied by Lu et al. [48]. Based on the results of previous experiments performed by Li et al., [47], DMA was performed to evaluate the structure–property relationship of polyurethane pads. The glass transition temperature is the temperature at which the macromolecules of the polyurethane break away from entanglements and undergo micro-Brownian motion. Investigation of glass transition temperature of the polyurethane pad is important as it
corresponds to a sharp decrease in the hardness and modulus of the pad which in turn affects the removal rate and global planarization.

**Non-uniformity of the polishing pad**

The pad might be directly responsible for several process defects like WTWNU where there is non-homogeneity of polishing when one wafer is compared to another or WIWNU where there is non-homogeneity of polishing at different areas of the same wafer. The scanning ultrasound transmission (UST) is a nondestructive technique developed that works on the principle of ultrasound permeability through absorbing visco-elastic medium [51, 52]. The difference in the ultrasound absorption in the areas of varying density and viscoelasticity is used to determine the non-uniformity within a single pad thereby giving an in-depth idea of the physical characteristics of the given pad. In order to improve the yield of the CMP process and, to get a highly planar, defect free uniform wafer surface with reduced overall manufacturing costs involved, there is a need to extensively study the fundamental properties of the CMP pads on the whole.

**CMP process in the future**

The semiconductor industry has effectively adapted its CMP technology for the 300 mm wafer [53]. Beyond the adoption of copper interconnects, several technologies are necessary to continue the shrinkage of device dimensions and the increase of packing density in ULSI manufacturing. The use of ultra low-k materials as interlayer dielectrics has been at the forefront for decreasing the “C” part of the “RC delay”. However, polishing of ultra low-k dielectrics which are mechanically weak (soft) is a daunting task. The single and dual damascene structures comprising of ultra low-k porous or polymeric materials are more prone to buckling and crushing failures due to significantly lower hardness and Young’s modulus as compared to silicon dioxide. According to the ITRS
roadmap [12] materials with dielectric constant 2.2 will be integrated in the IC by year 2007.

Also, there needs to be a marked improvement in slurry selectivity for accurate end point detection when the constituent layers of the damascene structure namely, metal, hard mask, cap layer, barrier layer and dielectric are polished. Furthermore, it is necessary to explore the niche of the CMP process in STI and other applications such as backside polishing and the fabrication of micro-electro-mechanical systems (MEMS). Here, the CMP process must be integrated horizontally and vertically to achieve high throughput and performance. In MEMS applications, horizontal integration ensures reliability and good performance of a specific CMP process run. The development of new polishing pads and pad architectures, novel slurries, new metrology techniques, etc. comes within the scope of horizontal CMP integration. Vertical integration ensures success of every successive CMP operation. This includes the integration of the upstream processes such as Cu/barrier deposition, etching and downstream processes such as ILD deposition and lithography. This opens a wide scope of research for further optimization and development in Cu CMP process. Due to intricate interactions of the processes involved in any CMP process, modeling efforts are needed to get a systematic understanding of the process and predictability of the CMP outputs.

**Scope of Dissertation**

The importance of polishing and the various available planarization techniques with emphasis on CMP have been discussed in Chapter 1. A general overview of the parameters involved in CMP are discussed with prominent interactions between them. Chapter 2 gives an overview of the CMP modeling efforts and the framework of the PERC I CMP model. The model’s challenges are described, which necessitated the
development of the PERC II model. The assumptions of the PERC I model are refined as necessary in chapter 3. Model development is also delineated with critical differences between PERC I and II models. Chapter 4 validates the PERC II model comparing to experimental literature data and comments on the predictive capability of the model. Chapter 5 summarizes the findings of the study along with the suggestions for future work.
Like most semiconductor manufacturing processes, simultaneous tuning of process variables is a formidable task due to many varying parameters. Unfortunately, much of the commercial CMP process development cycle is based on nearly blind design-of-experiment (DOE) testing. While these empirical tests help to optimize parameters in a very small range, the overall approach does not necessarily facilitate a deeper understanding of the process. Thus, CMP research has focused on developing process models in order to gain further insight on the process mechanics for better control over the CMP outputs.

The ultimate goal is to model at three scales of CMP, namely, the particle, die and wafer scales. The models at the particle scale are needed to understand the mechanisms involved in interactions between slurry particles, chemicals, polishing pad and wafer during blanket wafer polishing. The models at the feature or die scale are needed to address the topography evolution of integrated circuit (IC) chips as a function of pattern density, line width, pitch width and polishing time. The wafer-scale models also need to address the issues related to the material removal non-uniformity over the entire wafer surface. To satisfy the manufacturability requirements in the sub-micron IC fabrication, the margin of the profile non-uniformity is usually limited in the range of several hundred nanometers. The material removal non-uniformity across the wafer is rooted in the non-uniform distributions of pressure and velocity over the wafer surface, which in a large
degree, are determined by the configurations of CMP equipment. A model at this scale, therefore, needs to consider the equipment configurations appropriately.

**Literature Review**

In the last decade, various kinds of models have been proposed by researchers, concentrating on different aspects of CMP covering the particle, die and wafer scales. But the effectiveness of die and wafer scale models depends strongly on the robustness of the particle scale models, as the die and wafer-scale models are developed based on the particle scale models. Therefore, in order to understand and predict the performance of CMP in wafer scale, the current study aims to understand the behavior of CMP at particle scale. As the current work focuses mainly on particle scale modeling, only related models will be reviewed here.

**Preston Equation**

The Preston equation [54] as an empirical relation for optical glass polishing was substantiated by Cook [24] who considered that the interaction between abrasive particles and wafer surface leads to removal of atomic clusters due to bond breakage. Preston provided a simple model of material removal in glass polishing tools which is often adopted for CMP processes also. But, experimental results have shown that slurry and pad properties also significantly influence the process along with down pressure and platen speed [55, 56]. In order to explain the non-Prestonian behavior, many modeling efforts have been carried out in the past decade and is still a topic of considerable interest.

The three most important elements in the CMP material removal process are the wafer being polished, slurry and polishing pad. Researchers adopted two main approaches in modeling the CMP process. One approach is based on fluid hydrodynamics, where material removal is considered to be caused mainly by slurry
erosion. Slurry film thickness, hydrodynamic pressure and removal rate are generally calculated by solving Reynolds or Navier-Stokes equation [57-62]. The other approach uses the theory of contact mechanics, which assumes that the material removal occurs by particle indentation into the wafer surface. Pad surface roughness and hardness are generally used to calculate the particle indentation into the wafer thereby calculating the removal rate [27, 63-67]. Finite element modeling (FEM) has also been used to determine the pad asperity mechanical properties recently [68, 69].

It is generally agreed that the typical CMP process conditions lead to major material removal by particle abrasion rather than by slurry erosion [70]. Experimental observations of polished surfaces have revealed that material removal in CMP occurs as a consequence of a combination of chemical reaction of the slurry chemicals with the wafer surface and ploughing (repeated sliding, rolling, or indentation) of the abrasive particles against the wafer surface that is already modified. The chemical reaction results in formation of a thin layer on surface which is of different modulus than the substrate/wafer, aiding in material removal [71].

When particles are introduced between wearing surfaces, the wear is due to the particles embedded in one of the sliding surfaces. This process is referred to as three-body abrasion. In the CMP process, the particles acting as single point cutting tools, dig into the wafer and plow a series of grooves in it. Therefore, the CMP process is essentially a three-body abrasive wear. The analysis of the texture of polished surfaces also supports this observation. Fig. 2-1 illustrates typical post CMP surfaces as examined via an optical surface profiler. The surface texture consists of a number of randomly
oriented nano-grooves (Fig. 2-1). The width and depth of the grooves are determined by the penetration depth of the abrasive particles into the wafer surface.

Fig. 2-1. Polished wafer surfaces showing nanoscratch surfaces. (a) Silica wafer [72]. (b) Si wafer [73].

The material removal process in CMP can be considered as sliding of the indented slurry particles over a chemically influenced thin layer over the wafer surface. The
knowledge of critical factors that effect the CMP process and the interactions among them is critical to select the respective parameters and design an enhanced CMP process, predicting material removal rate, and surface quality (roughness, defects, etc.). The most popular material removal equation is the Preston’s Equation [54], which was initially introduced for glass polishing,

\[ \text{MRR} = K_e PV, \]  

(2.1)

where MRR is the material removal rate, \( K_e \) an all-purpose coefficient, \( P \) the down pressure and \( V \) the relative velocity over the wafer-pad interface. It demonstrates a linear dependency of material removal rate on the pressure and velocity. Not all experimental MRR data in CMP, especially in metal CMP, support the linear dependency between material removal and the product of pressure and velocity.

**Modifications to Preston Equation**

In order to explain the non-Prestonian behavior of the CMP process, revised Preston’s equations were proposed by several researchers. For example, considering that the material removal rate does not extrapolate to zero, Maury et al. [74] introduced a fitting parameter \( MRR_0 \) into Preston’s equation

\[ \text{MRR} = K_e PV + MRR_0 \]  

(2.2)

Later, the nonlinear equation

\[ \text{MRR} = K_e P^\alpha V^\beta, \]  

(2.3)

where \( \alpha \) and \( \beta \) are two fitting parameters, was proposed by Wrschka et al. [75] to get a better fit for the experimental data.

The major problem with Preston’s equation and its revisions are that they are empirical modeling efforts. Therefore, they do not shed any light on gaining insight into
the process especially regarding other process variables. By extending Brown’s model [76] of the metal polishing to the silicon polishing, Cook [24] developed a physical model to address this limitation. The interactions between the abrasive particles and the wafer surface is proposed as a Hertzian elastic penetration [77] of a spherical particle under uniform pressure $P$ into the wafer surface, sliding along the surface with a velocity $V$ and removing material volume proportional to the penetration. The MRR formulation was proposed as

$$\text{MRR} = \frac{PV}{2E}$$  \hspace{1cm} (2.4)

where $E$ is the Young’s modulus of the wafer material. This model can be taken as a theoretical equivalent to the Preston’s equation since it supports the linear dependency of MRR on pressure and velocity. The relationship between the wafer surface roughness $R_a$, the down pressure $P$, and abrasive size can also be obtained based on this model

$$R_a = \frac{3}{4} b \left( \frac{P}{2k_1E} \right)^{2/3}$$  \hspace{1cm} (2.5)

where $k_1$ is the particle concentration, which is unity for a filled close hexagonal packing [24] and $b$ the diameter of the slurry particles. A similar model was developed by Liu et al [78] based on the statistical method and Hertzian elastic penetration. Besides the wafer material properties including wafer hardness $H_w$ and wafer Young’s modulus $E_w$, this model included pad hardness $H_p$ and abrasive Young’s modulus $E_s$, resulting in

$$\text{MRR} = C_e \left( \frac{H_w}{H_w + H_p} \right) \left( \frac{E_s + E_w}{E_s E_w} \right) PV$$  \hspace{1cm} (2.6)

where $C_e$ is a coefficient to account for the effects of slurry chemicals and other consumable parameters. This model, similar to Cook’s model, suggests that the material removal is proportional to the applied pressure and relative speed.
The advantages of Cook’s and Liu’s models over Preston’s equation are that they provide more insights into the roles of the consumable parameters. The contributions of the slurry abrasives and pad, for example, have been attributed to their size and hardness. An additional benefit is that not only material removal rate, but also surface quality issues such as roughness (Eq. 2.5) can be addressed using these models.

