

Copyright © 2004, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

**INTEGRATED CMP METROLOGY
AND MODELING WITH RESPECT
TO CIRCUIT PERFORMANCE**

by

Runzi Chang

Memorandum No. UCB/ERL M04/11

21 May 2004

**INTEGRATED CMP METROLOGY
AND MODELING WITH RESPECT
TO CIRCUIT PERFORMANCE**

by

Runzi Chang

Memorandum No. UCB/ERL M04/11

21 May 2004

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Integrated CMP Metrology and Modeling With Respect To Circuit Performance

by

Runzi Chang

B.E. (Tsinghua University, Beijing, China) 1996

M.S. (University of California, Berkeley) 2001

A dissertation submitted in partial satisfaction of the requirement

for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:


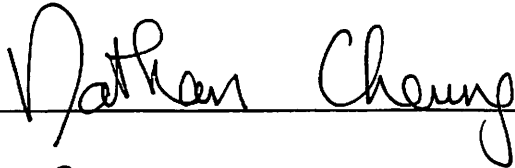

Professor Costas J. Spanos, Chair

Professor Nathan Cheung

Professor David Dornfeld

Spring 2004

The dissertation of Runzi Chang is approved:

 C. S. SPARROW	5/10/04
Chair	Date
 Nathan Chung	5/11/04
	Date
 Dennis Kifer	5/12/04
	Date

University of California, Berkeley

Spring 2004

Integrated CMP Metrology and Modeling With Respect To Circuit Performance

Copyright © 2004

By

Runzi Chang

Abstract

Integrated CMP Metrology and Modeling With Respect To Circuit Performance

by

Runzi Chang

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

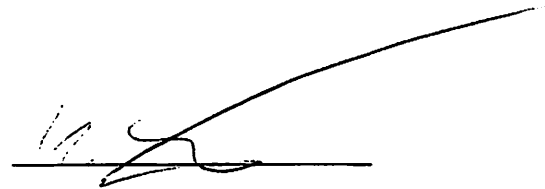
Professor Costas J. Spanos, Chair

As the semiconductor industry keeps its scaling efforts down to the sub-90nm technology node on the roadmap, the process and materials in the integration are being pushed to the limits. Integrated Circuits (IC) designers nowadays require more than 30% performance improvement from interconnects each technology generation. The introduction of copper and low-k materials in the last few years has been the key move toward that objective. Particularly Chemical-Mechanical Polishing (CMP) has been the enabling technique to planarize the metal surface and define metal layer thickness in contemporary copper Back-End-of-Line (BEOL) technology. However, CMP also introduces undesirable side-effects, including dielectric erosion and metal dishing, which degrade the process quality, cause significant yield losses in BEOL, and negatively affect interconnect performance.

The central theme of this thesis is the integrated metrology and modeling analysis in copper Chemical Mechanical Polishing (CMP) process towards optimizing interconnect and circuit performance. This work pursues special testing mask design and data analyzing techniques that are used to identify and model the sources of yield limiting factors in

copper CMP – namely oxide erosion and copper dishing. The main contributions of the thesis are the following: by applying special test structures design and data analysis, we developed and validated a model for copper dishing, and use that as the basis for process optimization and interconnect performance estimation. As the prerequisite and benefit of our effort, we applied the library-based scatterometry to monitor oxide CMP profile evolution; realized the model-based profile extraction using the e-test data in copper CMP. We set up the process optimization framework based on the contemporary models. Finally, we linked these CMP technology issues to circuit and interconnect design considerations through simulation work.

The validated dishing model integrated with other models and optimization frameworks serve the goal of design for manufacturability in the Back-End-of-the-Line Process. The process models and optimization framework developed in this thesis provide insight into the observables in the state-of-the-art CMP process. By transferring these principles into the realm of production, these building blocks can provide the opportunity for the process and integrated circuit designers to integrate and fuse information from both perspectives, thus improve the fabrication yield and circuit efficiency in the long term.

A handwritten signature in black ink, appearing to read 'C. Spanos', is written over a horizontal line. The signature is stylized and extends to the right beyond the end of the line.

Professor Costas J. Spanos
Committee Chairman

Table of Contents

Chapter 1. Introduction	1
1.1 Motivation	1
1.2 Thesis Organization	4
References	5
Chapter 2. Background	6
2.1 Background on Chemical Mechanical Planarization	6
2.2 Background on CMP Metrology	9
2.2.1 Spectroscopic Reflectometry	9
2.2.2 Spectroscopic Ellipsometry	10
2.2.3 Scanning Electron Microscopy (SEM)	12
2.2.4 Atomic Force Microscopy (AFM)	13
2.2.5 Scatterometry	14
2.2.6 Monitoring CMP Processes in a Production Environment	15
2.3 Background of CMP Process Modeling	17
2.3.1 Inter-Layer Dielectric (ILD) CMP Process Modeling	17
2.3.2 Copper CMP Process Modeling	22
References	24

Chapter 3. Integrated Characterization of Layout Dependency in Copper Damascene Process	27
3.1 Introduction	27
3.2 Test Pattern Design	30
3.3 Design of Experiments	35
3.4 Metrology	41
3.5 Copper CMP Process Modeling Using the Dishing Radius Concept	44
3.5.1 Erosion Modeling	45
3.5.2 Dishing Modeling	53
3.6 Model-Based 2D Profile Extraction	54
3.6.1 Key Ideas	55
3.6.2 Profile Modeling	56
3.6.3 Measurements and Validation.....	57
References	62
 Chapter 4. Model-Based CMP Process Optimization	65
4.1 CMP Process Performance Metrics	67
4.1.1 Material Removal Rate (MRR)	68
4.1.2 Selectivity	68
4.1.3 Inter-layer Dielectric (ILD) Erosion	70
4.1.4 Metal Dishing	70

4.2	Framework of Process Performance Optimization	71
4.2.1	Models Used in the Framework	71
4.2.1.1	MRR and Erosion Model	71
4.2.1.2	Metal Dishing Model	73
4.2.1.3	Selectivity Model	76
4.2.2	Weighting Coefficients and Optimization Discussion	77
4.3	Examples on Process Optimization	81
4.3.1	Linear Optimization	83
4.3.2	Application of Taguchi's Method	86
References	92

Chapter 5. Effects of CMP-Related Process Variations on

	Interconnect/Circuit Performance	96
5.1	Design for Manufacturability Overview	96
5.2	The Effect of Process Variations on Circuit Performance from CMP Process Technology Perspective	99
5.3	Model-Based Interconnect Performance Simulation Results and Discussion ...	106
References	118

Chapter 6. Conclusions and Future Work	123
6.1 Conclusions	123
6.2 Future Work	125
References	126

List of Figures

Figure 1.1	Typical chip cross section of hierarchical scaling of interconnects	2
Figure 2.1	A 90nm copper and low-k interconnect technology (source: NEC)	7
Figure 2.2	Typical components in a CMP system	8
Figure 2.3	Spectroscopic reflectometry measurements	10
Figure 2.4	Spectroscopic Ellipsometry Measurements	11
Figure 2.5	A cross-sectional view of a sample using SEM	12
Figure 2.6	A top-down view of an E-beam Lithography sample using SEM	13
Figure 2.7	MIT oxide CMP characterization mask set	20
Figure 2.8	Window used to calculate effective density	21
Figure 3.1	Complex copper interconnects fabricated with IBM's damascene process	28
Figure 3.2	Single (right side) and dual (left side) damascene process flow illustrations	29
Figure 3.3	Illustration of oxide erosion and copper dishing problems during copper damascene process	29
Figure 3.4	Cell design for the electrical characterization of dishing effect in copper damascene process	31
Figure 3.5	Schematic representation of the four point measurement method	32
Figure 3.6	Mask #1 layout	33
Figure 3.7	Mask #2 layout	34
Figure 3.8	Effective pattern densities of Mask #2 (defined by wider metal lines surrounding the E-test cells	34

Figure 3.9	Process flow chart of the single damascene fabrication	35
Figure 3.10	Surface of the perforated IC1000 pad	39
Figure 3.11	The standard IC1400 pad with K-groove	39
Figure 3.12	Comparison of the pre and post-CMP images evaluated by an optical microscope	42
Figure 3.13	Post-CMP images evaluated by an optical microscope for two different masks #1 (top picture) and #2 (bottom picture)	43
Figure 3.14	Calculation of the maximum deflection of the pad (at the center of the beam)	45
Figure 3.15	Topology scan of a commercially available polyurethane pad	47
Figure 3.16	Proposed wafer pad contact mechanism under typical Chemical Mechanical Polishing setup	47
Figure 3.17	Illustration of the Dishing Radius concept	49
Figure 3.18	Post-CMP cross-sectional pictures of copper lines (1.6 and 0.4 microns)	50
Figure 3.19	Surface profiling curves validate post-CMP surface dishing shape	51
Figure 3.20	Line profilometry scans of post-CMP features (Experiments No. 1 and 3)	52
Figure 3.21	The measured amount of oxide erosion as a function of copper pattern density on Mask #2	54
Figure 3.22	The difference between measured and expected R as a function of metal linewidth	55

Figure 3.23	The difference between measured and expected R as a function of metal linewidth normalized by R_expected	55
Figure 3.24	Profile modeling for post-CMP metal lines	57
Figure 3.25	Model parameters extracted from measurement	61
Figure 3.26	Residues plot after the extraction	61
Figure 4.1	Illustration of the basic components of the control chart using Oxide polish rate as an example: data points, center line, Upper Control Limit (UCL), Lower Control Limit (LCL) and Outlier	66
Figure 4.2	Illustration of the concept of selectivity – the scenario in plasma etching where high selectivity is desired	69
Figure 4.3	Illustration of the concept of selectivity – the scenario in the overpolishing step of the damascene process where low selectivity will be helpful in reducing metal dishing	69
Figure 4.4	During the overpolishing step in damascene process, the bulk ILD area has no erosion by definition	72
Figure 4.5	Illustration of the effective pattern density calculation for the point of interest (x,y) using a square window (size L) in damascene process	73
Figure 4.6	Cases when a three or more level becomes necessary: A two level study (the left) will give the illusion of no effect from the factor while a three level study (the right) will uncover the curvature effect	75
Figure 4.7	Flowchart of the CMP process optimization framework	80
Figure 4.8	The quadratic loss function	82
Figure 4.9	Data analysis using Taguchi’s quality metrics	90

Figure 5.1	A typical modern computer-aided integrated circuit design flow	97
Figure 5.2	Interconnect complexity increases in 0.25 μ m technology (right) compared with the 0.7 μ m technology (left)	100
Figure 5.3	Simulation results illustrate the increased crosstalk between neighboring wires as technologies shrink 0.7 μ m (left) to 0.25 μ m (right)	100
Figure 5.4	10% post-CMP ILD thickness variation observed	101
Figure 5.5	Pattern density difference on the mask between the logic and memory area	101
Figure 5.6	Systematic ILD thickness variations due to pattern density difference...	102
Figure 5.7	The truly random ILD thickness variations due to process uncertainty	102
Figure 5.8	128-stage inverter simulations with long interconnect	103
Figure 5.9	Simulation flowchart	103
Figure 5.10	Basic inverter layout (0.18 μ m TSMC technology)	104
Figure 5.11	Circuit delay for the ideal (no CMP process variation) case	105
Figure 5.12	Copper dishing on deep sub-micron metal lines (picture courtesy of Technical University of Dresden, Germany)	106
Figure 5.13	Overall process flow for the investigation of the metal dishing impact on interconnect performance	107
Figure 5.14	Interconnect structure used for the investigation of the metal dishing impact on performance	108
Figure 5.15	Metal dishing model used to generate the interconnect profiles in simulation	108

Figure 5.16	Dependence of metal line resistance on dishing radius as a function of metal line width	109
Figure 5.17	Dependence of metal line capacitance on dishing radius as a function of metal line width ($C_{total}=C_{ground}+2*C_{coupling}$)	110
Figure 5.18	RC delays as a function of line width with dishing (the optimal linewidth to achieve minimum RC delay is around 4 microns)	111
Figure 5.19	RC delay sensitivity (20% linewidth change) as a function of metal line width	112
Figure 5.20	Optimal linewidth as a function of dishing radius	113
Figure 5.21	Efficiency of process improvement for different metal thickness as a function of dishing radius	115
Figure 5.22	Illustration of the line-splitting idea ($W = W_{total}/N$, N is the number of lines; $s=s_{min}=0.5\text{micron}$)	116
Figure 5.23	RC delay gain and area penalty tradeoff	117
Figure 5.24	Optimization of RC delay using the line-splitting idea	117

List of Tables

Table 3.1	Conditions for Sputtering of Tantalum and Copper	36
Table 3.2	Typical Cu CMP process parameters	37
Table 3.3	Typical Cu CMP Conditioner settings (for step-1 and step-2)	38
Table 3.4	Input parameters, their variations, and coding	39
Table 3.5	Design of Experiments (DOE) for copper CMP processes	40
Table 3.6	Theoretical and measured copper line resistances	49
Table 3.7	Dishing values of 5 μ m wide lines for different experimental parameters	53
Table 4.1	Design of Experiments (DOE) in modeling the dishing effect in copper CMP process	74
Table 4.2	DOE measurement Results	75
Table 4.3	Copper CMP Process characterization Results	83
Table 4.4	Integrated CMP Optimization Results	85
Table 4.5	Experimental Design using Orthogonal Array L ₉ (3 ³)	87
Table 4.6	Control Factor Levels	88
Table 4.7	Experimental Results using Orthogonal Array L ₉ (3 ³)	89
Table 4.8	Optimization using the Taguchi's Method	91
Table 5.1	Circuit delays for three cases	105

Acknowledgements

First and foremost, I would like to express my deepest gratitude and appreciation to my research advisor, Professor Costas J. Spanos, the real professor in my life, for his superexcellent guidance and tremendous support, and for the opportunities he has created for me during my graduate study and research at Berkeley. His vision and leadership in the semiconductor manufacturing industry has been inspiring to both my research work and career development. I thank him for his consistent supervision and enlightenment in every detail of my research and education at Berkeley.

I would also like to thank Professor David Dornfeld and Professor Nathan Cheung for their insightful inputs to this work and for reviewing my dissertation. Their expertise in the field of semiconductor processing technology has provided me strong support throughout my graduate study. I thank Professor Kameshwar Poolla for teaching me interesting control theory and reviewing my master's thesis. I thank Professor Andrew Neureuther and Professor Borivoje Nikolic for serving as committee members in my preliminary and qualifying exams. I also would like to thank Professor Ronald Gutmann at Rensselaer Polytechnic Institute for his mentoring and support to some of the experimental work in this thesis.

Special thanks to the past and present members at the Berkeley Computer Aided Manufacturing (BCAM) group, Junwei Bao, Jason Cain, Mareike Claassen, Weng Loong Foong, Mason Freed, Paul Friedberg, Anna Ison, Nickhil Jakatdar, Michiel Kruger, Jae-

Wook Lee, John Musacchio, Xinhui Niu, Jiangxin Wang, Jing Xue, Haolin Zhang, Qiaolin Zhang, Dongwu Zhao, for their best friendship, help and support. I thank Dr. Frederick Dill from Hitachi Global Storage Technologies for his long-term mentoring and valuable brainstorming. I thank Professor Fiona Doyle at the Material Science department of UC Berkeley and Professor Jan Talbot from the Chemical Engineering program of UC San Diego for their long-term collaboration and mentoring to my research, which was part of the Small Feature Reproducibility (SFR) project led by Costas. Many thanks to the staff members of the Berkeley Microfabrication Laboratory, Katalin Voros, Sia Parsa, Yu Su, Kim Chan, Joseph Donnelly, for their support on some of the experimental work. I thank Dr. Anurag Jindal at the RPI Center for Integrated Electronics for his collaboration and discussions. I thank Dr. Bhanwar Singh at AMD, Professor Eray Aydil and Dr. Brian Thibeault at UC Santa Barbara for their help on accessing AFM and surface profiling tools. I also acknowledge the staff of the Berkeley EECS department and Electronics Research Laboratory, Ruth Gjerde, Dianne Chang, Vivian Kim, Linda Manly, Charlotte Jones, T.K. Chen, Tim Duncan, for their help and support.

I acknowledge many valuable discussions and collaborations with my former industry colleagues, Andreas Wiswesser, Lei Ping Lai, Quan Tran, Qing Ma, Tsung-Kuan (Allen) Chou, John Heck, Joseph Hayden, Dong Shim, Li-Peng Wang and Valluri Rao, for sharing their precious knowledge and experience with me during my internships at Applied Materials and Intel Corporation.

I deeply thank all my Chinese and international friends from different parts of the small world. They made my life in the United States colorful and enjoyable.

I would like to thank my mom and dad and other dear members in my big family for their love during my years in graduate school. Their care always provides the warmest support to my life and work, wherever I am.

Most importantly, I would like to thank my dear wife Jinghua, for her companion and love during the last a few years. Together we have managed to get lots of meaningful things done and overcome many difficulties. I thank her for her love, understanding and consistent support. Without her love and encouragement, this thesis wouldn't be possible. I look forward to enjoying a better and better life with her and our incoming children.

This work was funded by the State of California SMART program under research contract SM97-01, and by the following participating companies: Advanced Energy, ASML, Atmel Corp., Advanced Micro Devices, Applied Materials, Asyst Technologies Inc., BOC Edwards, Cymer, Etec Systems Inc., Intel Corporation, KLA-TENCOR, Lam Research Corp., Nanometrics, Inc, Nikon Research Corp., Novellus Systems Inc., Silicon Valley Group, Texas Instruments Inc., and Tokyo Electron America.

Chapter 1 Introduction

1.1 Motivation

The semiconductor industry is experiencing an unprecedented growth over the last 40 years. As the technology advances in deep-submicron process geometries, enabling companies to build smaller, faster and less-expensive transistors, interconnect delay has moved to the forefront as the limiting factor in IC performance, replacing a longtime concern with switching speeds. Today there are cases where interconnect delay accounts for more than 50% percent of total path delay [1-3]. Figure 1.1 illustrates the complexity of the interconnect layers for a state-of-the-art chip. Based on these observations, optimizing the interconnect performance from the processing technology's perspective becomes an indispensable component of the global efforts of advancing the Moore's Law even further.

Chemical mechanical polishing (CMP) is currently being used in the fabrication of state of the art integrated circuits, and has been identified as an enabling technology for the semiconductor industry in its drive toward multi-gigabit chips and sub-90nm feature sizes. At the present time, it appears that the global planarization necessary for establishing reliable multilevel copper interconnects can only be achieved by using CMP. As with many processes that stand in the critical path of IC development, this technology

has moved into production without the benefit of integrated and optimized models. In the long run, the availability of such models and optimization frameworks will help optimize the operation of CMP and permit the users to define the best operating conditions for each specific application.

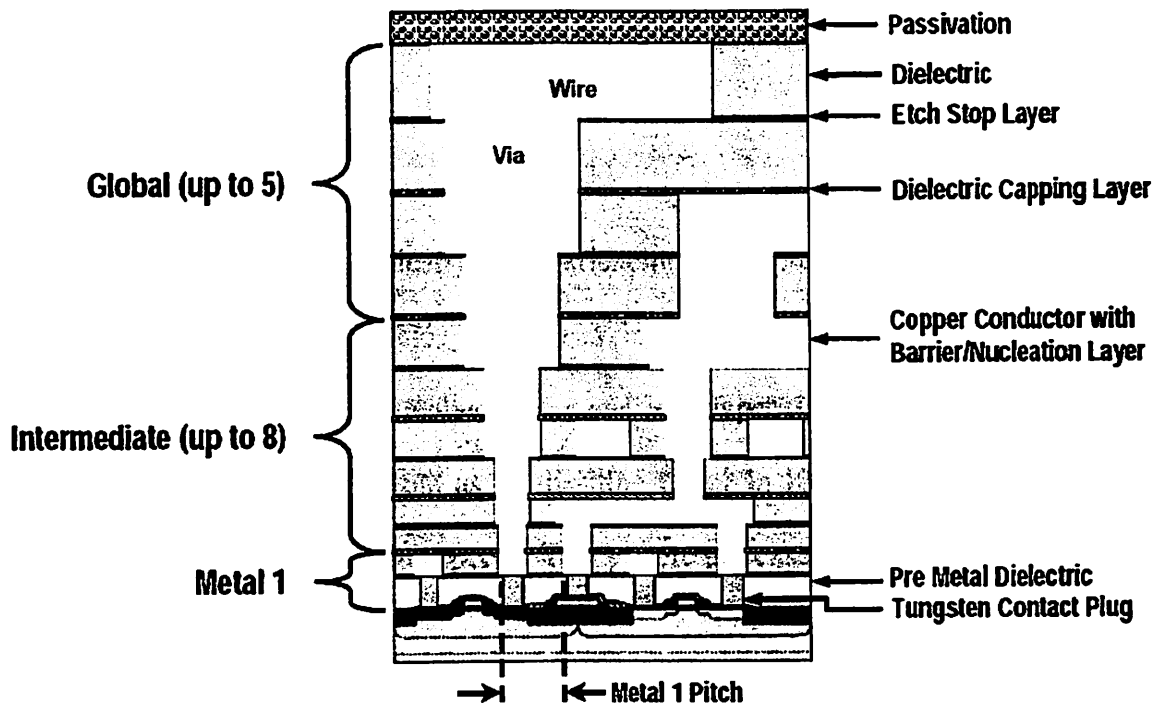


Figure 1.1 Typical chip cross section of hierarchical scaling of interconnect

This work focuses on the development of a fundamental understanding of the systematic variations mechanism during copper CMP, in both the feature-scale and wafer-scale level. This understanding is supported by careful experimental design and implementation. In the long term, this research will facilitate CMP technology development and Integrated Circuit designers in at least the following aspects:

- (1) It will aid in identifying the dominant relationships between the material and process parameters and the effectiveness (measured by quality and integrity of the

finished surface) and efficiency (measured by material removal rate) of a CMP process. Such a model will facilitate process optimization.

(2) In addition to providing a fundamental understanding of the CMP process, the value of the proposed experimental and modeling efforts lies in its enhanced capability for exploration of the "design space." Currently, many process design options (e.g., optimum selection of slurry and pad properties) remain trial-and-error propositions due to lack of reliable models depicting those effects. The proposed model will aid in identifying such unexplored process parameters and guide us toward new and novel avenues for designing CMP processes.

(3) To the circuit designers, these models will provide a realistic estimation of the interconnect performance, given the proper process input parameters and mask layout patterns. The motivation for this can be explained by the challenges of obtaining high yields at the 90nm node, concerns about reliability and potentially vastly differing yields for designs of similar size that are all design rule compliant for a particular process. The contributions from this work will be more precise than the traditional worst-case estimation.

To achieve the above objectives, we utilize a combination of experimental and analytical methods. The first phase of this work is to investigate the feasibility of developing a non-destructive, low-cost metrology method to monitor the profile evolution during the ILD CMP process. The second phase is to develop oxide erosion and copper dishing models by using specially designed masks and extensively explore the information contained in the collected data. We find that it is possible to extract the 2D profile for the post-CMP metal wires. This is noteworthy in that it provides the possibility

to develop inexpensive metrology techniques in copper CMP. We also develop the framework for the multiple objectives optimization in the CMP environment. Finally we link the findings from the experiments and analytical work to interconnect performance and explore the possibility of improving BEOL fabrication yield through decisions taken at both the process and design stages.

1.2 Thesis organization

This thesis presents an integrated framework to fuse the aspects of metrology, modeling and optimization for the state-of-the-art copper CMP process. We begin with Chapter 2 reviewing the modern metrology and modeling work in copper damascene process. Chemical mechanical polishing has shifted from “black art” to an “engineering science” by the continuing efforts from both academia and industry. This chapter will review the general trends in this field in recent years.

Chapter 3 focuses on the experimental work that we performed in order to model the origins of layout-dependent non-uniformity in a classical copper damascene process. Started from basic ideas applied in the mask design, this chapter elaborates the design of experiments, basic metrology tools used and the development of an empirical model for copper dishing. The chapter finishes with a demonstration of a model-based 2D profile extraction method, which has the potential of being utilized as non-destructive and fast metrology for BEOL process further down the roadmap.

Chapter 4 presents a framework for efficient multiple objective optimization in a copper damascene process. Processing engineers can only devote limited resources to take measurements, calibrate the tool and make decisions. This framework can thus

reduce the reliance on experimentation to some extent. After discussions on the process performance metrics, the chapter sets up the performance optimization framework based on classical multi-objective optimization theory. Experimental results from Chapter 3 are used to calibrate the model predictions from the optimizer.

Chapter 5 presents the effects of CMP-related process variations on interconnect and circuit performance, by performing extensive simulations and analysis. The simulation results show that dishing will be a concern for global layer interconnects if the dishing radius is less than 50 μ m. This chapter closes with a discussion on one application of the Design for Manufacturability (DFM) concept -- the tradeoffs between the die area, BEOL yield and interconnect performance.

Chapter 6 provides some concluding remarks for this thesis and future work in the areas of copper CMP process metrology, modeling, optimization and BEOL Design for Manufacturability.

References:

- [1] International Technology Roadmap for Semiconductors, International SEMATECH, Austin, TX, 2003.
- [2] J. Rabaey, A. Chandrakasan and B. Nikolic, "Digital Integrated Circuits: A Design Perspective", 2nd Edition, Prentice Hall 2003.
- [3] Z. Lin, C. Spanos, L. Milor, and Y.-T. Lin, "Circuit Sensitivity to Interconnect Variation," IEEE Transactions on Semiconductor Manufacturing, Vol. 11, No. 4, pp. 557-568, Nov. 1998.

Chapter 2 Background

The function of interconnect is to distribute clock and other signals, and to provide power/ground, to the various subcomponents of an Integrated Circuit. The fundamental development requirement for interconnect is to meet the high-speed transmission needs of chips despite further scaling of feature sizes. In the context of manufacturing these interconnects with a reasonably high yield, Chemical Mechanical Planarization (CMP) is used in-between critical patterning and deposition steps. CMP has proven to be indispensable, and it has been recognized as the enabling technology for Copper and Low-k based Back-End-of-the-Line (BEOL) processes.

2.1 Background of Chemical Mechanical Planarization

Rapid device scaling has been the factor governing the growth of the semiconductor industry, which has produced devices with ever-better performance characteristics in terms of high speed and low power. The semiconductor industry has been experiencing an average growth rate of 15% annually over the past four decades [1]. With this fast growth has come a reduction in average cycle time between introductions of new technologies from the traditional 3-year cycle towards an approximate 2-year cycle. Current (2004) production is being done employing 90nm technology, with 65nm technology expected to be introduced in the year 2005. Figure 2.1 shows a chip cross-

section using 90nm copper interconnect technology. Advances in integrated circuit (IC) manufacturing have also increased device density to about one billion transistors per cm^2 of chip area (for memory). The corresponding increase in circuit functionality requires many layers of metal interconnect to facilitate device and module communication. Up to ten levels of metal interconnect have been reported to date. It is note worthy that as late as 2001, industry experts thought that this number would not be achievable until the year of 2011 [2]. Now the experts predict 14 levels of metal interconnect by 2011. The ability to effectively and efficiently planarize the metal layers and the dielectric layers which are used to insulate these complex interconnect levels is indispensable for the realization of these ambitious objectives.

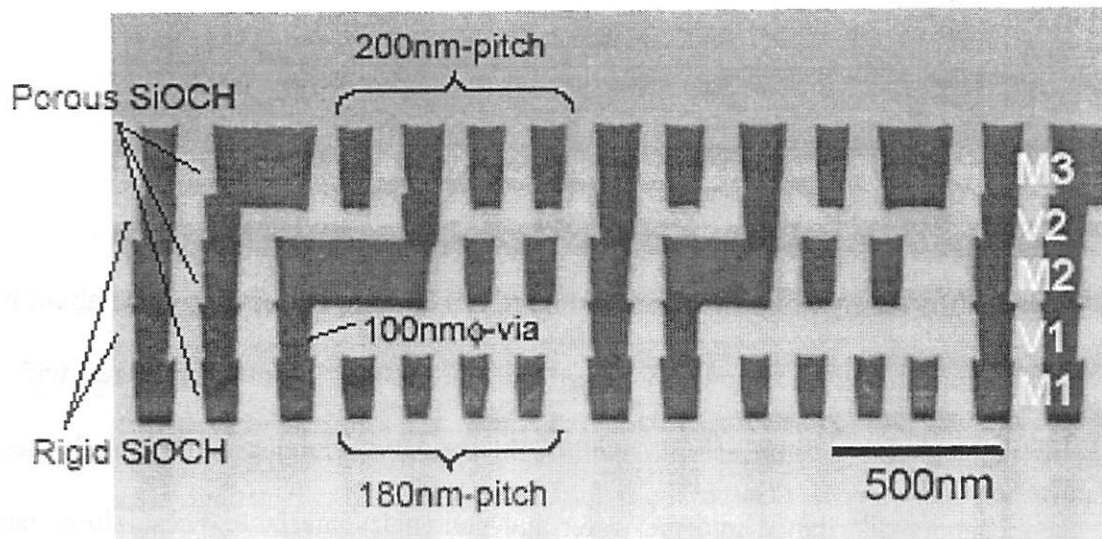


Figure 2.1 A 90nm copper and low-k interconnect technology (source: NEC)

While critical dimensions (CD) continue to shrink, modern photolithography tools continue to reduce their depth of focus. Especially for the future projected Extreme-Ultra-Violet (EUV) lithography where the wavelength is only about 13nm, the across-field depth of focus requirement will be at the nanometer level. Slight irregularities on the wafer surface—or on deposited films—can distort semiconductor patterns as they are

transferred by a lithographic process to the wafer surface. Chemical mechanical planarization has become the process of choice for preventing distortion, and it works by planarizing the wafer surface to a flat, uniform finish. To planarize the wafer, CMP systems use abrasive particles suspended in chemical slurry. Figure 2.2 shows the components in a simple rotary CMP system.

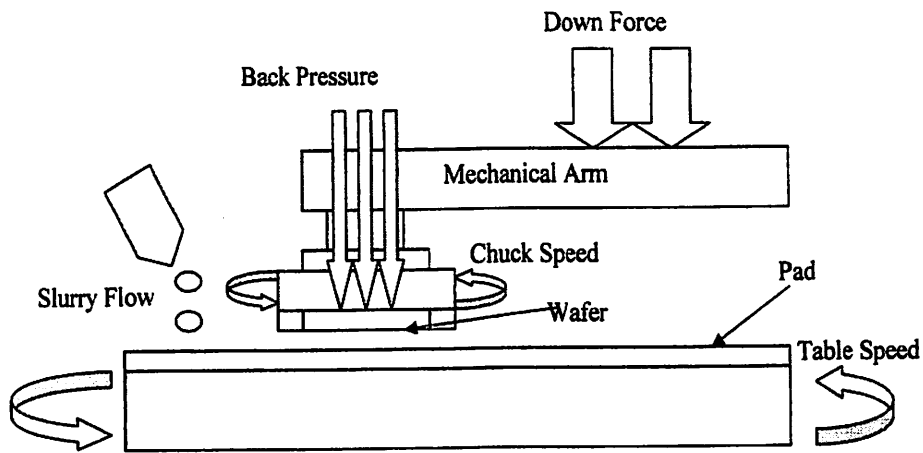


Figure 2.2 Typical components in a CMP system

In the configuration depicted in Figure 2.2, a silicon wafer is rotated about its axis while being pressed face-down by a carrier against a rotating platen covered with a soft polymer pad (belt). Slurries with nano-scale abrasive particles and specially designed chemicals are distributed into the wafer and pad (belt) interface. These slurry particles and chemicals work together with the pressures applied on the backside of the wafer and the relative movement between the wafer and the pad to remove some of the wafer materials and planarize its surface.

CMP is a technology that emerged during the last decade. It has become one of the most widely used planarization techniques in inter-level dielectric (ILD) planarization, shallow trench isolation (STI), and metal damascene processes. Because of

its high throughput and wide applicability, CMP is quickly replacing the traditional in-situ etch-back techniques and becomes one of the key fabrication processes in the manufacturing of advanced IC chips.

2.2 Background on CMP Metrology

Although CMP is a versatile process, it is often quite difficult to maintain without actively compensating for key parameter changes over time. Because of this, there is a strong interest to maintain good visibility of its progress, and this is typically done by measuring the status of the thin films and the evolving wafer topography before, during and after each CMP step. In this section, we introduce several types of metrology tools that are relevant to CMP and are widely used in either research or the industry.

2.2.1 Spectroscopic Reflectometry

This is a method used to measure the thickness and the optical properties of a thin, transparent film, such as oxides, nitrides, polysilicon, etc. In spectroscopic reflectometry, the reflected light intensities are measured in a broadband wavelength range. In most setups, non-polarized light is used at normal incidence. The biggest advantage of spectroscopic reflectometry is its simplicity and low cost. Figure 2.3 shows the setup of conventional spectroscopic reflectometry.

In reflectometry, only light intensities are measured, so only the amplitude of the complex *reflection coefficient* is of interest. $R = |r|^2$.

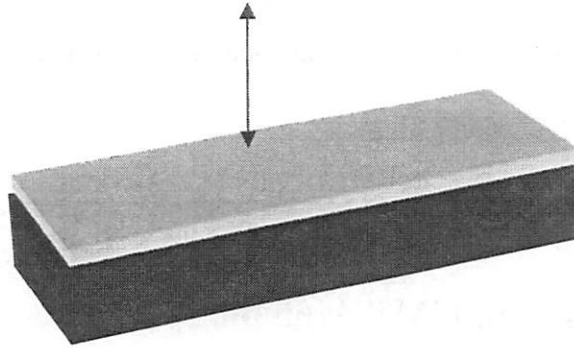


Figure 2.3 Spectroscopic reflectometry measurements.

