Modeling of Chemical Mechanical Polishing for Shallow Trench Isolation

by

Brian Lee

Bachelor of Engineering, Electrical Engineering, Cooper Union, 1996
Master of Science, Electrical Engineering and Computer Science, MIT, 1998

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2002

Author.....
Department of Electrical Engineering and Computer Science
May 24, 2002

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

Accepted by.....
Arthur C. Smith
Chairman, Department Committee on Graduate Students
Department of Electrical Engineering and Computer Science

Modeling of Chemical Mechanical Polishing for Shallow Trench Isolation

Brian Lee

Submitted to the Department of Electrical Engineering and Computer Science
on May 24, 2002, in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

Chemical mechanical polishing (CMP) is a key process enabling shallow trench isolation (STI), which is used in current integrated circuit manufacturing processes to achieve device isolation. Excessive dishing and erosion in STI CMP processes, however, create device yield concerns. This thesis proposes characterization and modeling techniques to address a variety of concerns in STI CMP. Three major contributions of this work are: characterization and modeling of STI CMP processes, both conventional and nonconventional; layout optimization to reduce pattern-dependent dishing and erosion; and modeling of wafer nanotopography impact on STI CMP yield.

An STI CMP characterization method is combined with a chip-scale pattern-dependent model to create a methodology that enables tuning of STI CMP process models and prediction of post-CMP dishing, erosion, and clearing times on arbitrary layouts. Model extensions enable characterization of STI CMP processes that use nonconventional consumable sets, including fixed abrasive pads and high-selectivity silica-based and ceria-based slurries. Experimental data validates the accuracy of the model for both conventional and nonconventional processes.

Layout optimization techniques are developed that reduce pattern-density dependent dishing and erosion. Layout design modification is achieved through the use of dummy STI active areas and selective reverse etchback structures. Smart algorithms allow for optimal density distributions across the layout.

The effect of wafer nanotopography (height variations that exist on unpatterned silicon wafers) is explored, characterized, and modelled. A diagnostic tool for examining the impact of nanotopography on STI device yields is developed, based on contact wear modeling. An aggregate estimator for the combined effect of wafer-scale nanotopography and chip-scale pattern-dependent dishing and erosion is developed.

The techniques developed in this thesis can be used both for process optimization and for diagnosis and correction of potential problems due to layout, wafer and CMP process interaction. The characterization and modeling methods create a comprehensive set of tools for process characterization and post-CMP erosion and dishing prediction in STI processes.

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

Acknowledgments

It's been a long and amazing ride from the time I first set foot on the MIT campus in 1996. Although I can barely remember my first days here, I can hardly believe that six years have passed in the blink of an eye. I've been fortunate enough to be able to study in an excellent institution surrounded by a group of extremely intelligent people.

This work was completed with the benefit of the assistance of many people. In no particular order, I would like to thank:

- My thesis advisor, Professor Duane Boning, who has provided me the opportunity to perform research work in his group. He has been a constant source of ideas and encouragement over the years.
- The members of my thesis committee, Professor Jung-Hoon Chun and Professor Akintunde Akinwande.
- Past and present research group members, who have helped with various discussions: Okumu Ouma, Charles Oji, Angie Nishimoto, Taber Smith, Vikas Mehrotra, Terence Gan, Tamba Tugbawa, Rhadika Dutt, Tae Park, and Xiaolin Xie.
- Others whom I've shared offices and have helped lighten the mood a bit: Abraham Kim, Aaron Gower, David White, H.-K. Chung, and Han Chen.
- Dale Hetherington and David Stein at Sandia National Laboratories, for helping me get started and providing support and guidance along the way. Much of the experimental work in this thesis could not have been accomplished without their help, and for that I am extremely grateful. Dale and David have both served as mentors, teachers, and friends.
- Jeff David and Benjamin Bonner at Applied Materials, for helping with some early wafer-scale modeling work.
- Noel Poduje, Pat Hester, and Yong Xia from ADE Corporation, and Winthrop Baylies from BayTech Group, for their collaboration on the nanotopography work.
- John Valley and Chris Koliopoulos at ADE Phase-Shift for their contributions to the discussions on nanotopography.
- Laertis Economikos at IBM's East Fishkill facility, for his assistance on the fixed abrasive experiments that provided useful data in the study on nonconventional pads.

- PDF Solutions, Inc., for use of their SiCat pattern density computation tool, which provided immense assistance in the pattern-density based modeling work.
- SKW Associates, for their contribution of patterned wafers for the nanotopography work.
- The laboratory staff at MIT's Microsystems Technology Laboratories for their help with experimental work performed at MIT.
- The MTL computer gurus for their advice and assistance on the various computer-related issues that came up.

A special thanks to my significant other, Maria Chan, who has been a constant source of encouragement over the last year, and who has injected countless moments of joy and hope into my everyday life. And of course, I would not be here if not for the most important people, my family. To my parents, I owe everything. To my siblings, I am grateful for their love and support. I dedicate this work to my family and my friends, all of whom have had parts in shaping the person that I am today.

This work was supported in part by Sandia National Laboratories, ADE Corporation, and the NSF/SRC Engineering Research Center for Environmentally Benign Semiconductor Manufacturing.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Chemical mechanical polishing (CMP) is the planarization technique of choice for current silicon integrated circuit (IC) fabrication processes. Only CMP satisfies the local and global planarity constraints imposed by current lithography methods [1]. CMP is used to planarize dielectrics, both for insulators between metal levels and for shallow trench isolation (STI) [2]. It is also used to polish back metals such as tungsten [3] and copper [4]. The CMP of copper in multilevel metal damascene processes is currently an area of active research and development as the drive to replace aluminum with the lower-resistivity of copper pushes on. Further advances in IC technology, such as metal gate CMOS devices, will also depend on CMP [14].

Despite the widespread use of CMP, a fundamental understanding of the physics behind the process remains elusive. Available models of the process tend to be incomplete or unable to extend to general conditions. Extensive experimentation is usually performed to find the optimal process and consumable selections for a particular layout pattern. However, optimal choices for one layout are not necessarily usable for a different layout.

In this work, methods are explored for characterization and modeling of CMP, specifically for shallow trench isolation (STI) processing. Closed form solutions for CMP removal in patterned wafers are presented, and extraction methods for model parameters are detailed. Experimental data verifies the validity of the modeling methodology. In addition, diagnosis methods to distinguish potential problems in STI CMP are discussed.

1.1 What is CMP?

Figure 1.1 shows a pictorial view of a CMP tool. A wafer is held on a wafer carrier via back pressure or surface tension such that the surface to be polished faces a porous polyurethane

polishing pad attached to a rotating table. The wafer carrier is rotated in the same direction as the pad, and lowered such that the wafer contacts the rotating pad. The carrier may exhibit lapping motion across the pad in addition to rotation. A slurry composed of particles suspended in a chemical solution is deposited on the pad during polish.

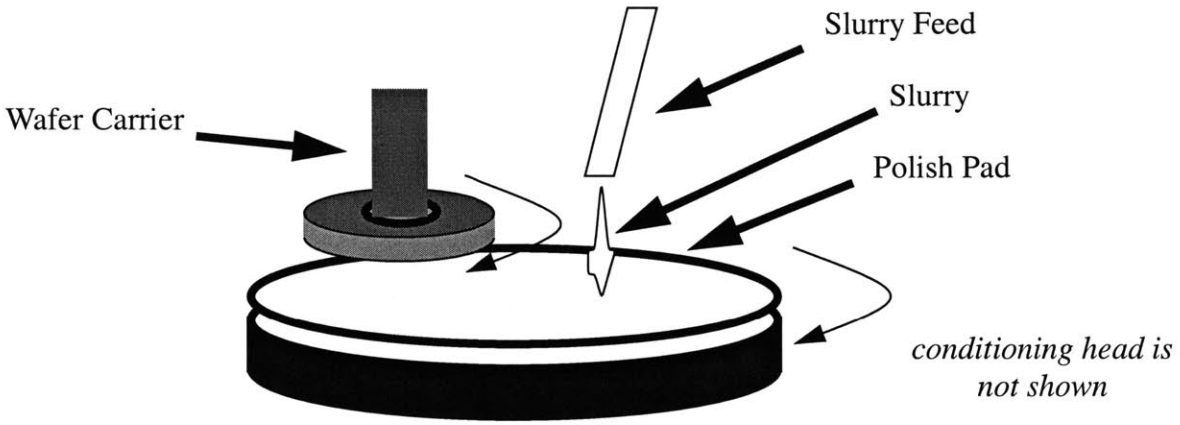


Figure 1.1: Pictorial view of a typical CMP tool.

The mechanism for material removal is similar to that which occurs in glass polishing: chemical degradation or weakening of the surface film followed by abrasion by mechanical interaction with particles [5]. Abraded material is flushed away by the slurry moving across the porous pad. Over time, the pad surface becomes glazed, resulting in a decrease in the polish rate. To keep desired asperities exposed on the pad surface, a diamond-tipped conditioning head is often used to scratch the surface of the pad to maintain the surface properties [6].

The term *CMP tool* refers to the machine used for the CMP process. The term *consumable set* refers to the pad and slurry (i.e., items that are consumed during the CMP process). In addition to rotary CMP tools, there are also linear CMP tools that use a rotating wafer carrier contacting a CMP pad moving on a linear belt [13]. However, the fundamental mechanism of removal (wafer surface contact with a CMP pad, aided by CMP slurry particles and chemistry) is similar.

1.2 Removal Mechanisms in CMP

The fundamental material removal mechanism in CMP of silicon dioxide has been

theorized to be similar to the removal found in glass polishing: a chemical reaction which softens the film surface, followed by a mechanical surface abrasion aided by slurry particles [5].

The chemical reaction involves the hydrolysis of the film surface. Hydroxyl radicals break the existing O-Si-O bonds on the film surface and form relatively weaker Si-OH bonds. This chemical process is limited by the diffusion of hydroxyl ions into the oxide surface. Surface hydrolysis has been experimentally determined to be a necessary step for film removal [5], and so typically the CMP slurry is set to be alkaline, in the regime of pH 10 or more. This ensures that the film surface is weakened and can be removed via mechanical abrasion.

The second step involves the removal of the weakened film surface through abrasion. The actual wear mechanism is not well understood; some speculate that a fluid layer exists which exerts the force necessary to remove the hydrolyzed film surface [7], and others speculate on a complex interaction of particle, fluid, and pad that results in the final film abrasion [8]. It is fairly clear that the abrasion removal mechanism is a dynamic process that depends on surface characteristics of the pad and slurry particles, although the exact contributions of these factors is not known.

The combination of the chemical and mechanical aspects of CMP make it a complex process to model based on physical principles. Typical characterization of a CMP process requires extensive experimentation that must be repeated for each particular CMP process (combination of tool, consumable, and process settings). There have been several reported modeling efforts based on various interpretations of the physics of the CMP process [5,7,8,9,10,11], with no common agreement; experimental data has shown that a variety of models of removal mechanisms in oxide CMP predict blanket wafer removal rates equally well [12].

1.3 Why use CMP?

The semiconductor manufacturing process requires the use of planarization in many steps. One such example is the metal layers. Aluminum metal layers are typically formed using a

subtractive metal process, where a blanket metal layer is deposited, patterned, and etched to form the metal lines. Dielectric is required to insulate the lines from each other and from other metal layers, and so an interlevel dielectric (ILD) is typically deposited on top of the patterned metal lines.

1.3.1 Planarization Requirements in Lithography

The existence of a nonplanar surface in ILD leads to several problems, one of the most serious of which is a lithography concern. Depth of focus (i.e., amount of surface height variation over which a lithography system can successfully resolve features) decreases with decreasing feature dimensions, and so as technology scales, the surface height variation budget becomes tighter, requiring very planar surfaces. It is possible to derive a formula to express depth of focus as a function of minimum feature size. Using the Rayleigh criterion [15] and depth of focus formula [16] gives an expression for depth of focus σ:

$$\sigma = 10.75\frac{b^2}{\lambda} \qquad (1.1)$$

b: minimum feature size
λ: wavelength of projection light

A typical lithography system with a mercury arc lamp using the I-line wavelength of 365 nm would produce a depth of focus of 498 nm for the current technology feature size of 130 nm. In order to increase depth of focus, it is necessary to move to decreasing wavelengths. However, for a given percentage reduction in minimum feature size, a larger percentage reduction in wavelength is required in order to maintain the same depth of focus range.

Table 1.1 lists some depth of focus numbers for a nominal projection system (I-line). As can be seen by the table, planarity requirements will only become more strict with future technologies. While innovations in lithography systems can result in decreasing wavelengths, planarization of film surfaces is also required to meet the depth of focus constraint [17].

Table 1.1: Depth of Focus for Decreasing Feature Size

Year	Minimum Feature Size (μm) [17]	Depth of Focus σ (μm)
2002	0.130	0.498
2003	0.115	0.389
2004	0.090	0.239
2005	0.080	0.188

1.3.2 Use of CMP to achieve planarity

The chemical mechanical polishing process has been shown to satisfy local and global planarity requirements. Figure 1.2 illustrates the difference between local and global planarity. The ideal goal for IC manufacturing is to obtain complete global planarization.

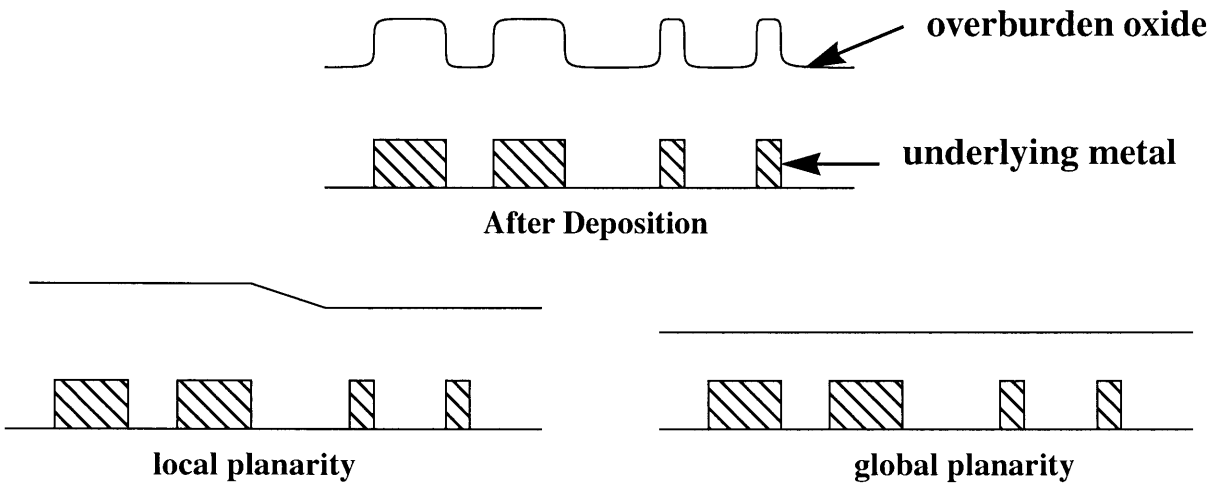


Figure 1.2: Planarization definitions.

Interlevel dielectric CMP has been studied in depth [29,30,33,35] and will be reviewed later in this chapter. This previous work will form the basis for the modeling advances introduced in this thesis.