**Slurry Erosion**

In Cook’s and Liu’s models, the mechanical removal by abrasive particles was assumed to be the dominant mechanism. However, some researchers instead assumed the material removal to be by “slurry erosion” approach, i.e., due to mechanical-enhanced erosion. Runnels et al. [79-81] developed an erosion-based model for CMP. They assumed that a fluid film exists between the wafer and pad interface, which affects the erosion/material removal rates at each single point through the fluid stress tensors.

Motivated by observations at feature scale, modification to Preston’s equation was proposed, with relative velocity $V$, replaced by the tangential stress on the wafer surface;

$$\text{MRR} = C_e \sigma_t \sigma_n$$

(2.7)

where $C_e$ is an all purpose coefficient, $\sigma_t$ is the shear stress due to the slurry flow and $\sigma_n$ the normal stress.

The model uses a steady state incompressible Navier-Stokes equation to model the slurry flow at the wafer-polishing pad interface. The wafer surface is assumed to be smooth and spherical with a large radius of curvature. The fluid layer thickness and the angle of attack between the pad and the wafer are obtained through an iterative procedure satisfying force and momentum balance. Runnels’ model has been integrated into several particle-scale models by researchers including Tseng and Wang [82] and Zhang and
Busnaina [83]. Tseng and Wang attributed the normal stress at the particle-wafer contact to the elastic indentation of the particle into the wafer surface, which is similar to that proposed by Cook and calculated the normal stress over the wafer-particle interface as

$$\sigma_n = \frac{F}{\pi r_c^2}$$  \hspace{1cm} (2.8)

where $F$ is the force acting on the spherical particles, which is proportional to the pressure $P$. The radius of particle-wafer contact, $r_c$ is given by

$$r_c^3 = \frac{3d}{8} F \left[ \frac{1 - \nu^2}{E} + \frac{1 - \nu'^2}{E'} \right]$$  \hspace{1cm} (2.9)

where $d$ is the diameter of particles, $\nu$ and $\nu'$ the Poisson’s ratios of wafer surface and the particle and $E$ and $E'$ the elastic modulii of the wafer and particles, respectively. The shear stress due to the slurry flow can be approximated as

$$\sigma_t = C_e \sqrt{\mu V A_0}$$  \hspace{1cm} (2.10)

where $\mu$ is the dynamic viscosity of the slurry and $A_0$ the area of wafer surface respectively. Substitution of equations (2.8) and (2.10) into (2.7) yields

$$\text{MRR} = K_e P^{5/6} V^{1/2}$$  \hspace{1cm} (2.11)

where $K_e$ is the parameter to account for material properties, slurry abrasive concentration and chemical processes. This model demonstrated a nonlinear relationship between the material removal and the product of pressure and velocity. Tseng’s model attempted to connect the elastic indentation to the erosion rate, whereas Cook’s and Liu’s models connect it to the mechanical abrasion. While the down pressure dependency ($P^{5/6}$) is quite close to linear dependency, the velocity dependency ($V^{1/2}$) however, is quite nonlinear. This is because the contribution of velocity has been attributed to the slurry
flow instead of abrasive sliding. Due to pad porosity and surface asperities, such a slurry film is bound to be interrupted by solid-solid contact. Thus hydrodynamic lubrication may be responsible for distributing the slurry, while the material removal occurs mainly through solid-solid contact.

Cook’s, Liu’s and Tseng and Wang’s models attributed the penetration of the abrasive particles to Hertzian elastic contact. Zhang and Busnaina [83] estimated the contact pressure between the particle and the contact surface in their model and found that it is larger than the yield stress of the polished materials. Therefore, they proposed that plastic deformation is a more likely deformation mechanism in polishing. The contact pressure over the particle-wafer interfaces is suggested to be equal to the hardness $H_w$ of the wafer materials. Replacing the normal stress (Eq 2 8) in Tseng and Wang’s model with the hardness $H_w$ yields the following material removal rate formulation

$$MRR = K_e (PV)^{1/2}$$

(2.12)

where $K_e$ is the parameter to account for materials properties, slurry abrasive concentration and chemical processes. Both Tseng and Wang; and Zhang and Busnaina’s models suggest a nonlinear pressure times velocity dependency of material removal rate. Moreover, it is noted that besides the external force applied on the particles from the pad, Zhang and Busnaina also proposed that an adhesion force, either van der Waals force or electrostatic force depending on the separation distance between the particle and the wafer, contributes to the indentation. This has been integrated into another particle-scale model by Zhang et al. [84]. In a series of articles by Ahmadi and Xia [85], and Mazaheri and Ahmadi [86, 87], a thermodynamic work parameter $W_a$ of adhesion was added to account for its effects on the indentation of abrasive particles. Mazaheri and Ahmadi [86]
introduced a double layer (dl) force into the indentation force, whose magnitude is a function of the zeta potential of the abrasives and the wafer. Beside the abrasive wear, Ahmadi and Xia [85] also considered the adhesion wear of the wafer in their model. Moreover, while most of the models treated the abrasives as having a spherical shape, Mazaheri and Ahmadi represented them as spheres with a number of hemispherical bumps around their surface. The penetrations of slurry abrasives are modeled using the penetrations of the bumps.

The above models imply that the abrasives are embedded in the pad and indent into the wafer surface. It should be noted that some modeling efforts also assume that the abrasive particles float in the slurry and impact the wafer surface from time to time causing material removal [88].

**Pad Considerations in CMP Models**

The pad is assumed to be smooth in the earlier particle-wafer interaction models. It has been observed that the pad topography and pad material play an important role in material removal processes. For example, the material removal rate has been reported to increase with the pad surface roughness [55]. Without conditioning of the polishing pad surface, the material removal rate decreases exponentially with polishing time [89]. In consideration of this, Yu et al. [63] developed a pad-based model. They approximated the peaks on the pad surface by hemispherical asperities with constant radius and the asperity heights to follow Gaussian distribution. Based on the model, the real contact area is smaller than the nominal (apparent) contact area and linearly proportional to the down pressure. Steigerwald et al. [19] proposed that the material removal rate is proportional to the number of abrasive particles over the contact area. Combining this argument with
Yu’s model yields a linear dependency of the material removal rate on the down pressure, which agrees with Preston’s equation.

Zhao and Shi [90] also proposed a model based on the wafer-asperity contact. Unlike Yu’s model [63], this model does not consider the Gaussian distribution of the asperity heights. The contact area between an asperity and the wafer is given by \( A_a \propto P^{2/3} \) based on Hertz elastic contact theory. By combining Steigerwald’s argument, the material removal rate formulation can be obtained as

\[
MRR = K(V)P^{2/3} \quad (2.13)
\]

where \( K(V) \) is a function of the relative velocity \( V \) and other CMP parameters. It is further considered by Zhao and Shi [90] that when the particles are rolling against the wafer surface, their contribution to material removal will be negligible. They argued that whether the particle is rolling or not is determined by the surface friction between the particles and the wafer, and only when the down pressure \( P \) is larger than a threshold down pressure \( P_{th} \), the pure rolling can be avoided. The material removal rate is given as:

\[
\begin{align*}
MRR &= K(V)(P^{2/3} - P_{th}^{2/3}) & \text{for } P \geq P_{th} \\
MRR &= 0 & \text{for } P < P_{th}
\end{align*} \quad (2.14a)
\]

The fundamental difference between the above pad-based models and the particle-based models by Cook and others is that pad-based models attribute the material removal rate to the number of abrasive particles captured by the polishing pad while the later attributes the material removal rate to the interaction between a single abrasive and the wafer.
A particle-pad interaction is critical to the particle scale models considering that the function of the pad is to trap the abrasive particles and transmit load to the particle-wafer interface. Cook’s model [24] suggested a close-packing of spherical abrasives onto the pad. It is assumed that the wafer and pad are separated completely by the abrasives and no direct pad-wafer contact exists. The force applied on a single abrasive under these assumptions is given by

\[ F = \frac{2\sqrt{3} Pr^2}{k_1} \]  

(2.15)

where \( P \) is the polishing pressure, \( r \) the abrasive size and \( k_1 \) the particle fill fraction on the pad. This particle-pad interaction model has been integrated into the material removal model of Cook. It is also used by Ahmadi and Xia [85] to evaluate the force on a particle in their case of a hard-pad and a larger concentration of abrasive particles. Later, Zhao and Shi [90] proposed that when the pad is soft enough, the abrasive particles will be embedded deeply into the pad and the force from the wafer is supported by the pad and abrasives together. This idea has been applied by Luo and Dornfeld [27, 91, 92] and Fu et al. [93] in their model. Luo and Dornfeld’s model [27, 91, 92] suggested that this force is proportional to the contact pressure multiplied by the abrasive size by assuming that the abrasives are closely packed to each other and these closely packed abrasives are enclosed by the pads so that the effective contact area between wafer and pad is equal to that without abrasives. Moreover, the size of the abrasives that may be captured by the pad is assumed to be a function of abrasive size distribution and pad properties. Their model, however, neglects the lower sizes of the particle size distribution in material removal calculation. Fu et al. [93] later assumed that the abrasives are dispersed evenly over the pad surface and used a beam model to evaluate the wafer-pad direct contact
between two single abrasives. The force supported by a single abrasive can be obtained from the beam model and is a function of abrasive size, down pressure and pad material properties [93].

Besides the force, the second purpose of an abrasive-pad interaction model is to evaluate the number of abrasives involved in the material removal process. Fu et al. [93] considered the number of trapped particles as an independent parameter in their model. This may be misleading considering that various parameters, for example, the abrasive weight concentration and abrasive size, have an influence on the number [27, 72, 91, 92]. Luo and Dornleld considered more complex scenarios and suggested that only a portion of abrasives are involved in the material removal process, which is not reasonable.

**Slurry Considerations in CMP Models**

The contribution of slurry chemicals to the material removal is either neglected in earlier models or represented by an all-purpose coefficient. Cook [24] suggested a complete but complicated scenario of the chemical effects. Besides the mechanical removal, it is proposed that the surface removal during polishing include the following five chemical processes (i) the slurry chemical diffusion into the wafer surface, (ii) the subsequent wafer material dissolution under the load imposed by the abrasive particles, (iii) the adsorption of the dissolution product onto the surface of the polishing grain, (iv) the re-deposition of the polishing materials back onto the wafer surface, and (v) the corrosion rate between particle impacts. These steps are not considered in most of the early models as they are complex to model quantitatively.

The major contribution of slurry chemicals on the material removal process has been attributed to the formation of a surface layer. This idea is well demonstrated by Kaufman et al. in their Tungsten CMP model [37].
Kaufman et al. proposed that this passivation layer is removed by slurry abrasives and the fresh tungsten surface is exposed, which is subsequently passivated and removed. This mechanism of continued passivation and removal can be used to explain the copper, aluminum and other metal CMP systems as well. A similar mechanism is supposed to work in silicon, silica and low-k CMP systems. Paul [94] and Zhao et al. [95] have proposed detailed surface kinetics models to connect the slurry chemical concentrations and fresh metal surface sites available to the formation rate of the surface layer. Their models can explain the material removal rate as a function of chemical and abrasive weight concentration. Borst et al. [62] have proposed a five-step model for CMP of SiLK dielectrics (i) mass transport of reactant from the slurry to the slurry/wafer interface, (ii) adsorption of reactant to available surface site, (iii) reaction between adsorbed reactant and specific wafer surface to form an altered wafer surface layers, (iv). mechanical removal of the altered wafer surface layer, and (v) mass transport of removed material to the bulk slurry. In their work, the mass transportation, slurry chemical concentration and reaction rate to the formation of the surface layer were incorporated. This detailed model was suggested to be able to be extended to the CMP of other materials.