This method is simple and easy for implementation. Its main application will be introduced in section 2.2.5 where we will also discuss its advantages and limitations.

2.2.2 Spectroscopic Ellipsometry

This is a method commonly used to analyze the optical properties and the thickness of one or multiple transparent films. This method is based on the characteristics of light upon reflection from multiple surfaces. The component waves of light, which are linearly polarized with the electric field parallel (p or TM) or perpendicular (s or TE) to the plane of incidence, behave differently upon reflection. The component waves experience different amplitude attenuations as well as different absolute phase shifts upon reflection; as such, the overall state of polarization changes. Ellipsometry refers to the measurement of the state of polarization before and after reflection for the purpose of determining the properties of the reflecting boundary. The measurement is usually expressed in the form

$$\rho = \tan\Psi e^{j\Delta} = \frac{r_p}{r_s}, \quad (2.1)$$

where r_p and r_s are the complex reflection coefficients for TM and TE waves respectively.

Ellipsometry derives its increased sensitivity over reflectometry from the fact that the polarization-altering properties of the reflecting boundary are modified significantly even when ultra-thin films are present. Further, unlike reflectometry, ellipsometry can derive both the thickness and the refractive index of the target film. Spectroscopic ellipsometry can do so over multiple wavelengths. Consequently, ellipsometry has become the rigorous method of characterizing thin films. An illustration of the basics of ellipsometry is presented in Figure 2.4 below:

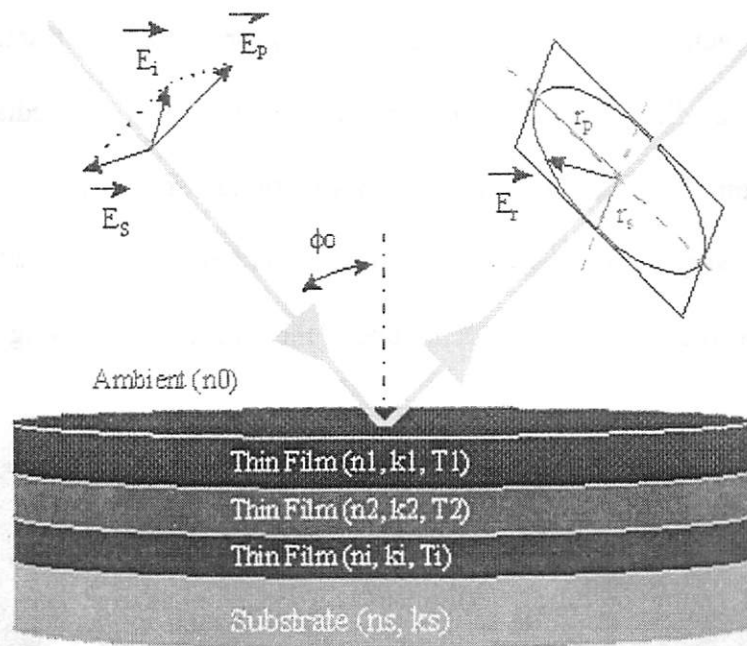


Figure 2.4 Spectroscopic Ellipsometry Measurements

The advantage of ellipsometry over reflectometry is better accuracy. As mentioned before, ellipsometry measures the polarization state of light by looking at the ratio of values rather than the absolute intensity of the reflected light. This property is especially useful in the DUV wavelength range, where very little light is typically available. Additionally, ellipsometry can gather the phase information in addition to plain

magnitude reflectivity information. Phase information provides more sensitivity to thin-film variations.

2.2.3 Scanning Electron Microscopy (SEM) [21]

The scanning electron microscope (SEM) is currently the main method used in production for measuring lateral sub-micron features because of its nanometer-scale resolution, precision as well as its relatively high throughput. SEM comes in many flavors, cross-sectional and top-down being the most common.

Cross-sectional SEM can provide profile information for structures on a wafer in the form of a direct image. This image can be used immediately for process characterization. However, obtaining a cross-sectional SEM image requires breaking a wafer and is very time-consuming; also there is the possibility of the presence of systematic profile errors dependent upon the image processing technique being employed.

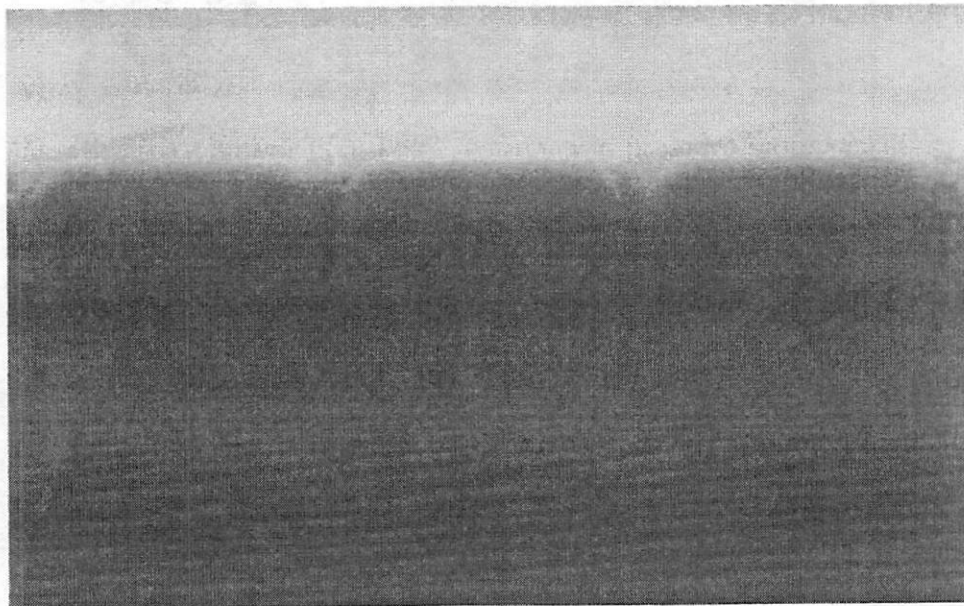


Figure 2.5 A cross-sectional view of a sample using SEM

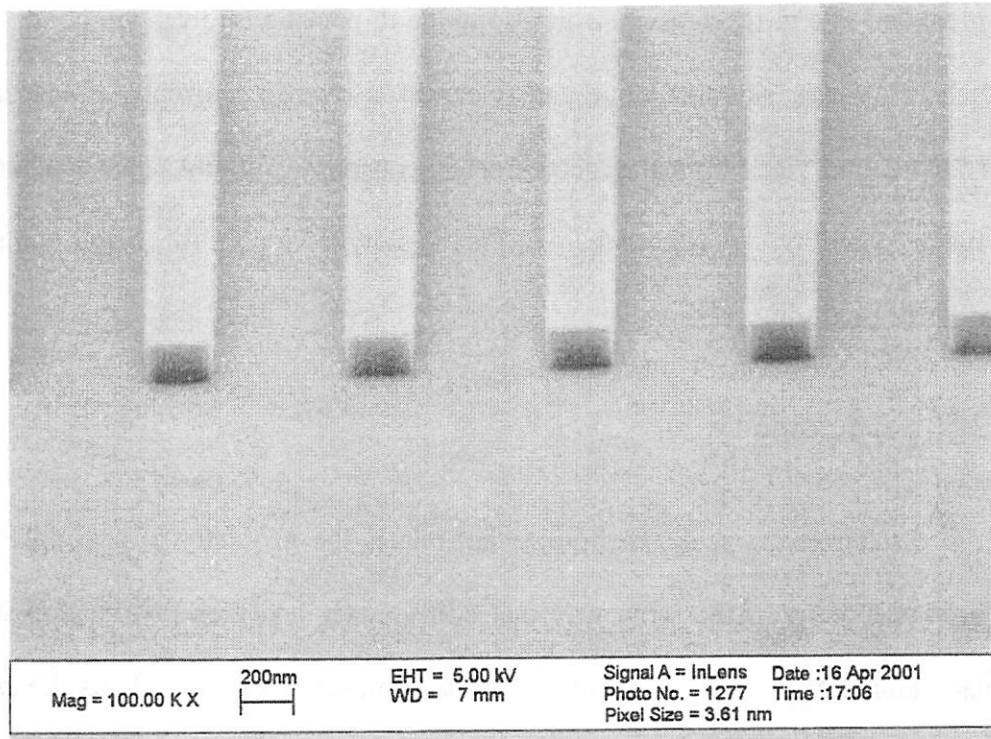


Figure 2.6 A top-down view of a E-beam Lithography sample using SEM

The top-down SEM, more commonly referred to as the CD-SEM, measures the CD of a profile at a somewhat arbitrary height and does not take into account the slope associated with the profile that results in a constantly changing profile CD. Another problem associated with this method is the build-up of charge in the sample under the electron beam. The CD-SEM, being a surface scanning technique, is also unable to provide information on underlying layers of the structure as well as undercut features. The state-of-the-art SEMs can measure sub-100nm CDs with a precision of about 2nm.

2.2.4 Atomic Force Microscopy (AFM) [22]

Atomic force microscopy (AFM) is a means for measuring nm-scale profiles with a resolution between 0.1nm and 5nm, depending on the hardness of the material being scanned. For typical semiconductor profiles, this translates into exceptionally high

vertical and lateral resolutions, which combine to provide information about a patterned structure's width, sidewall slope and thickness. However, current AFM scan rates are very slow, and measurement accuracy and precision are highly dependent upon the tip shape and stability. At present, the AFM is too slow to be used for imaging to support the CMP process during production

2.2.5 Scatterometry

Scatterometry is the metrology that relates the geometry of a sample to its light scattering effects. In the same way that ellipsometry analyzes polarization-state-in and polarization-state-out of light incident on thin film, scatterometry adopts the same theory and measures the polarization-state-in and polarization-state-out of light, incident not on a blanket thin film, but rather on periodic surface structures. The $\tan \Psi$ and $\cos \Delta$ values are measured after reflection and matched to the responses of known profiles.

Much work has been done on scatterometry in recent years. McNeil et al have explored the idea of variable-angle scatterometry, which uses angle-resolved diffracted light analysis to measure etched samples with line width dimensions as small as 150nm and poly-Si thickness on the order of 250nm [3].

Niu et al. have explored the idea of spectroscopic scatterometry, in which the responses for multiple wavelengths are taken into account [4]. This method consists of measurements taken at a fixed incident angle as opposed to variable-angle scatterometry. The lack of external moving parts, and hence simpler implementation, gives spectroscopic scatterometry the potential of being employed in in-line, in-situ process control. Niu et al also developed a simulation engine, known as the gtk (Grating Tool

Kit) based on well known method of Rigorous Couple-Wave Analysis [5, 6]. It was demonstrated that the simulated and measured diffracted light responses based on this technique correspond to profiles that closely match those obtained through the AFM method.

2.2.6 Monitoring CMP Processes in a Production Environment

Accurate thickness measurements in chemical mechanical polishing (CMP) are difficult to achieve consistently on complex state-of-the-art computer chips, so frequent monitoring is needed. This requires high-speed film thickness measurement of both thick films and ultra-thin films at multiple sites across the wafer surface. In order to make film removal uniformity measurements on product wafers, the metrology tool must have a small measurement spot and fast, robust pattern recognition to reliably guide the measurement spot to the film thickness test structures.

Spectroscopic reflectometers have been used to measure thick films because they can determine film thickness very rapidly. They also can operate with measurement spots as small as $\sim 50\mu\text{m}$ in diameter. However, reflectometer accuracy degrades significantly when measuring films thinner than 500-1,000 Å. The ability to measure very thin films is important in determining if over-polishing has occurred. On the other hand, measurement accuracy and reliability for thick films are challenging qualities in a production CMP metrology system. Inaccurate measurements resulting from the phenomenon known as "order skipping" for instance, lead to mis-processing and costly scrapping of damaged wafers.

Scatterometry is one of the few metrology candidates that have true *in-situ* potential for deep sub-micron oxide CMP profile analysis. In previous work, we have demonstrated the possibility of using a library-based scatterometry method to match the closest profile in oxide CMP [6]. In that project, the specular spectroscopic scatterometry is designed to measure the 0th diffraction order at a fixed angle of incidence and multiple wavelengths. The term “spectroscopic” means that multiple wavelengths are used simultaneously. Due to its fixed angle, specular spectroscopic scatterometry is easy to deploy. Specular spectroscopic scatterometry can make use of a conventional spectroscopic ellipsometer, and can be installed *in-line* or *in-situ*. Conventional spectroscopic ellipsometry equipment can be directly used in this type of metrology.

Specular spectroscopic scatterometry, when implemented with a library of generated profiles, is about 100 times faster than the SEM and the speed advantage is even more significant when compared with the AFM. It is non-destructive, inexpensive and easy to implement *in-line* in a production CMP system.

Metrology methods which can overcome the slow throughput and meet the requirements of state-of-the-art CMP process-control are needed. For research purposes, metrology methods which can monitor the profile evolution would be crucial for the development and verification of a rigorous, first principle model for the state-of-the-art CMP process for both oxide and copper.

2.3 Background on CMP Process Modeling

The main feature of CMP, namely the removal of the material, is described by the Preston's Equation [7]:

$$\frac{dT}{dt} = K \cdot \frac{N}{A} \cdot \frac{ds}{dt} \quad (2.2)$$

where T denotes the thickness of the wafer, N/A denotes the pressure caused by the normal force N on the area A. s is the total distance traveled by the wafer, and t denotes the elapsed time. This means that the material removal rate is proportional to the pressure and the velocity of the rotation. Any physical considerations are put into the Preston's Constant K, which is often considered the proportionality constant (i.e. independent of pressure and velocity), but may also contain chemical effects. However, the chemical reaction effects seem to dominate in real-life situations.

2.3.1 Inter-Layer Dielectric (ILD) CMP Process Modeling

Cook's model [8] is applicable to CMP for bare silicon wafers. However, many ideas are also applicable to the more general case of CMP for ILD. Cook starts from Preston's Equation. The slurry is assumed to be a viscous Newtonian fluid with a viscosity of around 10^9 P with particles in it. The mechanical part of the interaction between polishing particles and the wafer surface can be described by a model with a spherical particle of diameter Φ , which penetrates the surface with force F_s under the uniform load N. For a standard Hertzian penetration Preston's constant becomes $(2 \cdot E)^{-1}$, where the E denotes Young's modulus. The surface roughness is the penetration depth given by

$$R_s = \frac{3}{4} \cdot \Phi \cdot \left(\frac{P}{2 \cdot k \cdot E} \right)^{2/3}, \quad (2.3)$$

where k is the particle concentration (unity for a fully-filled closed packing) and $P = \frac{N}{A}$ is the pressure.

Impingement of particles carried in the turbulent liquid leads to Hertzian penetration of the surface, converting kinetic energy into strain energy. Local bonding during contact leads to weakening of binding forces at the surface, which allows atomic removal to occur without introducing lattice dislocations.

An extensive study of the chemical part is presented in Cook's paper. However, a discussion of it is beyond the scope of this thesis.

Cook's model is the most general modeling work for polishing so far. In particular, it deals with the mechanics of the polishing particles and with the chemical reactions. It covers almost all interesting topics and the method is explained by an example (SiO_2 polished by SiO_2 -particles). This model is based on a smaller feature length compared to the Preston's model, as it deals with the particles and the particle size in the slurry fluid. In the future, additional work is necessary to combine Cook's model with other models in order to get a sufficiently general model for the entire CMP process.

J. Luo and D.A. Dornfeld have explored the material removal mechanism in chemical mechanical polishing recently [9]. Based on the assumptions of plastic contact over wafer-abrasive and pad-abrasive interfaces, the normal distribution of abrasive size and the periodic roughness of the pad surface, a novel model were developed for predicting the material removal in CMP. The basic model predicts the material removal

rate (MRR): $MRR = \rho_w N \cdot Vol_{removed}$, where ρ_w is the density of the wafer, N the number of active abrasives and $Vol_{removed}$ is the volume removed by a single abrasive.

Compared with previous modeling approaches, such as Preston's equation, the model proposed by Luo and Dornfeld integrates not only the process parameters of pressure and velocity, but also other input parameters including the wafer hardness, pad hardness, pad roughness, abrasive size and abrasive geometry into the same formulation to predict the material removal rate. A link between the chemical and the mechanical effect has been captured through a fitting parameter in the model. It reflects the influence of chemicals on the mechanical material removal. The fluid effect in the current model is attributed to the number of active abrasives. The nonlinear down pressure dependence of material removal rate is related to a probability density function of the abrasive size and the elastic deformation of the pad. Compared with experimental results, the proposed model predicts the material removal rate fairly accurately.

At this point it is appropriate to mention the MIT work on pattern dependency effects [10-11]. Several models had been proposed to account for pattern effects in CMP before the MIT Model but their applicability had been limited [12]. The limitations range from being based on non-representative test structures to probing of small process windows which limit the utility of the models beyond the scope of the original experimental conditions. Most of the models before the MIT model, however, did not apply across a whole die but rather focused on individual features.

As a prelude to effective modeling of CMP for oxide planarization, the MIT metrology group performed polishing experiments for a wide range of die topography patterns. A set of four masks shown in Figure 2.7 was used to generate the die patterns.

Mask I explores the effects of area and consists of blocks of sizes ranging from $20\mu\text{m}$ to $3000\mu\text{m}$. It also contains blocks which mimic realistic circuit layouts. Mask II examines the effect of pitch. The pattern density – defined as the ratio of line width to pitch – is maintained at 50% while the pitch is varied from $2\mu\text{m}$ to $1000\mu\text{m}$ in the $2\text{mm} \times 2\text{mm}$ blocks. Mask III explores the effect of density which is increased from 4% to 100% in steps of 4%. Pitch is maintained at $250\mu\text{m}$ in each of the 25 $2\text{mm} \times 2\text{mm}$ blocks. Mask IV explores the effects of block perimeter. It consists of blocks of constant area ($1\text{mm} \times 1\text{mm}$) but with different perimeter/area ratios. The mask is divided into six sections and the spaces between the blocks are decreased from $60\mu\text{m}$ at the bottom to $10\mu\text{m}$ at the top.

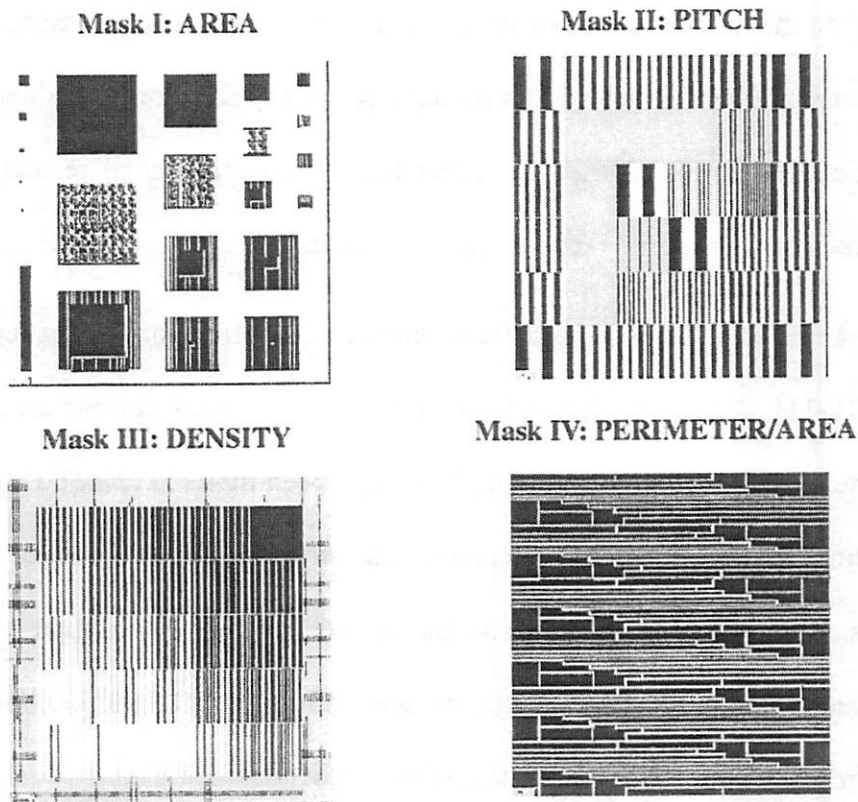


Figure 2.7 MIT oxide CMP characterization mask set

These masks were used in a single-mask fabrication process to generate surface topographies on 6" wafers to be planarized using CMP. The fabrication process consisted

of 1000nm LPCVD TEOS deposition, metal deposition, and pattern and etch followed by 2000nm TEOS deposition.

The experimental results led to two important conclusions which are becoming the basis for the later MIT CMP model: (1) the pitch (line width and line space), area and perimeter are all minor effects to the final oxide thickness; (2) effective density is the key layout parameter. That is, the oxide-polishing rate at each point is inversely proportional to the effective pattern density. The effective pattern density depends on the nearby topography and density. A certain window, whose side is so called the “planarization length”, can determine the pattern density. Please see Figure 2.8 for the illustration.

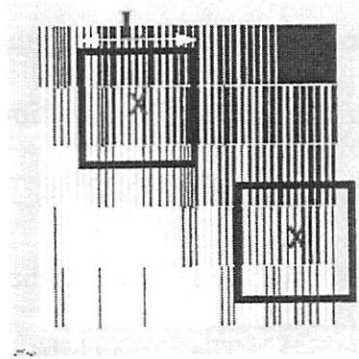


Figure 2.8 Window used to calculate effective density

Planarization length is the approximate length of the “ramp” joining areas of different removal rates, as determined by locally different pattern densities. However, the planarization length must be characterized for a given process. In figure 2.8, the effective density at X for a square constant weight window can be calculated as the following:

$$\text{effective - pattern - density} = \frac{\text{Raised - Area - in - the - window}}{\text{Total - Area - in - the - window}} \quad (2.4)$$

The long range “moving average” density calculation corresponds to a simple convolution picture:

$$d(x, y) = p(x, y) * l(x, y), \quad (2.5)$$

where $d(x, y)$ is the effective pattern density at (x, y) , $p(x, y)$ is the “planarization impulse response” (weighting function) to raised features, $l(x, y)$ is the local (feature scale) density.

2.3.2 Copper CMP Process Modeling

The mechanism of CMP of metal is less understood and more complex than that of oxide polishing. It has been conjectured that a metal polishing model should employ both chemical etching and a passivation mechanisms [13-14]. For metal CMP, the polishing slurry must contain three important constituents: the fine slurry particles, a corrosion (etching) agent, and an oxidant. Planarization is achieved by the mechanical rigidity of the polishing pad similar to oxide polishing.

Copper CMP processes have been studied at the MIT statistical metrology group led by Prof. Duane Boning [15-18]. The research has been helpful in establishing the pattern density and planarization length concept. They have been focusing on multi-layer interconnect stacks, chip or wafer level variations in copper CMP. One aspect that requires significant improvement however is the understanding of the feature-level, multiple process and layout input parameters interactions in damascene process.

Recently, Lakshminarayanan et al. proposed design rule modifications in order to improve the manufacturability by a reduction of the within-die resistance/thickness variation based on experimental data [19]. Smith et al reported the effect of barrier layer and dishing in copper interconnects using a fine Greek cross test structure [20]. Both

projects see the opportunity of improving manufacturability in the backend by a better understanding of the damascene process.

Copper metallization, which has replaced traditional aluminum technology for the state-of-the-art IC's, is expected to have ten levels of metal with as small as 50 Å of copper thickness loss for minimum feature arrays and less than 150 Å of wide copper line loss by year 2010. This is a challenging future requirement for the current chemical mechanical polishing process, especially when current planarization technology often exceeds 1000 Å of copper loss in dense regions. The constraint on copper planarization is made more difficult by the incoming integration difficulties coming from ultra-low-k materials. The integration will impact CMP performance directly by causing varying removal rates of copper on different pattern regions.

As we addressed before, while dielectric erosion has been extensively studied, practical and quantitative understanding of the dishing effect is still at the developing stage. The international technology roadmap for semiconductors calls for a full CMP model in the near future with 10% topography accuracy of specification limits. Thus, it is critical to have a systematic methodology for the characterization and modeling of pattern dependent issues and problems in copper CMP process. This characterization will enable an optimization framework to handle the CMP performance metrics in a balanced way. These are the objectives of this thesis.

References:

- [1] "International Technology Roadmap for Semiconductors", 2003 Edition, Semiconductor Industry Association (SIA).
- [2] <http://www.semiconductor.net/semiconductor/reference/reference.asp> "Editorial Archives", May 2001.
- [3] C. J. Raymond, S. S. H. Naqvi, J. R. McNeil, "Scatterometry for CD measurements of etched structures," Proceedings of SPIE, vol. 2725, 720-728, March 1996.
- [4] X. Niu, N. Jakatdar, J. Bao, C. Spanos, S. Yedur, "Specular spectroscopic scatterometry in DUV lithography", Proceedings of the SPIE - The International Society for Optical Engineering, vol.3677, pt.1-2, (Metrology, Inspection, and Process Control for Microlithography XIII, Santa Clara, CA, USA, 15-18 March 1999.
- [5] X. Niu, "An Integrated System of Optical Metrology for Deep Sub-Micron Lithography", Ph.D Dissertation, U.C. Berkeley, 1999.
- [6] R. Chang, "Full Profile Oxide CMP Metrology", Master's Thesis, University of California at Berkeley, May 2001.
- [7] F.W. Preston, "The Theory and Design of Plate Glass Polishing Machine," Journal of the Society of Glass Technology, Vol. 11, pp. 214-256, 1927.
- [8] L.M. Cook, "Chemical Processes in Glass Polishing", J. of Non-Crystalline Solids, vol. 120, pp152-171, 1990
- [9] J. Luo and D.A. Dornfeld, "Material Removal Mechanism in Chemical Mechanical Polishing: Theory and Modeling", IEEE Transactions on Semiconductor Manufacturing, vol. 14, no. 2, May 2001.

- [10] D. O. Ouma, B. Stine, R. Divecha, D. Boning, J. Chung, I. Ali, and M. Islamraja,,
"Using Variation Decomposition Analysis to Determine the Effect of Process on Wafer
and Die-Level Uniformities in CMP," First International Symposium on Chemical
Mechanical Planarization (CMP) in IC Device Manufacturing, Vol. 96-22, pp. 164-175,
190th Electrochemical Society Meeting, San Antonio, TX, Oct. 6-11, 1996.
- [11] Divecha, R., B. Stine, D. Ouma, J. Yoon, D. Boning, J. Chung, O.S. Nakagawa, S-Y
Oh, "Effect of Fine-Line Density and Pitch on Interconnect ILD Thickness Variation in
Oxide CMP Processes," 1997 Chemical Mechanical Polish for ULSI Multilevel
Interconnection Conference (CMP-MIC), p. 29, Santa Clara, February, 1997.
- [12] G. Nanz and L. E. Camilletti, "Modeling of Chemical-Mechanical Polishing: A
Review," IEEE Trans. Semiconductor Manufacturing, Vol. 8, No. 4, November 1995.
- [13] J.-Q. Lu, Y. Kwon, G. Rajagopalan, M. Gupta, J. McMahon, K.-W. Lee, R.P. Kraft,
J.F. McDonald, T.S. Cale, R.J. Gutmann, B. Xu, E. Eisenbraun, J. Castracane, and A.
Kaloyeros, "A Wafer-Scale 3D IC Technology Platform Using Dielectric Bonding Glues
and Copper Damascene Patterned Inter-Wafer Interconnects," in Proceedings of the 2002
IEEE International Interconnect Technology Conference (IITC), 78-80, San Francisco,
2002.
- [14] S.P. Murarka, I.V. Verner, and R.J. Gutmann, Copper Fundamentals for
Microelectronic Applications, John Wiley & Sons Inc., New York, 1997.
- [15] T. Park, T. Tugbawa, J. Yoon, D. Boning, J. Chung, R. Muralidhar, S. Hymes, Y.
Got-kis, S. Alamgir, R. Walesa, L. Shumway, G. Wu, F. Zhang, R. Kistler, and J.
Hawkins, "Pattern and Process Dependencies in Copper Damascene Chemical

Mechanical Polishing Processes,” Proc. VLSI Multilevel Interconnect Conference, Santa Clara, CA, June 1998.

[16] T. Tugbawa, T. Park, D. Boning, T. Pan, P. Li, S. Hymes, T. Brown, and L. Camilletti, “A Mathematical Model of Pattern Dependencies in Copper CMP Processes,” Proc. CMP Symposium, Electrochemical Society Meeting, Honolulu, HI, pp. 605-615, Oct. 1999.

[17] T. Park, T. Tugbawa, D. Boning, S. Hymes, T. Brown, K. Smekalin, and G. Schwartz, “Multi-Level Pattern Effects in Copper CMP,” Proc. CMP Symposium, Electrochemical Society Meeting, Honolulu, HI, pp. 94-100, Oct. 1999.

[18] T. Tugbawa, T. Park, and D. Boning, “Framework for Modeling of Pattern Dependencies in Multi-Step Cu CMP Processes,” SEMICON West, July 2000.

[19] S. Lakshminarayanan et al, "Electrical characterization of the copper CMP process and derivation of metal layout rules", IEEE Transactions on Semiconductor Manufacturing, vol. 16, no. 4, pp. 668-676, Nov. 2003.

[20] S. Smith, et al., “Evaluation of sheet resistance and electrical line width measurement techniques for copper damascene interconnect,” IEEE Transactions on Semiconductor Manufacturing, vol. 15, no. 2, pp. 214-222, May 2002.

[21] Banerjee, I.; Tracy, B.; Davies, P.; McDonald, B., "Use of advanced analytical techniques for VLSI failure analysis," Reliability Physics Symposium, pp. 61-68, 1990

[22] Neubauer, G.; Dass, M.L.A.; Johnson, T.J., "Imaging VLSI cross sections by atomic force microscopy," Reliability Physics Symposium International, pp. 299-303, April 1992

Chapter 3 Integrated Characterization of Layout Dependency in Copper Damascene Process

3.1 Introduction

For the sub-90nm CMOS nodes, Chemical-Mechanical Polishing (CMP) is widely used as the primary technique to planarize the Inter-Layer Dielectric (ILD) and metal surface. Since the introduction of copper metallization in 1998 by IBM (Figure 3.1), CMP has been the enabling technology for the copper damascene process with the high metal removal rate that is necessary in this kind of trench-first integration [1]. The damascene process derives its name from the ancient metal decorating art of the Middle East involving inlaying metal in ceramic or wood for decoration. Historically, the art of damascene was practiced for centuries by Egyptians, Greeks, and Romans. The modern damascene process was found to be a viable process to support the idea of using copper to replace aluminum around the 0.35 micron technology node. As shown in Figure 3.2, after the via plug process, the inter-level dielectric (ILD) is deposited without planarization, since the surface is already flat. Trenches for metal lines are then defined,

etched in the ILD, and filled with a metal such as copper. The excess metal on the surface is removed and a planar structure with metal inlays in the dielectric is achieved [2]. The damascene process eliminates the difficulty in filling small gaps between metal wires as well as in metal etching, especially for Cu and other hard-to-etch metals. A dual damascene process is demonstrated in the left side of Figure 3.2. In this process, vias and trenches are defined using two lithographic and RIE steps, but the via plug is filled in the same step as the metal line. Dual damascene minimizes the number of processing steps by reducing the barrier layer depositions from two to one and by eliminating the CVD W plug processes.

By replacing aluminum with copper, the metal wire resistance is reduced by about 30% and its electro-migration resistance also improves dramatically. Figure 3.2 shows the process flow for a single (the right side) and dual (the left side) damascene process, respectively.

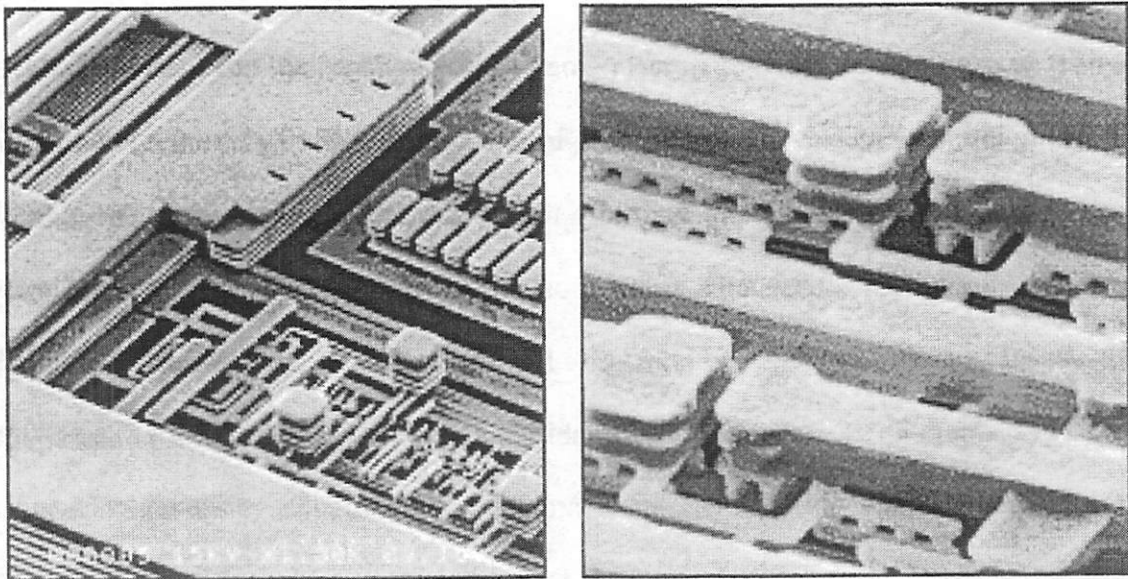


Figure 3.1 Complex copper interconnects fabricated with IBM's damascene process

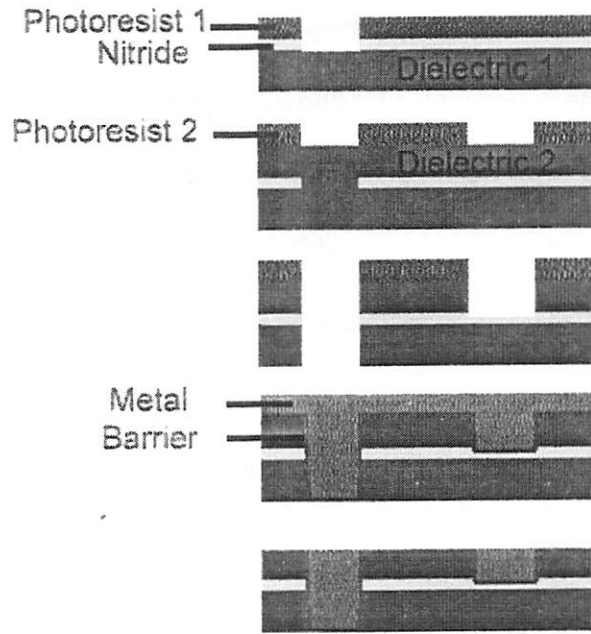


Figure 3.2 Single (right side) and dual (left side) damascene process flow illustrations

However, the CMP damascene process also introduces undesirable side-effects, including dielectric erosion and metal dishing. Fig. 3.3 illustrates their influences on metal line cross-sections after CMP. Both effects originate from the material property differences between dielectrics and metal under chemical and mechanical stresses. Both erosion and dishing degrade the process quality, cause significant yield losses in the Back-End-Of-the-Line (BEOL), and negatively impact interconnect performance, especially for very wide global interconnects and metal layers that have a wide distribution of pattern densities [3-5].