CMP is also useful for the removal of overburden material in damascene processes. Nominally, features are created by deposition of a blanket layer of material and direct patterning

of features (etching away non-feature regions). In the damascene approach, features are created by etching trenches where features are to be located, depositing material into the trenches, and then using CMP to remove the raised area overburden material. Damascene processes are useful when the feature material cannot be etched effectively (such as copper interconnects [4]), but are also used for other processes, such as shallow trench isolation [2], and tungsten interconnects [3].

CMP is a key step which enables viable damascene processing. While alternative planarization steps (such as spin etch [22] and electropolish [23]) are currently being explored in STI, CMP remains the commonly used planarization step in IC processing.

1.4 Modeling of CMP

The modeling of CMP can be based on a variety of approaches. One method, which will be focused on in this work, is the use of semi-empirical models derived from physical principles with model parameters that are calibrated using experimental data. There has been work performed using this modeling method in the field of dielectric CMP; this will be used as a basis for modeling of certain phases of STI CMP process. There also exist physically based methods which include computation of the fluid dynamics of the system [18,19] as well as the mechanical interaction of the slurry particles, pad, and wafer surface [20,21].

Physically based models seek to utilize the fundamental mechanisms behind the CMP process. Dornfeld reports recent work on modeling the CMP process based on physical relationships between the wafer, pad, and slurry [21]. He conjectures that CMP is based on hydro-dynamic contact of the pad, slurry fluid, slurry particles, and wafer surface and the solid-solid contact of the pad, slurry particles, and wafer surface. Dornfeld proposes a model for the material removal due to the solid-solid contact mode of the CMP process. Removal in this stage is due primarily to two-body abrasion, illustrated in Figure 1.3.

Two-body abrasion is modeled as a function of the force applied on a single abrasive, the size and geometry of the abrasive, the film material, and the relative velocity of the abrasives. The

amount of volume removal by a single abrasive particle is computed as a function of these four parameters [21]. In addition, Dornfeld notes that only a fraction of the abrasives (called the *active* abrasives) are involved in the material removal [21], and offers a method for computing this fraction. The total amount of material removed is then calculated by multiplying the volume removed by a single particle by the number of active abrasive particles. Dornfeld also offers a framework for modeling the hydrodynamic contact mode as a function of slurry chemistry, slurry particle properties, and CMP pad properties [21], which he proposes to use in conjunction with the solid-solid contact model to form a comprehensive integrated model for the CMP process.

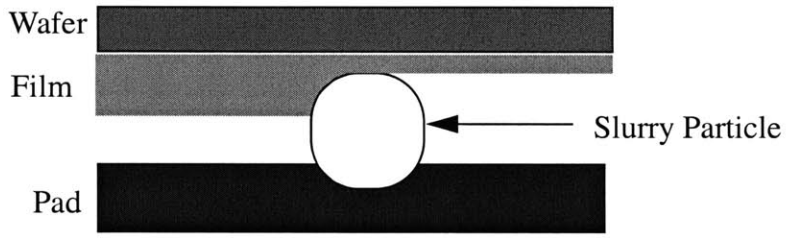


Figure 1.3: Model for two-body abrasion (from [21]).

Dornfeld's material removal model works well for CMP processes involving blanket wafers, but does not mention modeling of patterned wafers. In addition, the Dornfeld model is used as a predictor for single material systems, which excludes polishing of dual material systems such as copper or shallow trench isolation CMP processes. Another limitation of the Dornfeld model is that it requires a considerable amount of information about various aspects of the CMP process. The material removal rate model requires 17 parameter values based on various physical properties of the slurry, pad, and wafer [21].

In addition to physically based models, empirically based chip-scale CMP models have been reported. These models are useful because they provide phenomenological capturing of the behavior of the CMP process using a small set of characterization parameters. Work by Stine *et al.* [33] and Ouma [30] has demonstrated that pattern dependent effects are a large contributor to variation in post-CMP die-scale film thicknesses. Ouma [29] developed a methodology that can

be used for CMP characterization and modeling of interlevel dielectric (ILD) CMP processing. This methodology works well for characterization of a given oxide CMP process (with fixed tool, pad, slurry, and process parameters) where the pattern density effect dominates behavior, but the results cannot be used for other processes because the characterization parameters extracted for a particular CMP process are not applicable to, and cannot be extrapolated to, other oxide CMP processes. In addition, this approach requires extension for other CMP processes, including copper and STI CMP, where step height and selectivity effects are also important.

1.5 Importance of Pattern Density in CMP

The layout pattern density has been shown to be one of the dominant factors in determining the post-CMP thickness profile of a film [26,27,28,33]. Pattern density can be thought of as the area fraction of raised area that affects the CMP process at a particular point on the layout. Figure 1.4 illustrates the concept of pattern density in one-dimensional and two-dimensional cases.

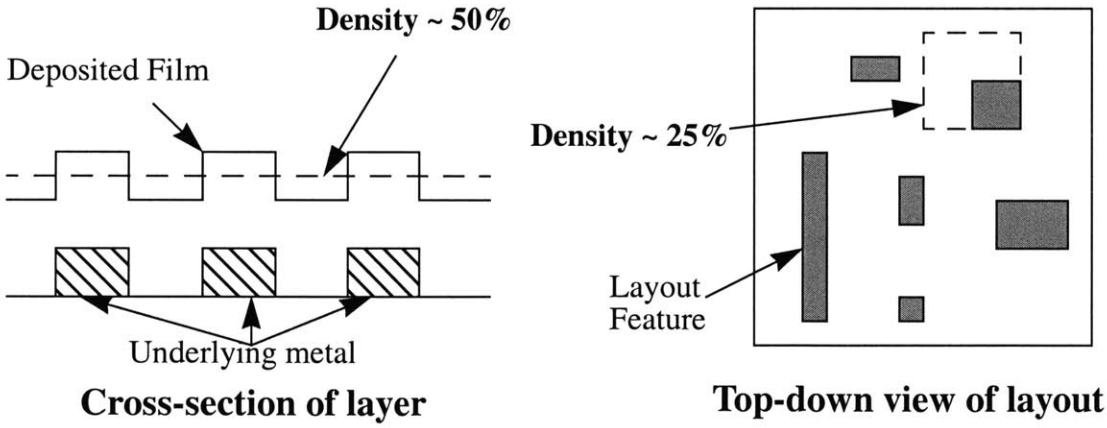


Figure 1.4: Illustration of pattern density.

The importance of pattern density in CMP can be explained by way of an intuitive argument. The higher the pattern density, the larger the contact area with the pad, and the lower the pressure on raised features. Worded differently, the removal rates at any particular point will vary inversely with the pattern density at that point (as will be seen later in Chapter 2). High density

regions polish slower than low density regions.

Variations in pattern density will result in locally planar but globally non-planar surfaces. Generally, the wider the pattern density range, the greater the post-CMP thickness variation. Therefore, narrow pattern density ranges in pre-CMP film surfaces are desirable so that post-CMP film profiles have minimal variation. This density variation effect can manifest itself in circuit layouts where the different regions on the chip can have large pattern density differences. For instance, a logic block of 30% density may be situated closely to a static memory block of 60% density. Reduction of pattern density variation is a critically important issue, and is one of the topics of this thesis.

1.5.1 Deposition Effects

The characteristics of the deposition can affect the pattern density of the film surface. Specifically, a deposition bias can be applied to an underlying layer to estimate the pattern density related to the post-deposition film surface [69]. This may result in a large difference in the computed pattern density of the film surface relative to the layout pattern or underlying patterned material.

Figure 1.5 illustrates the effect of deposition bias on film density, relative to underlying layout density. The type of deposition will affect whether there is positive, negative, or zero deposition bias. A typical value of deposition bias might be +0.5 μm. Chemical vapor deposition (CVD) processes tend to exhibit positive deposition biases to account for the lateral deposition in a conformal process, while high density plasma (HDP) depositions tend to exhibit negative deposition biases to account for the narrowing of deposited features. Deposition biases can be measured by taking SEM cross sections of a fine feature upon which the desired film has been deposited using the desired deposition process.

Deposition bias has the greatest impact on fine features, where the deposition bias relative to the size of the feature is considerable. Consider an HDP process (deposition bias of -0.5 μm

[69]) versus a CVD process (deposition bias of +0.5 μm [69]). An array of 0.5 μm features will have a film density of 0% using HDP, but a density of 100% using CVD. Clearly, the effects of deposition must be considered when computing pattern density.

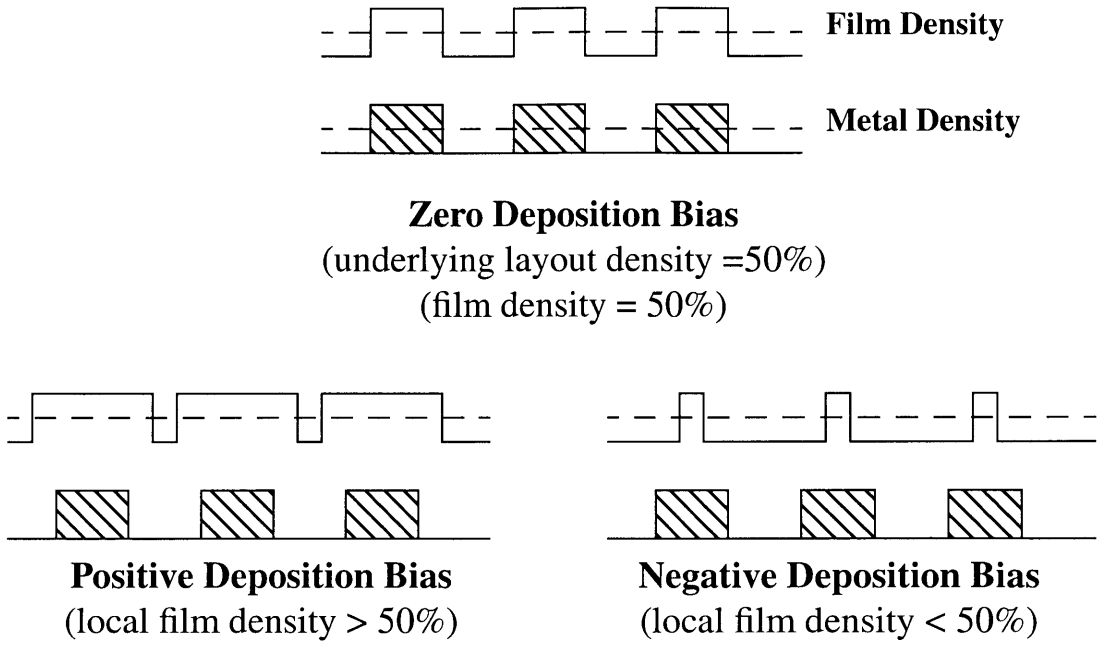


Figure 1.5: Deposition effects on pattern density.

1.6 Planarization Length

In previous CMP models [29], one crucial idea is that there exists a length scale over which the CMP pad and process can detect neighboring features, and that the averaging of local pattern densities over this length scale can be used to effectively predict the actual topography that influences the polish rate at any given point. It is this "effective density" that is used in the density-based models. In order to properly calculate density, then, it is essential to find this characteristic length (known as "planarization length") for a given process. When this length is known, it can be used to calculate effective densities for arbitrary layout, and used to predict post-polish film thicknesses for those layouts. Ouma defined a procedure to characterize the planarization length of the CMP process using special test masks with features specifically tailored for planarization length extraction [29]. The methodology was used with the basic density-based model

[33], but is extensible for use in other density based models which also address additional effects.

There is also an intuitive description of planarization length and how it can be used as a characteristic metric of a CMP process. CMP creates good local planarity but global nonplanarity due to pattern density mismatches which result in removal rate variation across the die. This results in planarization of local step heights, but nonplanarity over longer length scales across the die and wafer. The length scale over which this global nonplanarity exists is the planarization length, as illustrated in Figure 1.6.

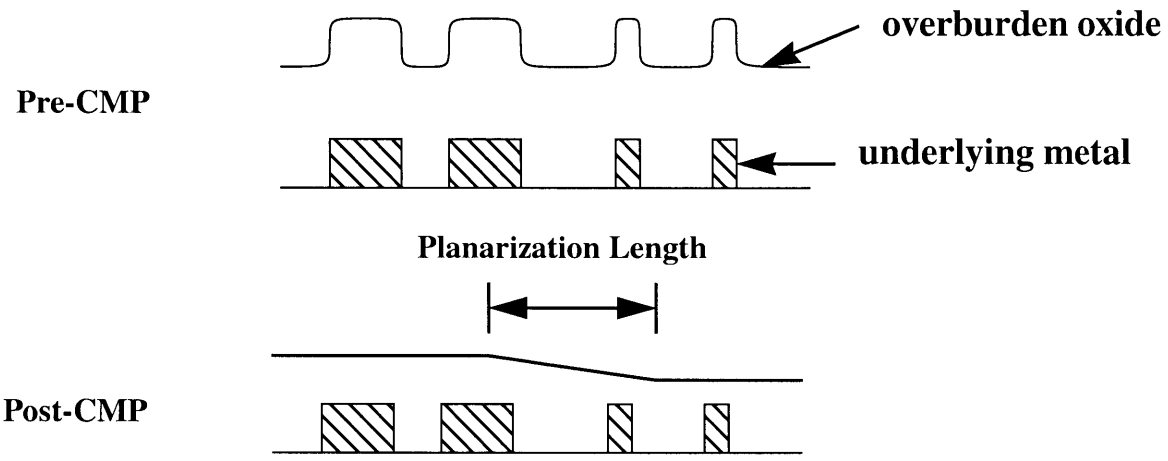


Figure 1.6: Illustration of planarization length.

1.7 Modeling and Characterization Methodology

Using the concept of effective pattern density, Ouma introduced a methodology (illustrated in Figure 1.7) which incorporates pattern density into a full characterization and modeling scheme which can be used to characterize a CMP process, and then to analyze an arbitrary layout to examine how pattern dependent effects factor into the post-polish film thickness [29,30]. In this way, it is possible to predict post-CMP thicknesses of films for an arbitrary layout. This methodology is based on Stine's oxide CMP model [33], which focuses on modeling of oxide CMP where pattern density is the primary effect.

The methodology consists of using special characterization masks in a short-flow CMP process, obtaining measurements from the resulting wafers, and using those measurements to

extract characterization parameters which describe the planarization capability of a CMP process. These characterization parameters can then be used in a model to predict the post-CMP thickness of arbitrary layouts.

The pure pattern density based model is reviewed in Chapter 2. Although this modeling methodology has been shown to be sufficient for characterization of conventional CMP processes, there are still much work to be done to improve the model itself. This process is best used for single material CMP processes, and does not handle reported step height effects on CMP behavior [26]. While pattern density is shown to be a strong effect in CMP film thickness evolution, for certain combinations of pads and slurries, Stine's pure pattern-density based model does not provide adequate modeling accuracy. It has also been shown that the simple relationships of removal rates to density are not applicable when using nonconventional consumable sets, such as the fixed abrasive pad [70]. Discussion of model improvements to incorporate these new effects is given in Chapter 3.