The idea of continuous surface modification and removal has been applied in the model developed by Qin and co-workers [96]. A similar approach has also been used in Luo and Dornfeld’s model [91]. They extended Kaufman’s model by proposing that the surface layer is a bi-layer structure, though in reality there might be a gradual transition between the layers. They did not cover the details on how the formation rate of the surface layer is affected by the slurry chemicals. But Luo’s model assumes that all the
asperities contact the wafer and that the particles fill the asperity-wafer contact spots, which is not necessarily true.

The slurry chemicals affect not only the wafer but also the slurry abrasives and polishing pad. Their contribution to material removal is therefore not only reflected through surface kinetics, but also through alteration of the abrasive and pad properties such as the abrasive shape, abrasive size and pad Young’s modulus. There are not many modeling efforts on these two interactions yet. A recent model by Castillo-Mejia et al. [97] tried to explain the effect of water-pad interaction in CMP. The water is proposed to plasticize the polishing pad and reduce its elastic modulus. A formulation on the ratio of the Young’s modulus of wet pad to the dry pad is suggested. This Young’s modulus is then used in a wafer-pad contact model to evaluate the material removal.

Many of these models were also extended to patterned wafers along with other modeling concepts including combination of the solid-solid contact and slurry erosion approaches [98-109]. Polishing mechanisms were also based on surface reaction kinetics [95, 110, 111] and modeled using molecular dynamics simulations [112-115].

**PERC I CMP Model**

Most of the earlier models have concentrated on one or two interactions of the CMP process parameters. They are useful for identifying the input parameters. However, they may not be sufficient for understanding the whole material removal process. The effects of the same input parameters may be contrary to each other when acting in different conditions. A comprehensive model of CMP integrating the main interactions together is therefore needed. For this purpose, a model was developed by Qin et al.[72], taking into consideration most critical CMP parameters. This section discusses the
assumptions on which the model is built, its effectiveness in predicting the CMP process variables and the model limitations.

**Model Development**

The chemicals in a slurry alter the properties of the wafer surface that is in contact with them. Since the chemical influence is restricted to a thin region near the wafer surface, a thin surface layer whose mechanical properties are different from those of the bulk wafer is assumed. This surface layer can be formed by chemical reactions at the surface or by the diffusion of slurry ionic species into the wafer. The surface layer is then removed by mechanical action of the slurry particles and/or by dissolution into the slurry through a convective diffusion process. Although there may be a few exceptional cases [Luo98], material removal by the dissolution is small compared to that by the mechanical action of the slurry particles, and it is not considered in this model. Fig. 2-2 describes schematically the physical situation that is considered in the present analysis. The pad asperities that are in contact with the wafer surface act as the anchoring sites for the slurry particles. The anchored particles then gouge the chemically altered wafer surface or the bulk wafer depending on the indentation depth resulting in material removal. In the following section, various parameters considered in the present analysis and the underlying assumptions are described followed by the details of the model development.

Topology of a pad surface has been characterized and described statistically by various researchers [63, 116] Their measurements indicated that pad asperity heights were in the order of tens of microns and appeared to have a normal (or Gaussian) distribution. In Fig. 2-2, the topology of the pad surface is shown schematically along with a statistical description characterizing the pad asperities. In this figure, $\beta$ is the
mean radius of curvature of the tip of the pad asperities, \( \phi \) is the distribution function for the height of the asperities, and \( \sigma \) the standard deviation of the pad asperity heights, respectively.

![Pad Asperity Diagram](image1)

**Fig. 2-2 Pad consideration in PERC I.** (a) Schematic of the pad surface that is in contact with the chemically modified layer formed on the wafer surface; (b) Distribution of the pad asperity heights (When the height of an asperity is greater than the average gap \( h \) between the pad and the wafer, that asperity is in contact with the wafer surface.) [72]
Since the characteristic length scale for the roughness of the pad surface (e.g., $\beta$) is much larger than the feature sizes on the wafer, it was assumed that the wafer surface is smooth and flat compared to the pad surface. This assumption makes it possible to use the Greenwood-Williamson model [117] that describes the contact area, $A_{\text{real}}$, between a flat surface and a randomly rough surface under a normal force (or load), $L$:

$$A_{\text{real}} = \pi N A_o \beta \int_h^\infty (z - h) \phi(z) \, dz$$  \hfill (2.16)

$$L = \frac{4}{3} N A_o E^* \beta^{3/2} \int_h^\infty (z - h)^{3/2} \phi(z) \, dz$$  \hfill (2.17)

where $N$ is the number density of the pad asperities (i.e., number of asperities per unit area), $A_o$ is the total area of the flat wafer surface, and $h$ is the average gap between the smooth wafer and the rough pad under the given conditions. Then, the down-pressure, $P$, applied to the wafer with an area $A_o$ is $L/A_o$ whereas the actual mean contact pressure, $P_r$, is $L/A_{\text{real}}$ if the normal force applied to the wafer is entirely supported by the pad asperities in contact with the wafer. That is

$$L = PA_o = P_r A_{\text{real}}$$  \hfill (2.18)

$E^*$ is the composite modulus between the two surfaces in contact, that is defined as

$$\frac{1}{E^*} = \frac{1 - \nu_p^2}{E_p} + \frac{1 - \nu_w^2}{E_w}$$  \hfill (2.19)

where $E$ and $\nu$ are the Young’s modulus and the Poisson’s ratio, and the subscripts $p$ and $w$ represent the pad and the wafer, respectively.

Equations 2.16, 2.17 and 2.18 provide an explicit relationship between the contact area ($A_r$) and the down pressure ($P$) that is needed in modeling the CMP process. From Eqs. 2.16 and 2.17,
Following Yu et al.[63], the normal distribution function may be used for the distribution of the asperity height, \( \phi \). When \( \phi(z) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right) \) is substituted into Equation 2.20, the integral cannot be obtained in an explicit form. Instead, a numerical integration should be performed. Johnson showed that the ratio of the two integrals in Eq. 2.20 is nearly a constant for a broad range of \( h/\sigma \) [77]. Thus,

\[
A_{\text{real}} = C^{-1} \left( \frac{\beta}{\sigma} \right)^{1/2} \frac{L}{E^*} = C^{-1} \left( \frac{\beta}{\sigma} \right)^{1/2} \frac{PA_o}{E^*}
\]

(2.21)

where \( C \) is a constant whose value is between 0.3 and 0.4 when \( h/\sigma \) is in the range of 0.5 and 3.0. Recalling that \( \sigma \) is the standard deviation of the asperity height distribution, \( h/\sigma=3.0 \) represents a situation where only 0.13% of the pad asperities are in contact with the wafer whereas about 30.85% is in contact when \( h/\sigma=0.5 \). Practical situation in a typical CMP process is well within this range of \( h/\sigma \) value. From Eqs. 2.18 and 2.21,

\[
P_r = C \left( \frac{\sigma}{\beta} \right)^{1/2} E^*
\]

(2.22)

It should be noted that the contact area \( (A_r) \) increases linearly with the down-pressure whereas the contact pressure \( (P_r) \) remains constant due to the linear increase of \( A_r \) with \( P \).

The underlying assumption for Eq. 2.18 is that the applied normal force on the wafer is supported entirely by the pad asperities. As is mentioned previously, experimental measurement on the friction coefficient indicates that the hydrodynamic
pressure is negligible under typical CMP operational conditions. Furthermore, commercial pads often have grooves or perforations to prevent hydroplaning (i.e., to prevent a large pressure development due to a lubrication flow) and to ensure uniform slurry distribution. Thus, the assumption that the hydrodynamic pressure is negligible is justifiable.

**Small Plastic Deformation of the Thin Layer on the Wafer Surface**

Considering the fact that the deformation of the surface layer by the slurry particles under the contact pressure is very small in the order of several nanometers, the mode of deformation can be assumed to be plastic deformation. In the case of oxide CMP, it is known that the silica surface is softened when it is in contact with an aqueous solution. A surface layer (or so-called gel layer) is formed due to the diffusion of water molecules into the silica surface altering the surface properties.

The situation may be different in the case of metal CMP. Since an oxidizer is added in most metal CMP processes, metal oxide is formed on the wafer surface creating a structure mismatch at the interface between the oxide layer and the bulk wafer underneath. Thus, delamination or microcracking of the oxide layer may occur at the interface during the metal CMP process. However, as evidenced in previous sections, it is likely that plastic deformation is still the dominant mode of deformation because the indentation depth by the slurry particles is small (e.g., less than several nanometers). Furthermore, the oxide layer may be soft due to hydration in an aqueous environment. Thus, the assumption that the chemically altered thin surface layer on the wafer exhibits plastic deformation is plausible. This assumption enables us to apply the abrasive wear theory in deriving the expression for the material removal rate. When the indentation depth is beyond the surface layer, the deformation of the unaltered wafer material may be
either plastic or elastic depending on the hardness of the wafer material and the slurry particles.

**Material Removal by a Single Particle**

In the present analysis, it is assumed that the slurry particles are monosized and of radius ‘R’. The overall material removal rate is assumed to be the multiple of the removal rate by a single particle and the number of particles in contact with the wafer:

\[(RR) = n (RR)_v/A_o \quad (2.23)\]

where \(n\) is the number of slurry particles anchored in the region of the pad that is in contact with the wafer, and \((RR)_v\) is the volumetric removal rate by a single particle. The overall removal rate, \((RR)\), is typically described in terms of the thickness removed per unit time; hence the division by the nominal wafer area \(A_o\).

Schematic illustration of single particle abrasion is given in Fig. 2-3. The material removal is viewed as a sliding-indentation process in which the volumetric removal rate by a particle is given as

\[(RR)_v = A_i V \quad (2.24)\]

where \(A_i\) is the cross-sectional area of the particle immersed in the wafer under pressure (hashed area in Fig. 2-3a), and \(V\) is the relative velocity of the pad to the wafer.

Assuming that the hard particle maintains its spherical shape,

\[A_i = -a(R - \omega) + R^2 \arcsin \frac{a}{R} \quad (2.25)\]

where \(\omega\) is the indentation depth and \(a\) the radius of the circular indentation on the wafer surface by the particle of radius \(R_i\).

Because \(\omega\) is very small compared to \(R\), \(a/R\) is also very small and Equation 2.25 can be linearized about \(a/R\) to give
\[ A_i = \frac{4}{3} a \omega \left( 1 + O\left( \frac{\omega}{R} \right) \right) \]  

Thus the volumetric removal rate is

\[ (RR)_v = \frac{4}{3} a \omega V \]  

(2.26) \hspace{1cm} (2.27)

Fig. 2-3. Schematic of single particle abrasion showing the removal of chemically modified surface layer. (a). \( \omega \leq t \). (b). \( \omega > t \) with soft wafer materials.

Because \( a^2 = (2R_\omega - \omega^2) \) and \( \omega \) is much smaller than \( R \), Equation 2.27 can be rewritten as

\[ (RR)_v = \frac{2a^3V}{3R} \]  

(2.28)

Once the indentation depth \( \omega \) (or \( a \)) is known, the volumetric removal rate by a single particle can be readily calculated from Eq. 2.28.