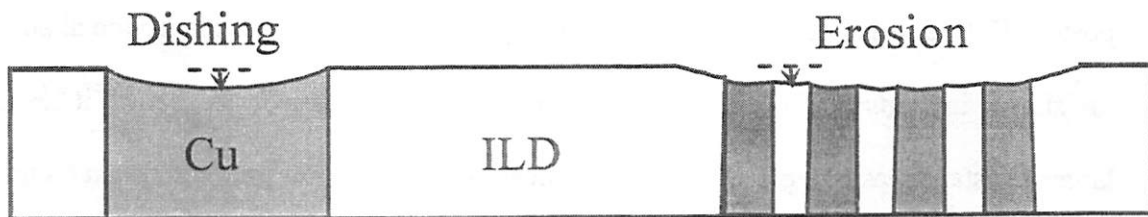


Figure 3.3 Illustration of oxide erosion and copper dishing problems during copper damascene process

In this chapter, we design and measure test patterns to investigate the correlations between the side effects existing in damascene process (e.g. oxide erosion and copper dishing) and the layout characteristics. We develop a model that well captures the correlation between the metal linewidth and the amount of dishing, given input parameters such as speed, pressure, chemicals and pad information. After model validation, we utilize the electrically tested conductance data to extract the useful parameters, such as oxide erosion and dishing radius.

3.2 Test Pattern Design

Characterization of the pattern dependent non-uniformities in copper damascene process is needed to understand the fundamental limitations of each process and assist in new process development efforts. As we discussed in Chapter 2, some work has already been done in this area [6-8]. In this work, we focused out study on one topic that people have limited knowledge on – metal dishing. Based on that, we integrate the effects of oxide erosion and copper dishing and attempt to present a general picture of the CMP performance in correlation with the special layout on the mask and process input parameters.

In our experiments, the impact of metal dishing is characterized by measuring post-CMP line resistance (R). Since the dishing effect causes a non-planar metal surface (as shown in Figure 3.3), and since it reduces the conductive cross-section it leads to larger resistance, as compared to the theoretical value of R for a line with planar surface. Therefore, in the design of test structures, the key consideration is its suitability for electrical testing of R (E-test). Furthermore, since metal thickness loss caused by dishing

exhibits strong correlation to metal width (e.g., wider lines suffers dishing more severely), the design of test cells particularly focuses on the relationship between the amount of dishing and line width (w).

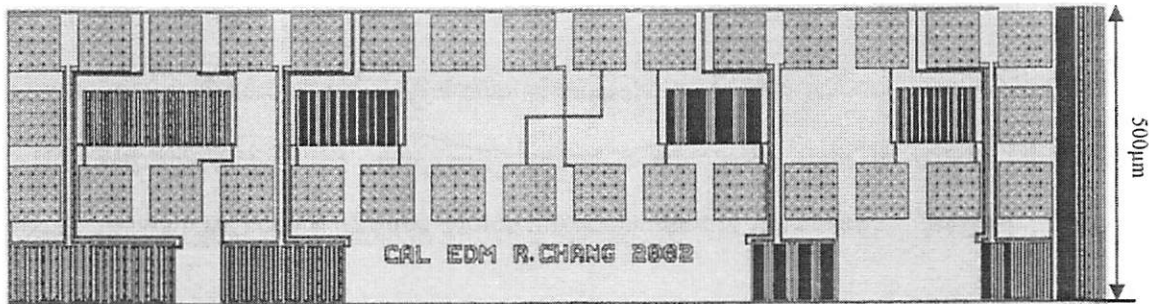


Figure 3.4 Cell design for the electrical characterization of dishing effect in copper damascene process

Figure 3.4 shows the layout of a test cell. Each cell contains eight serpentine-shaped metal lines that vary in line width ($0.4\mu\text{m}$, $0.8\mu\text{m}$, $1.2\mu\text{m}$, $1.6\mu\text{m}$, $2.0\mu\text{m}$, $3.0\mu\text{m}$, $4.0\mu\text{m}$, and $5\mu\text{m}$); the line length is 4mm ; and the target line thickness is $0.4\mu\text{m}$. These dimensions are chosen from typical global on-chip interconnect parameters. Depending on the linewidth, the estimated line resistance is within the range of 30Ω to 650Ω , so that R can be easily extracted by an automatic impedance test. For each line, there are two pads at each end that are used as probe contacts during E-test. Four-point measurement, which applies a known current through the two outer pads and measures voltage difference between the two inner pads, is employed to measure the resistances of the post-CMP lines (Figure 3.5). This enables reliable measurements of the resistance for the straight line between the two middle pads without suffering the parasitic noise coming from contact resistance variation.

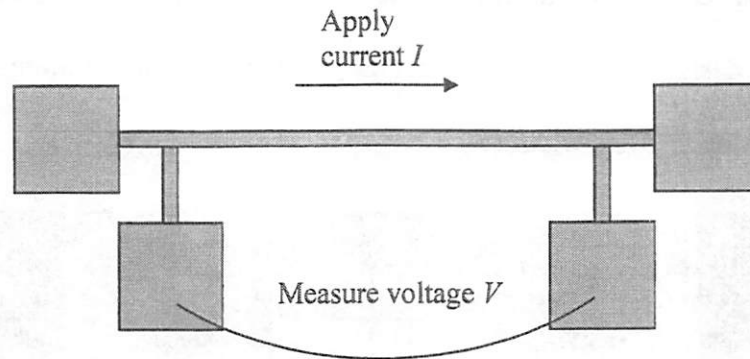


Figure 3.5 Schematic representation of the four-point measurement method

On the right side of the E-test cell, multiple copies of the eight long lines are laid out for the scanning electron microscopy (SEM) test, which can provide the cross-sectional view of metal lines after CMP. Overall, an E-test cell has a dimension of $1900\mu\text{m}$ by $525\mu\text{m}$. In order to decouple dishing effect from erosion, pattern densities are approximately uniform across the cell, so that metal thickness loss caused by erosion is similar to different test lines. Moreover, the cell footprint is small but arguably comparable to the characteristic length of ILD erosion during Cu Damascene polishing. To ensure that oxide erosion was about the same throughout the various electrical patterns, we established a wide area around the electrical patterns, where the pattern density was carefully controlled by means of a grating. In addition, we measured the post-CMP oxide erosion and found it to be constant within each group of electrical patterns. Thus, for test lines with different widths on the same cell, their differences in the increment of R after CMP are mainly caused by dishing. We expect at a large multi-millimeter die or wafer scale, the amount of oxide erosion is a strong function of the effective pattern density [9-10].

Based on the considerations above, we designed two masks. Mask #1 (as shown in Figure 3.6) has 7 by 7 arrays of the E-test cells. In this mask, the effective pattern density for each cell is approximately constant. Differences in the percentage resistance change will only come from dishing effect. Mask #2 (as shown in Figure 3.7), on the other hand, has 3 by 3 structures with different pattern densities set by the line widths and the interspaces.

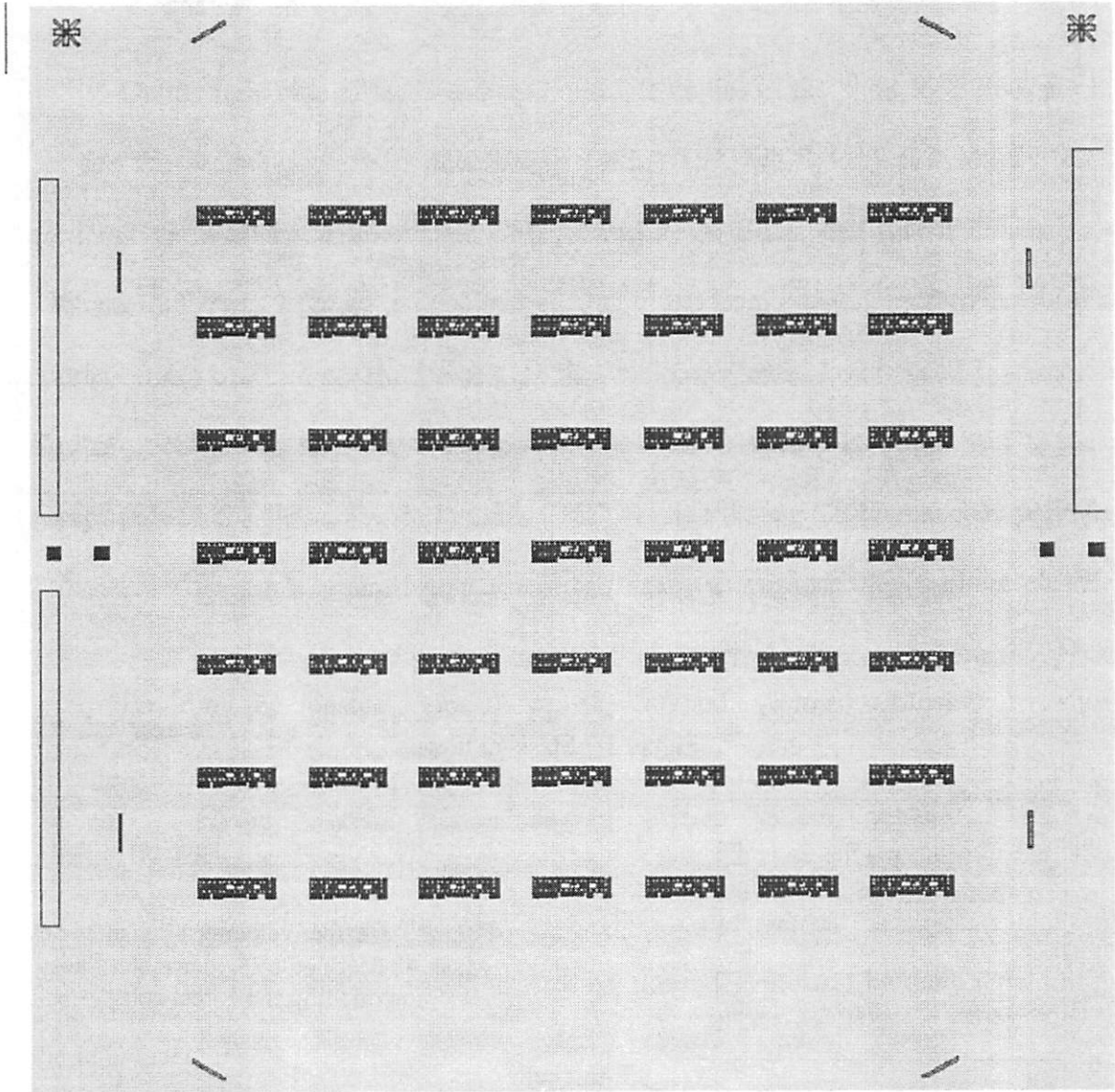


Figure 3.6 Mask #1 layout

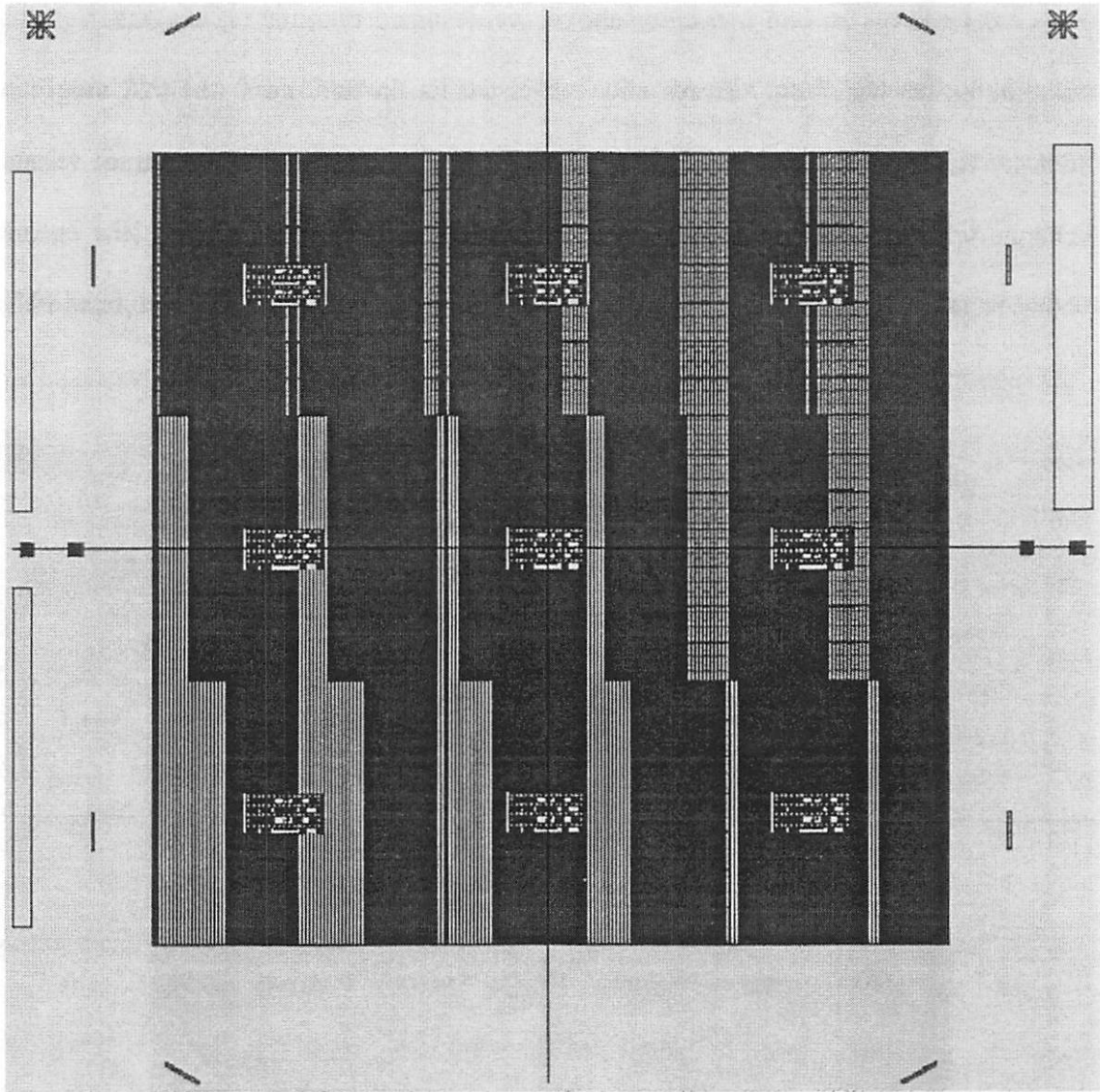


Figure 3.7 Mask #2 layout

□ 55%	□ 60%	□ 70%
□ 40%	□ 45%	□ 65%
□ 30%	□ 35%	□ 50%

E-test cell
 Pattern density

Figure 3.8 Effective pattern densities of Mask #2 (defined by wider metal lines surrounding the E-test cells)

On Mask #2, the pattern densities range from 30% to 70% with a 5% step. At the center of each structure, however, the pattern density remains constant as defined by the uniformly distributed wide metal lines (>20 microns). Two basic E-test cells were placed at the center of each of the structures. So the differences in the resulted sheet resistance change (after polishing) will come from both the erosion and the dishing effects. The equivalent pattern density of mask #2 is illustrated in Figure 3.8.

The pre-CMP test wafers using these masks are fabricated at the Berkeley and the RPI Micro-fabrication laboratories as we will present in the next section.

3.3 Design of Experiments

After the masks were ready, we started the preparation for CMP with 150 mm (6") prime silicon wafers. The process flow is schematically shown in Figure 3.9.

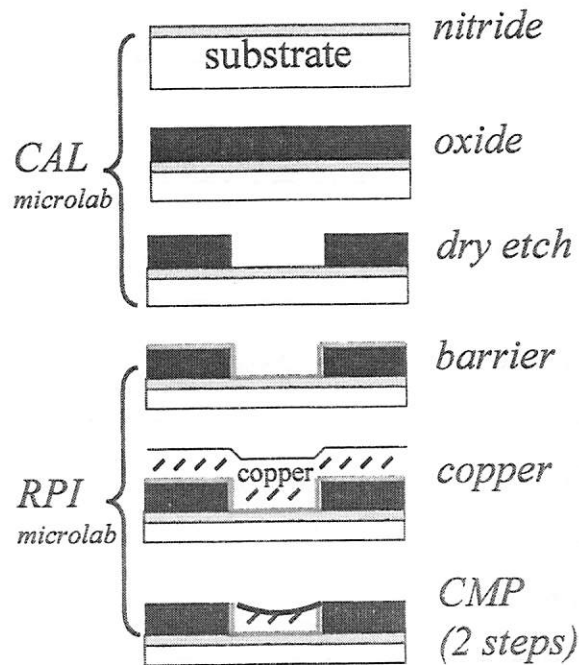


Figure 3.9 Process flow chart of the single damascene fabrication

As shown in the Figure 6, nitride and silicon dioxide chemical vapor deposition (CVD) was done in the micro-fabrication lab at University of California Berkeley. Photolithography was also performed at Berkeley using the ASML DUV Stepper Model 5500/90 which is capable of printing 0.35 μ m features. In the next step, dry oxide etching was done using the single wafer silicon oxide plasma etcher. Subsequently, the wafers were transported to the Micro-fabrication Clean Room (MCR) at the Center for Integrated Electronics, Rensselaer Polytechnic Institute (RPI). After proper process qualification, a thin barrier/adhesion promoter layer of Ta (~20nm) and bulk copper (~1.5 μ m) layer were deposited using CVC magnetron sputtering tool at RPI MCR. Sputtering conditions for the Ta and Cu layers are as given in Table 3.1.

Table 3.1 Conditions for Sputtering of Tantalum and Copper

Metal	Tantalum	Copper
Power (kW)	1.95	2.3
Base Pressure (Torr)	1×10^{-6}	1×10^{-6}
Operating Pressure (Torr)	5×10^{-3}	5×10^{-3}
Plasma Gas	Ar	Ar
Sputtering Time (minutes)	0.5	15
Sputtering Rate (nm/min)	4	100

The output of the process flow was 18 wafers (with 9 of each utilizing mask 1 and mask 2, respectively) available for the copper damascene CMP process.

The wafers thus prepared were polished on an IPEC-372M rotary polisher. A typical baseline process for step-1 and step-2 Cu damascene polishing is tabulated in Tables 3.2 - 3.3:

Table 3.2 Typical Cu CMP process parameters

Parameters	Step-1 Settings	Step-2 Settings
Down pressure	5 psi	5 psi
Back pressure	0.5 psi	0.5 psi
Platen speed	90 rpm	75 rpm
Carrier speed	90 rpm	75 rpm
Slurry	Commercial Step-1 slurry	Commercial non-selective step-2 slurry
Slurry flow rate	180 ml/min	180 ml/min
Pad type	Rodel IC-1400 with k grooves	Rodel IC-1400 with k grooves

Table 3.3 Typical Cu CMP Conditioner settings (for step-1 and step-2)

Parameters	Settings
Conditioner	automated 9" diamond grit
Conditioning pressure	0.1 psi
Conditioning Time	30 seconds
Conditioning sequence	ex-situ before polishing

The effect of four input parameters, namely, the pad, the polish pressure during step-2 polishing, the step-2 slurry, and the damascene patterns, on damascene CMP was studied in this work. Standard parameters as listed above were used during step-1 polishing; whereas these parameters were varied during step-2 polishing. One exception, however, is that the same kind of pad was used in step-1 as required for step-2 polishing. The variation in the parameters and their respective coding are listed in Table 3.4.

The pictures of the perforated IC1000 pad and the standard IC1400 K-groove pad are shown in Figure 3.10 and Figure 3.11, respectively.

Table 3.4 Input parameters, their variations, and coding

Parameters	Description	Code
Pad	Standard IC1400 with K-groove	A
	IC1000 perforated	B
Pressure/Back Pressure during Step-2 polishing	Standard 5/0.5,2.5/0.5psi	A
	Lower 4/0.5,2/0.5psi	B
Slurry	Step -1, followed by selective step-2	A
	Step -1, followed by non-selective step-2	B
Pattern	Uniform pattern density	A
	30%-70% pattern density with 5% step	B

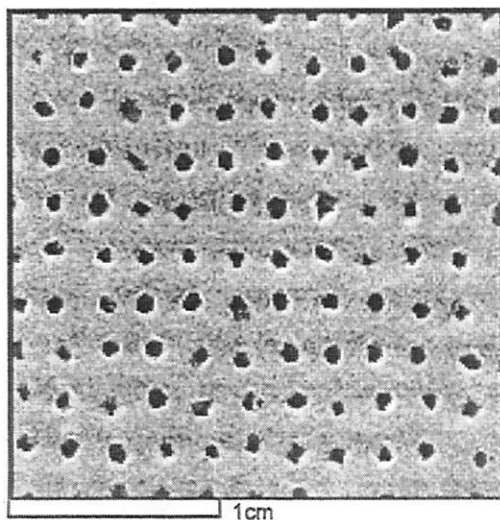


Figure 3.10 Surface of the perforated IC1000 pad

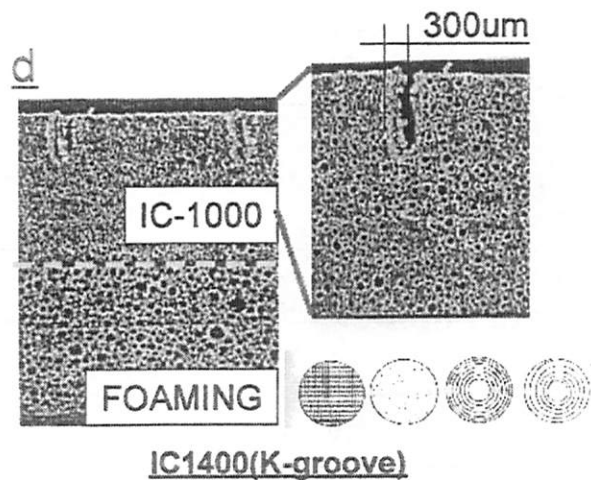


Figure 3.11 The standard IC1400 pad with K-groove

The design of experiment for exploring the effects of four different input parameters on dishing is shown in Table 3.5. Sixteen experiments (and the same numbers of total wafers) were needed for a complete permutation of the input parameters.

Table 3.5 Design of Experiments (DOE) for copper CMP processes

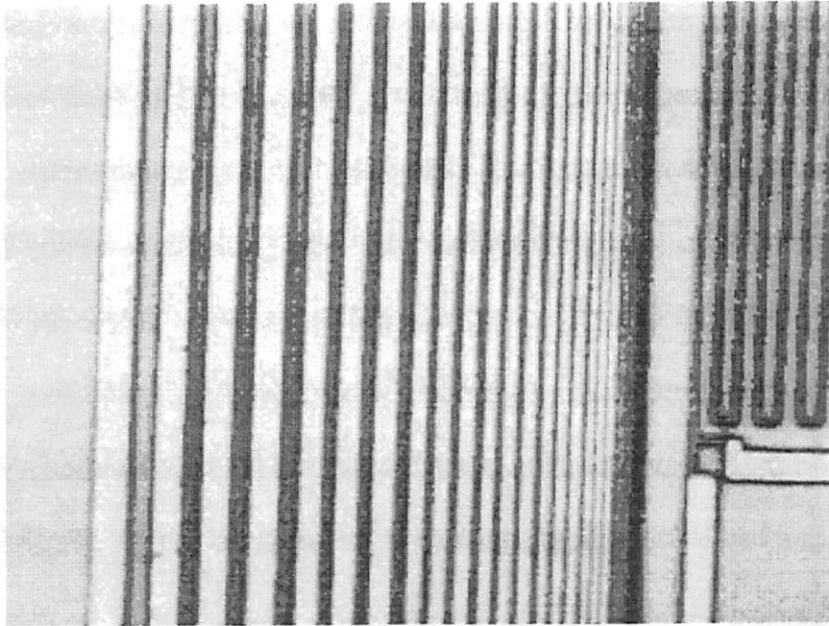
Experiment No.	Pad	Pressure	Slurry	Pattern
1	A	A	A	A
2	A	A	A	B
3	A	A	B	A
4	A	A	B	B
5	A	B	A	A
6	A	B	A	B
7	A	B	B	A
8	A	B	B	B
9	B	A	A	A
10	B	A	A	B
11	B	A	B	A
12	B	A	B	B
13	B	B	A	A
14	B	B	A	B
15	B	B	B	A
16	B	B	B	B

The polishing was performed in steps of 1 min. and 0.5 min. during step-1 and step-2 polishing, respectively. After each polish run, the wafers were cleaned with DI water using an OnTrak DSS200 double sided brush cleaner. These wafers were scanned using a stylus profilometer and looked under an optical microscope to detect the end point. Typically, step-1 polishing required a total of 3 minutes of polishing on IC-1400 with k groove pad and 4 minutes of polishing on IC-1000 perforated pad. Step-2 required 1 to 2 min. of polishing on IC-1400 with K-groove pad and 1.5 to 3 min. of polishing on IC-1000 perforated pad depending upon the polish parameters. It is clear that IC-1000 perforated pad had lower removal rate and required longer polish time both in step-1 and step-2 polishing.

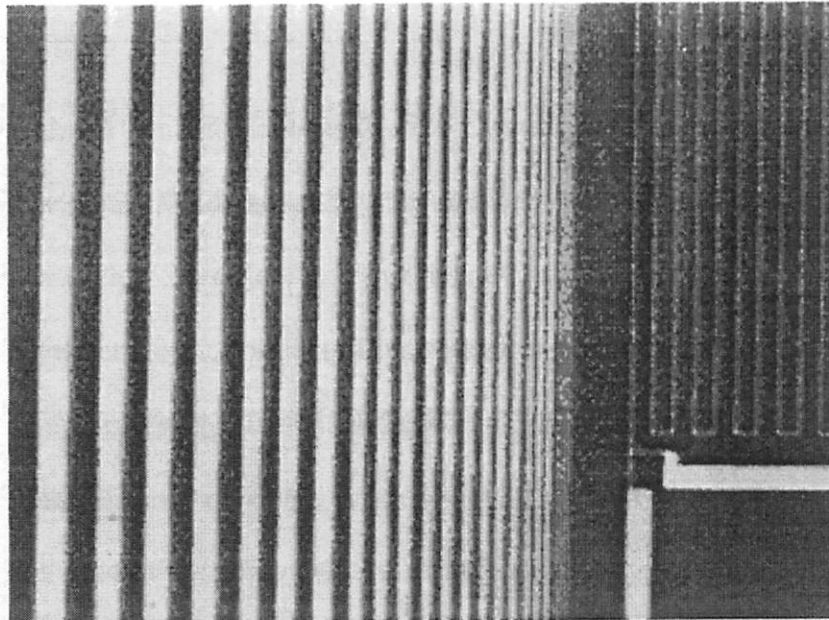
3.4 Metrology

As patterned wafers deposited with copper were polished using particular processes, the next step should be taking measurements on the finished structures and get the useful information in order to carry out successful process modeling. As the technology scaling proceeds, the cost of metrology during semiconductor manufacturing becomes more and more expensive. Having the ability to measure what's happening at the feature level is essential to quality control, yield improvement and process modeling.

The first kind of metrology that we applied was an investigation on the surface quality of the films after post-CMP cleaning evaluated by an optical microscope. A comparison of the pre and post-CMP images is shown in Figure 3.12. Please note the clear (blue) area in the post-CMP image was oxide and the opaque (yellow) area was copper metal line. These pictures were taken at the RPI Center for Integrated Electronics.



a. pre-CMP cell picture



b. post-CMP cell picture

Figure 3.12 Comparison of the pre and post-CMP images evaluated by an optical microscope

Figure 3.13 shows the post-CMP pictures for two different patterns (Mask #1 and Mask #2). The pictures indicate that the copper was cleared on the oxide surface after the two-step chemical mechanical polishing process.

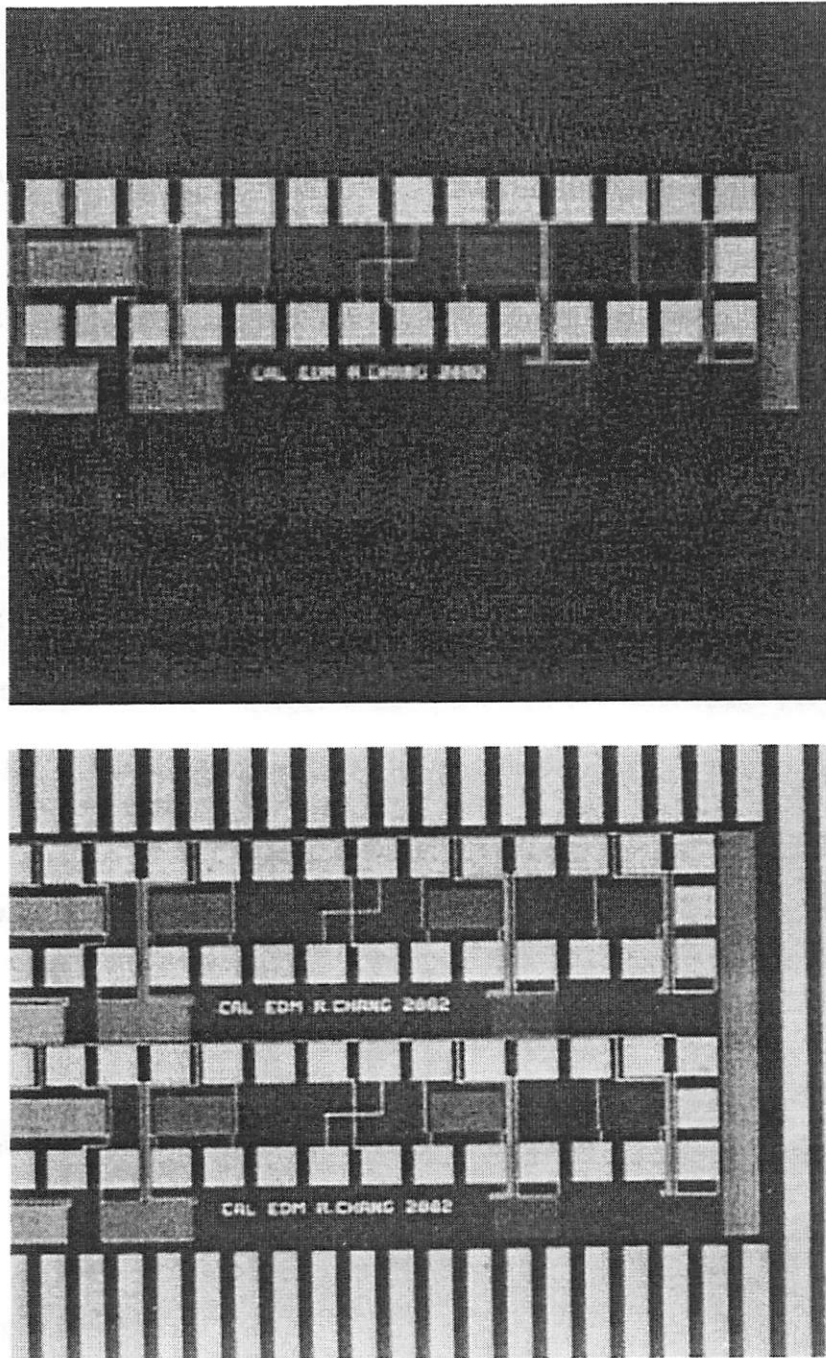


Figure 3.13 Post-CMP images evaluated by an optical microscope for two different masks #1 (top picture) and #2 (bottom picture)

These post-CMP wafers were then transferred to the UC Santa Barbara Nanofabrication Lab for some line profilometry measurements using their Dektak IIA tool.

In addition to these surface measurements, extensive data acquisitions were done at the Berkeley Microfabrication Lab: optical field area measurements of oxide thickness were done to obtain the absolute remaining film thicknesses at various locations around the serpentine shaped lines. Also, electrical measurements were performed to measure metal line resistance, which were later converted to oxide erosion and copper dishing information (refer to section 3.6) for comparisons with the surface profile data and optical dielectric film thickness. This electrical data analysis was particularly useful for multi-level CMP characterization where surface profile scan does not necessarily indicate the remaining copper thickness profile. We expected that the complementary use of both surface profiles and electrical measurements will be beneficial on obtaining a complete picture of the polished thickness/surface variations. All of these measurements are done across various pattern regions to ensure that we have a data set that covers wide range of pattern features.