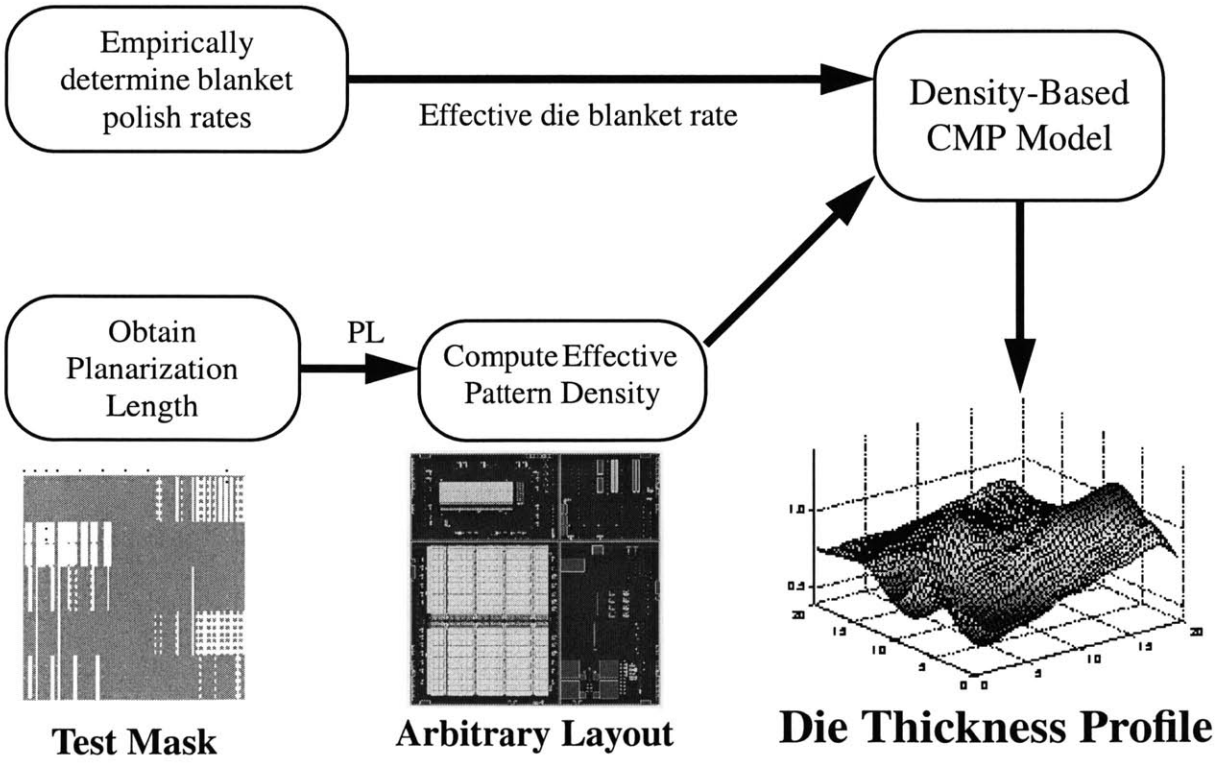


Figure 1.7: Dielectric CMP characterization modeling methodology (from [30]).

1.8 Shallow Trench Isolation CMP

Shallow trench isolation (STI) is the isolation technique of choice in today's integrated circuit technologies. The previously used isolation technique, LOCOS (local oxidation of silicon), suffers from lateral field oxide encroachment which leads to such problems as device leakage, high parasitic capacitances, and lower integration density [25]. A typical STI process flow is shown in Figure 1.8.

The typical STI process flow involves initially growing a thin (100 Å) pad oxide, and then depositing a blanket nitride film (1500 Å) on a raw silicon wafer. The isolation trenches are etched such that the desired trench depth (depth from silicon surface) is achieved (typical depth is 5000 Å). The CMP process is used to polish off the overburden dielectric, down to the underlying nitride, where the nitride serves as a polishing stop layer. After CMP, the nitride layer is then removed via etch, resulting in active area regions surrounded by field trenches. An SEM of a typical device manufactured using an STI process is shown in Figure 1.9.

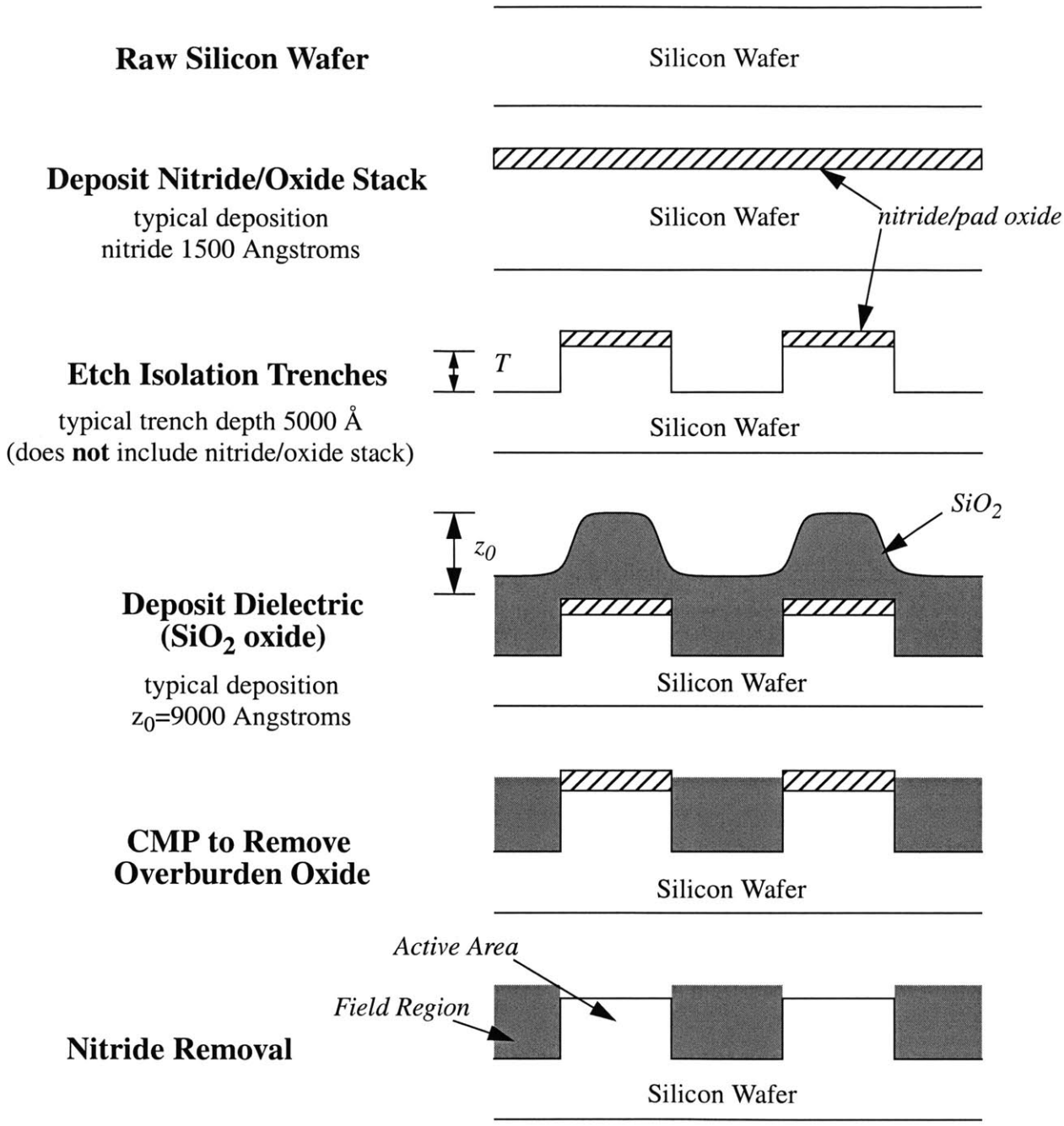


Figure 1.8: Illustration of a typical STI process - wafer cross section.

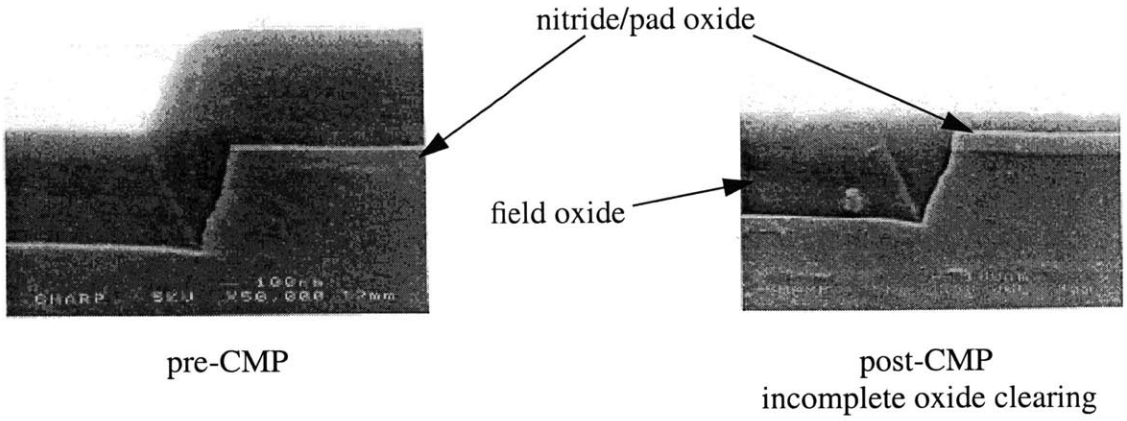


Figure 1.9: SEM of typical STI process pre- and post-CMP (from [24]).

The CMP process step used in STI leads to a number of nonplanarity issues. Ideally, the CMP process planarizes perfectly, and stops on the nitride layer. In reality, pattern effect results in nonuniform material removal. This means that the CMP process will clear overburden oxide and touch down on nitride at different times for different regions (low density areas clearing first). Thus, we have two potential problems: incomplete oxide clearing and nitride overpolish, as illustrated in Figure 1.10.

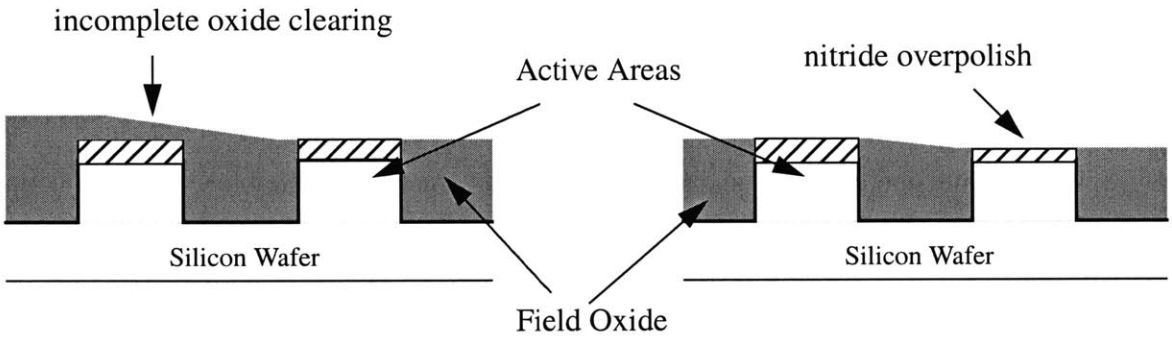


Figure 1.10: STI concerns: incomplete oxide clearing and nitride overpolish.

Creation of working devices requires clearing of the overburden oxide over all active area (else the nitride cannot be removed, and the device cannot be built properly). This mandates a certain required polish time to clear the densest areas. As a result, the less dense areas of the die will experience overpolish. Ideally, the relative polish rate of the nitride with respect to the oxide

(called the polish *selectivity*) is typically set to a small number to minimize the nitride loss. However, there are still implications in overpolishing regions on the die.

The overpolish manifests itself in two particular phenomena. The first is the erosion of the nitride. This is the thinning of the deposited nitride due to the overpolish. Erosion at a particular point is defined as the amount of nitride removed from that point (nitride thickness before CMP process minus nitride thickness after CMP). Positive erosion is defined as material removed. The second phenomena is dishing of the trench oxide. This is the creation of a step height between the nitride layer and the trench oxide. Dishing is defined as the step height from the trench oxide to the nitride active area. Positive dishing means that the trench oxide level is below the active area nitride. Erosion and dishing are illustrated in Figure 1.11.

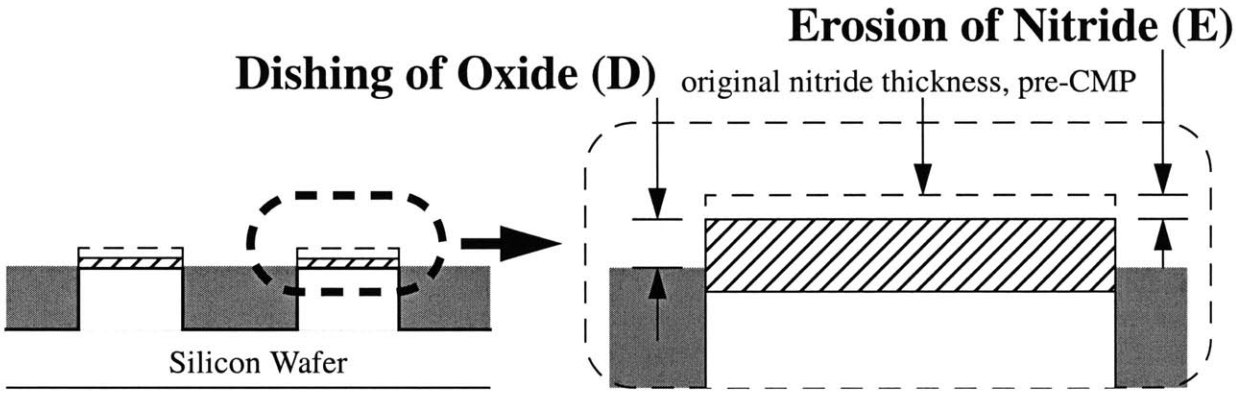


Figure 1.11: Illustration of problems in STI CMP: dishing of oxide and erosion of nitride.

Severe dishing and erosion can cause yield and performance problems. One major concern is dishing such that the field oxide is below the level of the silicon, resulting in the exposure of the side of the active area. This can lead to sidewall and edge-parasitic conduction as well as high electric fields in the gate oxide at the active-area edge [25]. The electric field concentration at the sharp corner region of the active area also serves to reduce the threshold voltage, and the lower the field oxide relative to the silicon height, the lower the threshold voltage [25].

1.9 Thesis Outline

The contributions of thesis can be divided into four categories: modeling and characterization of CMP for shallow trench isolation processes, modeling of effects of nonconventional CMP consumables in STI, methods for layout optimization to achieve reduction of pattern dependent dishing and erosion, and understanding wafer nanotopography and its impact on CMP of films.

1.9.1 Shallow Trench Isolation Modeling

A shallow trench isolation CMP model and characterization methodology is a critical tool for diagnosis of current and future STI processes. Development of such a model and methodology enables characterization of STI CMP processes followed by prediction of dishing and erosion on arbitrary layouts, thus providing a key diagnostic tool for chip designers.

The work performed here includes design of experiments to identify key process and pattern dependencies in STI, analysis of measurements and formulation of an empirical model, modification of the model to account for physically based understanding of the process, and development and verification of a characterization and modeling methodology for STI CMP. This work is based on the work done by Ouma [30] and others in CMP characterization.

1.9.2 Modeling Nonconventional STI CMP Processes

There has been extensive work on modeling dielectric CMP processes, using conventional consumable sets (pads and slurries) as typically used in production-level semiconductor manufacturing processes. Recently, new consumables, including both pads and slurries, have been introduced that may provide better planarization performance. This thesis contributes experimental and modeling results to study the effects of these nonconventional consumables on the CMP process.