Depending on the relative magnitude of the indentation depth (\( \omega \)) to the surface layer thickness (\( t \)), the material removal may be classified into two different regimes as depicted in Fig. 2-3; (a) \( \omega \leq t \) and (b) \( \omega > t \).
Indentation depth smaller than the surface layer thickness (i.e., $\omega \leq t$)

When the chemically modified surface layer is thick enough, the indentation depth $\omega$ will be no greater than the surface layer thickness and only plastic deformation within the surface layer occurs. In this case, $\omega$ (or $a$) can be obtained from the force balance imposed on the particle by the pad and the wafer that depends on the indentation depth of the particle into the wafer surface. The force imposed on a particle by the pad is given as

$$F_a = \pi b^2 P_r$$  \hspace{1cm} (2.29)

where $b$ is the radius of the circular indentation into the pad as described in Fig. 2-3a, and $P_r$ is the contact pressure given by Eq. 2.22. The force on the particle by the deforming wafer can be given as

$$F_a = \pi a^2 B_c$$  \hspace{1cm} (2.30)

where $B_c$ is the effective Brinell hardness of the chemically modified surface layer. The term “effective” is used because its value is influenced by the thickness of the chemically modified surface layer and the material property of the underlying substrate. The pad is supposed to hold the particles in place thereby transmitting the normal force to the particles that cause indentation of the wafer surface. The pad should be soft enough to provide stable anchoring of the particles on its surface. However, if it is too soft, the particles will be embedded too deeply into the pad diminishing the material removal capability of the particles. Thus, the pad should be sufficiently hard to make the value of $b$ in Fig. 2-3 to be close to the particle radius $R$. Then, from equations 2.29 and 2.30 with $b=R$,
This equation along with equations 2.22 and 2.28 gives the following expression for the volumetric removal rate by a single particle of radius $R$.

$$ (RR)_v = \frac{2}{3} C^{3/2} R^2 \left( \frac{\sigma}{\beta} \right)^{3/4} \left( \frac{E^*}{B_e} \right)^{3/2} V $$

(2.32)

**Indentation depth greater than the surface layer thickness (i.e., $\omega > t$)**

When the bulk wafer material beyond the surface layer is relatively soft (e.g., copper) compared to the hardness of the particles, it is possible for the particle to cause deformation into the wafer beyond the surface layer (Fig. 2-3b). As it was described in the previous section, the deformation is expected to be plastic for both the surface layer and the bulk wafer, and the force causing the deformation may be given as

$$ F_a = \pi \left[ 2tRB_e + (a^2 - 2tR)B_w \right] $$

(2.33)

where $B_w$ and $B_e$ are the Brinell hardness of the bulk wafer material and the chemically modified surface layer, respectively. Because this force is equivalent to the one imposed on the particle by the down-pressure, Eq. 2.33 is equated with Eq. 2.29 with $b=R$ to give

$$ a = \left[ \frac{R^2 P_e + 2(B_w - B_e)Rt}{B_w} \right]^{1/2} $$

(2.34)

Then, Eq. 2.28 for the volumetric removal rate by a single particle is reduced to

$$ (RR)_v = \frac{2}{3} \left[ C \left( \frac{\sigma}{\beta} \right)^{1/2} E^* R^2 \right] \frac{B_w}{B_e^2} R \left( 1 - \frac{B_e}{B_w - B_e} \right) \frac{V}{R} $$

(2.35)
While equation 2.35 describes a situation where the deformation reaches into the bulk wafer beyond the surface layer, deformation of the bulk wafer may not occur if the bulk wafer material is very hard. This limiting case corresponds to the situation of infinitely large $B_w$ in equation 2.35 that is reduced to

$$\frac{(RR)_v}{3} = \frac{4\sqrt{2}}{3} R^{1/2} t^{3/2} V$$  \hspace{1cm} (2.36)

Equations 2.32, 2.35 and 2.36 indicate that the volumetric removal rate by a single slurry particle depends on the geometry of the pad and the particle (i.e., $\beta$, $\sigma$ and $R$), the material properties ($E^*$, $B_w$ and $B_e$), the thickness of the surface layer ($t$), and the relative velocity of the pad (thus the particle) to the wafer ($V$). It may be interesting to note that $(RR)_v$ is independent of the down pressure, $P$. It is due to the fact that the contact pressure, $P_r$, is independent of $P$. As equation 2.21 indicates, the contact area of the pad with the wafer ($A_r$) increases linearly with $P$. Consequently, the mean contact pressure, $P_r$, which is $PA_o/A_r$, is independent of $P$. In the following section, the number of particles in contact with the wafer surface is estimated to determine the overall removal rate.

**Estimation of the Overall Removal Rate**

When the wafer is pressed against the pad and is in sliding motion relative to the pad, the particles can be entrapped only in the region where the local distance between the wafer and the pad asperity is smaller than the particle diameter, $2R$. Thus, the number of particles that can be trapped on the pad asperities is

$$n = 2R \cdot A_{real} \cdot \rho_n$$  \hspace{1cm} (2.37)
Here $A_r$ is the real contact area between the wafer surface and the pad asperities as it was described in previous section. $\rho_n$ is the number density of the particles in the slurry (i.e., number of particles per unit slurry volume), that can be calculated from the solid loading $C_a$ (wt%) of the slurry particles as

$$\rho_n = \frac{3}{4} \frac{C_a \rho_s}{\pi R^3 \rho_a} \quad (2.38)$$

where $\rho_a$ and $\rho_s$ are the densities of the particle and the slurry solution, respectively.

Substitute of Eqs. 2.38 into 2.37 gives

$$n = \frac{3}{2} \frac{C_a \rho_s}{\pi R^2 \rho_a} A_r \quad (2.39)$$

By combining Eqs. 2.39 and 2.18, the number of particles anchored on the pad asperities becomes

$$n = \frac{3}{2C} \frac{C_a \rho_s \left( \frac{\beta}{\sigma} \right)^{1/2} \frac{PA_o}{E^*}}{\pi R^2 \rho_a} \quad (2.40)$$

Now that $(RR)_v$ and $n$ are given in terms of various processing variables and material properties, the overall material removal rate in a CMP process can be estimated from Eqs. 2.32, 5.27 and 2.40 for the two different cases described above:

Removal Rate: *Indentation depth smaller than surface layer thickness (i.e., $\omega \leq t$)*,

$$(RR)_v = \left[ \frac{\sqrt{C}}{\pi} \frac{C_a \rho_s}{\rho_a} \left( \frac{\sigma}{\beta} \right)^{1/4} (E^*)^{1/2} B_e^{-3/2} \right] PV \quad (2.41)$$

Removal Rate: *Indentation depth larger than surface layer thickness (i.e., $\omega > t$)*,
As Eq. 2.23 indicates, the overall removal rate, \((RR)\), is the thickness removal rate which is the volumetric removal rate divided by the nominal wafer area \(A_o\). We may recall that \(C\) is a constant whose value is about 0.35 when \(h/\sigma\) is between 0.5 and 3.0 [Joh85]. The overall removal rate given above is reminiscent of Preston’s equation in that the material removal rate is proportional to the down-pressure, \(P\), and the relative velocity, \(V\). It should be noted that the Preston’s coefficient \(K_p\), whose value relies on empiricism, is now given as a function of various processing variables and material properties.

The model predicts the removal rate as a function of pressure and velocity product (PV) very well [72]. The removal rate is usually only considered as a function of PV in many modeling efforts, ignoring other processing parameters. But the PERC I model considers the particle size effect, which is crucial for any CMP system.

Equation 2.41 indicates that the overall removal rate is independent of the particle size \(R\) whereas Eq. 2.42 indicates a decrease in \((RR)\) with increasing particle size. When the indentation-depth is smaller than the surface layer thickness (i.e., \(\omega \leq t\)), the volumetric removal rate by a single particle \((RR)_v\) is proportional to \(R^2\) (Eq. 2.32). The number of particles in contact, on the other hand, is inversely proportional to \(R^2\) as the Eq. 2.40 indicates. Consequently, the particle size effect is negated.

Thus the model predicts as shown in Fig. 2-4 that the removal rate is either independent of particle size in the modified layer removal regime or decreases with particle size in the substrate indentation regime. But, the reported literature has shown
that removal rate also increases with particle size in a given range, and that there is an optimal particle size at which enhanced polishing can take place as shown in Figs. 2-5 and 2-6. To overcome these limitations, a (PERC II) model is developed by analyzing the crucial assumptions in PERC I, and refining them when necessary. The following chapter deals with assumption validation and PERC II model development.

Fig. 2-4. Schematic of PERC I predictions for increasing particle size.

Fig. 2-5. Polishing of silica by nanoslurry showing optimal particle size [118].
Fig. 2-6. Polishing of silica at various pressure, velocity and particle sizes showing optimal particle size [116].
The previous chapter highlighted the basic concepts involved in PERC I model development, along with various challenges faced by it. In this chapter, some of the critical model assumptions are verified and refined as necessary. The changes are incorporated into the basic framework of the PERC I model to develop the PERC II CMP model.

**Assumption Validation**

The critical components in the CMP process are the wafer, pad and slurry. As the wafer is fixed in a given CMP process, the contributions from pad and slurry are investigated to include their most crucial characteristics into the model. The main role of the pad in CMP is to trap the abrasive particles for material removal and transmit the load applied onto the trapped particles. In particle scale modeling, the function of the pad in global planarization need not be considered.

**Pad-wafer Contact Area**

The pad-wafer contact area strongly influences the number of particles that can be trapped under each pad asperity. The greater the contact area between pad and wafer, the higher the number of particles that can be trapped resulting in enhanced material removal. The changes in pad properties during polishing are important to model the CMP process. Often dry properties of the pad are investigated [119], rather than the more appropriate wet properties. The pad characterization investigations are also limited to the bulk [48,
68, 120] of the material rather than the surface. These bulk properties in dry state are usually employed in conventional CMP modeling efforts [68]. Since, polishing involves the interfacial behavior of pad, abrasive, and wafer, pad surface properties should be used rather than the bulk properties. Surface properties of the pad rapidly change due to immersion in slurry [121, 122], quickly reaching equilibrium values [46].

There is one fundamental question that needs to be answered. What is the real contact area between pad and wafer? And is the Greenwood-Williamson model [117], which states that “contact area is linearly proportional to the applied load” applicable in CMP systems? A basic assumption of the PERC I model is that removal rate is linearly dependant on the pad-wafer contact area. Abrasive particles trapped between pad asperities and wafer are considered to be responsible for material removal. As applied load on the wafer increases, the pressure at the pad asperity-wafer interface is supposed to remain constant due to elasticity of the pad, while the pad-wafer contact area increases. This behavior is predicted by the “Greenwood-Williamson” model, which assumes that the pad asperity tips contact the wafer under Hertz elastic theory conditions. Although a number of CMP models have employed this concept [63, 109], no systematic attempts are reported that provide validity of this assumption. Similarly, some researchers have assumed the contact area to be proportional to the power of two-thirds [90] of pressure without any experimental basis. In the past, pad-wafer contact area has been generally derived mathematically [27, 123] based on statistical surface roughness rather than actual experimental measurements. The pad-wafer contact area reported in literature including theoretical calculations varies from 0.0005% to 20% [27, 49, 63, 109, 124, 125]. Considering the importance of reliable value of the pad-wafer contact area and its
dependence on applied load in CMP modeling, a systematic attempt was made to measure it under CMP conditions. Kim et al. [124] have measured the pad-wafer contact area using optical microscopy and have reported a value of 5%. On the other hand, Basim [125] has reported the value to be 0.3% in dry state based on FTIR measurements. To investigate the pad-wafer contact area, a reliable technique is needed. To establish the reliability of optical microscopy and/or FTIR techniques, contact area measurements were performed on pad using optical microscopy.

**Optical microscopy**

The average pad asperity radius is around 30 μm for a typical CMP pad and the RMS roughness around 5μm. Therefore, an optical microscope can be used to study the contact spots, the majority of which can be expected to be over a micron in size. To simulate the wafer, a microscope coverslip was used. To minimize the bending of the coverslip under applied load, a fixture was fabricated as shown in Fig. 3-1 which enabled easy access to the microscope objective lens for translation in “in-plane” direction. The fixture consists of a base plate with a window hole of 10mm x 5mm to allow optical observation of the contact spots with minimum bending of the coverslip under applied load.