In the next step we did copper CMP process modeling using the ample amount of data.

3.5 Copper CMP Process Modeling Using the Dishing Radius Concept

In this section, we first decompose the systematic variations in copper damascene process to two effects namely copper dishing and oxide erosion. At the end of this section we will combine these two components together to relate layout to CMP performance.

3.5.1 Dishing Modeling

Since copper dishing becomes a concern in semiconductor manufacturing, people have had various conjectures on the origin of dishing. In other words, why dishing occurs during polishing?

One of the observations is that dishing is a consequence of over polishing. The over-polish step is necessary for the removal of copper and barrier layer over the whole wafer area to insure reliability and reasonable yield. And the polishing pad, which makes direct contact to the wafer surface, was conjectured as a possible cause for metal dishing because of its bending toward the metal surface over the pressure applied from the back side of the rotating table [11-13]. To test the validity of this hypothesis, we did a simple calculation based on bending beam theory. The setup of the wafer pad contact scenario is illustrated in Figure 3.14.

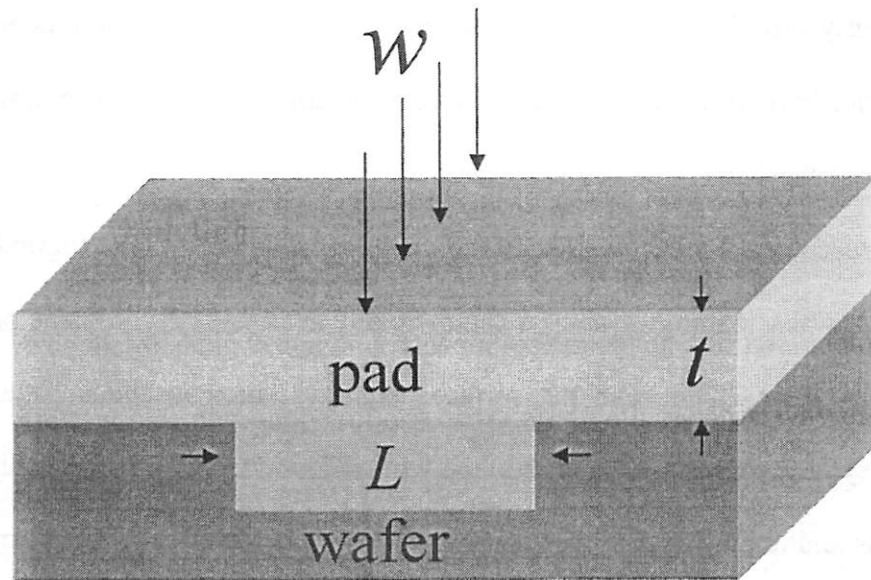


Figure 3.14 Calculation of the maximum deflection of the pad (at the center of the beam)

The maximum deflection of the pad (at the center of the beam) Δ is given by the following equation [14]

$$\Delta = \frac{wL^4}{384EI} = \frac{wL^3}{32Et^3} \quad (3.1)$$

where

E: modulus of elasticity of pad

L: width of the recess

I: moment of inertia = $t^3L/12$

t: pad thickness

w: load per length

If we use the numbers corresponding to the typical CMP process: $L=10\mu\text{m}$ (i.e. a 10-micron wide trench), $E=40\text{GPa}$, $t=1.25\text{mm}$ (based on the manufacturer's specifications), $w=450\text{KPa}\cdot\mu\text{m}$ (assuming 6psi down pressure) the calculation result gives a nearly zero deflection value. This strongly suggests that dishing is not a result of pad bending. In order to find out the dishing mechanism, we have to review the facts in the Chemical Mechanical Polishing system.

A recent study of the pad surface by Motorola showed the topography in Figure 3.15 [15]. It was a topology scan of a commercially available polyurethane pad. The scan area is $200\mu\text{m}$ by $200\mu\text{m}$ and the surface height variation is about $100\mu\text{m}$. The pad designers typically design a rough surface with random asperities. The main reason for this porous and rough felt-like surface is to provide slurry and by-product transportation during the CMP process. The following graph gives a general picture of the pad surface.

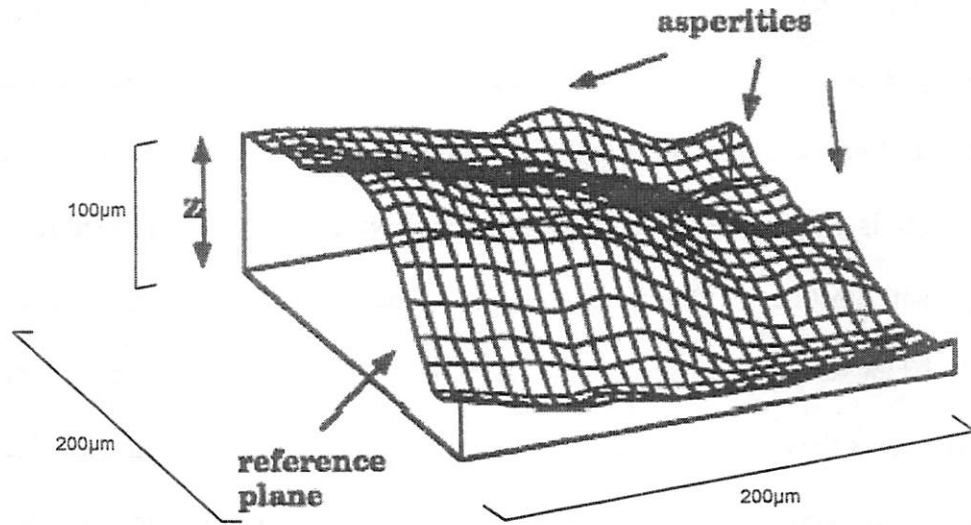


Figure 3.15 Topology scan of a commercially available polyurethane pad

Based on the nature of the polishing process, we should consider the wafer pad contact mechanics shown in Figure 3.16. In a typical CMP setup, the pad and the wafer are sliding with a relative speed v . The applied pressure is shared by pad asperities and the slurry film between wafer and the pad. For typical platen speed, the hydrodynamic pressure of the fluid film is minimal, so solid contact analysis is sufficient.

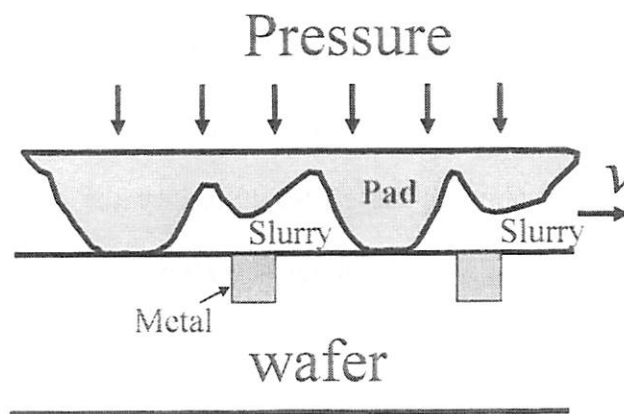


Figure 3.16 Proposed wafer pad contact mechanism under typical Chemical Mechanical Polishing setup

If we review the resistance data that we took using the electrical tests, we find interesting facts that relate resistance increase to linewidth. Table 3.6 lists the measured R for various lines and the corresponding values if without dishing. As expected, the measured R is larger due to dishing effect, and wider lines (that are subject to more dishing than thin lines) suffer more severe resistance increase. For instance, R increases by 9.4% when $w=5\mu\text{m}$, while it is only 1.1% larger than theoretical value if $w=0.4\mu\text{m}$. Fig. 3.18 shows the cross-sectional SEM pictures for $0.4\mu\text{m}$ and $1.6\mu\text{m}$ wide lines, which confirms the observation that wider lines experience more dishing than narrower ones. In addition, these pictures illustrate the non-planar shape of metal surface under dishing, which can be approximated by the segment of a circle [16-17].

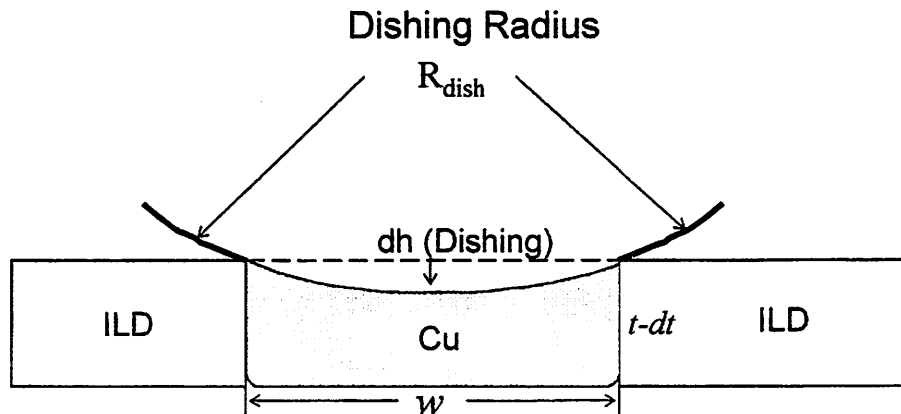
Such a concave-cylindrical surface is the result of both the overall mechanical polishing and the statistical behavior of the polishing pad asperities in a CMP system [18-20]. During chemical-mechanical polishing, material removal happens at the mechanical contact between the pad asperities and metal surface, and, at a relatively slower rate, on the dielectrics. Although the pad asperities have different sizes and heights, they come to contact with every line feature with equal probability. As a consequence, the concave-cylindrical topology forms during over-polishing. To empirically model this effect (in a manner consistent with its physical basis), we introduce the concept of dishing radius (R_{dish}) that captures the cylindrical metal surface after CMP (Figure 3.17). Under this model, line resistance under dishing is calculated as:

$$R = (\rho \cdot \text{length}) / \left[w(t - dt) + \frac{wR_{\text{dish}}}{2} \sqrt{1 - \left(\frac{w}{2R_{\text{dish}}} \right)^2} - R_{\text{dish}}^2 \sin^{-1} \left(\frac{w}{2R_{\text{dish}}} \right) \right] \quad (3.2)$$

where the parameters are defined in Figure 3.17 and ρ is the resistivity of copper.

Table 3.6 Theoretical and measured copper line resistances

Line w (μm)	Liner (μm)	Cu t (μm)	Line Length (μm)	Measured R (Ω)	Theoretical R(no dishing) (Ω)	Difference in R (%)
5	0.08	0.5	3961.0	35.18	32.16	9.39
4	0.08	0.5	3992.2	43.52	40.80	6.67
3	0.08	0.5	4023.4	58.10	55.47	4.74
2	0.08	0.5	4054.6	87.23	85.89	1.56
1.6	0.08	0.5	4067.1	111.69	109.69	1.82
1.2	0.08	0.5	4079.6	154.35	151.37	1.96
0.8	0.08	0.5	4092.1	246.63	243.27	1.38
0.4	0.08	0.5	4104.6	620.34	613.31	1.14

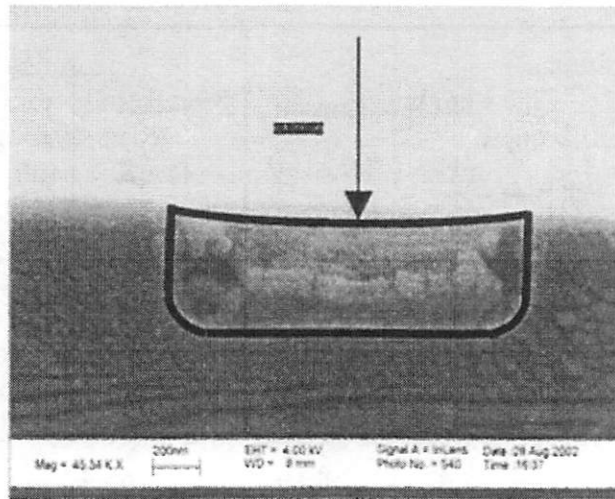


R_{dish} : dishing radius (the effective radius of pad asperity);

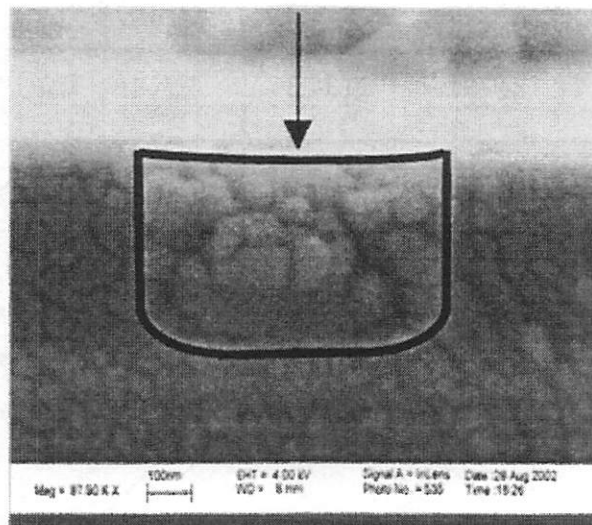
w : metal wire width

dh : metal non-planarity due to dishing;

Figure 3.17 Illustration of the Dishing Radius concept



1.6 μm line



0.4 μm line

Figure 3.18 Post-CMP cross-sectional pictures of copper lines (1.6 and 0.4 microns)

One of the reasons that make this dishing radius concept useful is the observation that it is independent of linewidth for typical on-chip interconnect structures contained in

our mask design. R_{dish} is independent of line dimensions, but is a function of process parameters (e.g., the pad asperity size, slurry chemistry, over-polish time, etc.). Once we know the dishing radius, we get a general estimation of the performance based on the following geometrical derivations (parameters are defined in Figure 3.17):

$$\frac{2R_{dish}}{\sqrt{\left(\frac{w}{2}\right)^2 + dh^2}} = \frac{\sqrt{\left(\frac{w}{2}\right)^2 + dh^2}}{dh} \quad (3.3) \quad \text{therefore:} \quad dh = \frac{2R_{dish} - \sqrt{4R_{dish}^2 - w^2}}{2} \quad (3.4)$$

When $w \ll R_{dish}$, we can multiply both the upper and lower side by $2R_{dish} + \sqrt{4R_{dish}^2 - w^2}$ to get the general relationship between the dishing radius and dishing height $h = \frac{w^2}{8R}$, which is a convenient way to get the dishing information based on linewidth.

Not only the SEM pictures qualitatively validate our dishing radius concept, but also the surface scanning profiles confirm the shape of the dishing after CMP. One of the surface profiling data is plotted in Figure 3.19.

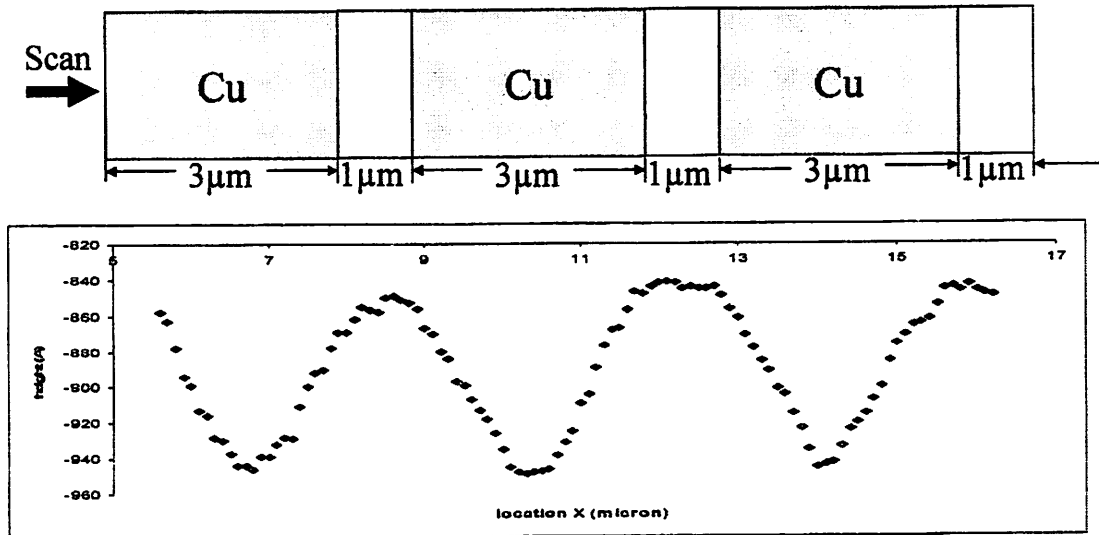


Figure 3.19 Surface profiling curves validate post-CMP surface dishing shape

Figure 3.20 shows a comparison between the line profilometry scans of 5 μ m wide lines on two post-CMP wafers (Experiment No. 1 and 3). As shown in the Figure, dishing values of 24nm and 3nm were obtained for the wafers polished with selective and non-selective step-2 slurries, respectively. This shows that the non-selective slurry effectively reduces dishing.

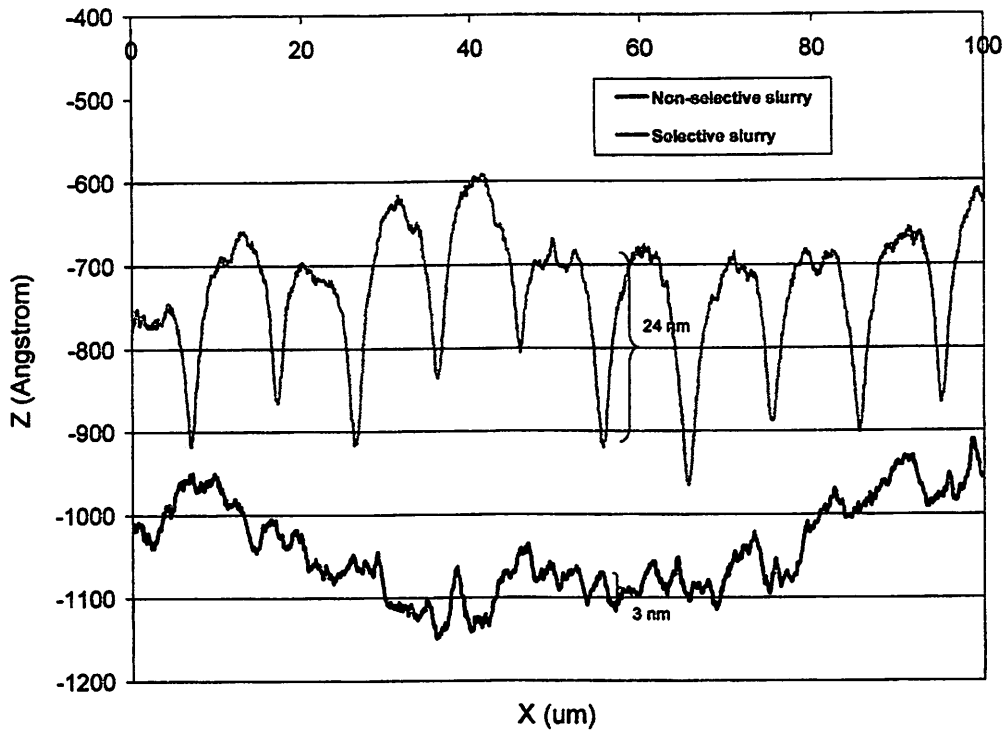


Figure 3.20 Line profilometry scans of post-CMP features
(Experiments No. 1 and 3)

Typical dishing values of the 5 μ m wide lines on other wafers are listed in Table 3.7. Various conclusions can be drawn from the data shown in the Table:

1. Least dishing was obtained when a non-selective step-2 slurry was used for Cu damascene polishing (Experiment No. 3 and 7).

2. Reducing pressure with non-selective slurry did not have an appreciable influence on the dishing behavior (Experiment No.3 versus 7).

3. The IC-1000 perforated pad yielded higher dishing than IC-1400 with k groove pad (Experiment No. 1 versus 9 and Experiment No. 5 versus 13).

4. The reduction of polish pressure while using an IC-1000 perforated pad had no effect on the dishing (Experiment No. 9 versus 13).

Table 3.7 Dishing values of 5µm wide lines for different experimental parameters

Experiment No.	Polish Parameters (Pad, Pressure, Slurry, Pattern)	Dishing Value (dh)
1	(A, A, A, A)	24 nm
3	(A, A, B, A)	3 nm
5	(A, B, A, A)	15 nm
7	(A, B, B, A)	3 nm
9	(B, A, A, A)	30 nm
13	(B, B, A, A)	30 nm

3.5.2 Erosion Modeling

As mentioned in Chapter 2, substantial amount of characterization work has been done in the field of ILD thickness variation during CMP process as a function of pattern

density. Here we just extend this concept to the context of copper damascene process and validate it using our measured data using NanoSpec™ Film Thickness Measurement Model 3000 in the UC Berkeley Micro-fabrication Lab.

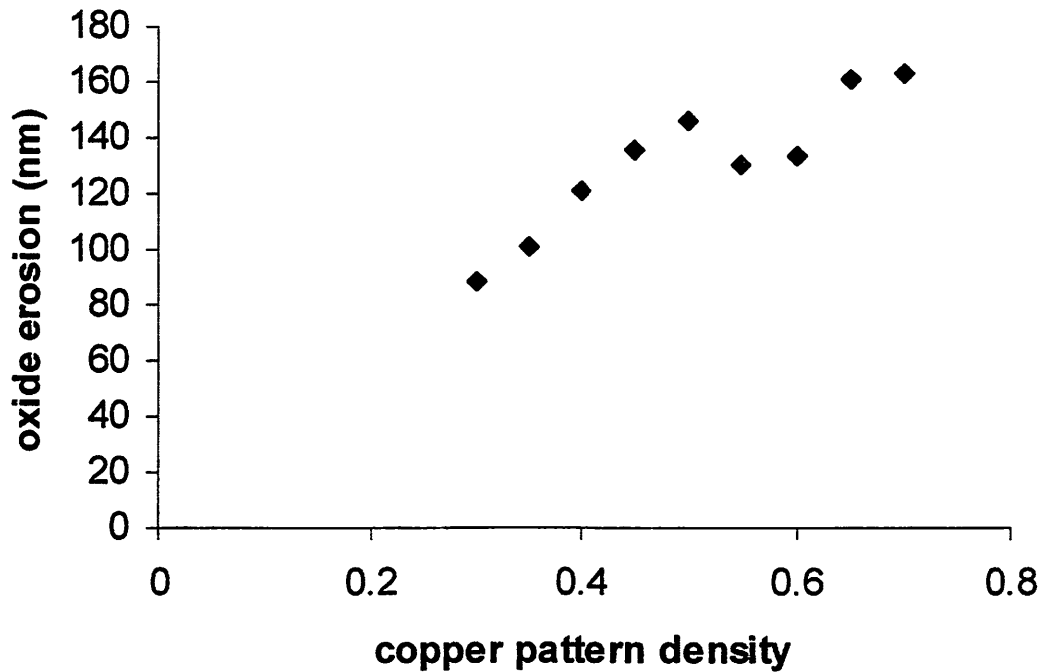


Figure 3.21 The measured amount of oxide erosion as a function of copper pattern density on Mask #2

This validates that the amount of oxide erosion is still approximately proportional to the copper pattern density as defined on Mask #2. It is clear that the concept of pattern density and planarization length holds in the context of copper CMP process.

3.6 Model-Based 2D Profile Extraction

We have qualitatively learned the shape of the post-CMP metal lines. Based on the data we have, we may be able to answer another question: Is it possible to extract more process-related information using the e-test data?

3.6.1 Key Ideas

Figure 3.22 shows the difference between the measured line resistance and the expected R with a perfect flat surface with no recession. Figure 3.23 shows the same metric normalized by the expected R. We find that for narrower lines, the R increase is mainly due to corner rounding and wider lines suffer more from copper surface dishing.

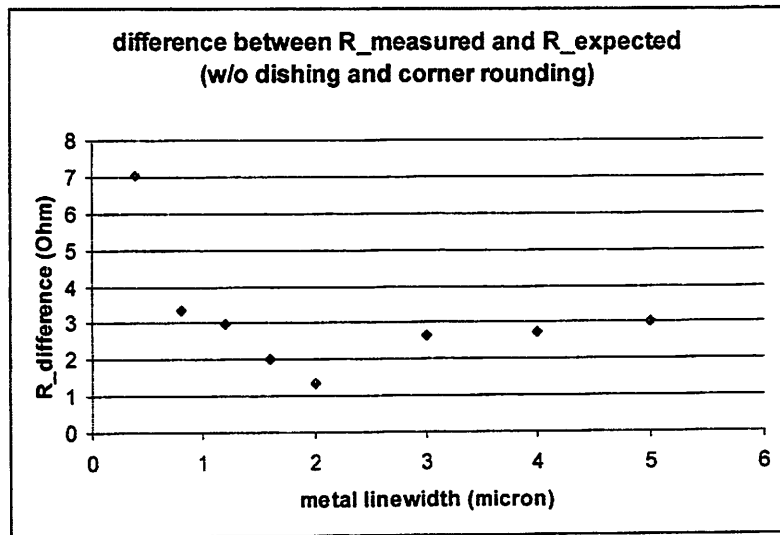


Figure 3.22 The difference between measured and expected R as a function of metal linewidth

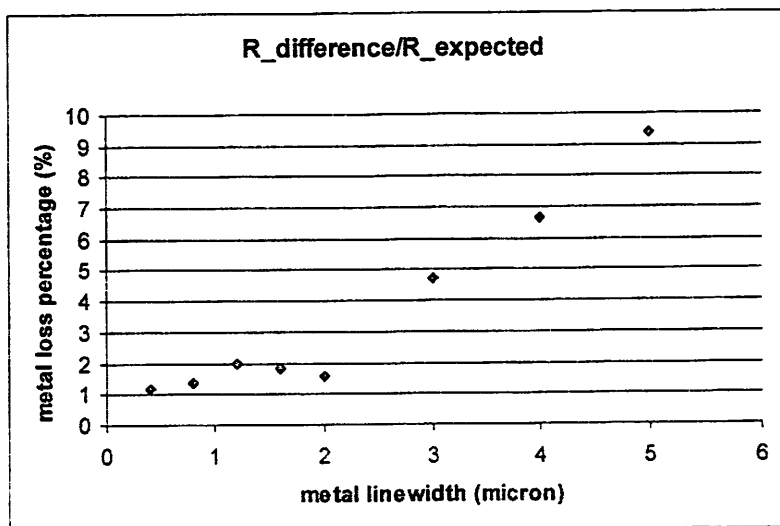
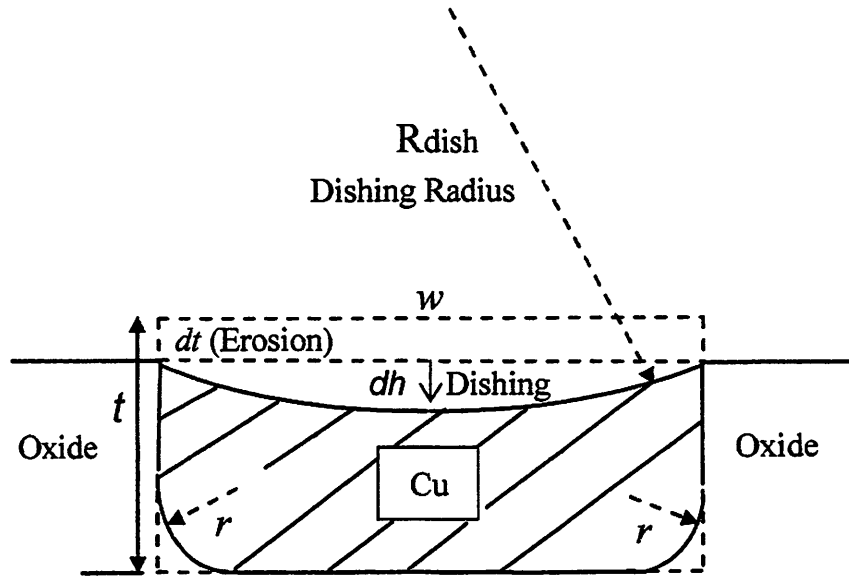


Figure 3.23 The difference between measured and expected R as a function of metal linewidth normalized by R_expected

3.6.2 Profile Modeling

We used electrical measurements to get the line resistance information after CMP; next objective is to find out what the shape of these damascene structured lines looks like. Based on some cross-sectional SEM pictures and analytical work, we model the profile mainly the following: The final cross-section is a confluence of a rectangular metal wire, some corner rounding at each lower side, and some dishing to the copper surface (as illustrated in Figure 3.24).

In Figure 3.24, w is the line-width; t is the initial total thickness; dt is the metal thickness loss due to oxide erosion; r is the radius of the corner rounding. R_{dish} is the dishing radius for the process. From the previous equation on dishing, we notice that the metal volume loss is a non-linear function of R_{dish} . In order to linearize this effect and without losing its physical integrity, we took the first order derivative of the dishing-related terms in equation (1) with regard to w . A dishing coefficient a is used to account for the dishing of different lines, replacing the R_{dish} term in our model. In other words, the dishing for a line with width of w is simplified to be $a.w$ as a first step. Notice that the volume loss due to dishing is roughly proportional to w^2 , however in terms of percentage of loss should be proportional to w . We also made the assumption that both the erosion and corner rounding are independent of linewidth. This is valid because we avoid using the lithography-limited features in our mask design (minimum linewidth on mask was $0.4\mu\text{m}$) and the lines were designed sufficiently close to each other in order to get the same amount of erosion, as stated in Section 3.2.



Rdish: effective radius of pad asperity
t: ideal metal thickness
dt: metal thickness loss due to erosion
r: corner rounding **dh:** metal dishing

Figure 3.24 Profile modeling for post-CMP metal lines

3.6.3 Measurements and Validation

Next we formulate the equations for N different lines. R is the resistance of the line and Y is the conductance of the line. We define y as the “normalized” conductance; y_0 as the “normalized” conductance with ideal rectangular cross section; y_e as the conductance loss due to erosion, y_r as the conductance loss due to corner rounding; y_d as the conductance loss due to dishing. The ideal resistance and conductance expressions are:

$$R = \frac{\rho \cdot L}{W \cdot t} \quad Y = \frac{1}{R} = \frac{W \cdot t}{\rho \cdot L} \quad (3.5)$$

both ρ and L are known numbers based on the metal film characterization and mask design. Define $y = Y \cdot \rho \cdot L$ we get

$$y = y_0 - y_e - y_r - y_d \quad (3.6)$$

therefore:

$$y_1 = W_1 t - W_1 dt - 2\left(1 - \frac{\pi}{4}\right)r^2 - (aW_1) \cdot W_1$$

$$y_2 = W_2 t - W_2 dt - 2\left(1 - \frac{\pi}{4}\right)r^2 - (aW_2) \cdot W_2$$

... ..

$$y_N = W_N t - W_N dt - 2\left(1 - \frac{\pi}{4}\right)r^2 - (aW_N) \cdot W_N$$

where N is the number of lines measured on one die (or wafer). Collecting the terms and rewriting the above equations into a matrix form, we get the following over-determined system:

$$\begin{bmatrix} y_1 \\ y_2 \\ \bullet \\ y_N \end{bmatrix} = \begin{bmatrix} W_1 & -2\left(1 - \frac{\pi}{4}\right) & -W_1^2 \\ W_2 & -2\left(1 - \frac{\pi}{4}\right) & -W_2^2 \\ \bullet & \bullet & \bullet \\ W_N & -2\left(1 - \frac{\pi}{4}\right) & -W_N^2 \end{bmatrix} \begin{bmatrix} t - dt \\ r^2 \\ a \end{bmatrix} \quad (3.7)$$

By defining

$$b = \begin{bmatrix} y_1 \\ y_2 \\ \bullet \\ y_N \end{bmatrix}, \quad A = \begin{bmatrix} W_1 & -2\left(1 - \frac{\pi}{4}\right) & -W_1^2 \\ W_2 & -2\left(1 - \frac{\pi}{4}\right) & -W_2^2 \\ \bullet & \bullet & \bullet \\ W_N & -2\left(1 - \frac{\pi}{4}\right) & -W_N^2 \end{bmatrix}, \quad x = \begin{bmatrix} t - dt \\ r^2 \\ a \end{bmatrix}$$

Equation (3.7) becomes $Y = Ax$. We then solve this and get the parameters we are interested in: dt is the erosion, r^2 is the corner rounding radius and a is the dishing coefficient by equation (3.8).

$$x = (A^T A)^{-1} A^T Y \quad (3.8)$$

To ensure that $(A^T A)$ is not singular, the basic requirement is that we use metal lines with different linewidths on our test structure. The number of data points can also determine whether the system is also relatively well conditioned.

We then plug in our electrically tested data points from one sample die and perform the parameter extraction. The plot of the original data points versus the model predictions is shown in Figure 3.25. The extracted parameters (after filtering out the shorted and broken lines) for a typical die are:

$$\begin{bmatrix} t - dt \\ r^2 \\ a \end{bmatrix} = \begin{bmatrix} 0.3026 \\ 0.0747 \\ 0.0062 \end{bmatrix}$$

Since the system is over determined, I was able to extract the confidence intervals for each of the parameters above. And the 95% confidence intervals are listed as follows:

$$t - dt: [0.2804, 0.3248]$$

$$r^2: [0.0540, 0.0954]$$

$$a: [0.0058, 0.0066]$$

These tight confidence intervals mean that the extraction tool filters out the signals (the parameters of interest) effectively.