For this work, two studies are shown. The first is the use of a fixed abrasive pad, in conjunction with a chemical slurry containing deionized water and KOH, but free of slurry

particles. The pad in this case contains abrasive particles that are embedded into a fixed resin matrix that is microreplicated across the top surface of the pad; the particles are believed to be exposed during the polish and interact with the wafer surface. The second study involves the use of two high-selectivity slurries, one based on silica particles and one based on cerium oxide particles. Conventional pads are used in conjunction with these slurries.

The modeling and characterization methodology for conventional STI CMP processes is extended to enable characterization and prediction for nonconventional CMP processes. Model accuracy is verified using experimental data from the studied nonconventional CMP processes.

1.9.3 Layout Optimization

Using the pattern density analysis techniques originally developed for CMP modeling and prediction by Ouma [30], it is possible to analyze a layout and generate an effective density map of the different layers of interest on a layout, such as the trench level (for STI processes), the polysilicon layer, and all metal levels. Modification of the layout design can potentially achieve significant reduction in pattern density dependent dishing and erosion. A methodology is needed whereby a layout can be both analyzed for potential pattern density variations, and also corrected such that the revised layout has an optimal pattern density distribution.

Two methods are typically used to optimize layout pattern density. One is a method known as dummy fill [40], which involves adding areas of material into the open (i.e., low density) regions of a particular layer, thus raising the pattern density in that region and making it more similar to the rest of the chip. A second method is known as blockout, or reverse etchback, which utilizes a second mask, similar to the original layer but with all features shrunk by a constant dimension. This mask is then used to etch away a substantial fraction of the raised areas on a surface film. Typically used for the trench level in STI processes, this method seeks to lower the pattern density in all regions. Use of selective etchback typically increases cost of ownership due to the additional mask and processing steps that are involved with implementation of this step.

Typically, dummy fill is used for both STI and metal layers, while etchback is typically only used for STI trench layers. While the use of etchback may prove to be sufficient to equalize the pattern density in an STI trench layer, dummy fill should still be used, so that the underlying nitride density is similar everywhere. This will avoid problems with nitride erosion and dishing. A methodology for incorporating dummy fill and reverse etchback schemes for arbitrary layouts is described, with an experimental verification performed on a typical SRAM chip.

1.9.4 Nanotopography Impact on CMP

The term nanotopography refers to nanometer scale height variations that exist on a lateral millimeter length scale on unpatterned silicon wafers [49]. The interaction of this height variation with the CMP process can result in thinning of the surface film, which is a critical concern in shallow trench isolation processes. The development of an accurate nanotopography CMP modeling and characterization procedure allows for the proper diagnosis of potential problems due to wafer nanotopography.

An extensive experiment is performed using sets of 200 mm epi wafers with distinct nanotopography signatures, on CMP processes with varying planarization lengths. Modeling of the interaction of the CMP process with initial wafer nanotopography is studied using several different methods.

1.10 Thesis Outline

This thesis is divided into several chapters describing the key contributions of this thesis. Chapter 2 describes a methodology that can be used for characterization of conventional STI CMP processes and prediction of post-CMP behavior of characterized STI CMP processes on arbitrary layouts. The extension of the model for nonconventional STI CMP processes is described in Chapter 3. Chapter 4 details the layout optimization techniques used to reduce pattern dependent dishing and erosion. Work involving nanotopography impact on STI CMP is described in Chapter 5. Conclusions and suggestions for future work are presented in Chapter 6.

Chapter 2
Shallow Trench Isolation Modeling

In this chapter a compact analytic STI CMP model is developed. Model equations for the dielectric overburden CMP phase, as well as the nitride overpolish phase, are developed, in order to predict erosion and dishing. An experiment is run to obtain data to extract model parameters for a variety of CMP processes, and the model is verified using measurements from wafers run under these CMP processes.

The modeling efforts capture the two intrinsic stages of the STI CMP process: removal of the oxide overburden, and polish of the combined nitride/oxide layer. The oxide overburden polish phase is modelled using ideas from previous work on dielectric CMP processes, while modeling of the dual material phase is developed using ideas and methods from other dual material (copper/oxide) CMP process modeling.

2.1 Pattern Density Based CMP Modeling

In this section, we first review existing models of the dielectric CMP process, as these are the basis for the STI CMP model. A pattern density based dielectric CMP model has been developed by Stine *et al.* [33]. Starting from Preston's glass polishing equation [34], which states that removal rate of a film being polished is proportional to the pressure-velocity product, Stine derives an equation to relate post-CMP film thickness to blanket polish rate (K) and underlying pattern density (ρ).

Equation 2.1 relates the blanket polish rate, pattern density, initial step height (z_1), initial oxide thickness (z_0), and polish time (t) to the final post-polish thickness, separating the polish process into two time regimes: before and after removal of the local step height. This model provides a reasonable prediction for the die-scale variation due to pattern density effects in a layout.

To incorporate wafer-scale CMP variation, it is possible to use different values of the blanket rate K to suggest polish rate variations in different dies across the wafer. The blanket rates may be determined by performing wafer-scale CMP analysis using blanket wafers polished under similar process conditions as the patterned wafers.

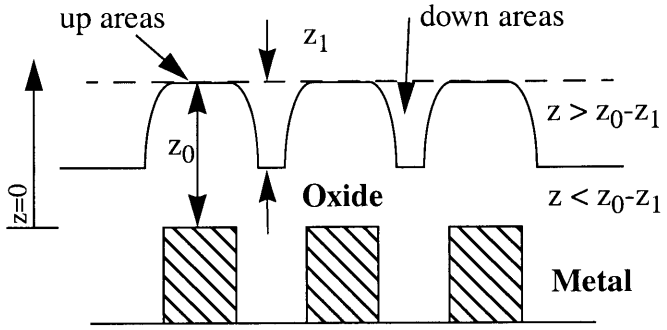


Figure 2.1: Definition of terms used in Equation 2.1.

$$z(t) = \begin{cases} z_0 - \left\{ \dfrac{Kt}{\rho_0(x,y)} \right\} & t < \dfrac{\rho_0(x,y)z_1}{K} \\ z_0 - z_1 - Kt + \rho_0(x,y)z_1 & t \geq \dfrac{\rho_0(x,y)z_1}{K} \end{cases} \tag{2.1}$$

Smith [35] later introduced an extension (known as the density-step height model) to the density-based CMP model, based on work performed by Grillaert *et al.* [38] and Burke [26]. The density-based model provides a good first-order prediction of post-polish film thicknesses, but less accurate predictions for low to medium density regions. The original pattern density based model assumes no down area polish until complete removal of the local step height, then removal of the down area at the blanket polish rate. The density-step height model provides better prediction for up and down-area polish for small step heights. The model is based on pad compressibility work by Burke [26] and Grillaert [38]. The Smith model for the amount removed in the up (AR_u) and down (AR_d) areas is summarized in Equation 2.2 and Equation 2.3, and introduces additional model parameters which are extracted via empirical fitting in a characterization meth-

odology [35] similar that proposed by Ouma [29].

$$AR_u = t_c\frac{K}{\rho} + K(t - t_c) + (1-\rho)\frac{h_1}{\tau}\left(1 - e^{-\frac{(t-t_c)}{\tau}}\right) \tag{2.2}$$

$$AR_d = K(t - t_c) - \rho\left(\frac{h_1}{\tau}\right)\left(1 - e^{-\frac{(t-t_c)}{\tau}}\right) \tag{2.3}$$

where ρ is pattern density, K is blanket removal rate, t_c is contact time, τ is a time constant, and h_1 is contact height. The model parameters here have physical meanings, which will be explored later in this chapter as the STI CMP model is derived.

2.2 Effective Pattern Density

Chapter 1 described the importance of pattern density (ρ) in CMP. The description of density that is used here refers to the *effective pattern density*. Ouma [30] shows that it is essential to weight the effects of neighboring features around a given point on a layout when computing the pattern density at that point. This takes into account the long-range pressure distribution of the polish pad as it deforms around regions on the film surface.

Calculation of effective density is performed as illustrated in Figure 2.2. First, a length scale parameter that is characteristic for a given CMP process (i.e., combination of tool, consumable, and process conditions) must be extracted. Next, the layout of interest is analyzed to create a local density map. This local density map is created by discretizing the layout into a grid of small cells, and computing the local pattern density (total up area within a cell divided by total area of cell) for each cell. Note here that the *total up area* is used rather than *total as-drawn feature area*; this local density map must take into account the effects of deposition on features, as

well as effects incurred if an etchback process is used. Finally, using an elliptical windowing filter with the characteristic length (*planarization length*), the local density map is weighted using FFT analysis to produce an effective density map [30]. This produces an effective density map.

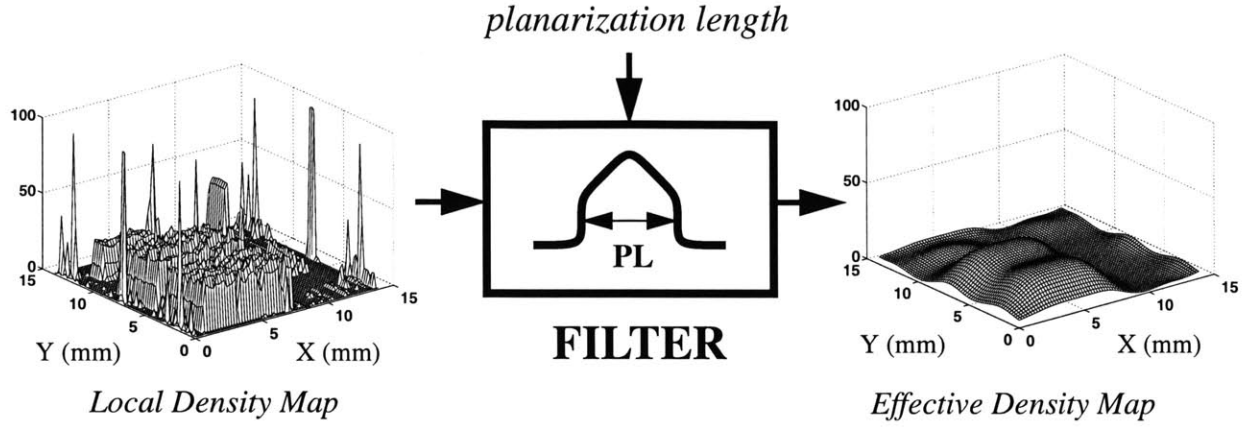


Figure 2.2: Using filtering to compute effective pattern density.

2.3 Existing STI Models

Shallow trench isolation CMP modeling has been the subject of much recent work. Several models have been proposed based on physical or semi-physical derivations of CMP polishing of a dual-material system. Grillaert [53] proposes a semi-empirical model for evaluation of the dishing of an STI structure. The model is based on pad compressibility and contains parameters that are derived from physical principles and values, but are usually extracted from measurement data. Grillaert's equation for dishing D as a function of time is:

$$D = D_0 e^{-\frac{\tau_{ox} - a\tau_{ox} + a\tau_{nit}}{\tau_{nit}\tau_{ox}}(t - t_c)} + \frac{\tau_{nit} - \tau_{ox}}{\tau_{ox} - a\tau_{ox} + a\tau_{nit}} r_{ox}\tau_{ox}\left[1 - e^{-\frac{\tau_{ox} - a\tau_{ox} + a\tau_{nit}}{\tau_{nit}\tau_{ox}}(t - t_c)}\right] \tag{2.4}$$

where D_0 is the initial dishing at nitride touchdown (empirically determined), τ_{ox} and τ_{nit} are empirically determined time constants, r_{ox} is the blanket oxide removal rate, and a is pattern

density. This model has been shown to agree with experimental results [53].

One major limitation of the Grillaert model is that it does not provide an equation to describe nitride erosion, which is a key metric in STI CMP processes. Indeed, Grillaert notes that nitride removal rates for the processes he studies are low enough that the nitride thickness is approximately constant for the length of the polish time [53]. While this may be true for certain STI CMP processes, it would be useful to be able to predict nitride erosion for those processes where there is appreciable nitride removal.

Tseng [54] also proposes a selective two-material CMP model that relates step height to various parameters, based on elastic theory, conservation of force, and polishing selectivity. Tseng's model assumes an exponential decay of step height with time for conventional (i.e., single-material) CMP, and then derives a model for dishing D in dual-material CMP:

$$D = \left[h_0 + \frac{(s-1)\delta}{1 + \frac{(s-1)A_p}{A_c + A_p}} \right] e^{-\left(1 + \frac{(s-1)A_p}{A_c + A_p} \right) Pt} - \frac{(s-1)\delta}{1 + \frac{(s-1)A_p}{A_c + A_p}} \qquad (2.5)$$

where h_0 is the initial step height, A_c is the high area, A_p is the low area, δ is the pad deformation from the polish of a blanket wafer, s is the polishing selectivity, and P is a constant dependent on consumable parameters. Tseng's model has several parameters that appear to require empirical determination (δ, s, P), although Tseng does not specifically describe methods for determining these parameter values. In addition, Tseng's model was not verified with experimental data.

Chiou [55] proposes an STI CMP model obtained by empirically fitting experimental step height data to obtain amount removed in up and down areas as a function of initial step height. This modelling approach requires the extraction of the empirical model parameters for arbitrary CMP processes. Chiou also notes that the model does not account for pattern density effects [55],

thus limiting the useful applications of the model for arbitrary layouts.

The goal of the modeling approach in the work presented in this thesis is to create a closed-form model that can be used for efficient prediction on arbitrary layouts, with only a few modeling parameters that can be extracted from a small number of experimental wafers.

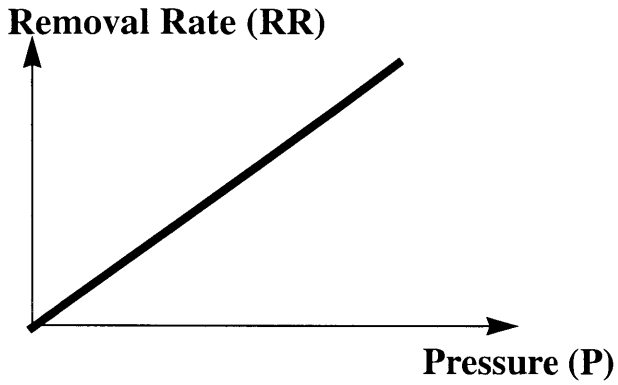
2.4 Removal Rate Diagram Analysis

Removal rate diagram analysis has been recently used to generate a polish evolution model for CMP processes involving copper-oxide systems [56]. The basic premise behind this approach is to derive from physical principles a relationship between the step height and removal rate of a particular system, for both up and down areas. From this relationship, it is possible to formulate a differential equation that can be solved to produce an expression of step height as a function of time. Further analysis of the system produces equations for removal rate and amount removed as a function of time, which can be used to derive erosion and dishing as functions of time. Model parameters such as pattern density and contact height are woven into the removal rate diagram.

Physical relationships form the basis of removal rate diagrams. To derive the removal rate diagram, it is necessary to decompose the removal rate vs. step height relationship into two other relationships: removal rate as a function of pressure, and pressure as a function of step height.

2.4.1 Removal Rate vs. Pressure

The removal rate vs. pressure equation is derived from Preston's equation [34], which states that the removal rate of a film is proportional to the velocity-pressure product. If Preston's equation is assumed to be valid to the CMP process of interest, then the removal rate vs. pressure curve is a simple linear curve, illustrated in Figure 2.3.