Optical measurements were performed in dry environment using Leika microscope equipped with frame grabber. One inch square coupons were punched out from IC1000 pad, which was used for polishing. The pad was ex-situ conditioned using a grid-abrade diamond pad conditioner. Minimal contrast between contact and non-contact regions was obtained, when the pad specimens were analyzed under the given load. This was not even sufficient for qualitative analysis. Measurable distinction between contact and non-
contact spots was achieved with coating of the pad surface with a 5-10 nm thick gold-palladium layer using the Denton sputter coater. The thin coated layer is not expected to affect the pad-wafer contact area results, as the attempt here is to measure the contact area under dry conditions.

Measurements were performed in two modes. Mode I consisted of measuring the contact area of selected asperities as a function of applied load and mode II consisted of measuring various contact spots under applied load by translating the microscope objective to cover the entire specimen in view. The optical micrographs were analyzed using Spot and Image-Pro softwares to calculate the percent contact area.

The micrographs obtained in mode I are shown in Fig. 3-2. The regions “P” and “N” indicate the pad asperities that are of interest in all three figures, though only Fig. 3-2a is labeled for clarity. In “P” regions, the asperity contact increases with applied load, whereas, in “N” regions, new asperity contacts are formed with applied load. The pad-wafer contact area values obtained from optical measurements are plotted in Fig. 3-3. The
results indicate that the contact area of a given pad asperity increases linearly with applied load and more importantly the magnitude of contact area obtained from optical technique is in the same range as obtained from the FTIR technique by Basim (Fig. 3-4). The discrepancy between this data and Kim’s optical microscopy result [124] could be due to the possibility that some non-contact areas were counted as contact spots due to poor contrast, as no pad surface modification or coating was mentioned in their study.

Fig. 3-2. Pad asperity contact spots seen through optical microscope. (a) 0.75 psi. (b) 5 psi. (c) 8.8 psi.

Fig. 3-3. Pad asperity contact area dependence on applied load. Each legend represents a selected asperity contact.
The main motivation behind investigating contact area measurements is to get a realistic value of pad-wafer contact area under CMP conditions. The optical and FTIR techniques are used in dry conditions, whereas the actual CMP is a wet system. Combining optical data from this study and FTIR data from Basim [125], it is established that FTIR with attenuated total reflection (FTIR-ATR) method can be successfully used to estimate the pad-wafer contact area. The same technique was employed in the present investigation to measure the contact area in wet condition simulating the CMP environment.

**Fourier transform infrared spectroscopy (FTIR)**

To investigate the pad-wafer contact area variation with applied load in various slurries, FTIR-ATR was used. The contact spots contribute toward the absorbance signal resulting in quantitative estimation of the pad-wafer contact area depending on signal
intensity. Maximum signal intensity is obtained when the pad surface is in full contact with the wafer surface. The contact area is obtained by normalizing the measured signal intensity with the maximum signal intensity [126]. The setup used for contact area measurements is shown in Fig. 3-5.

![Fig. 3-5 Schematic of the FTIR-ATR technique used to measure pad-wafer contact area. A silicon ATR crystal is used in contrast to the ZnSe crystal employed by Basim [125] to simulate the silica wafer system. A nitrogen purged Nicolet Magna 760 spectrometer equipped with liquid nitrogen chilled MCTA detector was used to carry out FTIR analysis for better accuracy. All the spectra were the result of 512 co-added scans at a resolution of 4 cm$^{-1}$. The background spectrum for all the experiments was the single-beam spectrum of the dry silicon crystal. The intensities of the CH$_2$ peaks of the polyurethane were recorded in the wavenumber range of 3000–2750 cm$^{-1}$. The asymmetric and symmetric stretches of CH$_2$ at wavenumbers of 2927 and 2855 cm$^{-1}$ respectively were analyzed. The contact area is expected not to depend on pad-wafer relative velocity, hence they are held stationary. The slurries were simulated by using solutions of varying pH with no other chemicals (passivating agents) or particles added. This is valid especially in silica CMP as no passivating chemicals are generally added in
slurry formulation. The absence of particles reduces the errors due to scattering. The water FTIR/ATR signal is subtracted from pad spectrum (both at same pH) to get the desired signature spectrum (Fig. 3-6). A representative piece of pad sample (IC-1000) soaked in given solution for one hour was placed on the (Si) ATR crystal and loaded at 4, 7 and 10 psi in the wet state under solution. The solutions studied are nanopure water adjusted to pH 4, pH 9 and pH10.5 with either 0.5M HNO$_3$ or 0.5M KOH.

Silica CMP is generally performed at slurry pH 10.5, as the silica solubility increases drastically after pH 10. Thus, there should be minimum silica wafer modification or “modified layer” formation at pH 4 and pH 9. Whereas, pH10.5 should lead to considerable surface dissolution of silica. These conditions are chosen, as all CMP process variables except pad properties are supposed to be constant in pH range of 4 to 9, and increased surface modification (dissolution) at pH10.5. The typical absorbance spectra obtained for a given specimen at different applied loads is shown in Fig. 3-7. The signal intensities at symmetric and asymmetric stretches are averaged to calculate the contact area. The results obtained for different pH and dry conditions indicate that contact area increases linearly with applied pressure for all conditions (Fig. 3-8). The normalized plot shows the accuracy associated with the measurement technique, which is consistent for all the conditions used. The linear dependence of pad-wafer contact area on applied load validates the use of Greenwood-Williamson theory in the CMP model. Pad-wafer contact area in the wet state was measured to be around 2-3 times the contact area in the dry state (Fig. 3-9). This implies that the removal rate calculated would be 2-3 times less if the dry contact area is used in the model. There is no discernible change in pad-wafer contact area under different pH conditions due to data scatter, possibly due to uncertainty
associated with subtraction of water spectrum. Utilizing the value of pad-wafer contact area obtained and the Greenwood-Williamson assumption, pad asperity modulus is derived which is used in calculation of pad asperity stress distribution.

Fig. 3-6. Analysis of the FTIR-ATR spectra.

Fig. 3-7. Absorbance spectra of pad (in pH 4) at different applied pressures.
Fig. 3-8. Contact area variation with applied load in dry and wet conditions.

Fig. 3-9. Contact area in dry and wet systems in various pH conditions.
Atomic force microscopy (AFM)

Contact area increase in the wet state is attributed to changes in the mechanical properties of the pad. Consequently, surface mechanical properties of pad were investigated using a colloidal probe atomic force microscope (AFM), where a particle is attached to AFM tip to measure pad properties. A Digital Instruments Dimension 3100 AFM equipped with a liquid cell adapter was used.

Particle indentation into pad was used to assess pad surface mechanical properties at pH levels of 4, 9 and 10.5. It was realized that a stiff cantilever is needed to probe the surface mechanical properties of the pad. Hence a tapping mode cantilever was modified by cutting the tip area with focused ion beam (FIB) yielding a cantilever with a stiffness of 75 N/m as compared to 30 N/m before the modification. A 7μm silica particle was attached to the tip of cantilever for measuring mechanical properties of the pad. Nanoindentation measurements were carried out using the modified AFM tip shown in Fig. 3-10. The pad surface was scanned with the colloid probe to find a relatively flat surface before indentation data was measured. This is expected to minimize the effect of surface asperities or roughness on the surface mechanical analysis. Typical indentation data are shown in Fig. 3-11. Only the approaching curve is used to investigate the pad surface mechanical properties. The results from AFM are summarized in Fig. 3-12.

The indentation data presented in Fig. 3-12 indicates that the pad surface mechanical response is markedly different in dry and wet systems. In dry pad case, the force on AFM probe increases with indentation depth at a constant pace without any change in stiffness. Whereas, for wet cases at all pH levels employed, a dramatic change in slope is observed. The indentation data is not reproducible over entire pad surface, as
the pad surface is not perfectly smooth due to asperities. But all the indentation data in wet state showed the stiffness change in contrast to the dry data. When the indentation data is aligned such that all curves match the dry pad behavior, it can be seen that the pad surface is affected by the solution pH which results in different mechanical response compared to the bulk material (Fig. 3-13). The moduli of pad surface and bulk shown in Fig. 3-12 were estimated utilizing Hertz theory. The modulus of dry pad is at least 3-5 times the pad modulus in dry state indicating pad swelling.

Fig. 3-10. Colloidal probe AFM tip (a) Modified tapping mode cantilever, dotted line shows the region of the cantilever removed by FIB. (b) Silica particle attached to the modified cantilever.
The surface modification of polyurethane is not entirely unexpected. The polyurethane polishing pads possess both -COC- and -COOC- functional groups in the main chain structure which are susceptible to hydrolytic attack [50], especially in alkaline media [50]. Disruption of hydrogen bonding is reported for pad samples immersed in DI water for 5 minutes [97]. Schematic of hydrogen bond disruption is shown in Fig. 3-14. This hydrolytic attack can explain the pad surface mechanical property behavior.

The presence of surface modification on pad surface or pad swelling could impact the abrasive trapping ability of the pad asperities along with the load transmitted to the particles during polishing depending on particle size. The change in contact area observed in wet state indicates that depending on slurry chemistry, the pad-wafer contact area could be effected, directly influencing the material rate. Hence, when different slurry chemistries are used in a CMP process, the CMP model needs to incorporate the slurry effect not only as a modified layer on wafer surface, but also as change in pad-wafer
contact area and abrasive trappability, which has not been accounted for in past CMP modeling efforts.

![AFM force-indentation curves](image)

Fig. 3-12. Force-indentation curves from AFM for different conditions. All curves are offset such that they match in the top portions to clearly show the evidence of a softer layer on pad in wet systems.

![AFM schematic](image)

Fig. 3-13. Schematic of pad surface as deduced from AFM data. (a) Dry state. (b) Wet state.
Fig. 3-14. Schematic of polishing pads (a) Structure showing hydrogen bonding between neighboring polymer molecules, (b) Water can form hydrogen bonds with polymer molecule by disrupting hydrogen bonds between the polyurethane molecules [127].

**Pad Asperity Height Distribution**

A quantitative understanding of the pad-wafer and pad-particle interface is fundamental to a comprehensive CMP model. The pad asperities are assumed to remain constant through the planarization of each wafer. It is known that the asperities involved in polishing wear out changing in height over time, while the shorter asperities, those that do not come into contact with the wafer, retain their original height. Thus the pad roughness or the asperity height distribution changes with polishing, if pad conditioning is not performed properly. The surface roughness of pad is known to be directly proportional to the polishing rate [128-130]. Different types of pads have different asperity height distributions, though they might have similar average surface roughness. In PERC I, pad asperites were assumed to be spherical and of same size with pad asperity height averaged out to obtain the pad asperity stress, and pad-wafer contact area. This implies that the pad asperities are not only assumed to be equal in size and shape, but also of equal height. This imposes the pad to have contact spots equal in size, experiencing
same contact pressure. Thus, all the particles trapped in these contact spots have same indentation into the wafer surface leading to equivalent material removal under each asperity. Thus the previous PERC model was equivalent to having a single asperity model with no effect of asperity height distribution. Depending on the CMP system and desired planarization criteria, different pads are used in practice. Therefore, it was considered necessary to include the asperity height distribution in the revised model (PERC II). Due to constant asperity height, only an average indentation depth was derived in the previous model (PERC I) to indicate surface roughness of the polished wafer. When asperity height distribution is used, asperities at different heights lead to different indentation depth, resulting in information about distribution of scratches (indentation depths). This data can be used to estimate the surface roughness as root mean square roughness (RMS).