In short, as a case study, we learn that the erosion for this die is about 0.1 μm ; the rounding radius is about 0.27 μm (the square root of 0.0751) and the dishing factor is 0.0062, which indicates that for 2 μm lines the dishing was around 12 nm (~3.1%) and for 5 μm lines the dishing was 31 nm (~7.7%). All of these extracted parameters are in good agreement to those in the SEM pictures and data from the surface profiler. A residual plot in this extraction is shown in Figure 3.26. The plot gets slightly noisier as it approaches the narrowest line on the chip. However, there is no obvious pattern left in the residual plot, which is a good indication that the model fits well with the measured data.

Since this is a well known least square estimation method for solving over-determined systems, its precision is a function the data sample size and the precision of the electrical test. Simulations show that this technique can be scaled down to sub-100nm features with scalable processing capabilities (lithography, etching and CMP). On the other hand, process engineers can assign longer lines in their test pattern design in order to achieve better sensitivity for the metal lines resistance. By putting these types of structures on the mask, erosion and dishing information can be brought to the front after few electrically tested data points. People can extract useful parameters at the die or wafer level non-destructively and readily. This can potentially be integrated into an automated probing tool and improve the metrology efficiency.

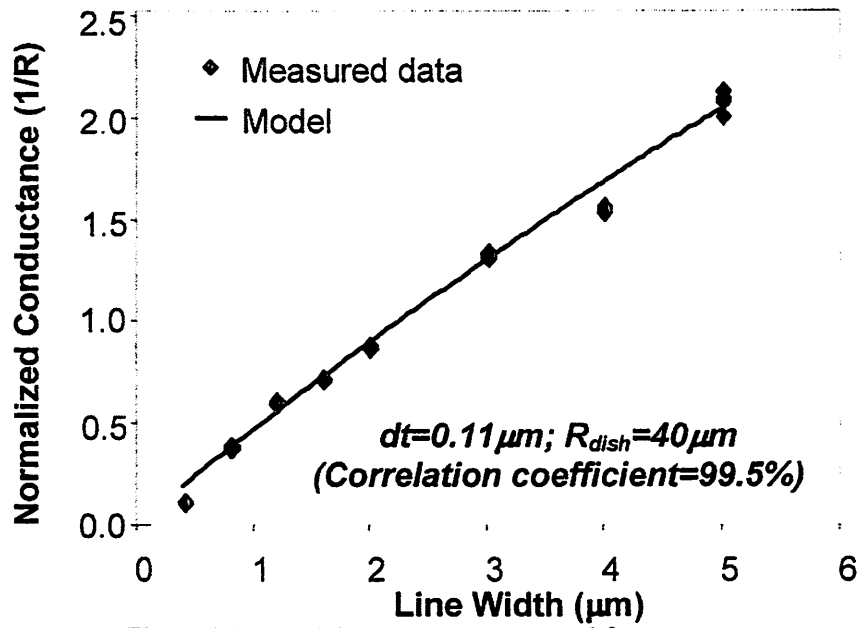


Figure 3.25 Model parameters extracted from measurement

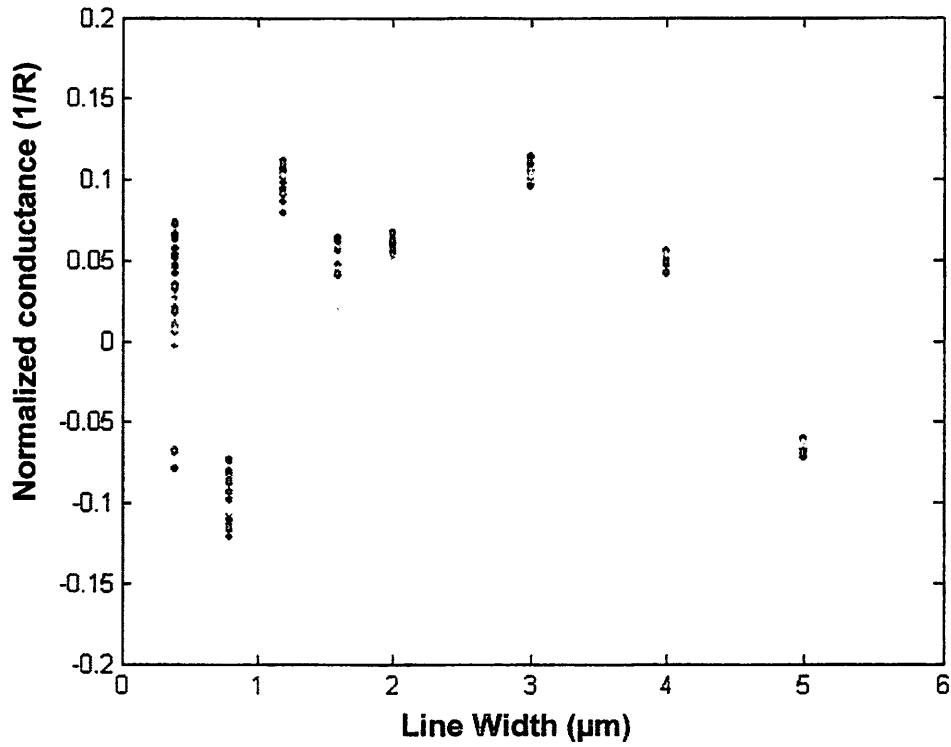


Figure 3.26 Residues plot after the extraction

References:

- [1] A.K. Stamper et al, "Sub-0.25-micron interconnection scaling: damascene copper versus subtractive aluminum," Proceedings of IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop, pp. 337 - 346, September 1998.
- [2] J. Heidenreich et al, "Copper dual damascene wiring for sub-0.25 μm CMOS technology," Proceedings of the IEEE International Interconnect Technology Conference, pp. 151 - 153, June 1998
- [3] H. Chen, et al, "Defect reduction of copper BEOL for advanced ULSI interconnect," IEEE International Interconnect Technology Conference, pp. 21-23, June 2001.
- [4] H. Goel, D. Dance, "Yield enhancement challenges for 90 nm and beyond," IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop, pp. 262 - 265, April 2003.
- [5] M. Gupta, et al., "Planarization yield limiters for wafer-scale 3D ICs," IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop, pp. 278 -283, May 2002.
- [6] B. Stine, et al., "The physical and electrical effects of metal-fill patterning practices for oxide chemical-mechanical polishing processes," IEEE Transactions on Semiconductor Manufacturing, vol. 45, no. 3, pp. 665-678, Mar. 1998.
- [7] D. Ouma, et al., "Characterization and modeling of oxide chemical-mechanical polishing using planarization length and pattern density concepts," IEEE Transactions on Semiconductor Manufacturing, vol. 15, no. 2 , pp. 232-244, May 2002.

- [8] R. Tian, D. Wong, and R. Boone, "Model-based dummy feature placement for oxide chemical-mechanical polishing manufacturability," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 20, no. 7, pp. 902-910, July 2001.
- [9] D. Ouma, et al., "An integrated characterization and modeling methodology for CMP dielectric planarization," *IEEE International Interconnect Technology Conference*, pp. 67-69, June 1998.
- [10] R. Tian, X. Tang, and D. Wong, "Dummy-feature placement for chemical-mechanical polishing uniformity in a shallow-trench isolation process," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 21, no. 1, pp. 63-71, Jan 2002.
- [11] T. Nishioka et al, "Modeling on hydrodynamic effects of pad surface roughness in CMP process," *IEEE International Conference on Interconnect Technology*, pp. 89 - 91, May 1999.
- [12] J. M. Steigerwald, R. Zirpoli, S. P. Muraka, and R. J. Gutmann, "Pattern geometry effects in the chemical-mechanical polishing of inlaid copper," *Journal of the Electrochemical Society*, Vol. 142, No. 10, pp. 2841-2848, 1994.
- [13] G. Fu et al, "A plasticity-based model of material removal in chemical-mechanical polishing (CMP)," *IEEE Transactions on Semiconductor Manufacturing*, Volume: 14 , Issue: 4, pp. 406 - 417, November 2001.
- [14] J. M. Steigerwald, S. P. Murarka, and R. J. Gutmann, *Chemical Mechanical Planarization of Microelectronic Materials*, John Wiley & Sons, Inc., 1997.

- [15] T.K. Yu et al, "A statistical polishing pad model for chemical-mechanical polishing," Technical Digest of International Electron Devices Meeting, pp. 865 - 868, December 1993.
- [16] V. Nguyen, et al., "Modeling of dishing for metal chemical mechanical polishing," IEEE International Electronic Devices Meeting, pp. 499-502, Dec. 2000.
- [17] R. Chang et al, "Copper Chemical-Mechanical Polishing Process Modeling using the Dishing Radius Concept," 8th International CMP Conference, September 2003.
- [18] Y. Zhao and L. Chang, "A micro-contact and wear model for chemical mechanical polishing of silicon wafers," *Wear*, Vol. 252, pp. 220-226, 2002.
- [19] J. Luo, D. Dornfeld and R. Chang, "Time Dependent CMP Model Based on Linear Visco-elasticity", 7th International CMP Conference, October 2002.
- [20] O. G. Chekina et al, "Wear-contact problems and modeling of chemical mechanical polishing," *Journal of the Electrochemical Society*, Vol. 145, No. 6, pp. 2100-2106, June 1998.

Chapter 4 Model-Based CMP Process

Optimization

The semiconductor industry is continuing its unabated growth over the last 40 years [1]. As the technology advances into the sub-100nm range of process geometries, the ability of IC manufacturers to build smaller, faster and less-expensive transistors and interconnects has become the critical issue in order to sustain the long term prosperity of the global semiconductor industry. A typical problem is due to process variability, to the extent that manufacturing system performance can not meet the stringent requirements imposed by Integrated Circuit designers.

In this chapter, we will utilize Chemical Mechanical Polishing as a case study in discussing the importance and feasibility of optimizing the process performance through Design of Experiments (DOE) and Statistical Process Control (SPC).

This issue predates the semiconductor revolution. From the beginning of the 20th century, quality, particularly the dimensions of component parts, became a very serious issue because no longer were the parts hand-built and individually fitted until the product worked. Now, the mass-produced part had to function interchangeably in every product built. Quality was initially obtained by inspecting each part and passing only those that

met specifications. This was true until 1931 when Walter Shewhart, a statistician at the Hawthorne plant at Western Electric, published his book *Economic Control of Quality of Manufactured Product* (Van Nostrand, 1931). This work is the foundation of modern statistical process control and provides the basis for the philosophy of total quality management or continuous process improvement. With statistical process control, the process is monitored through sampling. Considering the results of the sample, adjustments are made to the process before the process is able to produce defective parts [2-5].

The idea of control charts was based on the assumption that a normal process can have reasonable variations around the expected center point (based on the central limit theorem [6-7]). The center and spread of this fluctuation become the basics in a control chart (Figure 4.1). Any points outside the control limits need a closer look – they could be the results of normal fluctuation of the process or it can have an assignable cause, for example, the wear-out of the polishing pad or slurry particle degradation.

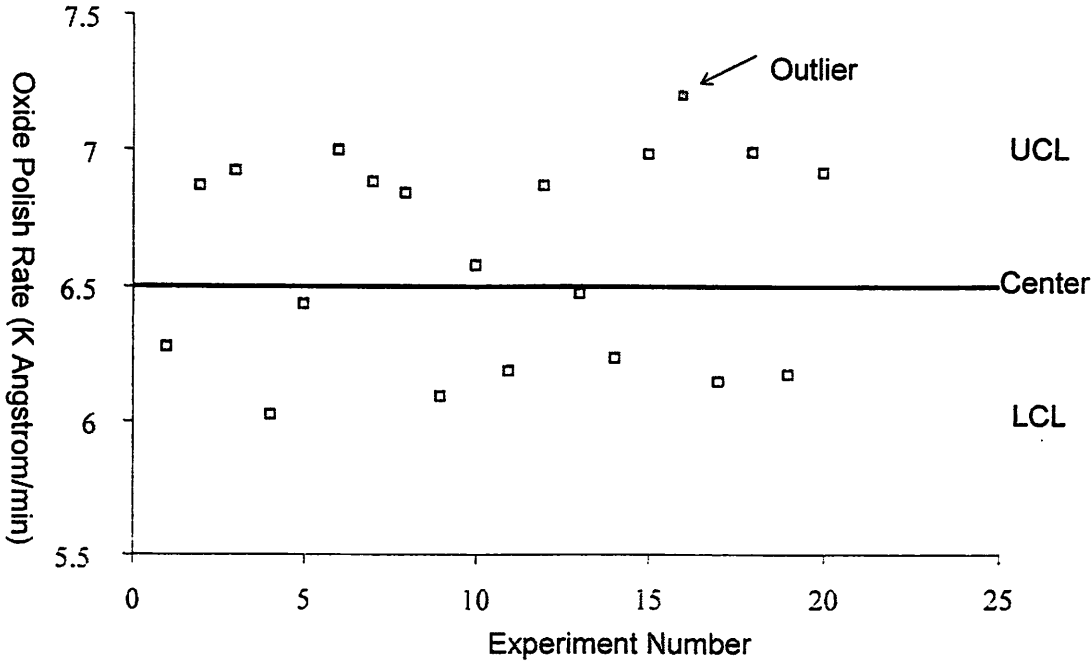


Figure 4.1 Illustration of the basic components of the control chart using Oxide polish rate as an example: data points, center line, Upper Control Limit (UCL), Lower Control Limit (LCL) and

Outlier

The paradox in real world often appears: on one side, people complain that they don not have enough resources (personnel, metrology tools etc.) to carry out the necessary experiments and obtain data in order to learn what is happening within the polishing process; on the other hand, people are puzzled by large amounts of data that is not properly or efficiently utilized. In this context, the idea of design of experiments comes into play. Engineers have tried “one-factor at a time” experiments to characterize and model the process, however, as the input parameters in a polishing system increase dramatically over the past few years, it becomes essential to efficiently plan and design the experiments before implementing them.

4.1 CMP Process Performance Metrics

There are several output metrics in a chemical mechanical polishing process, ranging from the waste volume to the flatness of the wafer surface. However, only a subset of these is critical for quality improvement purposes. In other words, it will be very costly to optimize everything because the problem will become too complex and eventually intractable. In this section, we attempt to identify and enumerate these key outputs.

4.1.1 Material Removal Rate (MRR)

The material removal rate (MRR) metric literally means how fast the polishing process can remove the film from the wafer surface. For blanket film, it can be defined as

$$MRR = -\frac{dh}{dt}, \quad (4.1)$$

where h is the thickness of the film and t is the polishing time. For patterned films, since the material removal rate at different sites can be different, MRR is thus defined as the removal rate of the bulk area, a large enough plain surface formed by the material of interest (typically more than 1mm^2).

The Material Removal Rate is just one of the key metrics in CMP because it directly relates to the throughput of the process. If we can fix other metrics, the higher MRR is desired because it will allow semiconductor manufacturers to process more wafers during the given amount of time.

4.1.2 Selectivity

Selectivity is defined as the ratio of the two different material removal rates:

$$S_{12} = \frac{MRR_1}{MRR_2} \quad (4.2)$$

With other metrics fixed, a higher selectivity might be desired because the material removal will prefer one over the other. A good case of this will be plasma etching of the trench in the back end of the process: higher selectivity means a more planar bottom surface as shown in Figure 4.2. However, in the copper damascene process, during the second step (over-polish), a lower selectivity might be preferred because it will cause less copper dishing, as shown in Figure 4.3.

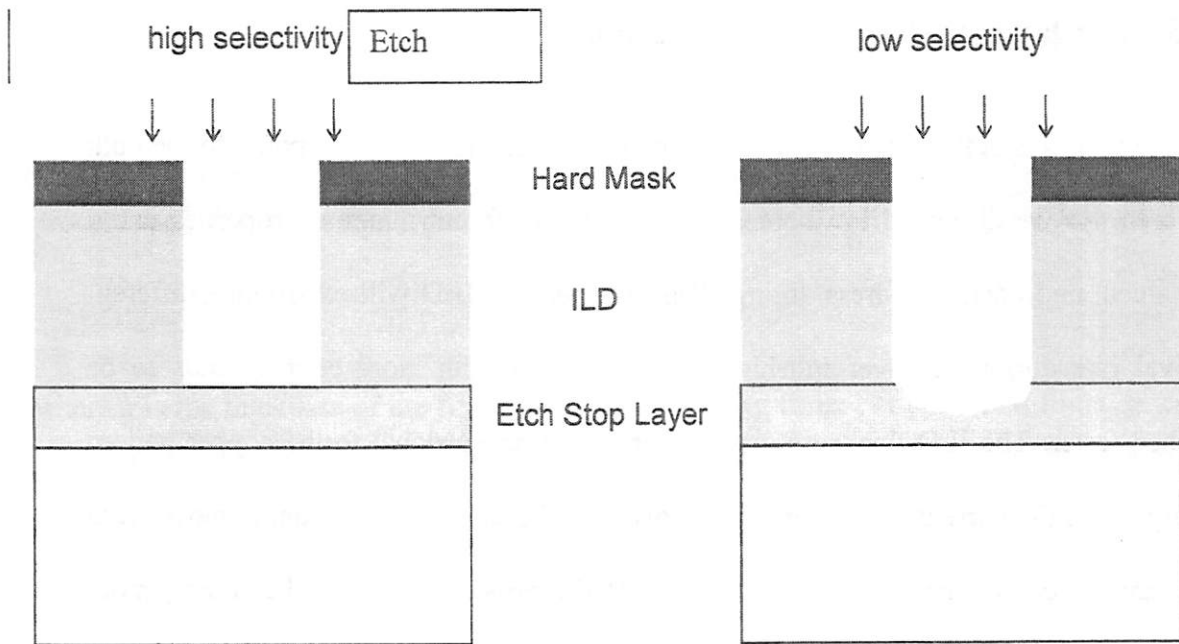


Figure 4.2 Illustration of the concept of selectivity – the scenario in plasma etching where high selectivity is desired

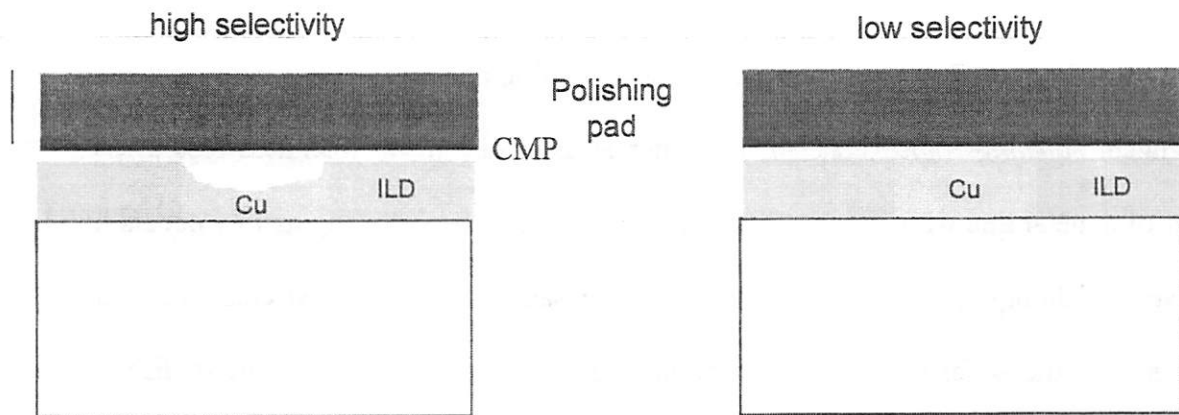


Figure 4.3 Illustration of the concept of selectivity – the scenario in the over-polishing step of the damascene process where low selectivity will be helpful in reducing metal dishing

The state-of-the-art semiconductor manufacturing process deals with several different materials in one step, which makes selectivity one of the critical metrics in process integration for the purpose of achieving higher yield.

4.1.3 Inter-layer Dielectric (ILD) Erosion

Erosion is defined as the inter-layer dielectric thickness loss compared to the bulk ILD area. As we discussed in Chapter 3, due to the different material properties under mechanical and chemical stress during CMP, copper and ILD will experience different removal rate during the over-polishing step. The area with more copper tends to be polished faster. The ILD thickness loss appears to be proportional to the copper pattern density. With the current scaling efforts of lowering the dielectric constant k , newer ILD films are softer and porous [8-10]; as a result the erosion metric is becoming more important in measuring the performance of the CMP process.

4.1.4 Metal Dishing

Based on the findings in Chapter 3, we can define metal dishing in two ways: one is using the *dishing radius* concept; the other is using the metal thickness loss at the center of a metal line with standard linewidth (e.g. 1 μm). Metal dishing mainly occurs in the over-polishing step [11-13]. This step is indispensable because it is extremely hard for every site on the wafer to reach the endpoint at the same moment. Some over-polishing will be needed for reliability considerations – if at any site the metal is not cleared the circuit will not function correctly, leading to yield loss.

All these four metrics (Material Removal Rate, Selectivity, ILD Erosion and Metal Dishing) are critical in the copper damascene process, since they capture the general performance of a copper CMP process. Other metrics, such as device scratching, etc, can be considered secondary. Optimizing the critical metrics by tuning the different

input variables, such as pressure and velocity hence becomes essential in improving the yield of copper damascene process.

4.2 Framework of Process Performance Optimization

The simultaneous control of the four critical performance metrics can be viewed as a multi-objective optimization problem. Some theoretical work has been done in this field [14-17]. J. Nash developed one of the most representative methods in the 1950s [18]. Other more modern approaches involve so-called “genetic” algorithms [19], but the application of those is beyond the scope of this work. We note that one key limitation for using any optimization technique in our CMP control problem is the lack of reliable models that link the input settings to the critical output metrics [20]. The models that we will use in this work are discussed next.

4.2.1 Models Used in the Framework

4.2.1.1 MRR and Erosion Model

From the characterization work in Chapter 3, we concluded that in the damascene process, the dielectric thickness variation is a function of the CMP process, the feature geometry, as well as the geometry of near by features. This observation is in general agreement to the conclusions drawn by the ILD CMP study by Ouma [21-22]. In copper CMP, the final ILD thickness variation model is fundamentally linked to the Preston’s equation, which relates the removal rate on a blanket wafer to the product of applied pressure and relative velocity. For the structure shown in Figure 4.4, the blanket (planar region) removal rate is given by:

$$MRR = -\frac{dh}{dt} = kPv, \quad (4.3)$$

where P is the pressure, v is the relative velocity between the pad and wafer, k is Preston's constant.

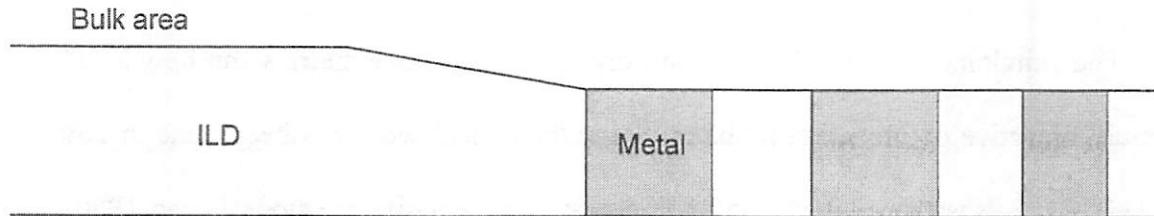


Figure 4.4 During the over-polishing step in a typical damascene process, the bulk ILD area suffers no significant erosion.

Differences in pattern density result in different post CMP ILD thickness across the chip since the ILD line sparse regions polish faster than dense regions. The effective pattern density is calculated as the amount of ILD in a given area divided by the total area in that region (see Fig. 4.5). The “given area” refers to the interaction distance, which is a function of the process, CMP tool, and consumables. The pressure can be represented as F/A , where F is the down force exerted on the wafer and A is the area of the oxide contacted by the pad. Copper also supports some of the down force but since it is much softer than oxide (the hardness of Copper is 3 while oxide is 5.5 on the Mohs scale) its share of down force support is relatively small. With L defined to be the interaction distance and $\rho(x,y)$ being the effective pattern density,

$$MRR = -\frac{dh}{dt} = \frac{-kFv}{L^2 \rho(x, y)} = \frac{-kP_{Nom}v}{\rho(x, y)}, \quad (4.4)$$

where P_{Nom} is the nominal pressure applied on the wafer.

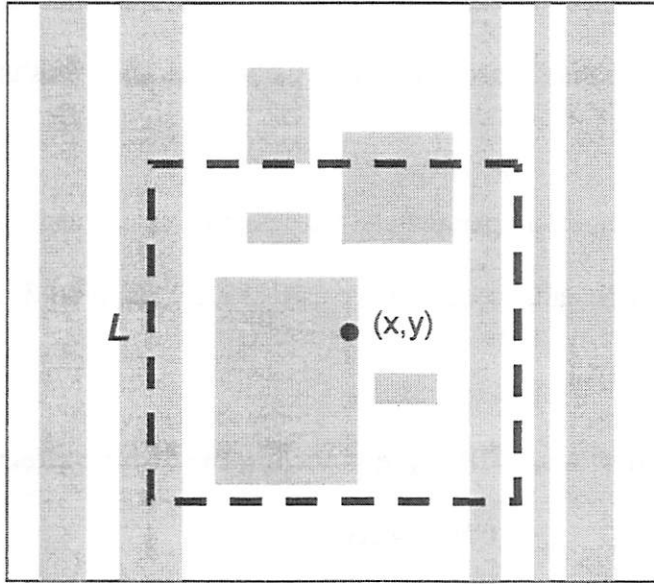


Figure 4.5 Illustration of the effective pattern density calculation for the point of interest (x,y) using a square window (size L) in damascene process.

This model captures the (ILD) material removal rate and erosion. For a given layout pattern on a die, the pattern density distribution is fixed; erosion is essentially a measure of the difference in polishing rate at different locations on the die. So the amount of erosion will be proportional to MRR in the bulk area.

4.2.1.2 Metal Dishing Model

Thanks to the dishing radius concept introduced in Chapter 3, we know that dishing and thickness loss at the center of a standard-width line are not independent. Moreover, once one parameter is known, the other can be calculated easily. In this study, we choose the latter as the metric (dishing thickness loss at the center of $5\mu\text{m}$ line). As we discussed before, dishing is a function of the consumables and process input parameters. Some of the properties, such as the pad roughness, are hard to quantify. In modern polishing systems it is not very common to change the pad type frequently. To

simplify the modeling problem, we assume that dishing is a strong function of the pressure, velocity and over-polish time; other input parameters such as slurry flow have minimal effect [23-24].

Once identified the important parameters, we designed experiments to model their effects on dishing. The experimental sequence we used is a 2^{3-1} factorial design as shown in Table 4.1.

Table 4.1 Design of Experiments (DOE) in modeling the dishing effect in copper CMP process

Experiment No.	Pressure (3psi-6psi)	Speed (60rpm- 100rpm)	Overpolish Time (15sec- 30sec)
1	-1	-1	1
2	-1	1	-1
3	1	-1	-1
4	1	1	1

Please note that in Table 4.1, each factor is exercised in two levels, and it is used an equal number of times at the high and low setting. The decision about number of levels will depend to some extent on prior knowledge (if any) about the factor. Three-level studies will be beneficial if curvature effects are likely to be present within the range of factor settings. This idea is illustrated in Figure 4.6.

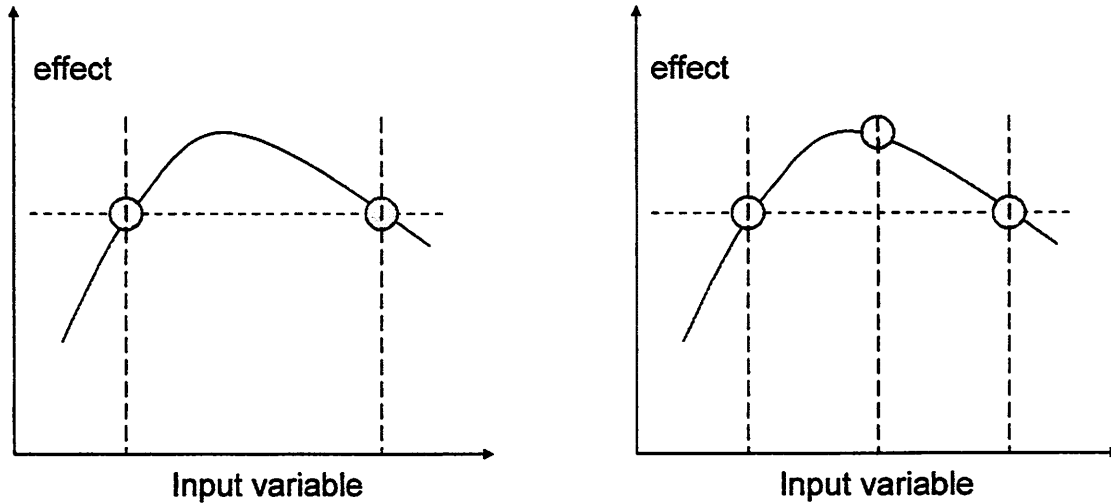


Figure 4.6 Cases when a three or more level becomes necessary: A two level study (the left) will give the illusion of no effect from the factor while a three level study (the right) will uncover the curvature effect

We had limited experimental resources, so we performed a two-level study assuming that the curvatures are second order effects (this assumption was justified in the case study in section 4.3). After randomizing the runs in Table 4.1, we carried out our experiments and built the linear model that relates the three parameters to dishing. The dishing measurements on the 5 μ m line using the Dektak Surface Profiler are listed in Table 4.2.

Table 4.2 DOE measurement Results

Run No.	Pressure	Speed	Overpolish Time	Dishing (5 μ m line) (unit: nm)
1	-1	-1	1	30
2	-1	1	-1	22

3	1	-1	-1	25
4	1	1	1	20

The model is then built by just using the main effects:

$$Dish = Dish_0 + pressure * Effect_P + speed * Effect_v + time * Effect_t + noise, (4.5)$$

where

$$Dish_0 = \text{average}(Dish_1, Dish_2, Dish_3, Dish_4) = (30+22+25+20)/4 = 24.25 \text{ (nm)}$$

$$Effect_P = \frac{\frac{(Dish_3 + Dish_4)}{2} - \frac{(Dish_1 + Dish_2)}{2}}{2} = \frac{\frac{(25 + 20)}{2} - \frac{(30 + 22)}{2}}{2} = -1.75$$

$$Effect_v = \frac{\frac{(Dish_2 + Dish_4)}{2} - \frac{(Dish_1 + Dish_3)}{2}}{2} = \frac{\frac{(22 + 20)}{2} - \frac{(30 + 25)}{2}}{2} = -3.25$$

$$Effect_OT = \frac{\frac{(Dish_1 + Dish_4)}{2} - \frac{(Dish_3 + Dish_2)}{2}}{2} = \frac{\frac{(30 + 20)}{2} - \frac{(25 + 22)}{2}}{2} = 0.75$$

please also note that in the model, the pressure, speed(velocity) and over-polish time should all be normalized within the [-1, 1] interval. The real values of these parameters were listed in Table 4.1. These calculated effects from different input parameters agree well with our findings in Chapter 3.

Based on the model that we just developed, it becomes possible to predict the dishing perspective of the CMP process performance by plugging in the appropriate normalized values of input parameters.

4.2.1.3 Selectivity Model

Limited by experimental and metrology resources, we were not able to get enough data points to develop a linear model between selectivity between copper and inter-layer

dielectrics. However if selectivity becomes more important in a copper CMP process, following the framework that we described in 4.2.1.2, it is straightforward to develop an empirical model between selectivity and process input parameters such as pressure and speed: the basic test patterns should include both materials (copper and oxide) on the same layer through damascene process. The metal line width should be in the millimeter or at least hundreds of microns range (to ensure it is much greater than the characteristic size of pad asperities which is around $50\mu\text{m}$). Then the input settings can be changed and the process can be fingerprinted in order to find out the optimum range for characterization. The rest is similar to 4.2.1.2 and a model can be set up in a few runs.

At this stage, we have got all the models that we are interested to discuss the optimization work.

4.2.2 Weight Coefficients and other Optimization Considerations

The process optimization methodology is based on the input and output parameters that have been selected. In general, a cost or objective function is created in terms of the outputs (and possibly the inputs), which defines the relative quality of a set of outputs (and inputs). Specifically, a lower cost implies a better set of process parameters, meaning that a globally minimum cost is the desired operating point. The following sections describe the strategies and issues involved in process optimization for the multi-input-multi-output (MIMO) chemical mechanical polishing control problem. There are two main steps involved: selection of an appropriate cost function, and performing a constrained optimization over the cost model to find the inputs that yield the lowest cost.

In order to make a decision in the face of multiple objectives it is necessary to know the relative importance of the different objectives. Yet, it is often very difficult to specify a set of precise weights before possible alternatives solutions are known. The weight selection process can be interactive and iterative before arriving at an acceptable solution. The decision maker gradually discerns what is achievable and adjusts the weights to enforce trade-offs between the objectives, as knowledge about these interactions increases. In our example, we initially assume that the process engineers know qualitatively what objectives weigh more and they have a tentative set of weights to start with.

Often the cost is based primarily on the outputs and only slightly, if at all, on the inputs. In this CMP performance optimization problem, a minimum cost solution should occur when all measurements have achieved the target values (MRR, erosion, dishing etc.), and all of the inputs are within the valid operating ranges. A constrained optimization capability eliminates the need to worry about input ranges, so the cost function is often based solely on the output optimality criteria.

Cost functions can take different forms, but a weighted sum of squared errors from the output targets is common and often analytically meaningful. In this study we apply a traditional quadratic form:

$$Cost = \sum_i [w_i (Target_i - Measured_i)]^2 \quad (4.6)$$

This function sums over all outputs (i), the product of a weighting term (w_i) and the squared difference between the output target ($Target_i$) and the actual (measured or predicted) output ($Measured_i$). The weighting terms scale the importance of meeting

various output targets. The weights could reflect trade-offs between comparing outputs of different types, or trade-offs between the relative importance of outputs that are the same type (e.g. at the different sites on the wafer).