Preston's Equation

$$RR = C_P \cdot V \cdot P$$

C_P: Preston's coefficient
V: velocity

Figure 2.3: Removal rate vs. pressure curve.

2.4.2 Pressure vs. Step Height Relationship

To derive the pressure vs. step height relationship, it is necessary to consider the mechanics of the pad-wafer interaction during the CMP process. In most previous CMP models [33], it has been assumed that the pad contacts only the up areas, and that down areas are not contacted until the step height is removed. This situation can be described as equivalent to polishing with an "incompressible pad" (i.e., the pad does not compress into the spaces between features). Grillaert *et al.* [38] assumed a finitely compressible pad, where the pad exhibits some ability to deform into the spaces between features. Grillaert conjectured that there exists a transition height (or *contact height*) at which the pad first contacts the down areas. This transition changes the pressure apportionment on features.

While the pad contacts only up areas, the pressure in the up areas depends on the pattern density (usually an inverse-density relationship), and the pad deformation is less than the step height. Once the transition height is reached, the pad contacts down areas and pressure distribution occurs across both up and down areas; the pad deformation is now equal to the step height. In this situation, Grillaert argues that Hooke's law applies, and that pressure on a feature is proportional to the pad deformation [38]. This implies a linear relationship between pad deformation and pressure, as illustrated in Figure 2.4. As step height decreases, pad deformation

also decreases proportionally. Therefore, there is also a linear relationship between step height and pressure, after the contact height is reached. As the step height decreases, the pressure on the up areas decreases, and the pressure on the down areas increases. This continues until the system reaches a step height of zero, where the pressure on the wafer is equal to the pressure on a blanket wafer. At this point there is no up area or down area, just a planar surface.

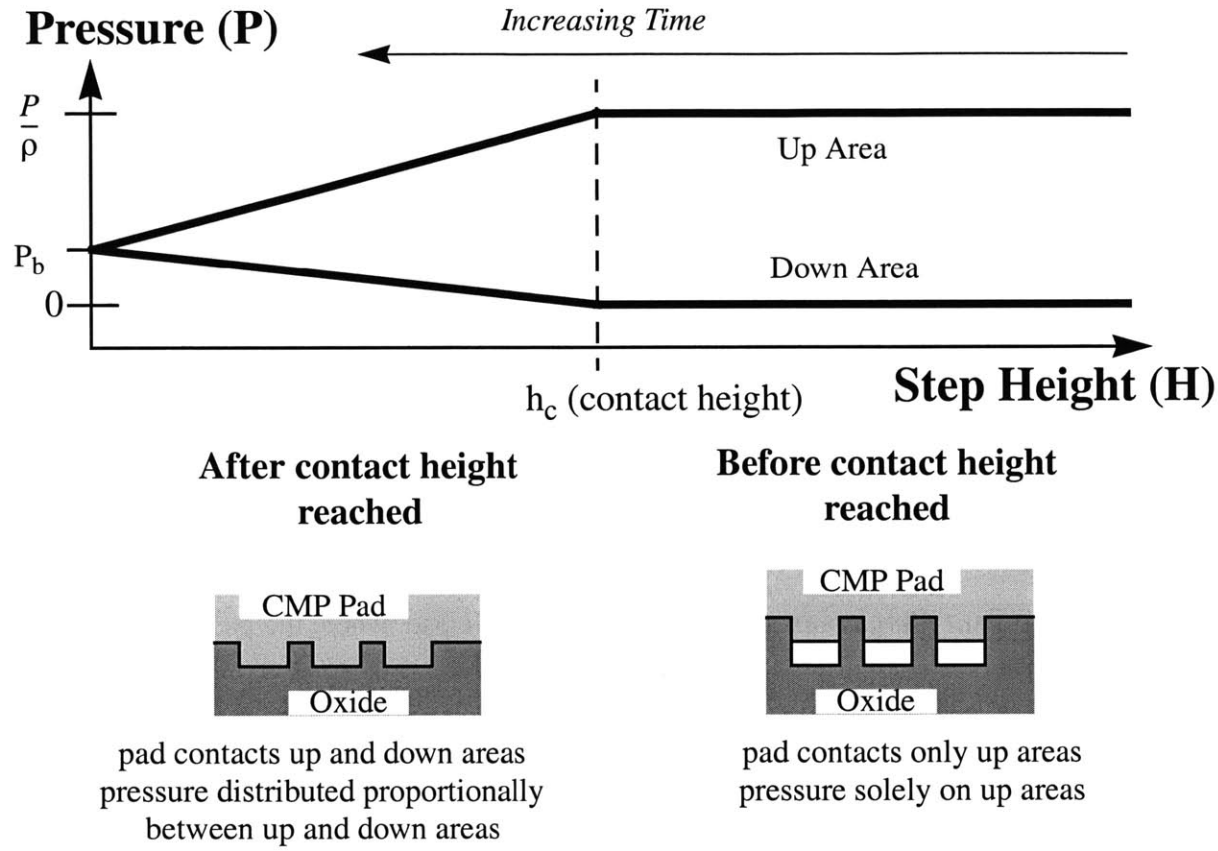


Figure 2.4: Pressure vs. step height curve.

2.4.3 Removal Rate Diagram

The final removal rate diagram is created by merging the removal rate vs. pressure and pressure vs. step height curves. In this work, the details of the removal rate vs. pressure curve and pressure vs. step height curve are not critical for the derivation of the STI CMP model; only the form of the final removal rate diagram (the removal rate vs. step height relationship) is important.

In the case where the CMP process removes only a single material, combining the two

curves is straightforward. Both the up and down area pressure dependencies of Figure 2.4 are combined with the removal rate vs. pressure dependence of Figure 2.3 to produce the removal rate diagram curve, as illustrated in Figure 2.5.

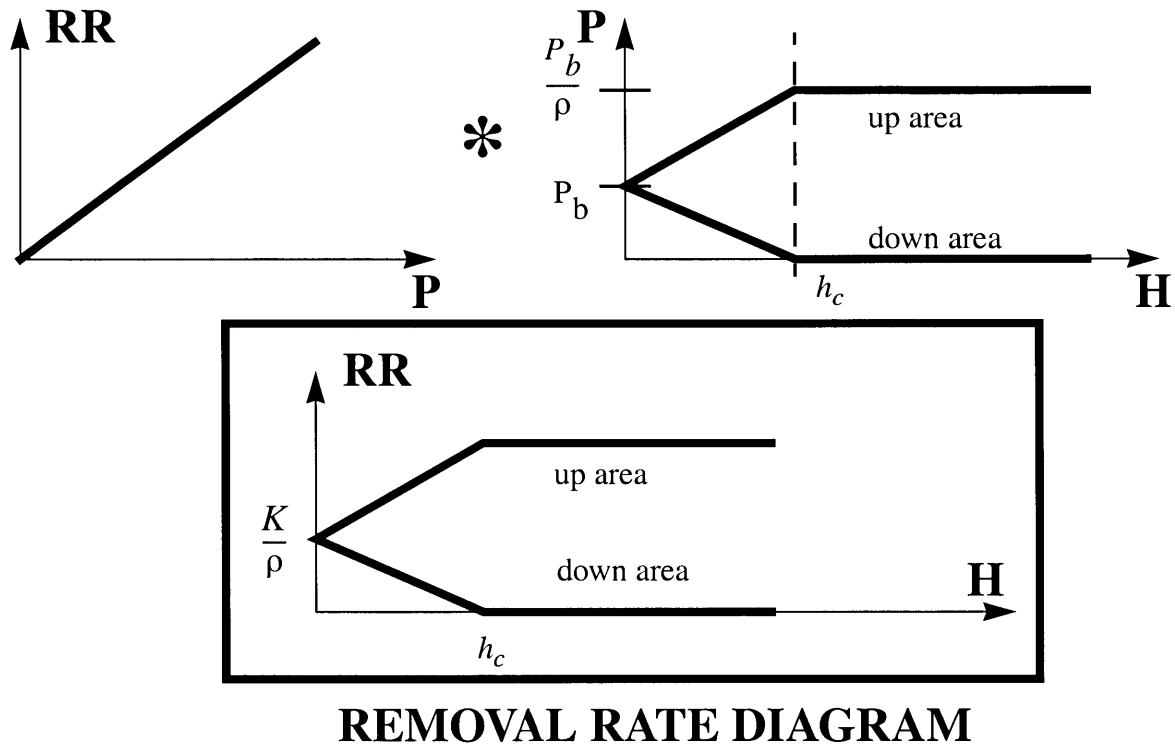


Figure 2.5: Deriving removal rate diagrams: single material case.

The final removal rate diagram illustrates the change in removal rates for up and down areas as functions of height. The removal rate will move towards a steady-state value, which it achieves at a step height of zero, as is illustrated in Figure 2.6. Note that the removal rate diagram is parameterized by both pattern density ρ and contact height h_c; these will vary depending on the pattern at different locations on the chip.

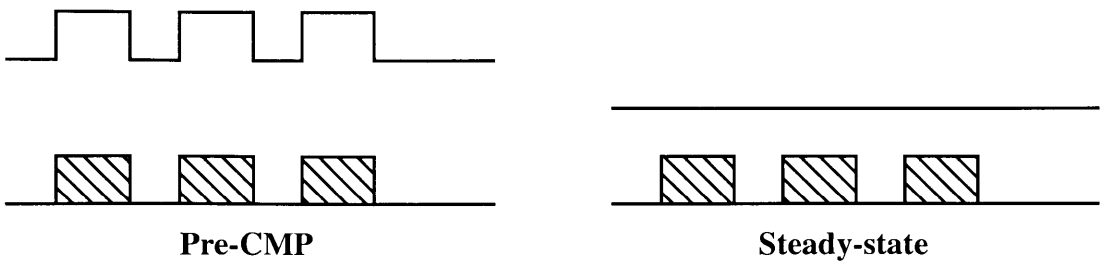


Figure 2.6: Single-material CMP.

In the case where the CMP process removes two materials, it is necessary to first draw the two removal rate vs. pressure curves that correspond to the two materials. The up area curve of Figure 2.4 is then transformed using the removal rate vs. pressure curve that corresponds to the material in the up area, and the down area curve of Figure 2.4 is transformed using the curve that corresponds to the material in the down area. This is illustrated in Figure 2.7.

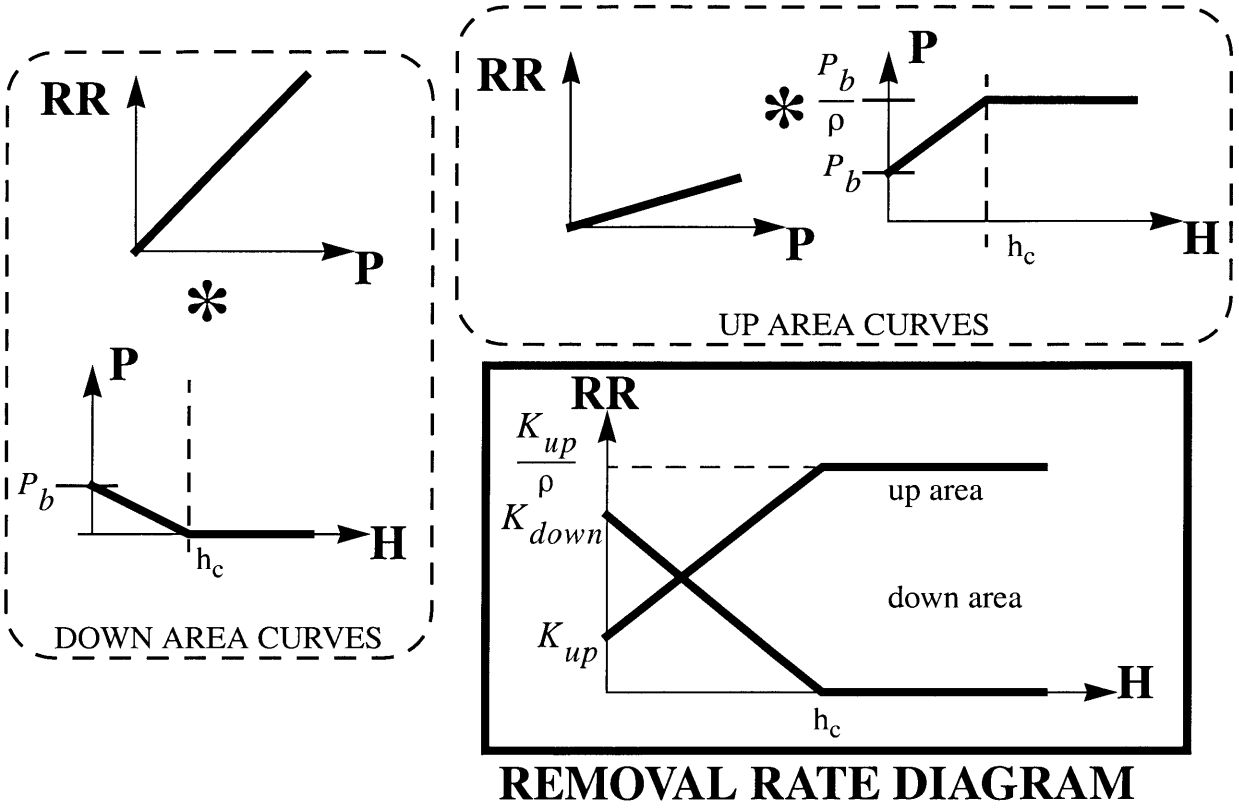


Figure 2.7: Deriving removal rate diagrams: two material case.

For CMP of a two-material system, the steady state that is reached is not at zero step height. This is because typically there is a removal rate difference between the two materials, causing one material to polish faster than the other. This removal rate difference is not due to pressure apportionment from height differences of up vs. down areas; it is a fundamental removal rate difference resulting from material properties. From a zero-step height initial situation, the removal rate difference creates a step height (this is the *dishing* phenomenon). The step height results in a pressure apportionment difference on the newly created up and down areas (higher

pressure in up areas, lower pressure in down areas). Eventually, this pressure apportionment balances out the material removal rate differences, resulting in a steady-state removal rate for both up and down areas, as illustrated in Figure 2.8. In this system, pad deformation is equal to step height, assuming zero initial step height.

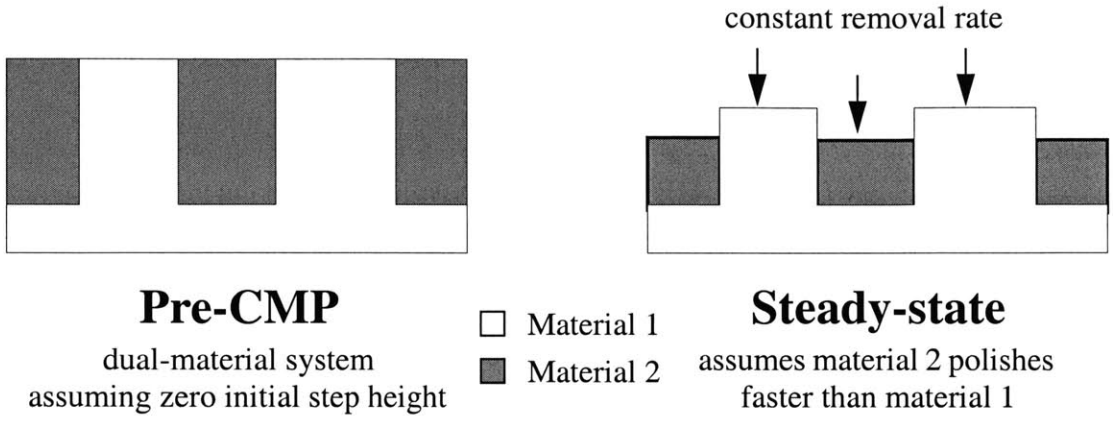


Figure 2.8: Two material CMP.