To incorporate the asperity distribution, Gaussian distribution of asperities is assumed, as the reported literature cites Gaussian distribution [116]. The pressure at contacting asperities were calculated individually to estimate the indentation depth of trapped particles.

**Abrasive Particle Size Distribution**

The role of slurry in CMP is two-fold. The chemicals in the slurry are needed to react with wafer surface resulting in formation of a surface layer which is amenable to enhanced material removal. The abrasives in the slurry on the other hand, remove this modified surface layer causing material removal. The effect of chemicals is attributed to surface layer of different mechanical properties in PERC I model. Whereas, the effect of abrasives is considered by assuming average size of the particles. The monosized slurry assumed in the PERC I model is an ideal case, since real CMP slurries are seldom
monosized (Fig. 3-15). Oxide polishing slurries are commonly purchased in concentrated form and diluted on-site to minimize shipping and labor costs. Some mult-component tungsten and copper polishing slurries must be combined shortly before use because of their short postmix lifetime. The slurry stability can change after it is combined with the oxidizer or due to dilution leading to polydispersity.

![Graph](image)

Fig. 3-15. Particle size distribution of silica slurry [118]

Sometimes mixed abrasive slurries are used [32], which invariably exhibit polydisperse particle size distribution. To account for the particle size distribution effect, PERC I model was modified to incorporate slurry abrasives with Gaussian size distribution.

**PERC II Model Development**

Conceptually, the difference between the PERC I and PERC II models can be illustrated as shown in Figs. 3-16 and 3-17. The PERC I model assumes spherical pad asperities of equal height all over the pad (Fig. 3-16a). As the slurry is assumed to be
monosized, each asperity traps equal number of particles in CMP resulting in equal removal rate by each asperity and each particle (Fig. 3-16b).

![Fig. 3-16. Schematic of pad, slurry and wafer in PERC I model. (a)Before contact. (b) During CMP.](image)

Whereas in the PERC II model (Fig. 3-17a), the spherical pad asperities are assumed to have a height distribution as observed experimentally [116]. As the slurry is assumed to be polydisperse, each asperity traps different number of particles of different particle sizes in CMP according to pad asperity height and particle distribution (Fig. 3-17b). Thus the PERC II model includes realistic contributions of pad and wafer characteristics compared to the PERC I model. The model formulation is detailed in the following section.
Fig. 3-17. Schematic of pad, slurry and wafer in PERC II model. (a) Before contact. (b) During CMP

Real contact area \( A_{real} \) between pad and wafer due to applied load \( L \) is given by:

\[
A_{real} = \pi N A_o \beta \int_h^\infty (z - h) \phi(z) dz \quad (3.1)
\]

\[
L = \frac{4}{3} N A_o E^* \beta^{\frac{3}{2}} \int_h^\infty (z - h)^{\frac{3}{2}} \phi(z) dz \quad (3.2)
\]

\[
\phi(z) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right) \quad (3.3)
\]

where \( N \) is the number density of the pad asperities (i.e., number of asperities per unit area), \( A_o \) is the total area of the flat wafer surface, and \( h \) is the distance between wafer
surface and the pad asperity average plane under the given conditions. $E^*$ is the composite modulus between the two surfaces in contact, defined as

$$
\frac{1}{E^*} = \frac{1 - \nu_p^2}{E_p} + \frac{1 - \nu_w^2}{E_w}
$$

(3.4)

where, $E$ and $\nu$ are the Young’s modulus and the Poisson’s ratio, and the subscripts $p$ and $w$ represent the pad and the wafer, respectively. Equations (3.1) and (3.2) are derived from Hertz theory as in Greenwood-Williamson model [77, 117], yielding contact area and load at a given asperity of height ‘$z$’ to be:

$$
A_{\text{real}}(z) = \pi \beta (z-h) \quad (3.5)
$$

$$
L(z) = (4/3)E^* \beta^{1/2}(z-h)^{3/2} \quad (3.6)
$$

where ‘$h$’ is calculated from Eq. 3.2, since applied load and pad properties are known. The underlying assumption for Eq. 3.2 is that the applied normal force on the wafer is supported entirely by the pad asperities. As was mentioned previously, experimental measurements indicate that the hydrodynamic pressure is negligible under typical CMP operational conditions [69].

From equations 3.5 and 3.6, contact pressure and contact area at a given asperity height can be calculated assuming that the pad asperity distribution is known. The pad asperity distribution can be determined from routine surface roughness measurement techniques like optical and force profilometers or atomic force microscopy (AFM). Contact pressure according to Hertz theory, on an asperity of height $z$ is given by:

$$
P(z) = \frac{4E^*(z-h)^{1/2}}{3\pi\beta^{2/3}} \quad (3.7)
$$
Plastic Deformation of Wafer Surface

When an abrasive particle is indented into the wafer surface, elastic and plastic deformation of wafer and abrasive particle occur. However, in CMP, slurries are designed such that the chemistry changes the modulus of the wafer surface for enhanced material removal. In the current model, particles are assumed to be rigid such that they do not deform during polishing. When the indentation of abrasive particle is beyond the modified surface layer, the substrate indentation could yield either elastic or plastic response depending on the substrate, for example in case of silica, the deformation will still remain plastic whereas for copper, it would be elastic deformation followed by plastic deformation. Therefore, in material removal calculation, indentation into ductile wafers is not considered to cause material removal, as it leads to change in shape only, whereas in the case of brittle material like silica, cracks are developed leading to material removal.

Slurry Characteristics – Particle Size Distribution

The abrasive particles in the slurry are assumed to be spherical in shape with a given size distribution $\psi(r)$. As mentioned earlier, particle size distribution $\psi(r)$ was assumed to be Gaussian, given by:

$$\psi(r) = \frac{1}{\sigma_r \sqrt{2\pi}} \exp\left(-\frac{(r - \mu_r)^2}{2\sigma_r^2}\right) \quad (3.8)$$

where $\mu_r$ is the mean particle radius and $\sigma_r$ the standard deviation.

Material Removal by Single Particle

The overall material removal rate (RR) is assumed to be the multiple of the removal rate by a single particle and the number of particles trapped between pad asperities and the wafer:
\[ RR = \left( \frac{1}{A_0} \right) \int n(r,z) \int_0^z RR_v(r,z) \, dr \, dz \]  
\hspace{1cm} (3.9)  

where \( n(r,z) \) is the number of slurry particles anchored between pad asperity and the wafer and \( RR_v(r,z) \) is the volumetric removal rate by a single particle of radius \( r \) trapped by asperity of height \( z \). The overall removal rate, \( RR \) is typically described in terms of the thickness of the wafer surface layer removed per unit time; hence the division by the nominal wafer area \( A_0 \).

Schematic illustration of single particle abrasion is given in Fig. 3-18. The material removal is viewed as a sliding-indentation process in which the volumetric removal rate by a single abrasive particle is given by

\[ RR_v(r,z) = A_i(r,z)V \]  
\hspace{1cm} (3.10)  

Where \( A_i(r,z) \) is the cross-sectional area of the particle indented into the wafer under pressure (hashed area in Fig. 3-18), and \( V \) is the relative velocity between the pad asperity and the wafer.

Fig. 3-18 Schematic of single particle abrasion showing the removal of chemically modified surface layer, (a) \( \omega \leq t \); (b) \( \omega > t \) with soft wafer materials.[72]
Assuming that the hard particle maintains its spherical shape,

\[ A,(r,z) = -a(r,z)(r - \omega(r,z)) + r^2 \arcsin \frac{a(r,z)}{r} \]  

(3.11)

where \( \omega(r,z) \) is the indentation depth and \( a(r,z) \) is the radius of the circular indentation on the wafer surface caused by the particle of radius \( r \) trapped under asperity of height \( z \).

Because \( \omega(r,z) \) is very small compared to \( r \), \( a(r,z)/r \) is also very small and Equation 3.11 can be linearized about \( a(r,z)/r \) as already discussed:

\[ A,(r,z) = \frac{4}{3} a(r,z) \omega(r,z) \left( 1 + O\left(\frac{\omega(r,z)}{r}\right) \right) \]  

(3.12)

Furthermore, since \( a^2(r,z) = (2r\omega(r,z) - \omega^2(r,z)) \) and \( \omega(r,z) \) is much smaller than \( r \), Equation 3.12 can be rewritten as

\[ A,(r,z) = \frac{2a^3(r,z)}{3r} \]  

(3.13)

Once the indentation depth, \( \omega(r,z) \) (or \( a(r,z) \)) is known, the volumetric removal rate by a single particle can be calculated from Eq. 3.13. It should be noted that \( \omega \) and \( a \) are dependant on particle radius and height of the trapping asperity. Removal rate per particle is given by:

\[ RR,(r,z) = \frac{2a^3(r,z)}{3r} V \]  

(3.14)

Depending on the relative magnitude of the indentation depth (\( \omega(r,z) \)) as compared to the surface layer thickness (\( t \)), the material removal may be classified into two different regimes as depicted in Fig. 3-18: (a) \( \omega(r,z) \leq t \) and (b) \( \omega(r,z) > t \) and can be calculated using force balance as follows.
Considering that the pad is sufficiently hard as compared to the modified layer, the value of \( b \) in Fig. 3-18 can be close to the particle radius \( R \),

\[
a(r, z) = r \left( \frac{P(z)}{B_e} \right)^{1/2} \quad \text{when } \omega(r,z) \leq t
\]

and

\[
a(r, z) = \left[ \frac{r^2 P(z) + 2(B_w - B_e)rt}{B_w} \right]^{1/2} \quad \text{when } \omega(r,z) > t
\]

where, \( B_e \) is the effective Brinell hardness of the chemically modified surface layer. In the following section, the number of particles in contact with the wafer surface is estimated to determine the overall removal rate.

**Number of Active Abrasive Particles**

When the wafer is pressed against the pad and is in sliding motion relative to the pad, the slurry particles can be entrapped only in the region where the local distance between the wafer and the pad asperity is smaller than the particle diameter, \( 2R \). For a slurry consisting of particles with Gaussian distribution, mean particle size \( (R_m) \) is used to calculate the number of particles that can be trapped on the pad asperity of height \( z \):

\[
N(z) = 2R_m A_{\text{real}}(z) \cdot \rho_n
\]

where \( R_m \) is the mean abrasive particle radius and \( A_{\text{real}}(z) \) is the real contact area between the wafer surface and the pad asperity as given by eq. 3.5. \( \rho_n \) is the number density of the particles in the slurry (i.e., number of particles per unit slurry volume), that can be calculated from the solid loading \( C_a \) (wt\%) of the slurry particles as

\[
\rho_n = \frac{3C_a \rho_s}{4 \pi \rho_s \int r^3 \psi(r) \, dr}
\]
where \( \rho_a \) and \( \rho_s \) are the densities of the particle and the slurry solution, respectively and \( \psi(r) \) the slurry particle size distribution. Substitution of Eq. 3.18 into 3.17 gives

\[
N(z) = \frac{3}{2} \frac{R_m C_a \rho_s A_{real}(z)}{\pi \rho_a} \int r^3 \psi(r) \, dr
\]  

(3.19)

From eq. (3.19), the number of trapped particles of radius \( R \) is given by,

\[
n(R, z) = N(z) \ast \psi(R) = \frac{3}{2} \frac{R_m C_a \rho_s A_{real}(z) \psi(r)}{\pi \rho_a} \int r^3 \psi(r) \, dr
\]

(3.20)

**Estimation of the Overall Removal Rate**

The overall material removal rate can be determined by following these steps:

- \( h \) is calculated using eq. 3.2.