Now we have all the available ingredients to carry out the optimization work. Let us assume the following models, which can be characterized by experimental designs, as described in section 4.2.1.2:

$$Dish = Dish_0 + pressure * Effect_P1 + speed * Effect_v1 + time * Effect_t1 + noise1 \quad (4.7)$$

$$Selectivity = Selectivity_0 + pressure * Effect_P2 + speed * Effect_v2 + time * Effect_t2 + noise2 \quad (4.8)$$

$$MRR = MRR_0 + pressure * Effect_P3 + speed * Effect_v3 + time * Effect_t3 + noise3 \quad (4.9)$$

$$Erosion = Erosion_0 + pressure * Effect_P4 + speed * Effect_v4 + time * Effect_t4 + noise4 \quad (4.10)$$

And our cost function is:

$$Cost = w_{dish} [(T \text{ arg et}_{dish} - Measured_{dish})]^2 + w_{selectivity} [(T \text{ arg et}_{selectivity} - Measured_{selectivity})]^2 + w_{MRR} [(T \text{ arg et}_{MRR} - Measured_{MRR})]^2 + w_{Erosion} [(T \text{ arg et}_{Erosion} - Measured_{Erosion})]^2 \quad (4.11)$$

By substituting equations (4.7) – (4.10) into equation (4.11), we get the cost function in the form of input parameters (pressure, speed, time). The optimization problem now becomes the minimization of the cost function

$$Cost = f(pressure, speed, time) \quad (4.12)$$

In order to get the minimum cost, we take derivatives on the cost function with respect to the input parameters and set them to zero:

$$\frac{\partial f(pressure, speed, time)}{\partial pressure} = 0 \quad (4.13)$$

$$\frac{\partial f(\text{pressure, speed, time})}{\partial \text{speed}} = 0 \quad (4.14)$$

$$\frac{\partial f(\text{pressure, speed, time})}{\partial \text{time}} = 0 \quad (4.15),$$

write the equations above into matrix form and solve them to get the best settings (pressure, speed, time). The settings should be examined using the second derivatives in order to eliminate the possibilities of getting the maximum or a saddle point in the cost function [34]. The overall process follow is shown in Figure 4.7.

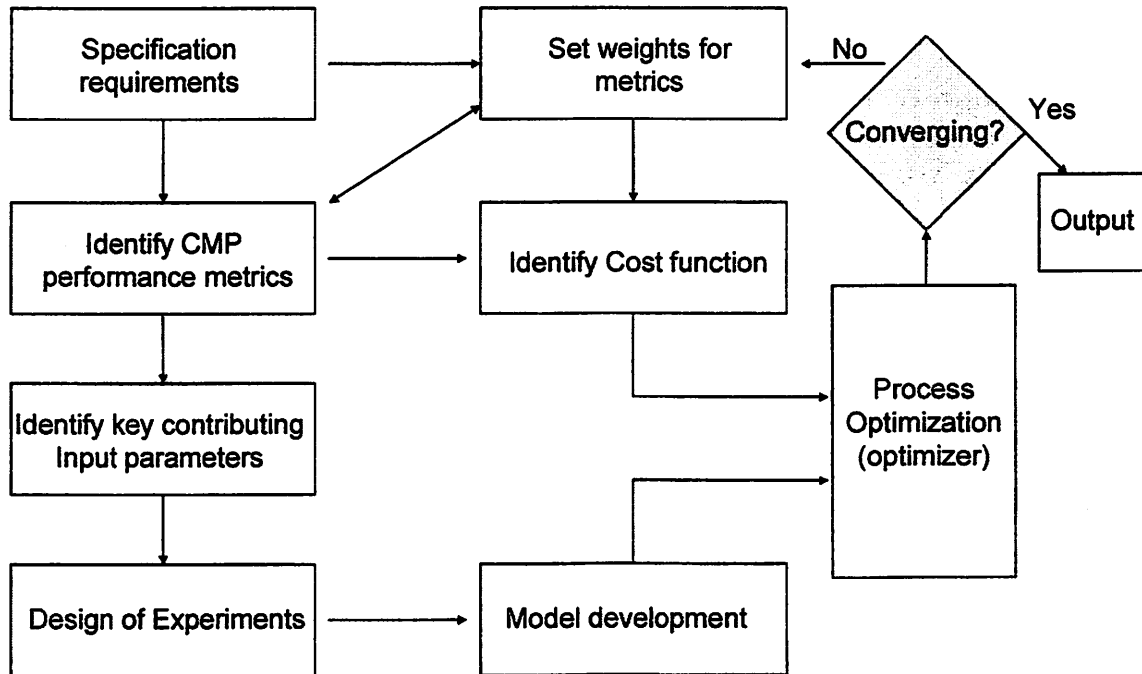


Figure 4.7 Flowchart of the CMP process optimization framework

We recognize that this method applies only in limited cases. In reality, the “zeroing” of the derivatives is often an implicit part of a more comprehensive minimization algorithm, which is more often than not, iterative in nature. Nevertheless, this optimization framework provides process engineers some opportunity of dynamic

Run-to-Run Control [25-29]. We will come back to this aspect in the “Future Work” session of this thesis.

4.3 Examples on Process Optimization

Now we have developed the basic optimization flow that has all the necessary building blocks in a semiconductor process development environment. Then it should be the time to bring in some numerical examples to see how these parameters interact and compromise under a certain preference. However, before we show that, it is useful to review the classic Taguchi approach in order to justify the weighting system that will be employed in this section.

The Taguchi approach has become increasingly popular as a method for developing engineered products [30]. It promises, and delivers, an ability to increase the quality of an engineered product via simple changes in the method by which engineers perform their usual design tasks. Given the bold claims, there has been relatively little research in the semiconductor process design community on Taguchi’s method: its assumptions, mathematics, techniques, and applications. In this section we will review the key concepts of Taguchi philosophy and apply a subset of them in the optimization case study.

Taguchi starts with a new definition of quality: Quality is related to the total loss to society due to functional and environmental variance. Taguchi’s method focuses on robust design through the use of S/N ratio to quantify quality and orthogonal arrays to investigate quality. It emphasizes that quality is not only meeting the specs but really hitting the target [31]. The quadratic loss function is illustrated in Figure 4.8.

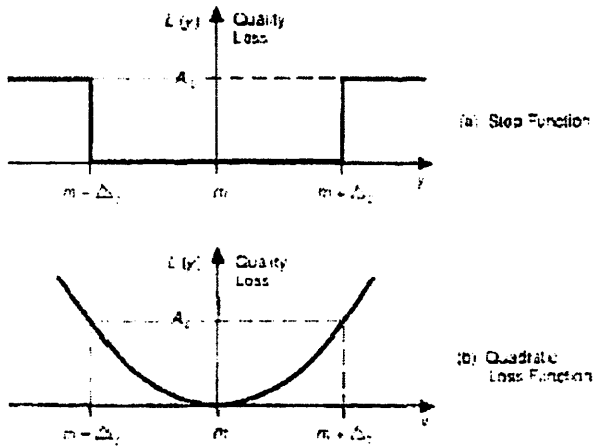


Figure 4.8 the quadratic loss function

The Taguchi concept of robust design is based on maximizing performance measures called signal-to-noise ratios by running a partial factorial set of experiments using orthogonal arrays. The signal-to-noise ratio is typically given by

$$S/N = -10 \log [\text{MSD}] \quad (4.16),$$

where MSD refers to Mean Square Deviation of objective function.

The S/N ratio aims at achieving the separability of design factors into control factors and signal factors. A robust optimum design is identified by finding the optimum setting of the control factors to reduce variation and then adjusting the signal factors to shift the mean. Additivity of factor effects is an important consideration in statistical design of experiments. It ensures that the performance measure is not adversely affected by the non-linearity of the objective function. Note that this form of the S/N ratio does not guarantee separability and additivity for all types of objective functions, and thus the use of transformations may become necessary [32].

If the additivity assumption holds, the S/N ratio in Taguchi's method is essentially exchangeable with the metric that we developed in previous section 4.2.2 of this chapter $L(y) = k(\text{target}-y)^2$. The classic Taguchi's concept provides a strong foundation for our optimization framework.

4.3.1 Linear Optimization

We assume that the key metrics have been identified as discussed in section 4.2 and data are obtained through basic experiments. The results are listed in Table 4.3. These results are conceptual and partially based on published data [33]; nevertheless they should be in good quantitative agreement with a typical copper damascene development process in the 0.13μm technology generation.

Table 4.3 Copper CMP Process characterization Results

Run No.	Pressure	Speed	Slurry particle size(nm)	Material removal rate (nm/min)	Erosion (50% area: nm)	Dishing (5μm line: nm)
1	-1	-1	1	490	28	30
2	-1	1	-1	525	23	22
3	1	-1	-1	546	34	25
4	1	1	1	503	31	21

The pressure and speed ranges are listed in Table 4.1. The (average) slurry particle size window is from 80nm to 150nm and it is normalized to the [-1, 1] scale. For dishing, we have built the model as discussed in 4.2. The only difference is now slurry particle size is used in place of over-polishing time as the input. Following similar procedures as in section 4.2, we obtain:

$$\begin{aligned}
 Dish &= Dish_0 + pressure * Effect_P + speed * Effect_v + size * Effect_s + noise \\
 &= 24.5 - 1.5 * pressure - 3.0 * speed + 1.0 * size
 \end{aligned}$$

Similarly, the erosion model can be set up as the following:

$$\begin{aligned} Erosion &= Erosion_0 + pressure * Effect_P + speed * Effect_v + size * Effect_s + noise \\ &= 29 - 3.5 * pressure - 2.0 * speed + 0.5 * size \end{aligned}$$

Finally, the bulk copper removal rate as a function of pressure, table speed and slurry particle size is modeled as

$$\begin{aligned} MRR &= MRR_0 + pressure * Effect_P + speed * Effect_v + size * Effect_s + noise \\ &= 516 + 8.5 * pressure - 2.0 * speed - 19.5 * size \end{aligned}$$

Let us assume that the targeted outputs are: MRR = 550nm/min; erosion = 20nm and dishing = 15nm. Using the models we just developed, the cost function is written as the following:

$$\begin{aligned} Cost(p, v, s) &= w_{dish} [(15 - 24.5 + 1.5 * p + 3.0 * v - 1.0 * s) / 15]^2 + w_{erosion} [(20 - 29 + 3.5 * p + 2.0 * v - \\ &0.5 * s) / 20]^2 + w_{MRR} [(520 - 516 - 8.5 * p + 2.0 * v + 19.5 * s) / 520]^2 \\ &= w_{dish} [(-9.5 + 1.5 * p + 3.0 * v - 1.0 * s) / 15]^2 + w_{erosion} [(-9 + 3.5 * p + 2.0 * v - \\ &0.5 * s) / 20]^2 + w_{MRR} [(4 - 8.5 * p + 2.0 * v + 19.5 * s) / 520]^2 \end{aligned}$$

Case 1: We want to put more weight on erosion and dishing (yield) in comparison to Material removal rate (throughput): 45% weight on dishing, 45% on erosion and 10% on MRR.

The cost function becomes:

$$\begin{aligned} Cost(p, v, s) &= 0.45 * [(-9.5 + 1.5 * p + 3.0 * v - 1.0 * s) / 15]^2 + 0.45 * [(-9 + 3.5 * p + 2.0 * v - 0.5 * s) / 20]^2 \\ &+ 0.10 * [(4 - 8.5 * p + 2.0 * v + 19.5 * s) / 520]^2 \end{aligned}$$

As discussed in section 4.2, in order to get the minimum cost, we take the partial derivatives on the cost function with respect to the input parameters and set them to zero we can get three equations. Write the equations into matrix form and solve to get the optimized settings (pressure, speed, size). By checking the second derivatives of each parameter to be positive, we did confirm that these settings minimize the cost function.

The results are listed in Table 4.4.

Case 2: We want to put more weight on Material removal rate (throughput) in comparison to erosion and dishing (yield): 5% weight on dishing, 5% on erosion and 90% on MRR.

The cost function becomes:

$$Cost(p, v, s) = 0.05 * [(-9.5 + 1.5 * p + 3.0 * v - 1.0 * s) / 15]^2 + 0.05 * [(-9 + 3.5 * p + 2.0 * v - 0.5 * s) / 20]^2 + 0.90 * [(4 - 8.5 * p + 2.0 * v + 19.5 * s) / 520]^2$$

Using the same procedure as in Case 1, we solve the partial derivatives matrix to get the optimized settings (pressure, speed, size). Again, the second derivatives of each parameter were checked to be positive in order to confirm the minimum claim. The results are also listed in Table 4.4.

Table 4.4 Integrated CMP Optimization Results

Case No.	Objective	Weights			Input parameter settings			Output performance		
		MRR	Dishing	Erosion	P (psi)	S (rpm)	Size (nm)	MRR (nm/min)	Dishing (nm)	Erosion (nm)
0	Center point	-	-	-	4.5	80	115	516	24.5	29
1	Minimal erosion and dishing (higher yield)	10%	45%	45%	4.65	76	128.4	498	21	15.6
2	Faster Material removal rate (Higher through-put)	90%	5%	5%	3.5	90	85.6	520	27.5	23.8

As shown above, the process input variables can be tuned to serve the purpose of optimizing the emphasized output performance. It can have different kinds of boundary

conditions (for example, erosion less than 25nm) and target values (e.g., the higher MRR the better), however, the key concepts (design of experiments, modeling, optimization) developed in this chapter will hold. After these numerical calculations, process developers can apply these settings to validate that optimum and feedback as shown in the flowchart Figure 4.7.

The previous example works well with linear models. In a state-of-the-art polishing process, due to the consumable complexity and interactions, the relationship between the inputs and outputs typically appears non-linear. In this case, the classical Taguchi's method using orthogonal arrays can be employed as a more powerful tool.

4.3.2 Application of Taguchi's Method

The method of investigating all possible combinations and conditions in an experiment (involving multiple factors) is traditionally known as factorial design. The factorial design is based on the theory, that for a full factorial design, the number of possible designs, N (number of runs), is

$$N = L * m, \quad (4.17)$$

where L = number of levels for each factor

m = number of factors involved

Orthogonal Arrays (OA) are a special set of Latin squares, initially constructed by Taguchi to lay out the product design experiments. During regression analysis, an orthogonal arrangement of the experiment gave us independent model parameter estimation. Orthogonal arrays have the same property: for every two columns all possible factor combinations occur equal times. By using the basic architecture as shown in Table

4.5, a 3-level orthogonal array of standard procedure can be used for a number of experimental situations in CMP process optimization.

Table 4.5 Experimental Design using Orthogonal Array $L_9(3^3)$

Run No.	Column (variables)		
	A	B	C
1	1	1	1
2	1	2	2
3	1	3	3
4	2	1	2
5	2	2	3
6	2	3	1
7	3	1	3
8	3	2	1
9	3	3	2

In our case study, three input variables are pressure, speed, slurry particle size. And the outputs are Material Removal Rate, Dishing ($5\mu\text{m}$ line) and ILD erosion (refer to Figure 4.3). The levels of inputs are shown in Table 4.6.

Table 4.6 Control Factor Levels

Input variable	Levels		
	1	2	3
Pressure	2	4	6
Speed	60	80	100
Slurry particle size	80	120	160

Let's assume that the objectives are: 1.) minimize dishing $n = -10 \log_{10} (\text{dishing})^2$
 2.) increase MRR $n' = 10 \log_{10} (\text{MRR}/\sigma)^2$, where σ is the experimental error which can be estimated from the ANOVA residuals 3.) reduce ILD erosion $n'' = -10 \log_{10} (\text{erosion})^2$.

After 9 runs, the experimental results using Taguchi's metrics are tabulated in Table 4.7. Please note that the order of the nine runs should be randomized to minimize the memory effects and create a better estimation of the experimental error through later Analysis of Variance (ANOVA).

Table 4.7 Experimental Results using Orthogonal Array L₉(3³)

Run No.	Column (variables)			Observations		
	A	B	C	Dishing n(dB)	MRR n'(dB)	ILD erosion n''(dB)
1	1	1	1	-30.9	52.0	-22.9
2	1	2	2	-20.0	50.6	-28.9
3	1	3	3	-18.1	48.1	-28.3
4	2	1	2	-29.8	55.7	-19.1
5	2	2	3	-26.0	53.7	-25.1
6	2	3	1	-32.0	48.9	-30.1
7	3	1	3	-21.6	55.1	-23.5
8	3	2	1	-28.9	51.2	-26.8
9	3	3	2	-24.6	53.0	-26.0

The subsequent step is the estimation of factor effects, or Analysis of Mean (ANOM), using the following equations for dishing effects:

$$m_D = [n_1+n_2+\dots +n_9]/9$$

$$m_{DA1} = [n_1+n_2+n_3]/3$$

$$m_{DA2} = [n_4+n_5+n_6]/3$$

$$m_{DB1} = [n_1+n_4+n_7]/3$$

$$m_{DC3} = [n_3+n_5+n_7]/3$$

and for MRR and erosion effects, the equations are similar, the only difference is that n_i ($i=1,2,\dots,9$) should be replaced by n_i' and n_i'' ($i=1,2,\dots,9$) in the equations, respectively.

The calculated effects for the three variables on the output metrics were illustrated in Figure 4.9.

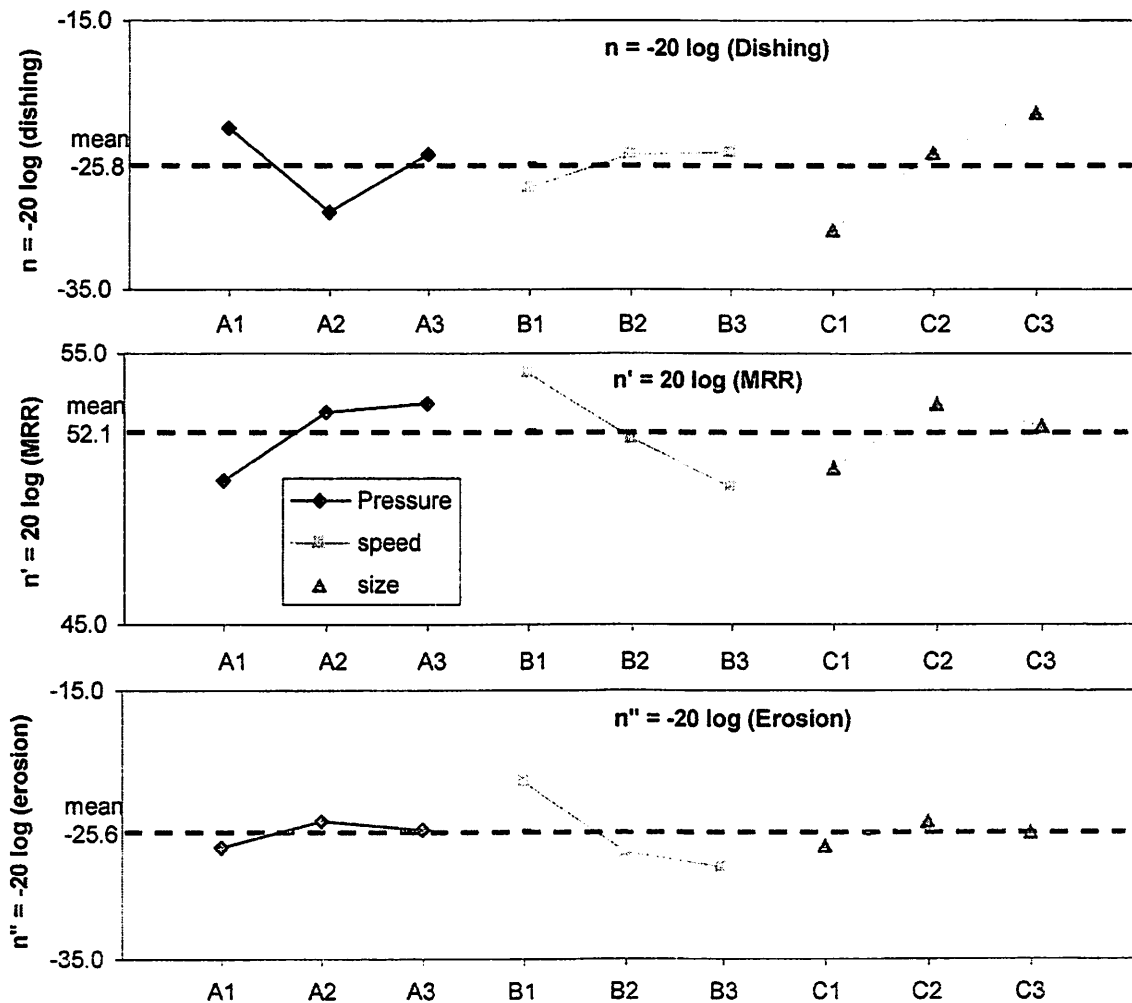


Figure 4.9 data analysis using Taguchi's quality metrics

Table 4.8 Optimization using the Taguchi's Method

Inputs	Starting Condition				Optimum Condition			
	setting	Dishing	MRR	ILD Erosion	setting	Dishing	MRR	ILD Erosion
A	A2	-29.3	52.8	-24.8	A3	-25.0	53.1	-25.5
B	B2	-25.0	51.9	-27.0	B1	-27.4	54.3	-21.8
C	C2	-24.8	53.1	-24.7	C3	-21.9	52.3	-25.6
Total Mean		-26.4	52.6	-25.5		-24.7	53.3	-24.3
Real Output		20.8 (nm)	426 nm/min	18.8 (nm)		17.3 (nm)	459 nm/min	16.4 (nm)

Figure 4.9 provides an intuitive, comprehensive view of the process. By tuning the different knobs (input variables), the process can be optimized within few iterations. Table 4.8 shows the initial setup and optimized inputs and outputs. The optimization process improved the Materials Removal Rate by about 10%, and reduced the erosion and dishing by more than 15%. This process can also be programmed to be an automatic routine, enabling the implementation of Advanced Process Control (APC) including automatic recipe generation in the state-of-the-art semiconductor production environment.

References:

- [1] International Technology Roadmap for Semiconductors, International SEMATECH, Austin, TX, 2003.
- [2] G.E.P. Box, W.G. Hunter, and J.S. Hunter. Statistics for Experimenters, John Wiley and Sons, 1978.
- [3] D. M. H. Walker, Yield Simulation for Integrated Circuits, Norwell, MA: Kluwer, 1987.
- [4] Seber, G. A. F. Linear Regression Analysis. Wiley, New York, 1976.
- [5] Douglas C. Montgomery, "Design and analysis of experiments", J. Wiley & Sons, 1996.
- [6] A.R. Barron. Entropy and the central limit theorem. Ann. Probab., 14:336-342, 1986.
- [7] A. Araujo and E. Gine, The central limit theorem for real and Banach valued random variables, Wiley, 1980.
- [8] C. Chang et al, Interconnection challenges and the National Technology Roadmap for Semiconductors, pp. 3-6, Proceedings of the IEEE 1998 International Interconnect Technology Conference, June 1998
- [9] M. Rasco et al, Packaging assessment of porous ultra low-k materials, pp. 113-115, Proceedings of the IEEE 2002 International Interconnect Technology Conference, June 2002
- [10] K. Mosig et al, Integration of porous ultra low-k dielectric with CVD barriers, 2001 IEDM Technical Digest, International Electron Devices Meeting, Dec. 2001

- [11] J. Pan et al, Copper CMP integration and time dependent pattern effect, Proceedings of the IEEE 1999 International Interconnect Technology Conference, pp. 164 - 166, May 1999
- [12] G. Fu et al, An analytical dishing and step height reduction model for chemical mechanical planarization (CMP), IEEE Transactions on Semiconductor Manufacturing, Vol. 15, No. 3, pp. 477-485, Aug. 2003
- [13] S. Li et al, A low cost and residue-free abrasive-free copper CMP process with low dishing, erosion and oxide loss, Proceedings of the IEEE 2001 International Interconnect Technology Conference, pp. 137-139, June 2001
- [14] Matsuyama, Y.; Nakayama, H.; Sasai, T.; Yuh Perng Chen, Penalized learning as multiple object optimization, IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on Neural Networks, Volume: 1, June-2 July 1994, Pages:187 - 192 vol.1
- [15] J. S. Hunter, "The Exponentially Weighted Moving Average," Journal of Quality Tech. , Vol. 18, No. 4, October 1986.
- [16] G. E. P. Box and T. Kramer, "Statistical Process Control and Automated Process Control - A Discussion," Technometrics, Vol. 34, No.3, pp. 251-267, 1992.
- [17] S. Crowder, "Design of Exponentially Weighted Moving Average Schemes," Journal of Quality Tech., Vol. 21, No. 3, July 1989.
- [18] J.F. Nash, "Equilibrium points in n-person games," Proceedings of National Academy of Science, USA, 36:46-49, 1950.

- [19] Sefrioui, M.; Perlaux, J., Nash genetic algorithms: examples and applications, Proceedings of the 2000 Congress on Evolutionary Computation, Pages:509 - 516 vol.1, July 2000
- [20] Nanz, G.; Camilletti, L.E., "Modeling of chemical-mechanical polishing: a review," IEEE Transactions on Semiconductor Manufacturing, Volume: 8, Issue: 4, Nov. 1995, Pages:382 - 389
- [21] D. O. Ouma, B. Stine, R. Divecha, D. Boning, J. Chung, I. Ali, and M. Islamraja,, "Using Variation Decomposition Analysis to Determine the Effect of Process on Wafer and Die-Level Uniformities in CMP," First International Symposium on Chemical Mechanical Planarization (CMP) in IC Device Manufacturing, Vol. 96-22, pp. 164-175, 190th Electrochemical Society Meeting, San Antonio, TX, Oct. 6-11, 1996.
- [22] Divecha, R., B. Stine, D. Ouma, J. Yoon, D. Boning, J. Chung, O.S. Nakagawa, S-Y Oh, "Effect of Fine-Line Density and Pitch on Interconnect ILD Thickness Variation in Oxide CMP Processes," 1997 Chemical Mechanical Polish for ULSI Multilevel Interconnection Conference (CMP-MIC), p. 29, Santa Clara, February, 1997.
- [23] R. Chang et al, "Copper Chemical-Mechanical Polishing Process Modeling using the Dishing Radius Concept," 8th International CMP Conference, September 2003.
- [24] J. Luo, D. Dornfeld and R. Chang, "Time Dependent CMP Model Based on Linear Visco-elasticity", 7th International CMP Conference, October 2002.

- [25] E. Sachs, R. Guo, S. Ha and A. Hu, "Tuning a Process While Performing SPC: An Approach Based on the Sequential Design of Experiments," Proc. of IEEE/SEMI ASMC, 1990.
- [26] E. Sachs, R. Guo, S. Ha and A. Hu, "Process Control System for VLSI Fabrication", IEEE Trans. on Semi. Manuf., Vol. 4, 1991.
- [27] A. E. Gower-Hall, Integrated Model-Based Run-to-Run Uniformity Control for Epitaxial Silicon Deposition, Ph.D. thesis, MIT EECS, 2001.
- [28] E. Sachs, A. Hu, and A. Ingolfsson, "Run by Run Process Control: Combining SPC and Feedback Control," IEEE Trans. Semi. Manuf., vol. 8, no. 1, pp. 26-43, Feb. 1995.
- [29] T. Smith and D. Boning, "A Self-Tuning EWMA Controller Utilizing Artificial Neural Network Function Approximation Techniques," IEEE Trans. on Comp., Pack., and Manuf. Technol. Part C, Vol. 20, No. 2, pp. 121-132, April 1997.
- [30] Byrne, D.; Quinlan, J., "Robust function for attaining high reliability at low cost," Annual Reliability and Maintainability Symposium, 1993. Pages:183 - 191, Jan. 1993
- [31] C. Spanos and K. Poolla, EE290H class notes, UC Berkeley, Fall 2003.
- [32] Box, Hunter, and Hunter, Statistics for Experimenters, Wiley, 1988.
- [33] N. Ohashi, "Improved Cu CMP process for 0.13 μ m node multilevel metallization," International Interconnect Technology Conference, 2001
- [34] M. Abramowitz, "Handbook of Mathematical Functions", Harri Deutsch Verlag, 1984

Chapter 5 Effects of CMP-Related Process Variations on Interconnect/Circuit Performance

5.1 Design for Manufacturability Overview

The development of integrated circuit technology has spawned a rapid growth in personal conveniences such as personal computers with wireless fidelity, cell phones with digital imaging and global positioning systems. As the technology scaling keeps its fast pace, the IC industry has been transformed into two main sectors during the last few years: fab-less design house and design-less foundry. In this context, Design for Manufacturing (DFM) is being defined as the missing link between integrated circuit design and manufacturing [1-5]. Increasing semiconductor manufacturing expenses continue to represent real economic burdens measured in billions of dollars every year. The chance for first silicon success is getting smaller as design capabilities continue to outpace manufacturing realities in the newer technology nodes. In order to validate this statement, I review a typical integrated circuit design flow [6] as shown in Figure 5.1.

As shown in Figure 5.1, designers must first draft a series of specifications (design specs) and define the performance of the circuit to be developed. This includes defining the inputs and outputs required. In many cases, the specifications include a series of mathematical algorithms needed for the circuit to function properly. Sometimes a mathematical model is used to help create a schematic (electronic diagram) of various functions, using standard cells from the library. Next, using commercially available software such as the Simulation Program with Integrated Circuit Emphasis (SPICE),

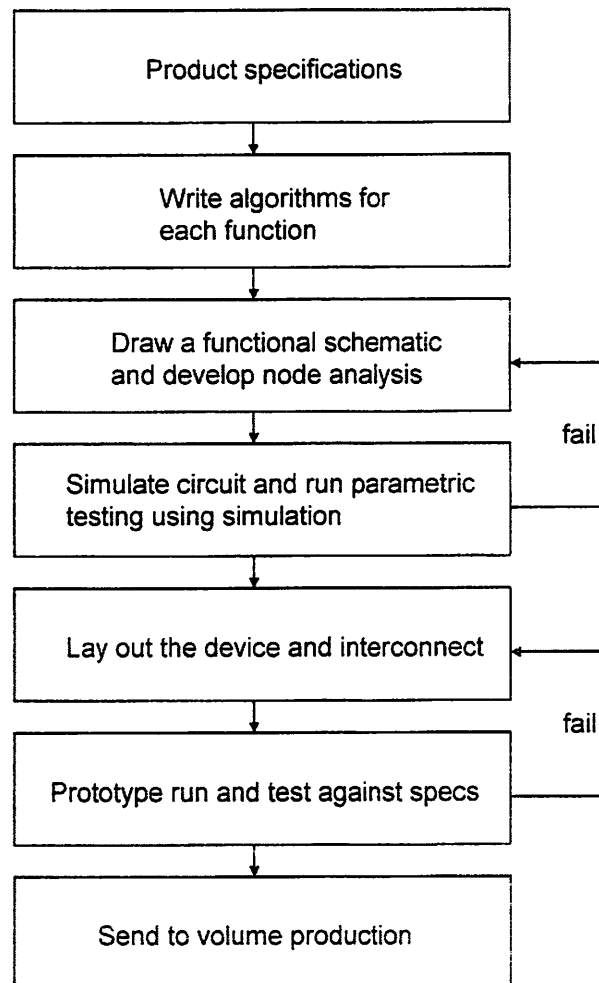


Figure 5.1 A typical modern computer-aided integrated circuit design flow

designers can simulate the circuit's performance to check the outputs under various input conditions. This does not completely eliminate and hand calculations or schematic

drawings as compared in the traditional design flow, but greatly reduced the time to test the performance of the design. Simulation programs help identify problems, such as stray capacitance, inductance, timing issues, propagation of gate delays and a host of things that much be considered for a successful outcome.

Once the simulations are successfully completed at both the schematic cell level and device level, the circuit maybe laid out for manufacturing. This is done by the engineers using Computer Aided Design (CAD) Software. When all the data from the schematic is loaded into the program, the program generates the physical layout for each mask required to manufacture the circuit. Quality design rule checking and further simulations may be required before the circuit goes to the next step.

If the circuit appears to be a solid functional design, it is then prototyped in a short run fabrication process and tested. Once the circuit is proven to work functionally and meet specifications, it is then sent to an in-house fab or foundry for mass production. The prototype fabrication process becomes costly as some 90-nanometer mask sets rate multi-million dollars, and 65-nanometer masks are expected to cost even more [7-9].

In short, the design-to-manufacture chasm is growing at an alarming rate as chip size and complexity increase and process geometries decrease, and as product life-cycles decrease. The emerging state of semiconductor production demands that designers take manufacturability into consideration up front, rather than leaving it as an expensive and time-consuming afterthought.

In this chapter, we first discuss the effect of process variations on circuit performance from the chemical mechanical polishing technology perspective; the second part will show the importance of decoupling the variations into systematic and random

components in order to better understand and ultimately narrow down the distribution of circuit performance; the last part of this chapter focuses on the interconnect performance simulation based on the dishing models developed in the previous chapters. We conclude the chapter by proposing and evaluating the line-splitting technique at the design stage to mitigate the effect of dishing.

5.2 The Effect of Process Variations on Circuit Performance from CMP Process Technology Perspective

The continuously increasing scale of integration used in the design and manufacturing of integrated circuits has drawn special attention toward interconnect effects. As the minimum feature size in Ultra-Large-Scale Integration (ULSI) systems drops to 90nm and below, interconnect characteristics are becoming limiting factors on performance, since the time constant associated with interconnect is scaled by a smaller factor compared to those of devices [10-12]. Figure 5.2 shows the rapid interconnect architecture complexity increase from the 0.7 μm to 0.25 μm technology generation. Future chip complexity and speed advances will depend on the ability to model the electrical behavior of interconnect in an accurate and efficient fashion. Critical path delays in circuits depend upon interconnect as well as on device parameters. The effects of device parameter variations have been widely studied [13-16]. However, these simulations currently do not take into account the effects of interconnect parameter variations. As a result, the yield estimation and circuit optimization based on these studies may not be able to provide accurate results in current and future technologies, where more and more significant portions of path delays will result from interconnect.