2.4.4 Deriving Step Height vs. Time Relationship

Based on the removal rate diagram, closed-form equations for up and down area removal rates as functions of step height (i.e., $RR_u(H)$ and $RR_d(H)$) can usually be derived. One additional equation is then utilized to derive the CMP model equations we use in this work: the rate of change of the step height with time is equal to the difference of the up and down area removal rates. Also, if the up area removal rate RR_u is greater than the down area removal rate RR_d, then the step height should decrease in time. Putting this into equation form:

$$RR_u - RR_d = -\frac{dH}{dt} \tag{2.6}$$

Replacing RR_u and RR_d in with the functional dependencies on step height obtained from the removal rate diagram results in a differential equation that can be solved for H as a function of time. The fully specified solution to this differential equation also requires known boundary conditions. Once this equation is known, then RR_u and RR_d as functions of time can be computed

by substituting $H(t)$ into the removal rate dependencies on step height obtained from the removal rate diagram. The resulting removal rate equations can then be integrated to obtain equations for amount removed in the up and down areas (AR_u and AR_d) as functions of time.

2.5 Modeling the STI CMP Process

There are two major phases in the shallow trench isolation CMP process. The first phase is the polish of the overburden oxide. The second phase is the overpolish into the nitride. There is a third phase that might exist in STI CMP polish (the overpolish through the nitride into the silicon). However, this third phase is an undesirable regime since polishing into the silicon results in likely device failure. Therefore, the third phase will not be modelled in this work.

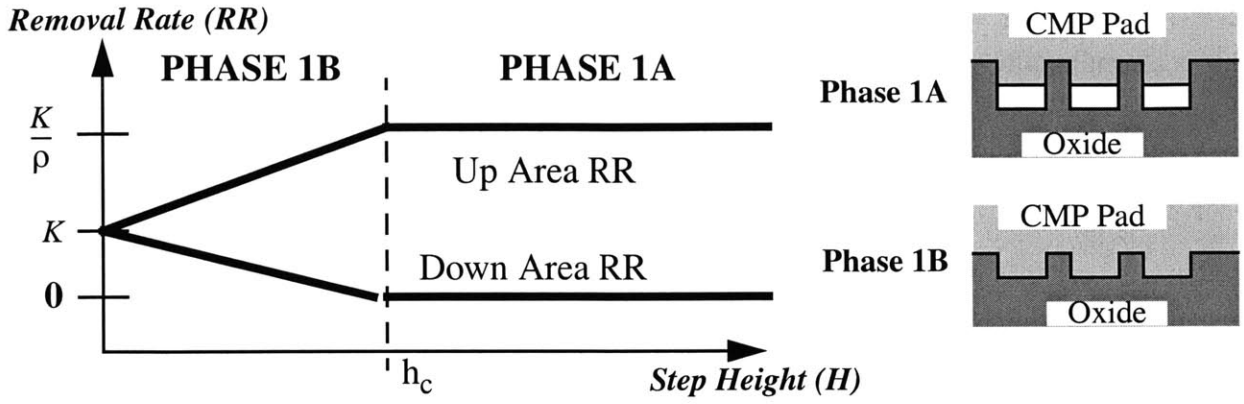
The first phase (polish of overburden oxide) can be broken down into two subphases. The first occurs between the start of the polish and the time before the CMP pad contacts the down areas. The second subphase exists from the time the CMP pad contacts the down areas to the time the up area overburden oxide has been cleared to the nitride.

2.5.1 Modeling Oxide Overburden Phase (Phase 1)

The first phase can be expressed in the removal rate diagram for the polish of a single material. This analysis assumes that the initial starting point is a dielectric film with a fixed step height and feature densities which vary based on location on the die. Thus any given removal rate diagram represents a point of fixed or known effective pattern density on the die. The removal rate diagram for the first phase is shown in Figure 2.9.

For a sufficiently large step height, pad compressibility is not an issue. The pad solely contacts up areas, and the down area removal rate is zero. This is Phase 1A. The up areas polish at a particular patterned removal rate, indicated by K/ρ on the removal rate diagram. As the CMP process progresses, the step height is reduced and the contact height is eventually reached. At this point, the pad contacts the down areas, and down area removal begins. This is Phase 1B. The up and down area removal rates linearly approach each other as the step height decreases until zero

step height is reached, after which the entire film polishes at the blanket film removal rate K.



Phase 1A indicates polish before the CMP pad contacts the down areas.
Phase 1B indicates polish after down area has been initially contacted.

Figure 2.9: Removal rate diagram for STI CMP polish, Phase 1 (oxide overburden).

There are initial conditions from which model equations can be derived. The first case occurs when, starting at time $t = 0$, the step height is sufficiently large that $h_c < z_1$. In this case, it is possible to use Figure to formulate removal rate vs. step height equations for Phase 1A and Phase 1B. Subscript "u" is used for up area, and subscript "d" for down area.

Phase 1A:

$$RR_u = \frac{K}{\rho} \tag{2.7}$$

$$RR_d = 0 \tag{2.8}$$

$$RR_u = \left(\frac{K}{h_c}\right)\left(\frac{1-\rho}{\rho}\right)H + K \tag{2.9}$$

Phase 1B:

$$RR_d = -\left(\frac{K}{h_c}\right)H + K \tag{2.10}$$

Phase 1A equations are independent of step height and can be solved to produce equations for removal rate, amount removed, and step height as functions of time. Phase 1B requires more algebra. Substituting Equations 2.9 and 2.10 into Equation 2.6 yields:

$$\frac{dH}{dt} = -\left(\frac{K}{\rho h_c}\right)H \tag{2.11}$$

This is a simple differential equation with a closed form solution:

$$H = h_c e^{\frac{-(t-t_c)}{\tau_{ox}}} \tag{2.12}$$

with the boundary condition that at the time t_c (the contact time), the step height is h_c (the contact height). This is only valid for the case when the polish situation is in Phase 1A at time $t=0$. For simplicity, the quantity $\rho h_c/K$ is defined as the constant τ_{ox}.

Equation 2.12 can be substituted into Equations 2.9 and 2.10 to produce expressions for removal rate as a function of time.

$$RR_u = K\left(\frac{1-\rho}{\rho}\right)e^{\frac{-(t-t_c)}{\tau_{ox}}} + K \tag{2.13}$$

$$RR_d = -Ke^{\frac{-(t-t_c)}{\tau_{ox}}} + K \tag{2.14}$$

The removal rates as a function of time can then be integrated to compute closed form expressions for amount removed as a function of time. Taken along with the amount removed equations from Phase 1A, Equations 2.15 through 2.21 form the model for the oxide overburden phase of the STI CMP process.

Phase 1A Equations ($t < t_c$)

$$AR_u = \frac{K}{\rho}t \tag{2.15}$$

$$AR_d = 0 \tag{2.16}$$

$$RR_u = \frac{K}{\rho} \tag{2.7}$$

$$RR_d = 0 \tag{2.8}$$

$$H = z_1 - \frac{K}{\rho}t \tag{2.17}$$

Phase 1B Equations ($t_c < t < t_n$)

$$AR_u = \frac{K}{\rho}t_c + K(t - t_c) + h_c(1-\rho)\left[1 - e^{\frac{-(t-t_c)}{\tau_{ox}}}\right] \tag{2.18}$$

$$AR_d = K(t - t_c) - h_c\rho\left[1 - e^{\frac{-(t-t_c)}{\tau_{ox}}}\right] \tag{2.19}$$

$$RR_u = K\left(\frac{1-\rho}{\rho}\right)e^{\frac{-(t-t_c)}{\tau_{ox}}} + K \tag{2.13}$$

$$RR_d = -Ke^{\frac{-(t-t_c)}{\tau_{ox}}} + K \tag{2.14}$$

$$\tau_{ox} = \frac{\rho h_c}{K} \tag{2.20}$$

$$t_c = \frac{\rho(z_1 - h_c)}{K} \tag{2.21}$$

The second case occurs when the polish situation begins in Phase 1B. This is the situation

where the contact height is larger than the initial step height. In this case, the initial condition of Equation 2.12 has changed. Using the new initial condition that at $t = 0$, the step height is z_1, we find:

$$H = z_1 e^{\frac{-t}{\tau_{ox}}} \tag{2.22}$$

$$\tau_{ox} = \frac{\rho h_c}{K} \tag{2.20}$$

It is then possible to derive new equations for removal rate and amount removed, as given in Equations 2.23 through 2.26.

Polish Situation Includes Only Phase 1B ($h_c > z_1$)

$$RR_u = \frac{Kz_1}{h_c}\left(\frac{1-\rho}{\rho}\right)e^{\frac{-t}{\tau_{ox}}} + K \tag{2.23}$$

$$RR_d = -\frac{Kz_1}{h_c}e^{\frac{-t}{\tau_{ox}}} + K \tag{2.24}$$

$$AR_u = Kt + z_1(1-\rho)\left[1 - e^{\frac{-t}{\tau_{ox}}}\right] \tag{2.25}$$

$$AR_d = Kt - z_1\rho\left[1 - e^{\frac{-t}{\tau_{ox}}}\right] \tag{2.26}$$

It will be assumed that contact height varies inversely with increasing density. This implies that τ_{ox} is constant across all densities. This is reasonable; Smith proposes an exponential relationship for h_c that is quite similar to an inverse-density relationship [35], and Grillaert [38] suggests that τ_{ox} only has a weak dependence on density, if any. Effective density can be fully specified by a planarization length for the oxide phase (PL_{ox}). Thus, there are three key Phase 1

modeling parameters: K, τ_{ox}, and PL_{ox}.

For the pure dielectric CMP model, it is possible to compute the amount removed for up and down areas for Phase 1 from blanket removal rate (K), effective density (ρ), and time constant (τ_{ox}). The Phase 1 STI CMP model (which also involves only single material dielectric CMP) is similar to the Smith model [35], as it has been derived from the same concepts.

2.5.2 Modeling Overpolish Phase (Phase 2)

The second STI CMP phase is expressed in a separate removal rate diagram to account for a polish of two separate materials of silicon dioxide and silicon nitride. The removal rate diagram is constructed by considering the two removal rate vs. pressure curves for nitride and oxide. In simplest form (assuming a Prestonian relationship), these are linear curves as shown in Figure 2.10.

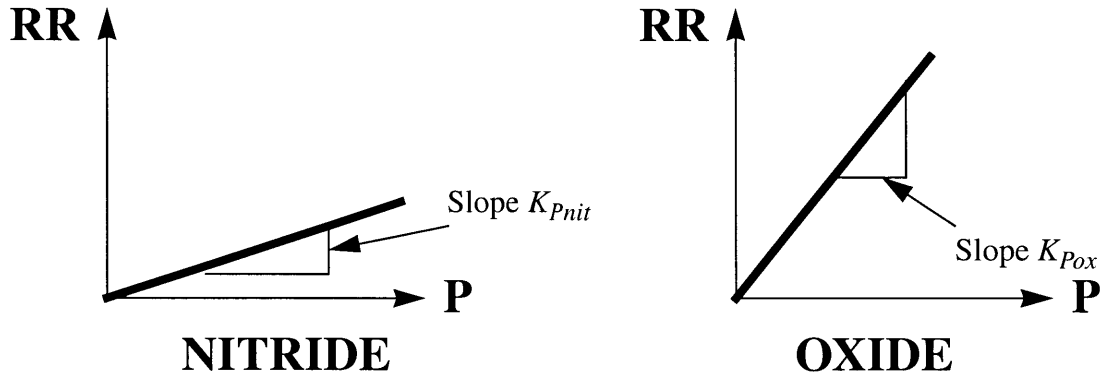


Figure 2.10: Removal rate vs. pressure curves, for nitride and oxide.

Assuming Prestonian relationships allows us to formulate equations for the curves in Figure 2.10, as well as a definition of a polish selectivity s that is valid for all pressures.

$$RR_{nit} = K_{Pnit} \cdot P \tag{2.27}$$

$$RR_{ox} = K_{Pox} \cdot P \tag{2.28}$$

$$s = \frac{RR_{ox}}{RR_{nit}} = \frac{K_{Pox}}{K_{Pnit}} \tag{2.29}$$

Having defined removal rate vs. pressure equations, it is possible to combine Figure 2.10

with Figure 2.4 to produce the resulting removal rate diagram as shown in Figure 2.11.

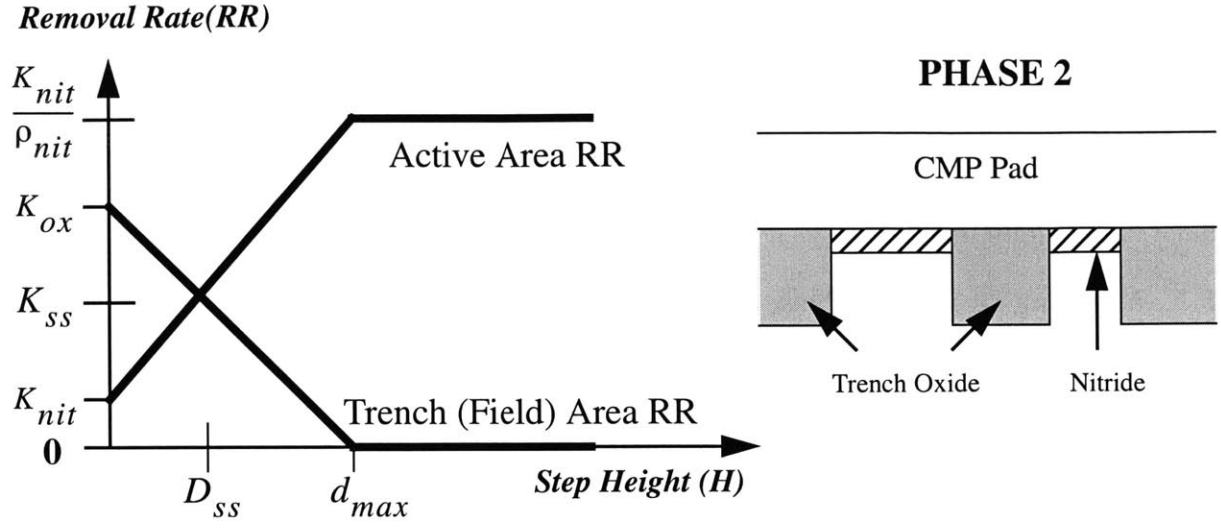


Figure 2.11: Removal rate diagram for STI CMP polish, Phase 2 (overpolish into nitride).

To understand the diagram, the different potential step height situations that occur in this dual material polish system are considered. First, the initial condition of zero step height is considered. At zero initial step height, the nitride polishes at a single removal rate K_{nit}, while the oxide also polishes at a different, higher removal rate K_{ox}, since blanket oxide removal rate is larger than the blanket nitride removal rate for most CMP processes. As the polish progresses, the differences in removal rates will result in an increase in the step height. As the step height increases, the nitride (active area) region becomes an up area, and the trench oxide (field area) becomes a down area. The resulting pressure distribution causes an increase in the up area (nitride) removal rate, and a corresponding decrease in down area (oxide) removal rate, with the result being an increase in the step height (dishing). This change in removal rate is modelled as a linear dependence on the step height. Eventually a point is reached where the two removal rates are equal. This point is the steady-state point of the polish, where further polishing does not change the step height. At this point, the step height is the steady-state dishing D_{ss}, and the removal rate is the steady-state removal rate K_{ss}.