- Contact pressure \( P(z) \) at a given asperity is calculated along with the real contact area \( A_{real}(z) \) of that asperity with wafer.

- Total number of particles \( N(z) \), that are trapped are determined, from which number of particles of given radii, \( n(r, z) \) are derived.

- From \( P(z) \), removal rate by single abrasive of radius \( r \) is calculated which is multiplied with \( n(r, z) \) to give removal rate by that particle size. This process is repeated to account for all the particles of different radii trapped on that asperity yielding material removal rate by particles trapped on a given asperity.

- It should be noted that there are \( \phi(z)N_{A_0} \) asperities at a given height which will each contribute the same material removal.
The above step is repeated for all the pad asperities of various heights that are in contact with wafer (determined by \( h \)) to give overall removal rate by the following equation:

\[
RR = \frac{1}{A_0} \int \int n(r,z) \mathcal{R}_v(r,z) dr dz
\]

\begin{align*}
\Rightarrow RR &= \frac{V R_m C_a \rho_a N}{\pi \rho_a} \int_{r_0}^{\infty} \mathcal{A}_{real}(z) \phi(z) \left\{ \int_{r}^{\infty} \frac{\psi(r) a^3(r,z)}{r} dr \right\} dz \\
&= \text{(3.21)}
\end{align*}

where, \( a(r,z) \) varies as given in eqs.3.15 and 3.16 depending on particle radius and asperity height.

The indentation depth, \( \omega(r,z) \) is used to calculate the surface roughness RMS of the polished wafer according to the following equation:

\[
RMS = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \omega_i^2}
\]

\text{(3.22)}

where \( \omega_i \) is the indentation depth.

The overall removal rate predicted by Eq. 3.21 and surface roughness given by Eq. 3.22 depend on various parameters that represent the geometry of the pad and the particle (i.e., \( \beta, \sigma \) and \( R \)), material properties (\( E^*, B_w \) and \( B_e \)), thickness (\( t \)) of the modified surface layer on wafer, and the relative velocity of the pad (thus the particle) to the wafer (\( V \)). The PERC II model will be validated with respect to critical CMP parameters in the next chapter.
CHAPTER 4
MODEL ASSESSMENT

The parameters governing the material removal calculation consist of CMP operating parameters, slurry parameters, pad and wafer parameters, as listed in Table 4-1. However, some of these parameters, especially the sub-nanometer thick modified layer hardness and thickness can not be determined experimentally due to equipment limitations. The remaining parameters are relatively easy to investigate experimentally for many CMP systems. Hence, modified layer hardness and thickness are used as the fitting parameters in the model.

Table 4-1. Parameters included in the revised PERC model.

<table>
<thead>
<tr>
<th>Operating Parameters</th>
<th>Pad properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure</td>
<td>Pad modulus</td>
</tr>
<tr>
<td>Velocity</td>
<td>Asperity height distribution</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Slurry parameters</th>
<th>Wafer properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particle size distribution</td>
<td>Modified layer thickness</td>
</tr>
<tr>
<td>Solids loading</td>
<td>Modified layer hardness</td>
</tr>
<tr>
<td>Abrasive density</td>
<td>Wafer hardness</td>
</tr>
<tr>
<td>Solvent density</td>
<td></td>
</tr>
</tbody>
</table>

Model Validation

In this section, PERC II model predictions are compared with experimentally measured values to assess the reliability of the proposed CMP model. The experimental literature data is also compared with PERC I predictions to assess the advantage of PERC II over PERC I. Though PERC II is a refined model compared to PERC I, the model is still a simplified version of real CMP, mainly in estimating the number of trapped
particles and indentation depth considering an abrupt transition between modified layer and wafer substrate. The removal rate values for both models are normalized by a constant to have reasonable modified layer thickness values. Surface roughness data is as produced by the models without any normalizing factor. The size distribution of slurries given in the respective literature are fit to the closest Gaussian distribution, which is used in the model.

**Silica CMP: Micron Sized Slurries**

The polishing results of silica by silica abrasives as reported by Choi et al. [131] are considered here. Micron sized CMP slurries were formulated by dispersing sol-gel silica particles of different sizes from Geltech Corporation in deionized water and adjusting pH values to 10.5. The average diameter of the slurry particles were 0.2, 0.5, 1.0 and 1.5 μm. The size distribution of the slurries is plotted in Fig. 4-1. A 2 μm thick silica layer was deposited on p-type silicon wafer by plasma-enhanced chemical vapor deposition. 1.5 x 1.5 in silica samples were used to mimic the wafers. A Struers RotoPol-31 tabletop polisher was used with IC 1000/Suba IV stacked pad. The following conditions were applied: down pressure of 3.5 psi, slurry flow rate of 100 ml/min, and rotation speed of 150 rpm (110 cm/s). The polishing rates were determined by measuring oxide film thickness using variable angle spectrometry ellipsometry and surface quality was assessed by atomic force microscopy (AFM).

The polishing rate was investigated as a function of particle size and solids loading. The removal rate decreased with increasing particle size (Fig. 4-2). The parameters used are: wafer Hardness=8 GPa, modified layer Hardness = 450 MPa, modified layer thickness = 7.2 Å and Normalization factor = 100. The PERC II model correlates well
with the experimental values than the PERC I model. The reported literature belongs to the substrate indentation regime. The model projects that enhanced removal rate can be obtained by using smaller sized slurries. The surface roughness on the other hand, increases with particle size on an average for all the slurries considered (Fig. 4-2b). This is not surprising as the larger size particles have deeper indentation depths increasing surface roughness. Both the models predict the average trend in surface roughness with increasing particle size.

![Fig. 4-1. Slurries used in oxide polishing [131].](image)

**Silica CMP: Nanosilica slurries**

The slurry sizes in the earlier study spanned from 0.2 to 1.5 μm, which is well above the normal slurry size range (<300 nm). Considering the nanosize slurries of 50-150nm mean particle sizes, CMP behavior of silica polishing by silica abrasives is investigated further [116]. Slurries with mean particle diameters of 50, 80, 110 and 140nm were used to polish 100mm diameter, p-type polished silicon wafers that contained a thickness of 1.3 μm thermally grown oxide. The silica particles were
precipitated from a colloidal solution. The polishing was done at three different levels of pressure and relative pad-wafer velocities: 18kPa, 0.4m/s; 32kPa, 0.6m/s; and 45kPa, 0.9m/s. The parameters used in the model are: wafer Hardness=8 GPa, modified layer Hardness = 250 MPa, and modified layer thickness = 9.3 Å. The removal rates for the given set of pressure and velocity with different particle sizes is shown in Fig. 4-3. The material removal rates with slurry mean sizes of 10 and 20nm were ignored, as the authors [116] noticed considerable particle staining on the polished surfaces indicating agglomeration.

The PERC II model corresponds with the experimental results better than the previous PERC I model. Each set of modeled removal rates for given pressure and velocity were normalized individually with normalizing factor ranging from 185-200. The discrepancy in the removal rate at particle size 50nm between experimentally observed values and modeled values can be attributed to agglomeration, which is suggested by the authors for that size [116].

![Graph showing removal rate vs. mean particle size](image)

Fig. 4-2. Silica polishing by slurries in the micron range. (a) Removal rate. (b) Surface quality
The surface roughness of wafers polished was reported to be 6 Å on an average. The PERC II model predicts the surface roughness to be from 5.8 – 6.2 Å, which is in good agreement with the experimental results. The PERC I model predicts an average indentation depth of 5.5 to 6 Å which is also in good correlation.
As the model correlated well with the micron and submicron sized slurries, nanosized slurries in the low-end regime (10-100nm) are considered further. Colloidal particles of size 12, 25, 45, 75nm were used to polish thermal oxide (silica) wafers with potassium iodate (KIO₃) in the slurry [118]. The removal rate was reported to increase with particle size, reach an optimum and then decrease with particle size (Fig. 4-4). The parameters used are: wafer Hardness=8 GPa, modified layer Hardness = 200 MPa, modified layer thickness = 5.1 Å, and Normalizing factor=330. The PERC I model predicts the removal rate to remain constant with particle size, whereas the PERC II model predicts the optimal particle size at which enhanced material removal can take place. The surface roughness predicted is higher than the reported values.

![Fig. 4-4. Silica polishing by nanosized slurries. (a) Removal rate. (b) Surface quality.](image)
Copper CMP

As the model is found to correlate well with the brittle wafer polishing, metal CMP process is now considered to review the comprehensive nature of the model. Copper wafers were polished with particles of size 100, 200 and 300 nm [132]. The chemicals used were 5 wt% H₂O₂ and 1 wt% glycine with pH adjusted to pH 4.

The parameters used are: wafer Hardness=1.75 GPa, modified layer Hardness = 300 MPa, modified layer thickness = 15 Å and Normalizing factor=7. The PERC I model predicts removal rate independent of particle size, while the PERC II correlates well with the observed values (Fig. 4-5). The opposing trend of surface roughness between experimental and observed values is surprising. The discrepancy in surface quality predictions could indicate presence of few bigger sized particles or agglomeration which can result in high surface roughness than predicted [125]. The presence of larger size particles or slurry agglomeration, is however not mentioned by the authors.
Tungsten CMP

Bielman et al., [133] have conducted CMP of tungsten with alumina slurries of size distributions shown in Fig. 4-6. The slurry contained K$_3$Fe(CN)$_6$ as the chemical additive with pH adjusted to pH 3. The removal rate at 15wt% solids loading is shown in Fig. 4-7. The parameters used are: wafer Hardness=6.5 GPa, modified layer Hardness = 1.5 GPa, modified layer thickness = 9Å and normalization factor = 23. The PERC II model correlates well with the experimental removal rate values compared to the PERC I model (Fig. 4-7a). The surface roughness values predicted by PERC I and II are both higher compared to the observed values, but the surface roughness trend with particle size matches very well.

Fig. 4-5 Copper CMP as a function of particle size. (a) Removal rate, (b) Surface quality.
Fig. 4-5 Continued.

(b)

Fig. 4-6. Alumina slurry size distributions [133].
Fig. 4-7. Effect of particle size on tungsten CMP. (a) Removal rate, (b) Surface quality.

**Model Predictions**

The better match of PERC II model predictions than the PERC I model indicates that the refinements made to the PERC I model assumptions are indeed significant and add to the predictive capability of the model. In the following sections, predictive capability of PERC II model is explored, comparing with the PERC I model. The
enhanced performance of the proposed model (PERC II) is compared with PERC I predictions considering the case of a brittle wafer as in Silica CMP. The parameters used for comparison are listed in Table 4-2.