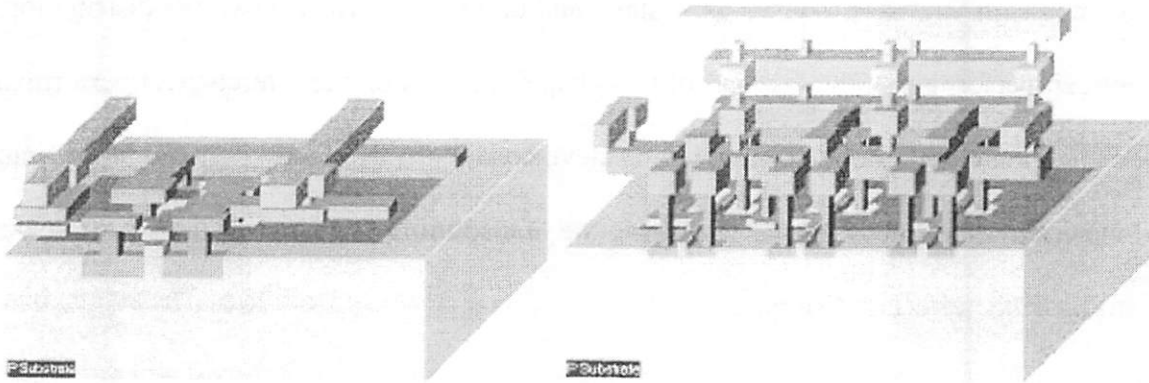


Figure 5.2 Interconnect complexity increases in 0.25µm technology (right) compared with the 0.7µm technology (left)

Some previous simulation results in Figure 5.3 [17] demonstrate the increase in crosstalk for the neighboring wires from 0.7µm technology to 0.25µm technology. As the technology moves down to the nanometer era, the process variations in BEOL will further affect the whole circuit performance. From the current technology library, these effects are treated as a worst case such as “10% variation”. This will inevitably cause

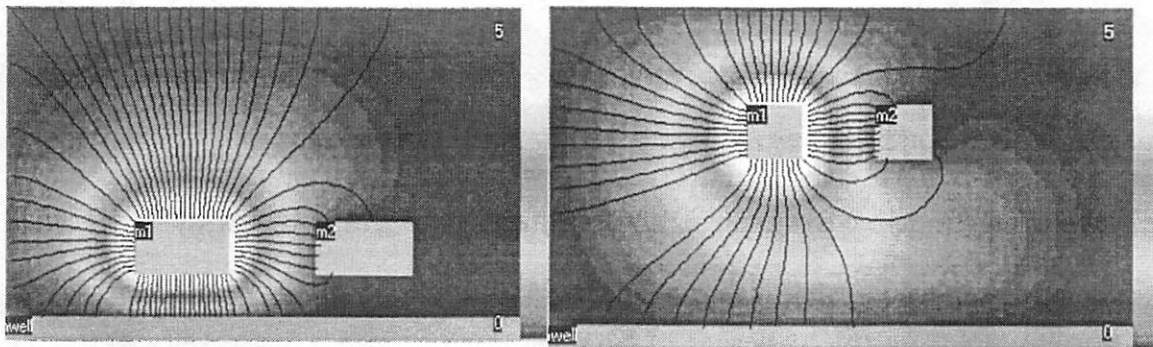


Figure 5.3 Simulation results illustrate the increased crosstalk between neighboring wires as technologies shrink 0.7µm (left) to 0.25µm (right)

The idea of process variation decoupling is illustrated the following figures: the post-process ILD thickness measurement might be noisy and have about 10% variation, as shown in Figure 5.4. Currently circuit designers will take the message that the process

has 10% uncertainty so they need to design the circuits to make the timing work at this uncertainty level. This objective will obviously cost more time and die area to the designers.

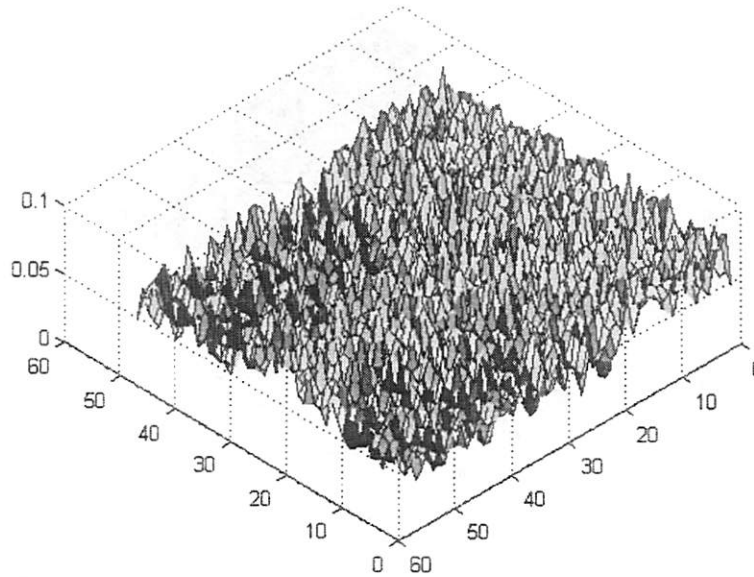


Figure 5.4 10% post-CMP ILD thickness variation observed

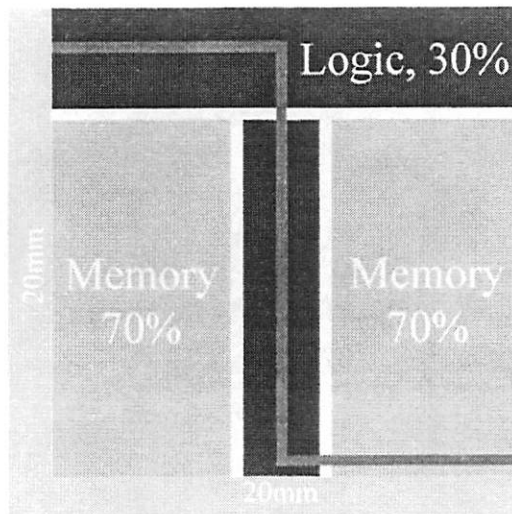


Figure 5.5 Pattern density difference on the mask between the logic and memory area

However, if the process engineers take a closer look into the patterns on the mask illustrated in Figure 5.5, the variations may be found to be mostly layout-dependent.

Eighty percent of the variations are systematic in the ILD CMP process due to the pattern density difference on the mask (Figure 5.6) while only a small part of the variation is really random (Figure 5.7).

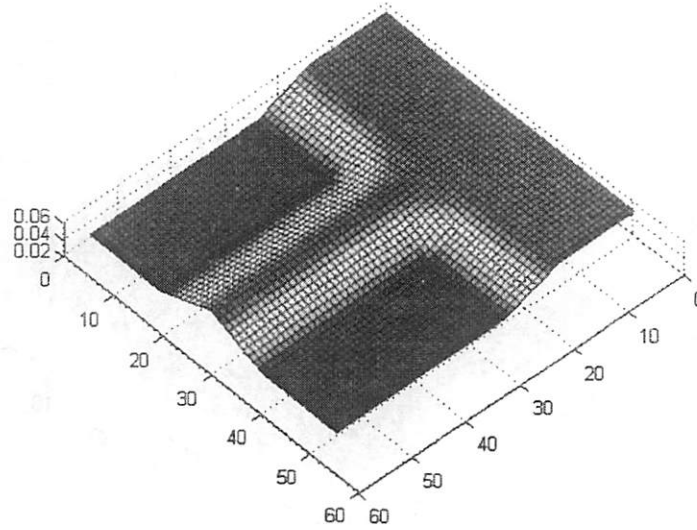


Figure 5.6 Systematic ILD thickness variations due to pattern density difference

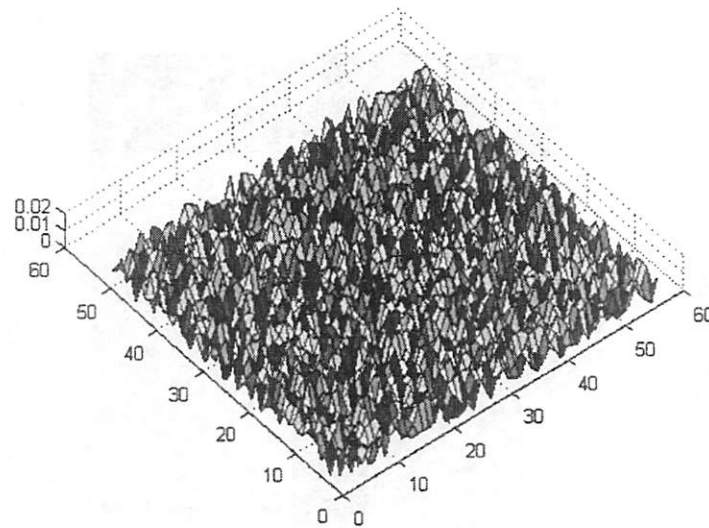


Figure 5.7 The truly random ILD thickness variations due to process uncertainty

The systematic component of the variation is repeatable and predictable in the process, which has the potential to be transformed into information to facilitate the circuit

designers. In this work, I carried out some simulations with 128-stage inverters (Figure 5.8) using the Taiwan Semiconductor Manufacturing Company (TSMC) 0.18 μ m technology device model and technology library. The simulation flowchart is shown in Figure 5.9.

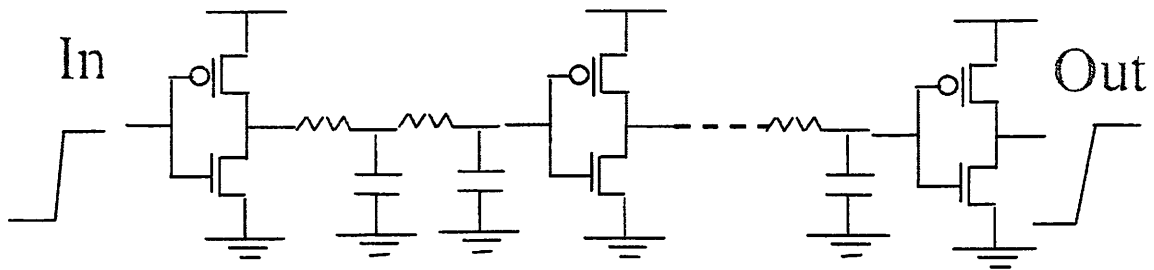


Figure 5.8 128-stage inverter simulations with long interconnect

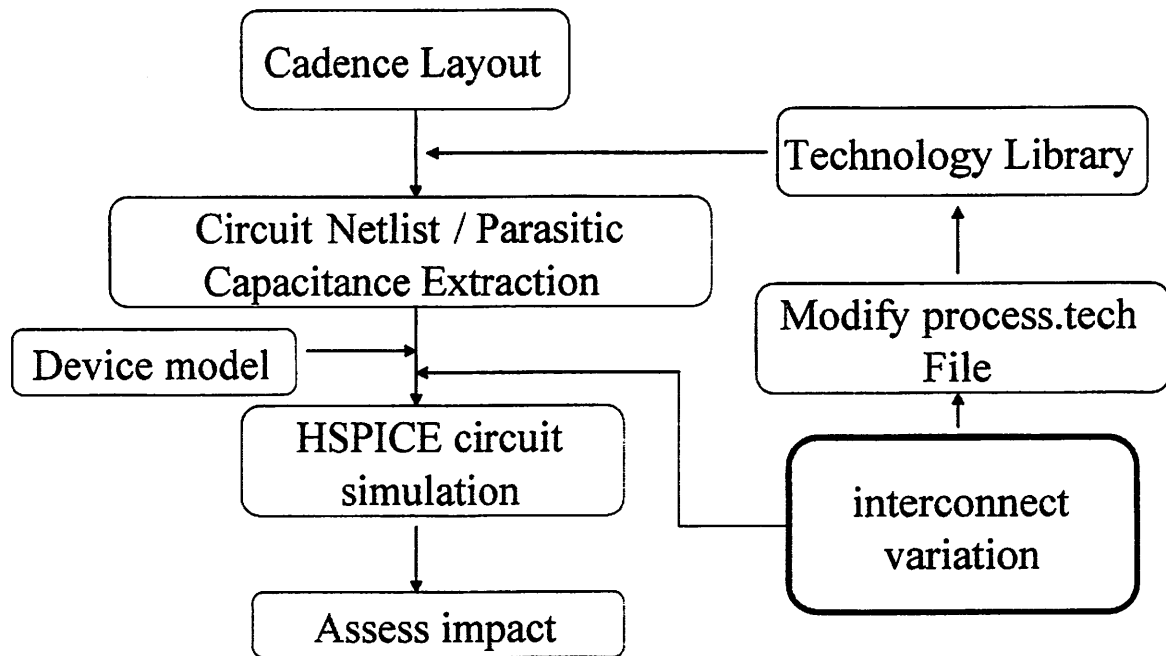


Figure 5.9 Simulation flowchart

The detailed layout of one inverter cell is shown in 5.10. The poly gate widths of the inverters were carefully sized to better deliver the signal from one end to the other.

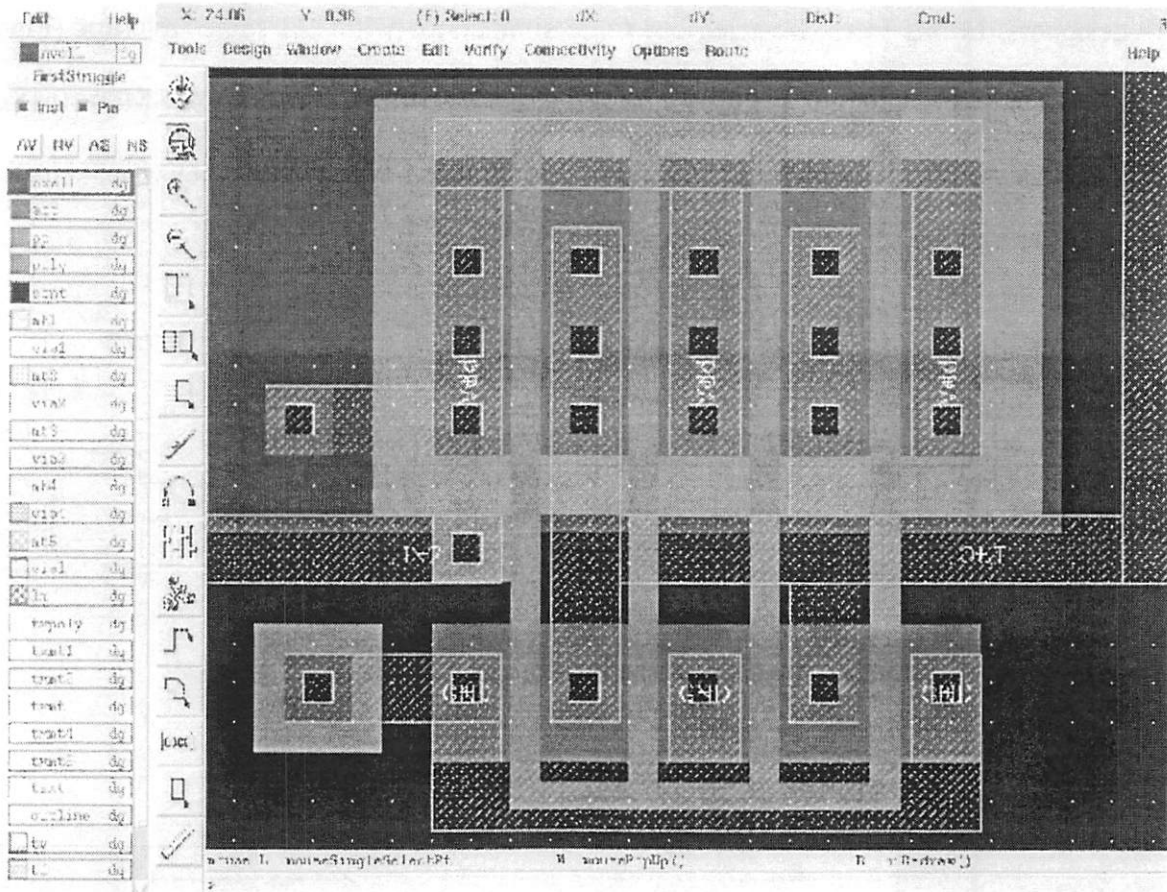


Figure 5.10 Basic inverter layout (0.18µm TSMC technology)

The process variation information was integrated into the simulation spice deck in two ways: first is to modify the “process.tech” file, which will allow users to change the oxide/metal thickness and dielectric constant information for the whole layer, second is to change the extracted netlist based on the ILD/metal thickness predictions from the CMP models developed in previous chapters. Three total cases were studied: first is the “ideal case” which means there is no ILD/metal thickness variation in the process; second is the ‘worst case’ scenario which assumes that every metal wire will experience 10% loss due to the variations in CMP process; the last is the “reality case” when the spatial layout-dependent prediction information for the wires on the critical path was integrated

to the simulation netlist. Figure 5.11 showed the circuit delay for the rising signal (the solid line) in the first case and the delays for the three cases are tabulated in Table 5.1.

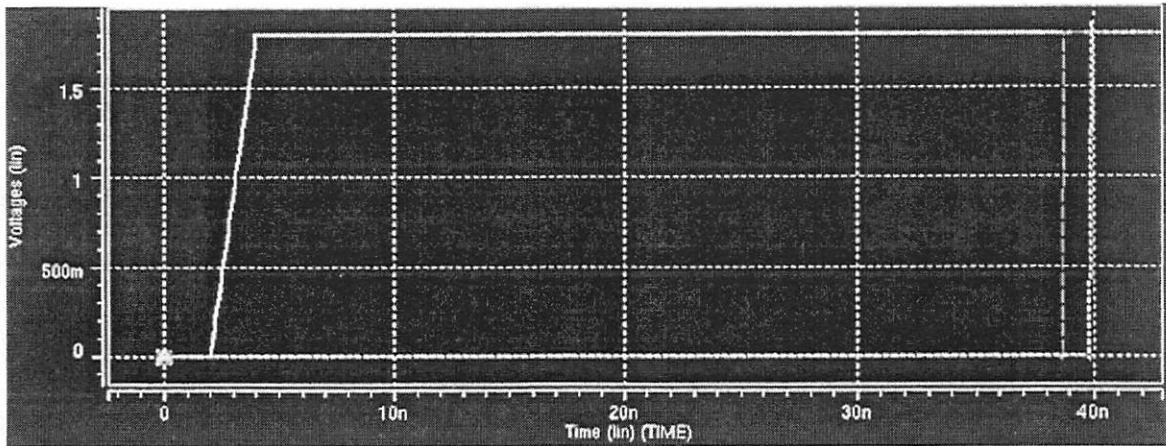


Figure 5.11 Circuit delay for the ideal (no CMP process variation) case

Table 5.1 Circuit delays for three cases

Case Name	Circuit delay-Rising	Circuit delay-Falling	Average Delay
Ideal (no loss)	35.9ns	36.2ns	36.05ns
Worst case (10% random)	38.7ns	38.9ns	38.8ns
Real case (8% systematic + 2% random)	36.5ns	36.6ns	36.55ns

The Real case which decouples the systematic component has a more accurate prediction over the traditional worst case prediction and it allows designers to push the circuit performance to a higher limit. The integration of the process information into a modern CAD tool is beyond the scope of this thesis. The bottom line, however, is to provide precise interconnect performance information based on the CMP process, which is the focus of the next section in this chapter.

5.3 Model-Based Interconnect Performance Simulation Results

According to the 2003 International Technology Roadmap for Semiconductor, the average active wire length on a typical microprocessor chip has reached the 10km/cm² mark [18]. For the two problems existed in copper damascene, while ILD erosion is widely studied and well understood, the modeling and impact assessment of metal dishing has not been done in a quantitative and practical way. As shown in Figure 5.12, copper dishing costs significant amount of metal volume loss even for narrow sub-micron lines [19]. Based on the modeling work presented in Chapters 3 and 4, in this session I carry out extensive simulations for copper interconnect in the presence of copper dishing and discuss its impact on design for manufacturability in BEOL process. To illustrate this connection, the overall flow is shown in Figure 5.13.

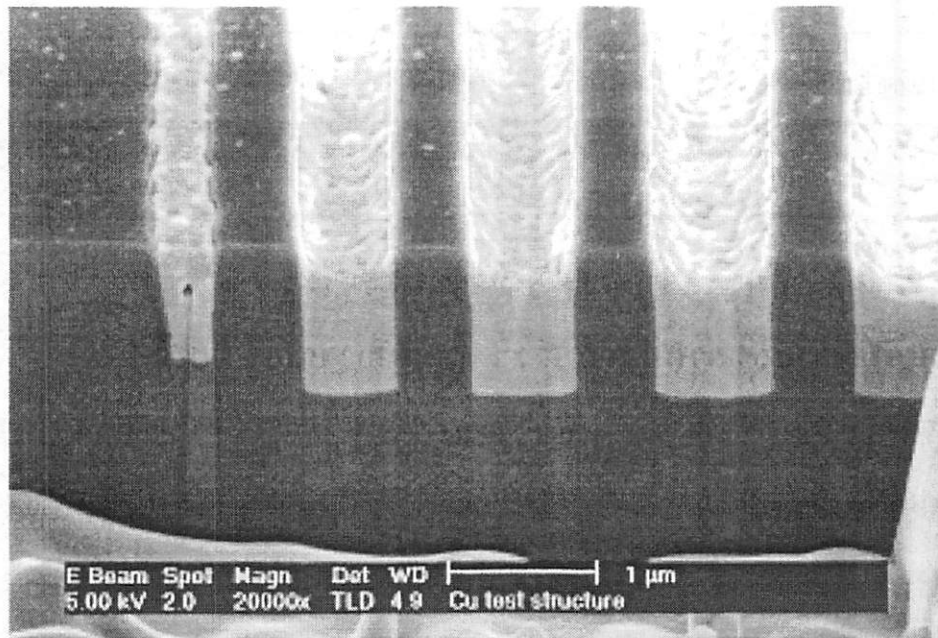


Figure 5.12 Copper dishing on deep sub-micron metal lines (picture courtesy of Technical University of Dresden, Germany)

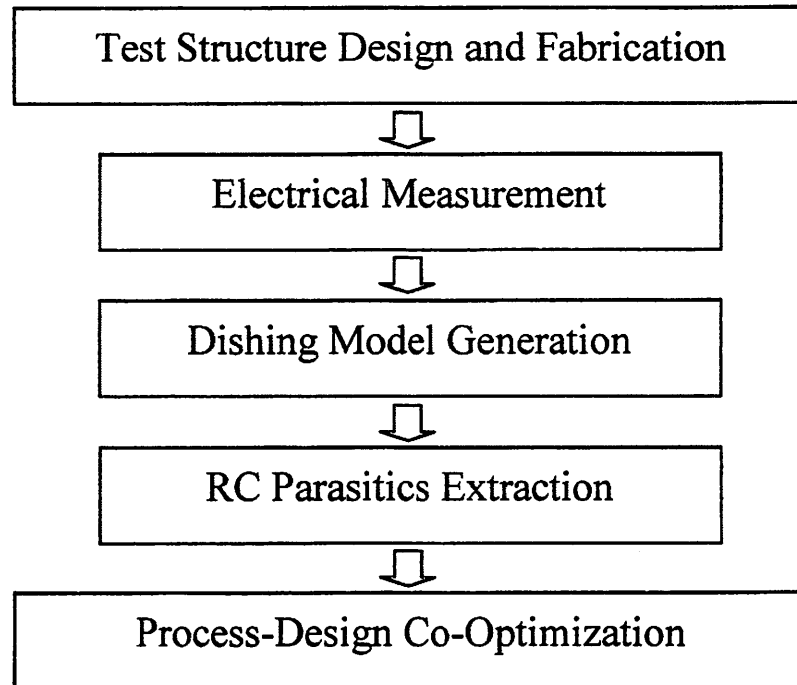
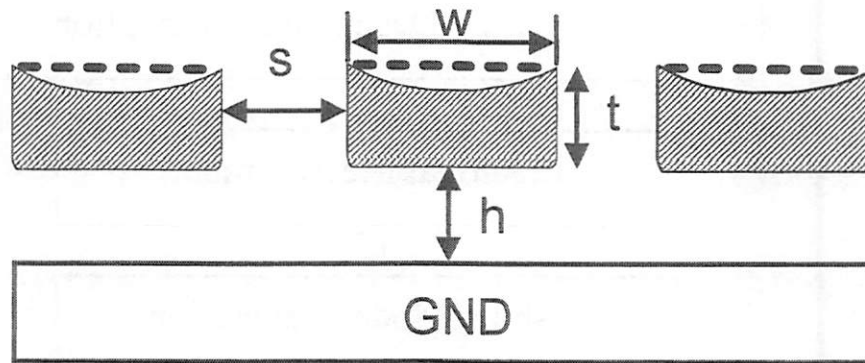


Figure 5.13 Overall process flow for the investigation of the metal dishing impact on interconnect performance

The simulations are done with the commercially available tool Raphael™. It is essentially a collection of 2D and 3D field solvers and interfaces providing the ability to obtain interconnect models for the designers to achieve on-chip signal integrity [20]. It is a slow simulator but has the merit of rigorously simulating the resistance, inductance and capacitance for modern interconnect structures.

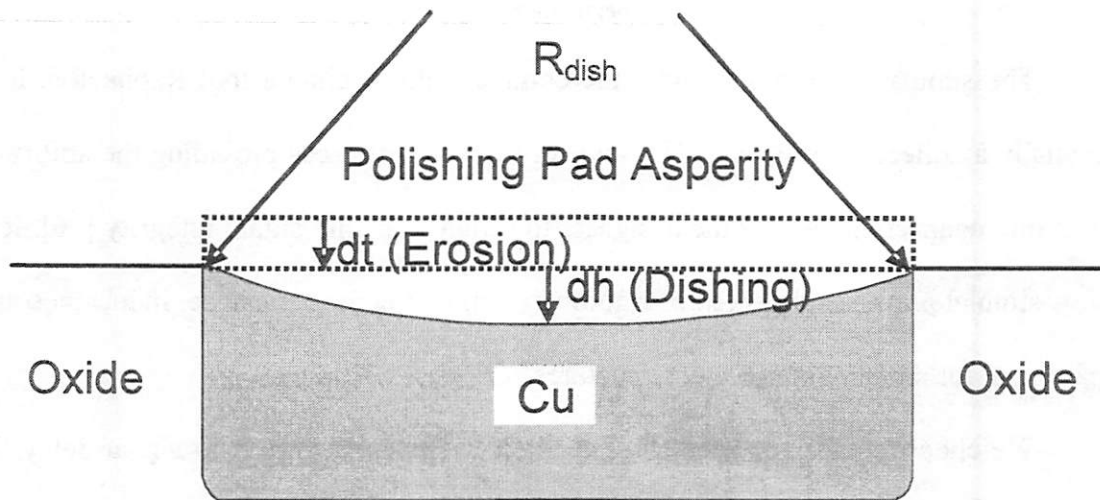
We choose to focus on the assessment of the dishing impact to signal delay for global interconnect, since global layers typically involve wider metal lines for power and clock distribution. These wide lines will suffer more dishing based on the modeling work that was presented in the previous chapter. The basic global interconnect structures are shown in Figure 5.14 and the dishing model that I used to generate the simulation profiles are shown in Figure 5.15.



- w: metal line width;
- t: metal line thickness (default=0.5 μm);
- s: metal line spacing (default=0.5 μm);
- h: dielectric thickness(SiO₂, default=0.5 μm);

Figure 5.14 Interconnect structure used for the investigation of the metal dishing impact on

performance



- R_{dish} : effective radius of pad asperity;
- dt: metal thickness loss due to erosion;
- dh: metal nonplanarity due to dishing;

$$\frac{\rho \cdot length}{R} = w(t - dt) + \frac{wR_{dish}}{2} \sqrt{1 - \left(\frac{w}{2R_{dish}}\right)^2} - R_{dish}^2 \sin^{-1}\left(\frac{w}{2R_{dish}}\right)$$

Figure 5.15 Metal dishing model used to generate the interconnect profiles in simulation

In order to evaluate the impact of metal dishing, the RC product is employed as a performance metric that can be easily related to interconnect layout specifications, such as line width and space. Although more complicated metrics, such as RLC delay or full waveform models may be more accurate, the simple RC product precisely predicts the correct performance scaling with line size tuning and is suitable for design optimizations at the early stage [21-22]. In terms of scalability, the RC and RLC metrics are similar for typical circuits operating in the Giga-Hertz range [23-25].

Based on R_{dish} model in Figure 5.15. Figure 5.16 shows the Raphael simulation results for resistance, given various dishing radius and line width values. The simulations showed that the resistance increases significantly for wider lines and smaller R_{dish} . Therefore, it is necessary to incorporate the dishing effect in resistance calculation (refer to the equation in Figure 5.15).

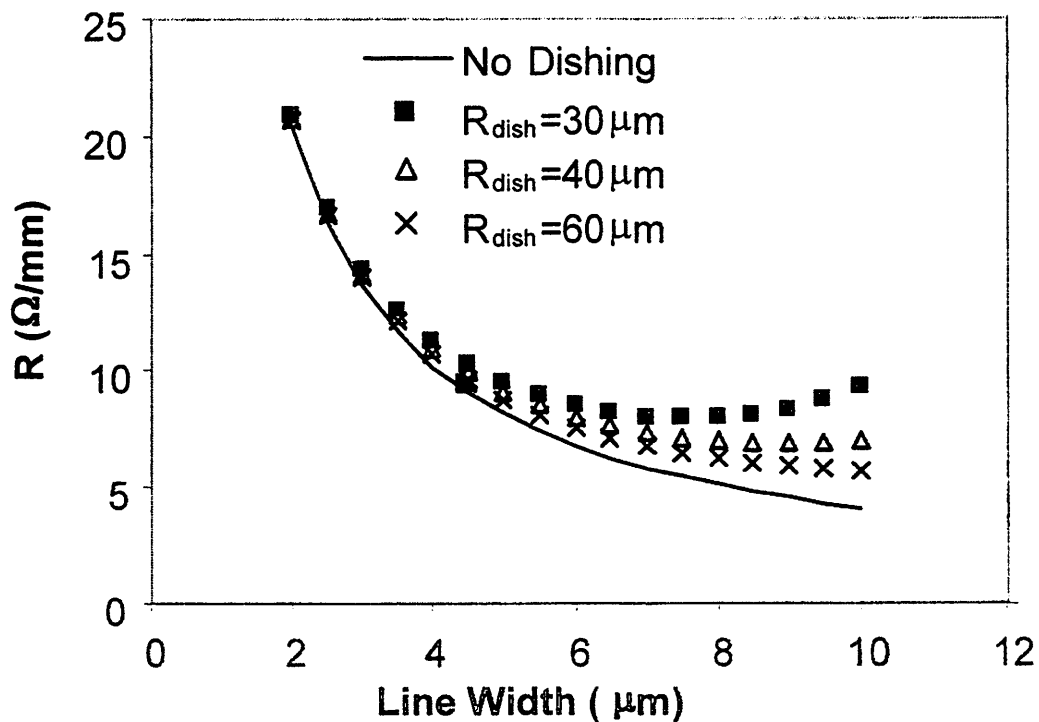


Figure 5.16 Dependence of metal line resistance on dishing radius as a function of metal line width

Based on the dishing model in Figure 5.15, Figure 5.17 shows the Raphael simulation results for capacitance, given various dishing radius and line width values. The dishing radius is chosen to be $30\mu\text{m}$, $40\mu\text{m}$ and $60\mu\text{m}$ and line widths are ranged from $2\mu\text{m}$ to $10\mu\text{m}$. The simulations show that C_{total} is relatively insensitive to the dishing conditions in comparison to the resistance case. This is because the neither the parallel plate capacitance to the ground plate, the coupling capacitance to the neighboring lines on the same layer, nor the fringing capacitance to the ground plate is sensitive to the surface shape change due to dishing.