Next, the case of extremely large initial step heights is considered. For a sufficiently large initial step height, the CMP pad does not contact the trench oxide down areas, and so the trench oxide removal rate is zero. At this same point, the pad only contacts the up areas of nitride, and so the removal rate of the nitride is the patterned removal rate of nitride K_{nit}/ρ. The step height past which this occurs is denoted d_{max} (maximum dishing), which is the step height at which the pad no longer contacts the trench oxide. At this initial point, the differences in removal rates causes the step height to decrease. Once the step height reaches d_{max}, the pad begins to contact the down areas, and the resulting redistribution of pressure causes the removal rate of the down areas to increase and the removal rate of the up areas to decrease. This change in removal rates is modelled as a linear dependence on the step height towards the steady state. The step height further decreases until steady state is reached, at which the step height is D_{ss} and the removal rate of both up and down area is K_{ss}.

In Figure 2.11, the following relationships also relate blanket rates and selectivity:

$$K_{ox} = K \tag{2.30}$$

$$K_{nit} = \frac{K_{ox}}{s} = \frac{K}{s} \tag{2.31}$$

where K is the oxide blanket removal rate parameter from the Phase 1 STI CMP polish, and s is the selectivity parameter, defined by Equation 2.29. Also, it is possible to express D_{ss} and K_{ss} as functions of the other variables in Figure 2.11:

$$D_{ss} = \frac{d_{max}\rho_{nit}(s-1)}{\rho_{nit}(s-1)+1} \tag{2.32}$$

$$K_{ss} = \frac{K}{\rho_{nit}(s-1)+1} \tag{2.33}$$

Equations for the amount removed for Phase 2 can be obtained from the removal rate diagram in a similar manner as that of Phase 1. First, the removal rate vs. step height equations are formed. At the end of Phase 1, the step height is normally small, and so the assumption that the initial step height at Phase 2 is less than the maximum dishing (d_{max}) can be made. This leads to the following formulas for removal rate, where subscript "*a*" denotes active area (up area), and subscript "*f*" denotes field area (down area):

$$RR_a = \left(\frac{\frac{K_{nit}}{\rho_{nit}} - K_{nit}}{d_{max}}\right)H + K_{nit} = \left(\frac{K}{s}\right)\left(\frac{1}{d_{max}}\right)\left(\frac{1-\rho_{nit}}{\rho_{nit}}\right)H + \frac{K}{s} \tag{2.34}$$

$$RR_f = -\left(\frac{K_{ox}}{d_{max}}\right)H + K_{ox} = -\left(\frac{K}{d_{max}}\right)H + K \tag{2.35}$$

Substituting into Equation 2.6 results in a differential equation for H:

$$\frac{dH}{dt} + \left(\frac{K}{d_{max}}\right)\left(\frac{1+\rho_{nit}(s-1)}{s\rho_{nit}}\right)H - K\left(\frac{1-s}{s}\right) = 0 \tag{2.36}$$

This equation can be solved to find H as a function of time:

$$H(t) = A_1 e^{-\frac{t}{\tau_{nit}}} + \frac{d_{max}\rho_{nit}(s-1)}{\rho_{nit}(s-1)+1} \tag{2.37}$$

$$\tau_{nit} = \left(\frac{d_{max}}{K}\right)\left(\frac{s\rho_{nit}}{1+\rho_{nit}(s-1)}\right) \tag{2.38}$$

Solving this completely requires setting a boundary condition. This can be done by

defining a time t_n, at which the step height is h_n. The parameter t_n is called the nitride touch-down time, and is defined as the time it takes to clear all overburden oxide from a particular location. The value of t_n can be obtained if Phase 1 model parameters are known, as described in Section 2.5.3.The value of h_n can be solved for if t_n is known by using Phase 1 equations (specifically, Equation 2.12 or 2.22):

Polish Situation

$$h_c \le z_1 \qquad h_n = h_c e^{\frac{-(t_n - t_c)}{\tau_{ox}}} \tag{2.39}$$

$$h_c > z_1 \qquad h_n = z_1 e^{\frac{-t_n}{\tau_{ox}}} \tag{2.40}$$

Using the boundary condition that at time t_n, the step height is h_n, and noting that the second term in Equation 2.37 equals D_{ss}, we find that the step height goes as:

$$H = (h_n - D_{ss}) e^{\frac{-(t - t_n)}{\tau_{nit}}} + D_{ss} \tag{2.41}$$

Using Equation 2.41 with Equations 2.34 and 2.35, it is now possible to derive removal rate equations:

$$RR_a = \left(\frac{K}{s}\right)\left(\frac{1}{d_{max}}\right)\left(\frac{1-\rho_{nit}}{\rho_{nit}}\right)(h_n - D_{ss}) e^{\frac{-(t - t_n)}{\tau_{nit}}} + K_{ss} \tag{2.42}$$

$$RR_f = -\left(\frac{K}{d_{max}}\right)(h_n - D_{ss}) e^{-\frac{(t - t_n)}{\tau_{nit}}} + K_{ss} \tag{2.43}$$

Integrating these equations and setting boundary conditions gives equations for amount removed:

$$AR_a = z_0 + K_{ss}(t - t_n) + \left(\frac{h_n - D_{ss}}{1 + \frac{s\rho_{nit}}{(1-\rho_{nit})}} \right) \left(1 - e^{\frac{-(t-t_n)}{\tau_{nit}}} \right) \tag{2.44}$$

$$AR_f = z_0 - z_1 + h_n + K_{ss}(t - t_n) - \left(\frac{h_n - D_{ss}}{1 + \frac{(1-\rho_{nit})}{s\rho_{nit}}} \right) \left(1 - e^{\frac{-(t-t_n)}{\tau_{nit}}} \right) \tag{2.45}$$

Note that the equations for amount removed include the aggregate amount removed during both Phase 1 and Phase 2. These equations make up the Phase 2 STI CMP model. It is also possible to derive equations assuming other starting conditions (e.g., h_n larger than d_{max}). These equations are not presented in this thesis, since they are not normally applicable in STI CMP processes.

It remains to determine the key parameters for this model. For the single material oxide polish phase, it is conjectured that the contact height h_c has an inverse-density relationship. This conjecture will be extended for the dual nitride/oxide system, but here an inverse-density relationship will be assumed for d_{max}, which allows for the definition of a "pure nitride time constant" τ_2:

$$\tau_2 = \frac{d_{max}}{\left(\frac{K_{nit}}{\rho_{nit}} \right)} = \frac{s\rho_{nit} d_{max}}{K} \tag{2.46}$$

The parameter τ_2 is the time constant that would result from deriving CMP system equations for a single material nitride polish situation, equivalent to the τ_{ox} constant derived for the single material oxide system.

It is now possible to redefine other parameters in the system in terms of τ_2:

$$d_{max} = \frac{K\tau_2}{s\rho_{nit}} \tag{2.47}$$

$$\tau_{nit} = \frac{\tau_2}{1+(s-1)\rho_{nit}} \tag{2.48}$$

$$D_{ss} = K\tau_{nit}\left(\frac{s-1}{s}\right) \tag{2.49}$$

This gives the final equations used for the nitride overpolish phase (Phase 2) of the CMP STI process. There are three key model parameters for Phase 2: planarization length for nitride (PL_{nit}), selectivity (s) [or equivalently, K_{nit}], and time constant (τ_2).

2.5.3 Finding nitride touch-down time

The nitride touch-down time t_n is defined as the time it takes to remove all the overburden oxide from a given point. It will vary by location on the die (or more specifically, by effective pattern density), and can be found by solving Equations 2.18 and 2.25 for z_0, resulting in Equations 2.50 and 2.51. All parameters in these equations are known (or can be directly computed) if the Phase 1 parameters are known.

$$h_c \le z_1 \qquad z_0 = \frac{K}{\rho}t_c + K(t_n - t_c) + h_c(1-\rho)\left[1 - e^{\frac{-(t_n - t_c)}{\tau_{ox}}}\right] \tag{2.50}$$

$$h_c > z_1 \qquad z_0 = Kt_n + z_1(1-\rho)\left[1 - e^{\frac{-t_n}{\tau_{ox}}}\right] \tag{2.51}$$

Equation 2.50 and Equation 2.51 can be rearranged to obtain equations for t_n:

$$h_c \le z_1 \qquad t_n = \frac{z_0 - z_1}{K} + \frac{z_1}{\left(\frac{K}{\rho}\right)} + \left[\frac{1}{K} - \frac{1}{\left(\frac{K}{\rho}\right)}\right] h_c e^{\frac{-(t_n - t_c)}{\tau_{ox}}} \tag{2.52}$$

$$h_c > z_1 \qquad t_n = \frac{z_0 - z_1}{K} + \frac{z_1}{\left(\frac{K}{\rho}\right)} + \left[\frac{1}{K} - \frac{1}{\left(\frac{K}{\rho}\right)}\right] z_1 e^{\frac{-t_n}{\tau_{ox}}} \tag{2.53}$$

These equations can be viewed as breaking the nitride touch-down time into three components: the time to remove remaining oxide after initial step height removal (first term), the time to remove the step height assuming no pad compressibility (second term), and an adjustment for pad compressibility (third term). Equation 2.52 and 2.53 are transcendental equations that can be solved by assuming an initial guess for t_n, and iterating the equation until the value of t_n converges.

2.6 Modeling Dishing and Erosion

Dishing and erosion equations can be derived from the amount removed equations. These equations are more useful since it is the dishing and erosion phenomenon that is of most interest in shallow trench isolation CMP. The dishing and erosion equations are also more useful because they isolate key model parameters, making simpler equations from which to extract out model parameters. Dishing is simply the step height as a function of time:

$$D(t) = H(t) = (h_n - D_{ss}) e^{-\frac{(t - t_n)}{\tau_{nit}}} + D_{ss} \tag{2.54}$$

This equation makes intuitive sense, since dishing at steady state (infinite time) is equal to

the steady-state dishing parameter D_{ss}. Erosion can also be computed as the amount of nitride removed:

$$E(t) = AR_a(t) - z_0 \tag{2.55}$$

$$E(t) = K_{ss}(t - t_n) + \left(\frac{(h_n - D_{ss})(1-\rho_{nit})}{1 + \rho_{nit}(s-1)} \right) \left(1 - e^{-\frac{(t-t_n)}{\tau_{nit}}} \right) \tag{2.56}$$

Therefore, dishing and erosion can be fully specified and predicted if the Phase 1 and Phase 2 STI CMP model parameters are known. These model parameters are characteristic of a given CMP process (tool, consumable set, process parameters), and the model equations can be used to prediction dishing and erosion on wafers patterned with arbitrary layouts that are subjected a particular characterized CMP process.

2.7 Model Parameters

There are three primary model parameters for the oxide overburden phase (PL_{OX}, K, τ_{ox}) and three primary model parameters for the nitride overpolish phase (PL_{NIT}, s, τ_2). In addition, there are other parameters in the model that are derived from the primary model parameters that have physical meanings.

2.7.1 Phase 1 Model Parameters

The oxide planarization length PL_{OX} is used to compute the effective density of the oxide overburden film. It is a measure of the planarization capability of the oxide film. The blanket removal rate K describes the average blanket removal rate for a particular die on the wafer. The blanket rate typically varies across the wafer, so this parameter should be extracted from measurement data from a single die. The oxide time constant τ_{ox} describes how fast the oxide step height will reduce once the down area is contacted by the pad. It is independent of density, but

decreases with increasing pressure.

2.7.2 Phase 2 Model Parameters

The nitride planarization length PL_{NIT} is used to compute the effective density of the underlying nitride film. It is a measure of the planarization capability of the combined active areas nitride/trench oxide film.The selectivity s describes how fast a blanket oxide film will polish relative to a blanket nitride film. A higher selectivity decreases the steady-state removal rate and increases the steady-state dishing. The parameter τ_2 is a time constant that describes how fast the step height of a single film nitride layer will decrease once down area contact has been achieved. It is not strictly relevant to the dual nitride/oxide phase, but is used to derive other parameters. The nitride time constant τ_{nit} (derivable from the other model parameters via Equation 2.48) describes how fast the dishing approaches steady-state. It is density and selectivity dependent.

2.8 Model Calibration and Characterization Methodology

The model calibration methodology used here (summarized in Figure 2.12) is similar to the methodology proposed by Ouma [29]. Test wafers are run, and measurements are taken to obtain post-CMP thickness data. Model parameters for Phase 1 and Phase 2 are extracted. At this point, prediction may be performed on any arbitrary layout using the extracted model parameters. The local density map for the arbitrary layout Phase 1 (oxide overburden) is computed, using the deposition bias associated with the film used. An effective oxide density map can then be computed using the PL_{OX} model parameter. The local density map for Phase 2 (nitride overpolish) is computed directly from the active area layer of the arbitrary layout, and the effective density map for Phase 2 can be computed using the PL_{NIT} model parameter. The effective density maps for Phase 1 and Phase 2 can then be used in conjunction with the other Phase 1 and Phase 2 modeling parameters to predict the post-CMP dishing and erosion of this layout for an arbitrary polish time. Examples of prediction are presented in Section 2.12.

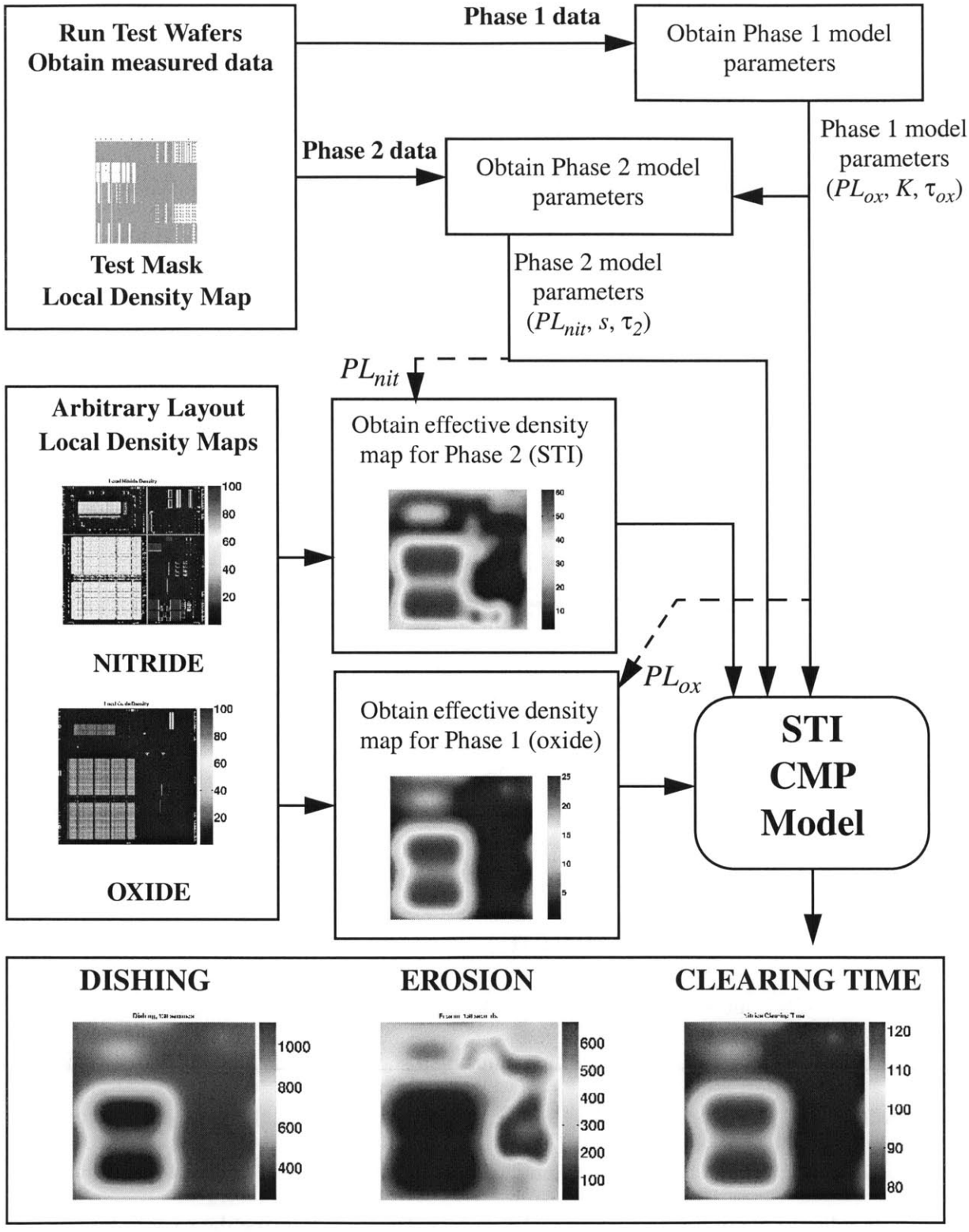


Figure 2.12: Illustration of characterization and modeling methodology for STI CMP.