Table 4-2. Parameter values used for comparing PERC I and II models.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pad modulus</td>
<td>60 MPa</td>
</tr>
<tr>
<td>Pad asperity density</td>
<td>$2 \times 10^8$/m$^2$</td>
</tr>
<tr>
<td>Mean pad asperity radius</td>
<td>50 μm</td>
</tr>
<tr>
<td>Pad asperity height distribution σ</td>
<td>6.7 μm</td>
</tr>
<tr>
<td>Pressure</td>
<td>6.5 psi</td>
</tr>
<tr>
<td>Velocity</td>
<td>1 m/s</td>
</tr>
<tr>
<td>Particle size</td>
<td>200 nm</td>
</tr>
<tr>
<td>Particle size distribution σ$_r$</td>
<td>16 nm</td>
</tr>
<tr>
<td>Abrasive density</td>
<td>2.2 gm/cc</td>
</tr>
<tr>
<td>Solids loading</td>
<td>10 wt%</td>
</tr>
<tr>
<td>Wafer hardness</td>
<td>8 GPa</td>
</tr>
<tr>
<td>Layer hardness</td>
<td>600 MPa</td>
</tr>
<tr>
<td>Layer thickness</td>
<td>10 Å</td>
</tr>
</tbody>
</table>

**Removal Rate as a Function of Particle Size for a Given Solids Loading:**

The comparison of the two models with varying particle size is shown in Fig. 4-8. As discussed earlier (in chapter 3), material removal rate as per PERC I model is independent of particle size. The PERC II model, however, predicts an increase in removal rate with particle size during modified layer removal, and as the particle indentation proceeds into wafer substrate, the removal rate decreases smoothly with particle size, as also observed experimentally. Decrease in removal rate with particle size in wafer indentation is caused by the substrate being harder compared to the modified layer. The PERC II model differs with the PERC I model mainly in estimation of the modified layer removal and transition regimes and not considerably in the substrate removal regime. The smooth transition from modified layer removal regime to substrate removal regime is attributed by inclusion of pad asperity distribution into the model. Whereas, the removal rate increase with particle size in the modified layer removal
regime is due to the inclusion of slurry particle size distribution and strongly depends on it.

Fig. 4-8. Comparison between PERC I and PERC II models with constant solids loading of different particle sizes. (a) Removal rate. (b) Surface roughness.

The other feature of the PERC II model is the estimation of surface quality as root mean square surface roughness and maximum scratch depth. The PERC I model on the other hand, gave an estimation of surface roughness through average indentation depth.
The PERC II model, however gives the scratch (indentation) depth under each contacting pad-asperity due to each trapped particle size. This information is used in calculation of the surface roughness (RMS) as described in the previous chapter. The PERC I model yielded an average scratch depth whose dependence on particle size suddenly changed when the particle was large enough to indent past the modified layer into the substrate. PERC II also reveals that particle indentation into the substrate takes place at a smaller mean particle size as compared to that predicted by PERC I model.

This is expected, as some of the larger particles in the slurry indent into the substrate, while most others indent into the modified layer alone keeping the average particle size lower. Indentation in to the wafer substrate below the modified layer is not a desirable feature, as it negates the purpose of CMP. But, as long as the substrate removal is minimal with low indentation depths, it can be tolerated.

**Effect of Modified Layer Thickness**

The chemical reaction between slurry chemicals and wafer surface leads to the modified layer, whose thickness and mechanical properties are strongly dependent on the type of chemicals present in the slurry. The thickness of the modified layer in an optimized CMP process should be such that the particle indents completely into the modified layer and not into the substrate below. However, in a slurry with particles of various sizes, not all particle sizes can indent/remove the entire modified layer. From Figs. 4-9 a and b, the transition from modified layer to substrate indentation regime occurs around 8 Å which corresponds to 200nm particle size. Therefore, as the modified layer thickness decreases below 8 Å keeping all the other parameters constant, the slurry particles indent into the substrate and the removal rate decreases. Whereas, when the modified layer thickness increases beyond 8 Å, the indentation becomes more and more
limited to modified layer alone. Thus, after a critical thickness, the removal rate remains constant, even if the modified layer thickness increases. This indicates that the chemical aggressiveness of the slurry should not focus on forming a thicker layer, but on forming modified layer faster of desired thickness.

Fig. 4-9. Model comparison with varying modified layer thickness. (a) Removal rate, (b) Surface quality.
**Effect of Modified Layer Hardness**

As the modified layer hardness increases, less and less indentation of the particles take place, resulting in lower material removal rates (Figs. 4-10 a and b). Thus, if there is no surface modification due to slurry chemicals in silica, minimal removal rate will be observed. As 200nm particle was in the transition region, with increasing modified layer hardness, most of the particles indent only the modified layer. The transition from modified layer to substrate indentation is depicted more clearly by the PERC II model than by the abrupt transition predicted by the PERC I model.

Combining the model predictions and validation of metallic and dielectric CMP systems, the PERC II model can be used as a CMP parameter selection guide for emerging CMP systems and to optimize the current CMP processes. The fitting parameters, modified layer hardness and thickness are in the range of expected values for the respective CMP systems. The thickness of the modified layer determines the optimal particle size predicted, whereas the modified layer hardness in comparison to wafer hardness determines the sharp decrease in removal rate after transition from modified layer to substrate removal. The increase in removal rate with particle size in the modified layer removal regime is influenced solely by the abrasive particle size distribution.

The normalizing factor is found to be specific to the CMP system and strongly dependent on the toolset used. The normalizing values of 185-200 for three sets of pressure and velocities used in silica CMP by nanosilica slurries [116], indicate that the normalizing value is influenced strongly by the toolset used. This is expected, as the ability of pad to trap the particles is strongly dependent on the toolset configuration. The normalizing factor is also expected to depend on type of wafer being polished and the slurry used. For example, if particles are attracted to wafer due to electrostatic attraction,
the number of trapped particles will be much less compared to the system, where particles are repelled by the wafer surface charge. The next chapter summarizes the findings in this study and future work is suggested based on the insights obtained through the study.

Fig. 4-10. CMP performance comparison with varying modified layer hardness. (a) Removal rate, (b) Surface quality.
CHAPTER 5
SUMMARY AND SUGGESTIONS FOR FUTURE WORK

The device shrinkage in microelectronics industry necessitates chemical mechanical polishing and continues to demand stringent performance of CMP. The material removal mechanism needs to be understood with the relative contributions from particle-wafer, pad-wafer, and pad-particle interfaces and the reaction of slurry with wafer and pad materials. In this study, a model is developed to predict process parameters for designing an optimal CMP process.

Most of the CMP modeling efforts concentrated mainly on explaining the Preston’s equation. Therefore, most of them were focused on product of pressure and velocity (PV) as the operating parameter and not on other important parameters like slurry and pad variables. The PERC II model is developed based on PERC I model [72]. The assumptions in PERC I model were refined to include particle size distribution of slurry and pad asperity height distribution. Though recently a model developed by Luo [27] incorporated particle size distribution, part of the distribution was neglected and hence doesn’t incorporate the process variables appropriately.

The pad-wafer contact area dependence on applied load is generally assumed in many CMP models with no basis of a systematic study. The material removal rate is directly proportional to the pad-wafer contact area. Therefore, the magnitude of pad contact area is significant in modeling the process. The earlier studies focused on contact area in dry state and the results reported were not consistent. The contact area was investigated in dry and wet states in the current study. Contact area in dry state was
studied using optical microscopy. The change in contact area due to presence of aqueous environment and different pH conditions is investigated using Fourier transform infrared spectroscopy. The contact area was found to be 2-3 times the contact area in dry state. The pad surface mechanical properties were also investigated at different pH conditions using Atomic force microscopy. The slurry was found to react with pad surface layer (~50nm thick), modifying its mechanical properties. Both the pad-wafer contact area and contact stresses are strongly dependent on pad asperity height distribution. To incorporate the effects of contact area and contact stresses, Gaussian asperity height distribution of asperities is incorporated into the PERC II model.

The slurry used in commercial CMP is seldom monosized. The size distribution of the particles can strongly affect the process. Hence, the slurry abrasive sizes were assumed to be Gaussian in distribution were incorporated into the PERC II model. The developed model correlates well with experimental data for oxide and metal CMP systems. The model can be used as a selection tool for choosing the proper parameter values optimizing the CMP process. For optimization of the process, a sensitivity analysis can be performed to narrow down the parameters that are of utmost importance. An example is shown here of silica CMP system with the parameter ranges shown in table 5-1. SOBOL analysis was performed with 4608 data sets (using SimLab software) assuming particle size distribution to be Gaussian with standard deviation equal to the one-fourth the mean particle size. The results revealing the first order and interaction effects of parameters on removal rate are shown in Fig. 5-1. The sensitivity analysis indicates that the particle size has the significant effect on removal rate compared to all the other parameters. The particle size determines if the material removal will be in
modified layer regime or will extend to substrate indentation regime. Depending on the indentation regime, the hardness of the material being removed strongly effects the material removal.

Table 5-1. Process parameter ranges used in sensitivity analysis.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure</td>
<td>2-8 psi</td>
</tr>
<tr>
<td>Mean Particle Size</td>
<td>10-500 nm</td>
</tr>
<tr>
<td>Asperity Radius</td>
<td>20-100 μm</td>
</tr>
<tr>
<td>Asperity Height S.D</td>
<td>1-20 μm</td>
</tr>
<tr>
<td>Asperity Density</td>
<td>0.02- 20 x 10⁹/m²</td>
</tr>
<tr>
<td>Pad Modulus</td>
<td>20-200 MPa</td>
</tr>
<tr>
<td>Modified Layer Hardness</td>
<td>200-500 MPa</td>
</tr>
<tr>
<td>Modified Layer Thickness</td>
<td>5-10 Å</td>
</tr>
</tbody>
</table>

Fig. 5-1. Sensitivity analysis of silica CMP system.

The reported findings in this study provide insight into the fundamental understanding of the polishing mechanism and can give guidelines for optimizing any given CMP system.
Suggestions for Future Work

The insight provided by the current study into the CMP process in general can be used to conduct further studies to better understand the process from various perspectives. The study focuses on slurry with rigid abrasive particles. With the advent of low-k dielectric polishing, softer abrasives are being used. Inclusion of non-rigid abrasives into the model will facilitate better exploitation of the process potential and faster adaptation of emerging CMP systems to the industry requirements.

Recently, many commercial slurries are designed with surfactants or particle surface modification for slurry stability. Addition of these extra factors can significantly effect the CMP process. For example, it is shown by Basim [125] that polishing performance is degraded with addition of certain surfactant concentration to the slurry. These steric effects or electrostatic effects due to ionic species in the slurry should be included in the model for comprehensive understanding of the effects involved. Similarly, effects of pH and ionic species on CMP processes [134-137] can also be included.

The current model assumes that there are no stresses involved in the wafer, which could affect the CMP process. But, it has been shown by McGarth et al. [138] that if the stresses induced in the wafer due to material deposition and other processes are not relieved, the CMP process could be affected. Inclusion of such a stressed wafer could help in understanding the effect of deposition process on CMP process performance.

One of the experimental challenges faced in this study was the measurement of the modified layer properties. Nanoindentation and AFM were used to investigate the modified layer on the silica and copper surfaces. But, the instruments were not sensitive
enough to detect the modified layer properties quantitatively. It is suggested that a
diamond tipped AFM cantilever with a narrow needle type tip be used to sense the
modified layer properties, without any interference from the substrate. This data can be
used to incorporate gradual transition between modified layer and substrate compared to
the abrupt transition employed now.

The current model assumes formation of a modified layer on the wafer surface
which is critical in material removal process. Based on the AFM data of pad surface,
even a modified layer on the pad can be assumed to assess the trapping capability of
abrasives and load transmittal of pad onto the wafer through the abrasives. Developing a
model including such a layer on pad will result in better predictive capability when
different slurry chemicals are used that effect pad properties.

The model developed shows predictive capability in the particle scale. Therefore, it
can be extended to die and wafer scales to predict the topography evolution during
polishing. Such a model extension would also involve including the selectivity aspect of
the CMP slurry, dishing and erosion.
LIST OF REFERENCES


BIOGRAPHICAL SKETCH

Suresh Babu Yeruva was born on Feb 18th, 1978, in Machilipatnam, Andhra Pradesh, India. He finished his undergraduate education at the Indian Institute of Technology, Madras (IITM), India. He pursued his Master of Science at Washington State University in the Department of Materials Science and Engineering. In 2001, he started his Ph.D. program with Professor Brij Moudgil at the University of Florida.