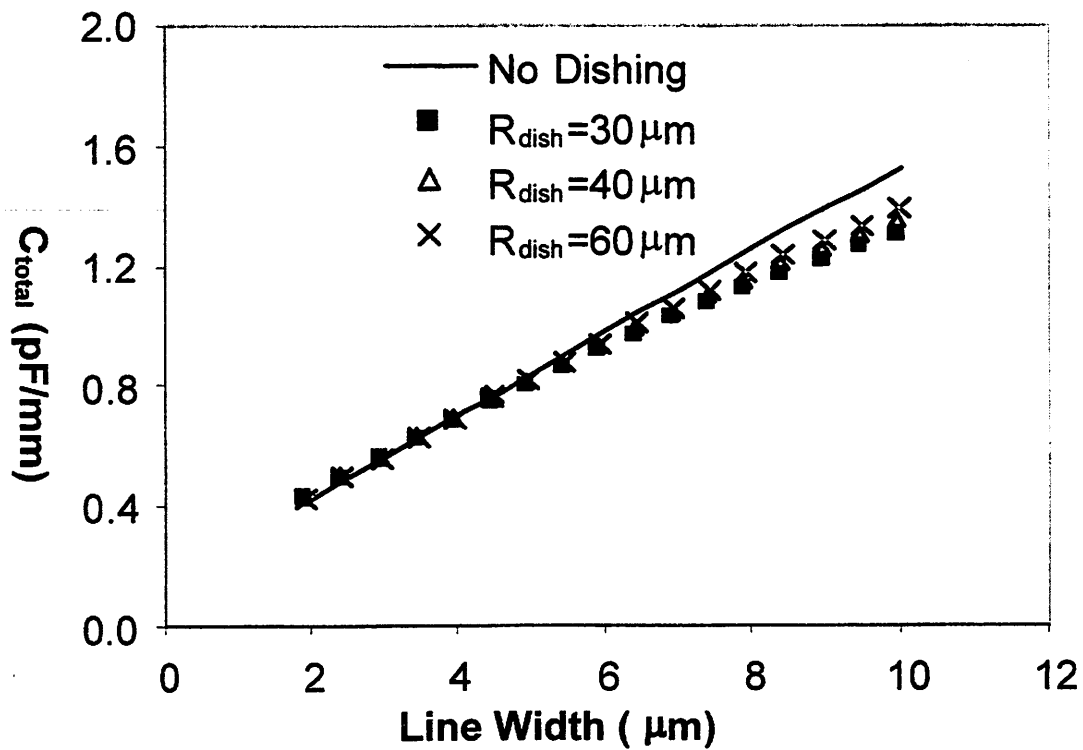


Figure 5.17 Dependence of metal line capacitance on dishing radius as a function of metal line width ($C_{\text{total}} = C_{\text{ground}} + 2 * C_{\text{coupling}}$)

Figure 5.18 shows the RC delay as a function of dishing conditions and linewidth. In the case there is no dishing (a perfectly flat post-CMP metal surface), RC delay

monotonically decrease as a result of widening the linewidth. However, when there is severe dishing in the CMP process, say the dishing radius is $60\mu\text{m}$ or $30\mu\text{m}$, increasing the metal linewidth may lead to more RC delay because of the rapid increase of resistance as a consequence of dishing. In our experiments, the minimum delay is achieved at the optimal width $w_{\text{opt}} \sim 4\mu\text{m}$ (note that for an ideal chemical mechanical planarization process without dishing, w_{opt} is infinite).

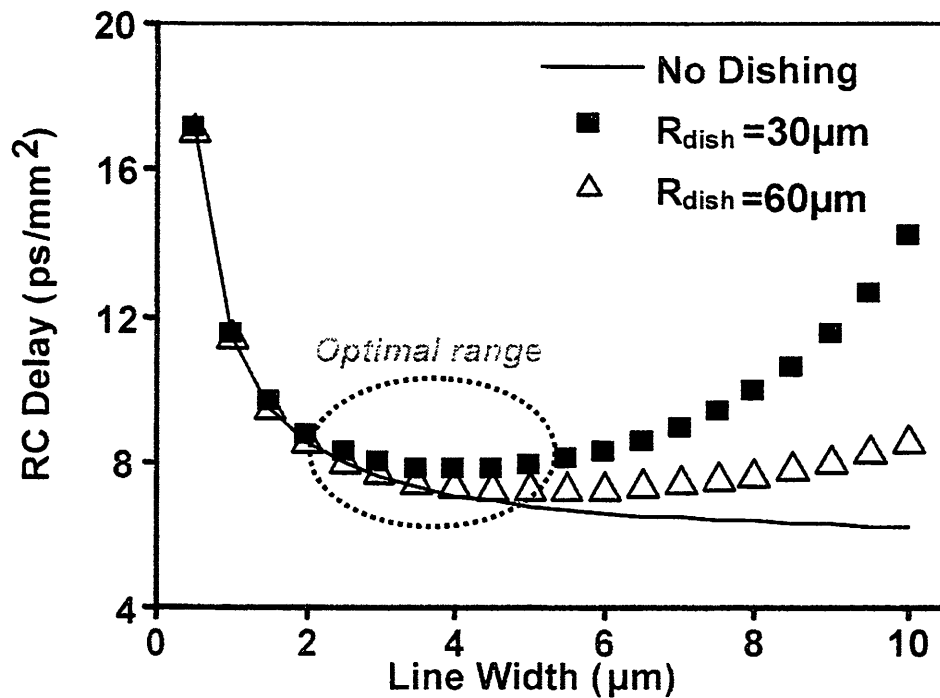


Figure 5.18 RC delays as a function of line width with dishing (the optimal linewidth to achieve minimum RC delay is around 4 microns)

Furthermore, I implement the sensitivity simulations in order to investigate the effect of the line width variations on RC delay. The copper line widths are varied by plus and minus 20% and the average of $|\Delta\text{RC}|$ are plotted in Figure 5.19. As we should expect, the RC delays appear to be insensitive to metal line width variations. This supports the optimal linewidth range as shown in Figure 5.18. In the case of no dishing, $|\Delta\text{RC}|$

becomes smaller as line width gets larger. This is intuitive because the change in resistance will be countered by that in parallel plate capacitance to the ground plane. However, when dishing becomes an issue (in this case when $R_{\text{dish}} = 40\mu\text{m}$), from the circuit designers perspective, using the optimum line width (in this case $w_{\text{opt}} = 4\mu\text{m}$) will make the circuit more robust against line width variations due to lithography, etching or other steps in mask pattern transfer.

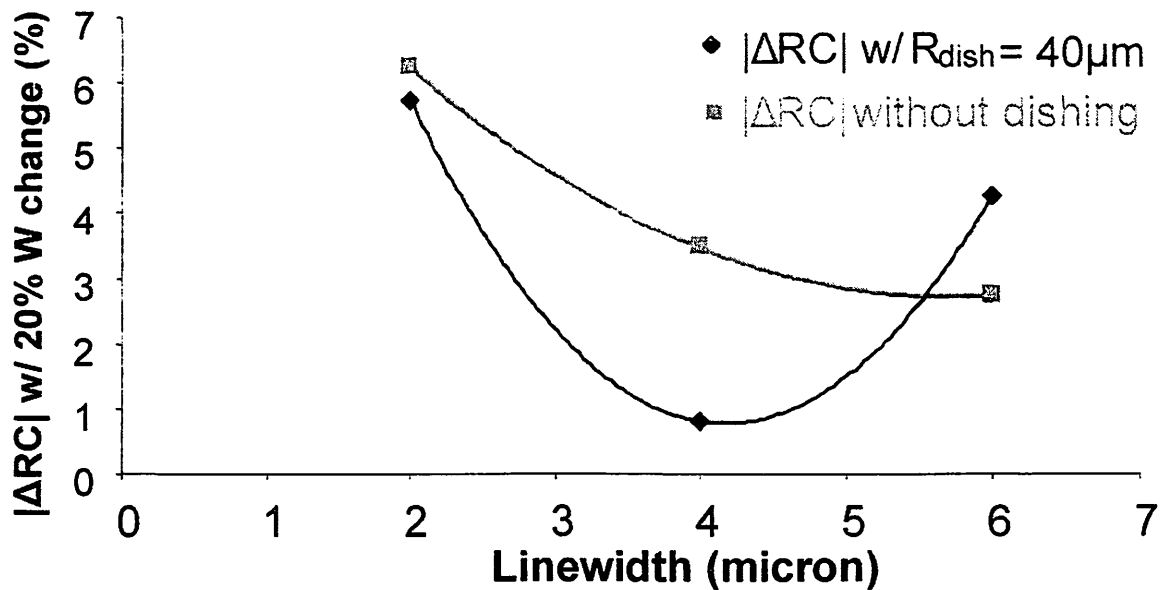


Figure 5.19 RC delay sensitivity (20% linewidth change) as a function of metal line width

The dependence of the optimal linewidth on dishing conditions is further studied and the results are illustrated in Figure 5.20. The better dishing condition (larger dishing radius R_{dish}) will increase the optimal linewidth to achieve the minimum signal delay. The effect of dishing on the optimal linewidth can be mitigated by using thicker interlayer dielectric or increase the metal layer thickness, as shown by the square box and cross lines in Figure 5.20. The latter method is straightforward because it simply decreases the

percentage of dishing in metal volume loss, which allows designers to use wider lines to achieve lower signal delay. However, increasing the metal layer thickness may introduce integration issues [26-28] in terms of uniform barrier layer deposition, etch micro-loading and step coverage problems due to the higher aspect ratio (AR) especially for the narrower lines on the same metal layer.

Overall the inclusion of the dishing effect restricts the design space and increases the complexity of the interconnect performance analysis. However, based the significance of dishing effect that we observe in these simulations, adding dishing effect into consideration at the design stage has the potential to provide more accurate performance results.

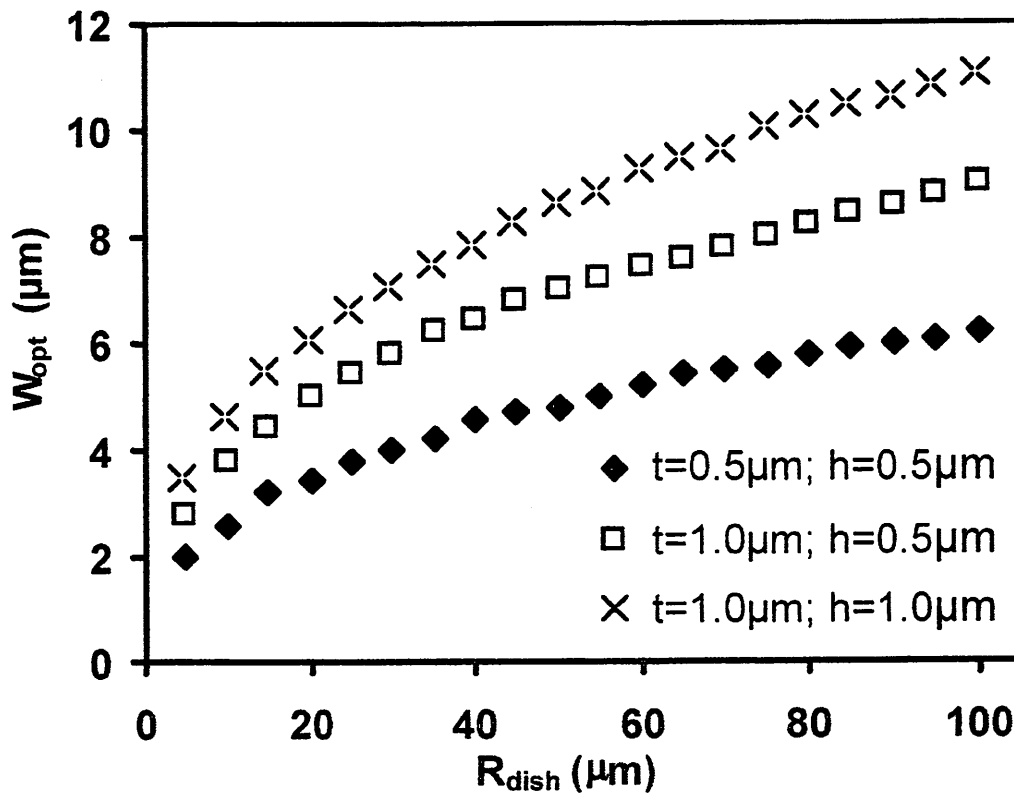


Figure 5.20 Optimal linewidth as a function of dishing radius

The delay penalty from dishing can be mitigated either by improving the CMP process or by physical layout techniques at the design stage. These solutions are more essential at future technology nodes as the timing budget becomes tighter. For instance, metal-filling is a commonly used technique to reduce the intra-die variations in metal planarity. By inserting small metal islands in a blank area, metal pattern density can be balanced, and thus dielectric erosion is mitigated.

To suppress metal dishing, various attempts have been made to improve the quality of the CMP process (i.e., increase R_{dish}). One of the methods developed in recent years focuses on better slurry design: after the first-step slurry, which has high copper removal rate, another type of slurry is used when polishing reaches the barrier layer. The second-step slurry is selective in that it has higher removal rate for the barrier layer, while the polishing rate is much lower for copper and dielectrics. In future advanced BEOL technologies, where more complex materials will be involved, this multi-step slurry strategy is crucial in controlling the metal planarity. Besides slurry optimization, other methods, including spindle engineering and polishing pad design, are also helpful in reducing the variation of metal thickness.

With better process control, R_{dish} can be increased, reducing the impact of dishing on signal delay. On the other hand, the efficiency of these approaches diminishes when R_{dish} goes further up. As illustrated in Figure 5.21, I simulate the difference in RC delay between an ideal process (without dishing) and the realistic case (with dishing) under various dishing conditions for 4-micron metal lines, which we have shown in Figure 5.8 to be the center of the optimal linewidth range. For a variety of line thicknesses, the results show that it is only favorable to suppress dishing via process control when R_{dish} is

less than 50 μm . Beyond that point, the gain of dishing radius increase becomes negligible. Thus, for practical concerns, about 50 μm dishing radius is the upper-limit of CMP process enhancement from the design perspective.

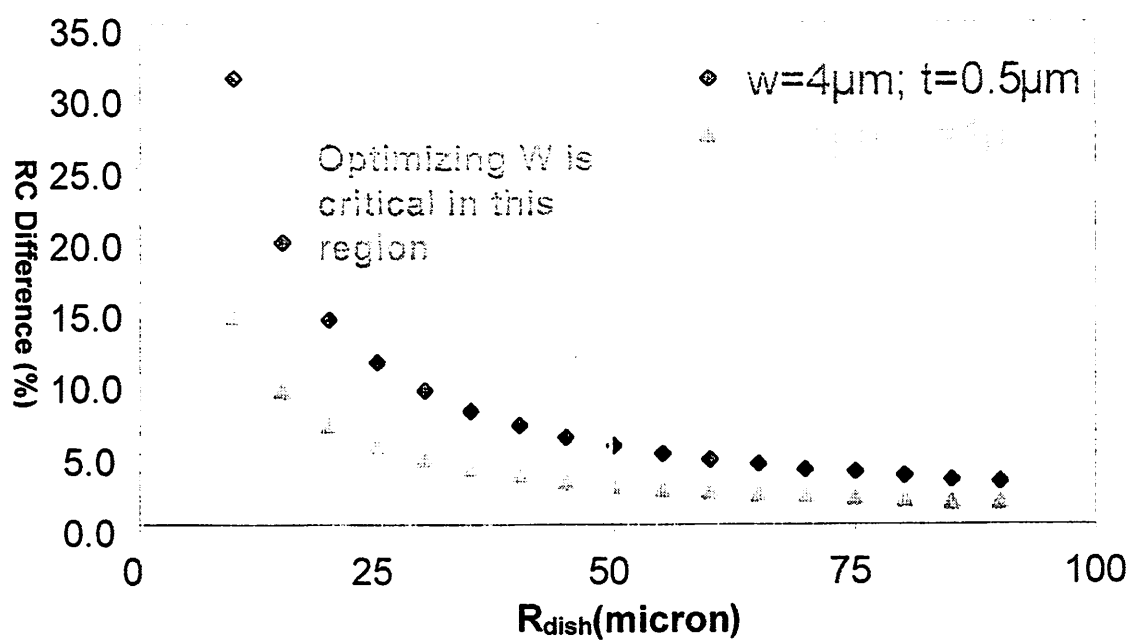


Figure 5.21 Efficiency of process improvement for different metal thickness as a function of dishing radius

Besides process improvement, the dishing effect can also be mitigated by layout design techniques, such as inserting holes into a wide line (metal-drilling) or dividing a wide metal line into narrower lines (line-splitting, as shown in Figure 5.22). These techniques can alleviate the performance penalty due to dishing, since narrower metal segments are more robust to dishing (Figure 5.18). Furthermore, the application of narrower lines also enhances thermal robustness of high-speed interconnects [29-31]. In the case of wide copper lines, either metal-drilling or line-splitting increases the aspect ratio of the wire. Hence, the surface-area-to-volume ratio is improved, resulting in more efficient heat dissipation.

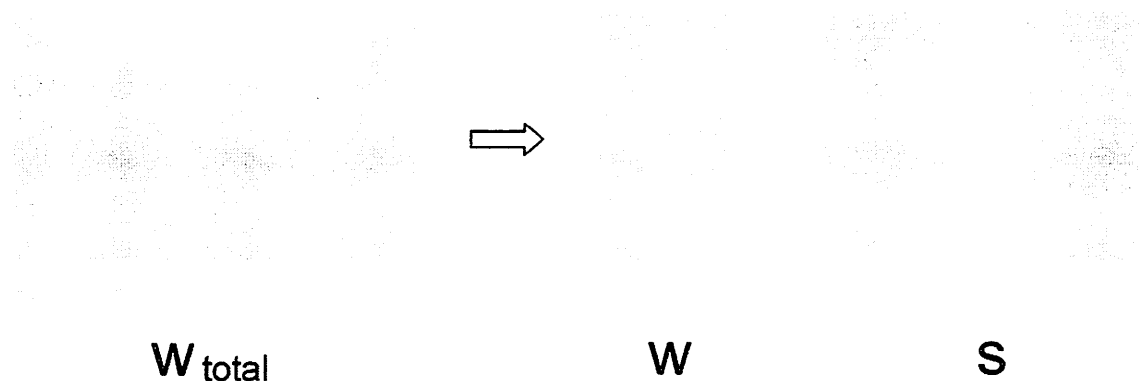


Figure 5.22 Illustration of the line-splitting idea ($W = W_{total}/N$, N is the number of lines; $s=s_{min}=0.5\text{micron}$)

In terms of the impact on interconnect performance, metal-drilling and line-splitting are electrically similar to each other, and thus we only focus on line-splitting in this thesis. For a fixed pattern density (hence the same expected erosion), the larger the number of split lines (N), the smaller the impact of dishing, as illustrated in Figure 5.23. The tradeoff of this approach is the extra cost of chip area: area cost goes up linearly with increasing N , assuming minimum space is applied between narrower lines. Figure 5.23 further indicates that the performance gain via line-splitting drops to a negligible level when N exceeds five. Therefore, considering the area cost penalty during splitting, a practical approach may only require N values between 2 and 4. In addition to area concerns, line-splitting also introduces extra fringing capacitances.

Figure 5.24 demonstrates that when the line-splitting number N is more than 5, further splitting leads to a delay penalty from the larger C_{total} (mainly due to the rapid increase in fringing capacitance due to more sidewall area), which almost equals the RC penalty due to the dishing of the original wide line.

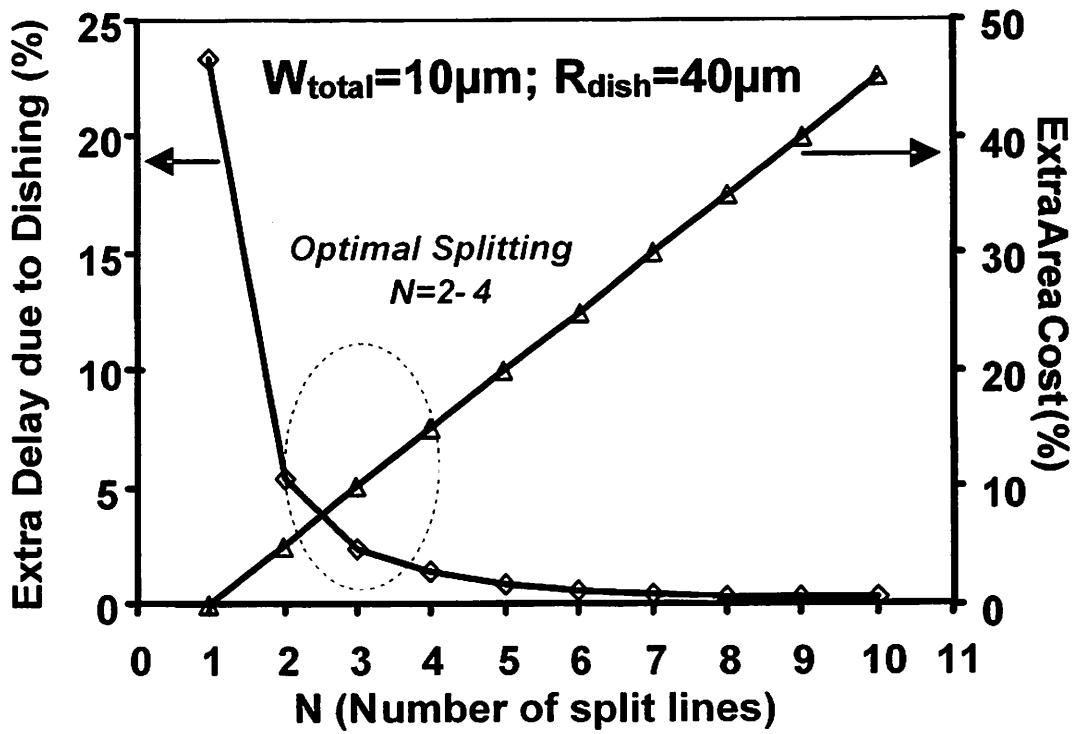


Figure 5.23 RC delay gain and area penalty tradeoff

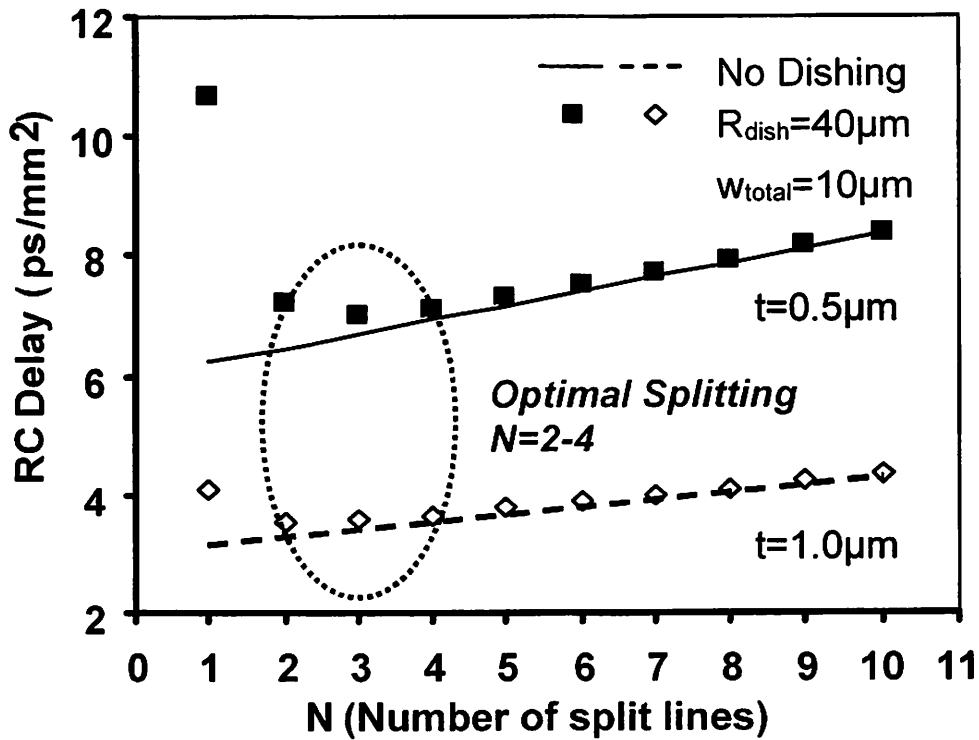


Figure 5.24 Optimization of RC delay using the line-splitting idea

In conclusion, based on the results from extensive interconnect performance simulations, from both efficiency and performance considerations, line-splitting number N between 2 and 4 is optimal for typical on-chip interconnects. This assumes that systematic layout-dependent CMP process variations and tight aspect ratio control will continue to be the concerns in the semiconductor scaling roadmap. I recognize that CMP process technologists and slurry chemistry scientists manage to find ways to mitigate dishing and erosion in the last few years. However, with the aggressive scaling of low- k dielectrics, the back-end-of-the-line integration process will be demanding down the road. The methodologies presented in this thesis work will be valid and useful for the process and design trailblazers in the future years.

References:

- [1] J. Khare, D. B. I. Feltham and W. Maly, "Accurate Estimation of Defect-Related Yield Loss in Reconfigurable VLSI Circuits," *IEEE Journal of Solid State Circuits*, vol. 28, no. 2, pp. 146- 156, February 1993.
- [2] T. L. Michlaka, R. C. Varshney and J. D. Meindl, "A Discussion of Yield Modeling with Defect Clustering, Circuit Repair, and Circuit Redundancy," *IEEE Transactions on Semiconductor Manufacturing*, vol. 3, no. 3, pp. 116- 127, August 1997.
- [3] P. Mullenix, J. Zaloski and A. J. Kasten, "Limited Yield Estimation for Visual Defect Sources," *IEEE Transactions on Semiconductor Manufacturing*, vol. 10, no. 1, pp. 17- 23, February 1997.

- [4] R. K. Nurani, A. J. Strojwas, W. P. Maly, C. Ouyang, W. Shindo, R. Akella, M. G. McIntyre and J. Derrett, "In-Line Yield Prediction Methodologies Using Patterned Wafer Inspection Information," IEEE Transactions on Semiconductor Manufacturing, vol. 11, no. 1, pp. 40-47, February 1998.
- [5] C. Ouyang and W. Maly, "Efficient Extraction of Critical Areas in Large VLSI ICs," Proc. Int. Symp. Semiconductor Manufacturing (ISSM 96), Tokyo, Japan, 1996, pp. 301-304.
- [6] <http://www.mosis.org/design/flows/design-flow-digital.html>
- [7] P. J. Silverman, "Who can afford advanced lithography," Solid State Technology, Nov. 2003.
- [8] R. Wilson, "Chip industry tackles escalating mask costs," EE Times, June 17 2002.
- [9] C. R. Helms, "Semiconductor Technology Research, Development, and Manufacturing: Status, Challenges, and Solutions," April 2003.
- [10] S. Wong, G. Lee, and D. Ma, "Modeling of interconnect capacitance, delay, and crosstalk in VLSI," IEEE Transactions on Semiconductor Manufacturing, vol. 13, no. 1, pp. 108-111, Feb. 2000.
- [11] A.K. Stamper et al, "Sub-0.25-micron interconnection scaling: damascene copper versus subtractive aluminum," Proceedings of IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop, pp. 337 - 346, September 1998.
- [12] Lin, Z.; Spanos, C.J.; Milor, L.S.; Lin, Y.T., "Circuit sensitivity to interconnect variation," IEEE Transactions on Semiconductor Manufacturing, Volume: 11 , Issue: 4 , Nov. 1998, Pages:557 - 568

- [13] E. D. Boskin, "A methodology for modeling the manufacturability of integrated circuits," Ph.D. dissertation, Univ. California, Berkeley, 1995.
- [14] D. E. Hocevar, P. F. Cox, and P. Yang, "Computing parametric yield accurately and efficiently," in Proc. ICCAD, 1990, pp. 116–119.
- [15] J. C. Zhang and M. A. Styblinski, "Design of experiments approach to gradient estimation and its application to CMOS circuit stochastic optimization," in Proc. ISCAS, Singapore, 1991, pp. 3098–3101.
- [16] Z. Daoud, "DORIC: Design of optimized and robust integrated circuit," M.S. thesis, Univ. California, Berkeley, Dec. 1993.
- [17] R. Streiter, H. Wolf, U. Weiss, X. Xiao, T. Gessner, "Optimization of ULSI interconnection systems including aerogels by thermal and electrical simulation," Proc. Advanced Metallization Conference 1999 (AMC 1999), Orlando, Florida, Sept. 28-30, 1999
- [18] The International Semiconductor Roadmap for Semiconductors, International SEMATECH, 2003.
- [19] C. Wenzel, "Diffusion Barriers and Copper Line Processes," 2000 TU Dresden IHM-Workshop.
- [20] http://www.synopsis.com/products/mixedsignal/raphael_ds.html
- [21] R. Chang, Y. Cao, and C. Spanos, "Modeling metal dishing effect in chemical-mechanical polishing process for on-chip interconnect optimization," submitted to IEEE Transactions on Electron Devices.

- [22] Y. Cao, X. Huang, D. Sylvester, T.-J. King, and C. Hu, "Impact of on-chip interconnect frequency-dependent $R(f)L(f)$ on digital and RF design," submitted to IEEE Transaction on VLSI Systems.
- [23] T. Sato, Y. Cao, K. Agarwal, D. Sylvester, and C. Hu, "Bi-directional closed-form transformation between on-chip coupling noise waveforms and interconnect delay change curves," IEEE Transaction on Computer-Aided Design of Integrated Circuits and Systems, vol. 22, no. 5, pp. 560-572, 2003.
- [24] Y. Cao, R. A. Groves, N. D. Zamdmer, J. Plouchart, R. A. Wachnik, X. Huang, T. King, and C. Hu, "Frequency-independent equivalent circuit model for on-chip spiral inductors," IEEE Journal of Solid-State Circuits, vol. 38, no. 3, pp. 419-426, Mar. 2003.
- [25] X. Huang, P. Restle, T. Bucelot, Y. Cao, and T. King, "Loop-based interconnect modeling and optimization approach for multi-GHz clock network design," IEEE Journal of Solid-State Circuits, vol. 38, no. 3, pp. 457-463, Mar. 2003.
- [26] C. H. Stapper, "Modeling of Integrated Circuit Defect Sensitivities," IBM J. Res. Develop, vol. 27, no. 6, pp. 549- 557, November 1983.
- [27] Dennis J. Ciplickas, Xiaolei Li, Rakesh Vallishayee, Andrzej J. Strojwas, Randy Williams and Raman Nurani, "Predictive Yield Modeling for Reconfigurable Memory Circuits," Proc. Of ASMC '98, pp. 1- 6, Boston MA, September 1998.
- [28] C. Hess, D. Stashower, B. E. Stine, G. Verma and L. H. Weiland, "Fast Extraction of Killer Defect Density and Size Distribution Using a Single Layer Short Flow NEST Structure," Proc. of the 2000 ICMTS, pp. 57- 62, Monterey, CA, March 2000.
- [29] H. B. Bakoglu, Circuit, Interconnections, and Packaging for VLSI, Addison Wesley, 1990.

- [30] Y. Cao, X. Huang, N. Chang, S. Lin, O. S. Nakagawa, W. Xie, D. Sylvester, and C. Hu, "Effective on-chip inductance modeling for multiple signal lines and application on repeater insertion," *IEEE Transaction on VLSI Systems*, vol. 10, no. 6, pp.799-805s, Dec. 2002.
- [31] T. Sato, D. Sylvester, Y. Cao, and C. Hu, "Accurate in-situ measurement of peak noise and signal delay induced by interconnect coupling," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 10, pp. 1587-1591, Oct. 2001.

Chapter 6 Conclusions and Future

Work

6.1 Conclusions

Moore's Law has been described the growth of the semiconductor industry for more than 35 years [1-2]. An exponential scaling factor in most aspects of integrated circuits performance keeps the industry on its historic 25%/year reduction in the cost/function figure, despite the escalating factory costs (>20%/year). However, this aggressive scaling relies on an effective and reliable back-end-of-the-line integration of the inter-layer dielectric and metal layers. Advanced CMP process modeling and affordable metrology at and below 65nm now become the next big challenge. Greater synergy must be developed between the areas of metrology, modeling, simulation and process optimization.

This thesis has presented a framework that integrates the metrology of copper damascene process observables with physical and analytical models for the same, in order to facilitate the multiple objective optimization of a copper CMP process as well as to achieve enhanced predictive interconnect performance.

The Dishing Radius is a novel performance metric that is first introduced in this thesis. With specially designed patterns, copper damascene processes were completed at the Berkeley and RPI micro-fabrication laboratories. Multiple metal lines with typical on-chip interconnect widths are polished with randomized design of experiment sets. The dishing radius concept is found to precisely link the amount of dishing with the metal linewidth. Model-based metrology using electrically tested data points was developed for the copper damascene process. The extracted profiles matched well with those from cross-sectional SEM and high-resolution surface profiler.

This thesis also presents an efficient methodology for multiple objective optimizations in copper CMP process. First we develop the scheme for building process models to link the key process inputs and outputs. By applying the classical Taguchi philosophy, an optimization structure was set up to balance the different performance metrics. A couple of cases were studied using the theory developed on this aspect. This optimization framework provides a hierarchical method for determining the tradeoffs in the ramping-up stage state-of-the-art semiconductor fabrication processes.

Finally, the impact of copper CMP process variations on interconnect performance is quantified in this thesis through extensive simulations. Moreover, techniques that can be implemented at the design stage in order to mitigate the implications of dishing are proposed based on simulation results. This work attempts to provide remedies for the layout dependent process variations such as dishing in copper damascene process. Both the theoretical analysis and simulation results support our conclusions.. The simulation results show that dishing will be a concern for global layer interconnects if the dishing radius is less than $50\mu\text{m}$. This thesis closes with a discussion

on one application of the Design for Manufacturability concept -- the tradeoffs between the die area, BEOL yield and interconnect performance.

6.2 Future Work

This work has focused on the building blocks for the integrated metrology and modeling of the copper damascene process. Future research topics can be the following:

As the technology requires about 30% performance improvement from interconnect for each new generation, the introduction of low-k dielectric materials (<2.5) becomes an enabling factor. This will certainly bring more research topics on the integration of low-k and copper CMP process which will be featured by ultra-low down force (less than 2psi) and high speed (>100 rpm).

In terms of optimization, to achieve more predictable interconnect and circuit performance, the effect of via resistance variation and the shape imperfectness introduced by the dielectric etch can not be ignored. Also the increasing percentage of the barrier layer in the metal line cross-section should be taken into account for sub-90nm technology generations.

There will be more opportunities in the BEOL Design for Manufacturability. Computer aided integrated circuit design experts are required to learn more on the suppression of the process variability, which provides a unique opportunity for wider application of the methodologies developed in this thesis.

References:

[1] International Technology Roadmap for Semiconductors, International SEMATECH, Austin, TX, 2003.

[2] Chenming Hu, "Scaling CMOS Transistors For Another 25 Years," UC Berkeley EECS Department Solid State Seminar, February 25th, 2000.