2.9 Parameter Extraction

Model parameters are extracted from test wafer data run on wafers with specific characterization patterns. At a minimum, two test wafers should be run: one that polishes for a time before complete oxide overburden removal (Phase 1) and one that polishes for a time after complete clearing (Phase 2). From the Phase 1 data, model predictions are fitted to amount removed in both up and down areas to find the model parameters PL_{ox}, K, and τ_{ox}. From the Phase 2 dishing and erosion data and the extracted Phase 1 parameters, Phase 2 model parameters PL_{nit}, s, and τ_2 are extracted. The general model parameter extraction strategy, illustrated in Figure 2.13, is an iterative procedure that finds the optimal model parameters by minimizing RMS error between the predicted model fit and measured data.

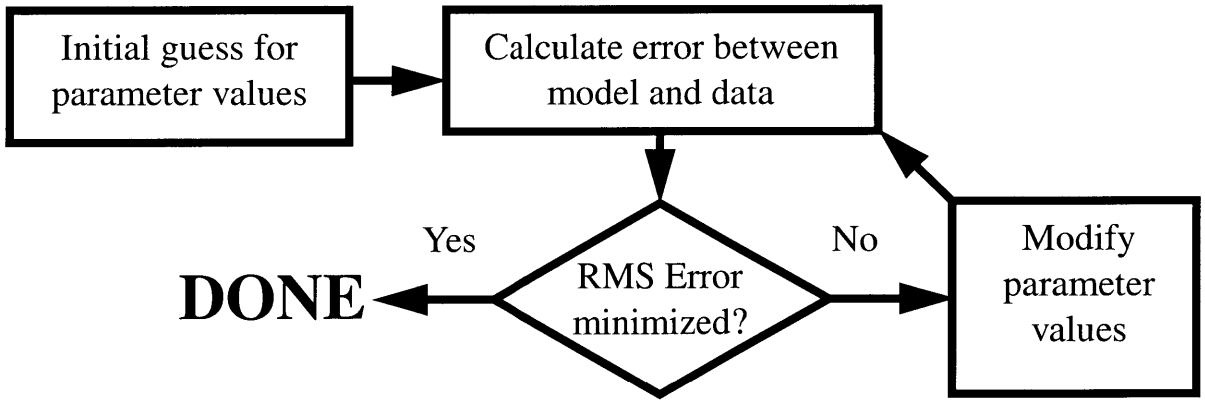


Figure 2.13: Optimization strategy for model parameters.

2.10 Experiment Verification

The STI CMP model is verified by running experimental wafers using the STI characterization test pattern. The wafers are run using a standard CMP process that is typically used in production. The density-based STI CMP model derived in this chapter is expected to yield reasonably accurate predictions, since a standard Prestonian relationship between removal rate and pressure should apply for a standard CMP process. Here we describe the experimental conditions; Section 2.11 presents model fit results for this experiment.

2.10.1 Experiment Description

Wafers are run on an IPEC 472 rotary polisher, using five different combinations of pressure and speed, and two different pads (IC1400 and IC1000 solo). The IC1400 pad is a stacked pad composed of an IC1000 surface pad on top of a softer subpad. The IC1000 solo pad contains simply the top surface pad, which is stiffer than the combined stacked pad.

Eight or nine wafers are run for increasing time steps for each pad/process combination. Time steps are targeted to include both Phase 1 and Phase 2 regimes of the polish process. The process combinations are listed in Table 2.1.

Wafers are patterned with the STI CMP characterization test mask described in Section 2.10.2. A nominal STI process is used to produce patterned CMP wafers with the following structure: nitride thickness is 1500 Å, pad oxide thickness is 100 Å, and the STI trench depth was 5000 Å, resulting in an initial step height of 6600 Å.

Table 2.1: Experimental Process Combinations

Process ID	Pad	Pressure (psi)	Speed (rpm)	# Wafers
A	IC1400	3	30	7
B	IC1400	3	90	7
C	IC1400	9	30	7
D	IC1400	9	90	7
E	IC1400	6	60	8
F	IC1000 solo	3	30	7
G	IC1000 solo	3	90	7
H	IC1000 solo	9	30	7
I	IC1000 solo	9	90	7
J	IC1000 solo	6	60	8

2.10.2 Test Mask Description

A special test mask is designed and used here for general STI CMP characterization. The design is similar to previous designs used for dielectric characterization [30] and is shown in Figure 2.14. Table 2.2 describes the various structures on the test mask and their usefulness in CMP characterization.

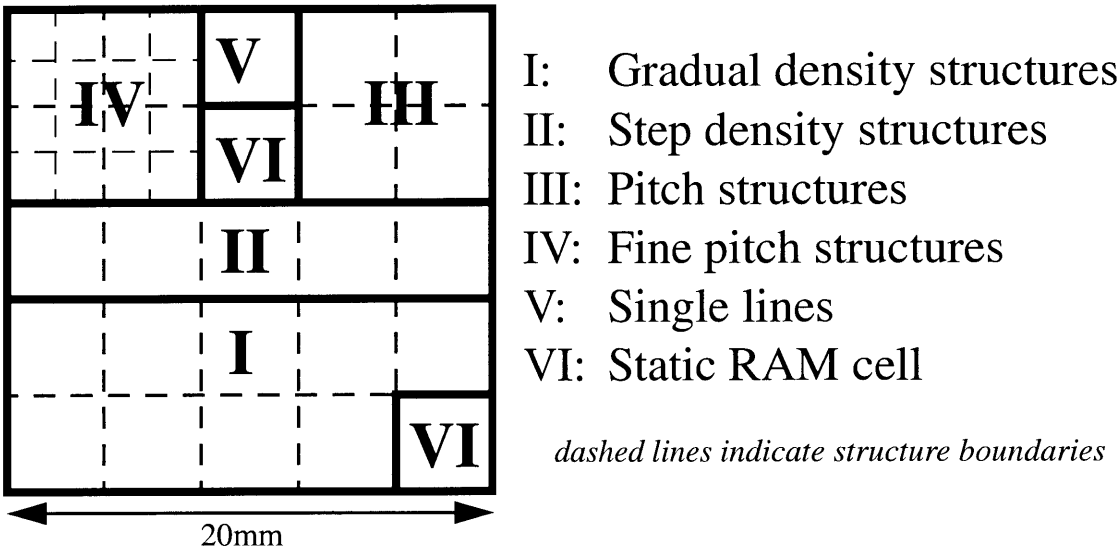


Figure 2.14: Floor plan of STI CMP characterization mask.

Table 2.2: Description of Structures on Test Mask

Region	Description	Use
I	Nine 4mm x 4mm density structures, gradually varying from 10% (lower left) to 90% (upper right). Fixed 100 μm pitch.	Analysis of CMP effects across a wide range of pattern densities
II	Five 4mm x 4mm structures with densities of 20%, 90%, 30%, 70%, and 50%. Fixed 100 μm pitch.	Characterization of planarization length
III	Four 4mm x 4mm pitch structures of fixed 50% density.	Analysis of line width/line space dependencies
IV	16 2mm x 2mm structures of fine line widths and spaces.	Characterization of deposition characteristics (e.g., deposition bias)
V	Single isolated lines	Analysis of CMP of isolated lines
VI	Static RAM cell taken from a typical cell library	Evaluation of CMP of a region resembling an actual chip

2.10.3 Experimental Data

For analysis, 33 up area points and 24 down area points on the test die are measured, as shown in Figure 2.15, both before and after CMP. The measurement locations focused on the step density region of the test mask, and included points on the SRAM cells, high and low density structures of the graduate density region, and structures in the pitch region. In certain up area locations, the corresponding down area locations are excluded due to difficulty in obtaining reliable measurement data in these locations. Measurements are taken from five dies on each wafer, using a KLA-Tencor Prometrix (model UV1250 or UV1280) thin film thickness measurement system.

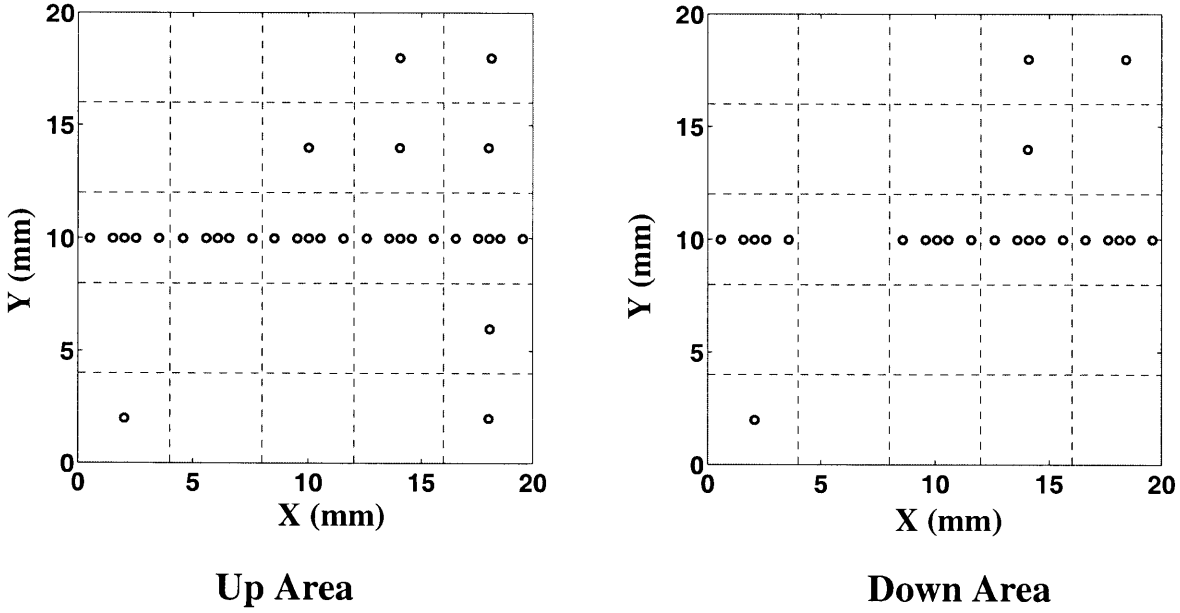


Figure 2.15: Measurement locations for conventional STI CMP experiment.

2.11 Results

Wafer data is gathered for each of the ten process conditions shown in Table 2.1, and used to extract model parameters for the STI CMP model. Phase 1 and Phase 2 data are extracted from time splits from the appropriate phases of the STI CMP process.

2.11.1 Parameter Extraction - Phase 1

Phase 1 extracted model parameters are shown in Table 2.3. Extracted model parameters are obtained by minimizing the RMS error between the model prediction and measured data for amount removed in the up and down areas.

Table 2.3: Parameter Extraction Results - Phase 1

Process ID	Polish Time (seconds)	Planarization Length (mm)	K (Å/min)	τ_{ox} (secs)	RMSE (Å)
A	240	3.9	880	80	488
B	120	4.1	1910	30	361
C	120	2.7	2490	20	391
D	40	2.3	7048	8	359
E	80	3.0	2833	20	472
F	240	6.9	1002	78	609
G	80	7.5	2318	36	513
H	90	5.3	2676	20	416
I	45	6.3	6663	4	359
J	60	6.2	3385	17	84

RMS error for all processes is approximately 600 Å or lower. The planarization length for the stacked pad process is between 2.3 and 4.1 mm, while the stiff pad process exhibits a planarization length between 5.3 and 9.2 mm, which agrees with the commonly held idea that stiffer pads exhibit longer planarization lengths. Planarization length also decreases with increasing pressure. Extracted blanket film removal rates increase with increasing speed and pressure, which agrees with the assumed Prestonian relationship. The oxide time constants are fairly similar when comparing the two pad processes with similar process conditions. The time constant τ_{ox} decreases with increasing pressure, which is also in agreement with previous assumptions.

2.11.2 Parameter Extraction - Phase 2

Using the extracted parameters from Phase 1, Phase 2 data is used to calibrate the model parameters for the nitride overpolish step. Extracted model parameters are obtained by minimizing the RMS error between the model prediction and measured data for dishing and erosion (computed from up and down area amount removed data from 24 common location sites). The results are given in Table 2.4.

Table 2.4: Parameter Extraction Results - Phase 2

Process ID	Polish Time (seconds)	Planarization Length (mm)	s	τ_2 (secs)	RMSE (Å)
A	540	4.3	7.2	297	218
B	210	19.9	2.7	320	286
C	210	14.8	38	115	136
D	75	4.3	4.2	34	127
E	150	9.8	3.1	101	160
F	420	3.0	10.3	170	103
G	150	19.9	2.8	125	1006
H	150	5.9	8.9	58	206
I	65	13.4	4.4	40	101
J	120	11.2	8.6	66	133

RMS error for most processes is 300 Å or lower. Planarization lengths for the nitride overpolish phase tend to increase with increasing speed, and decrease with increasing pressure. When compared to the planarization lengths for the oxide phase, the planarization lengths of the nitride phase are equal or larger for stacked pad processes, but have no discernible relationship to oxide phase planarization lengths for the stiff pad processes. The selectivity also appears to change with changing process conditions.

CMP behavior can be predicted as a function of site number by computing the pattern

density of particular locations (i.e., x, y coordinates) on the die using a die local density map and the planarization length of the process. Site pattern densities are then used along with the model equations and model parameters to compute amount removed, step height, dishing, and erosion values.

Plots of Phase 1 and Phase 2 data for various sites are shown for Process E and J in Figure 2.16 to Figure 2.19. The locations that are plotted are common locations in up and down area coordinates on the die. Amount removed and step height is plotted for Phase 1, and dishing and erosion are plotted for Phase 2. The figures illustrate good agreement between predicted quantities and the actual measured data.

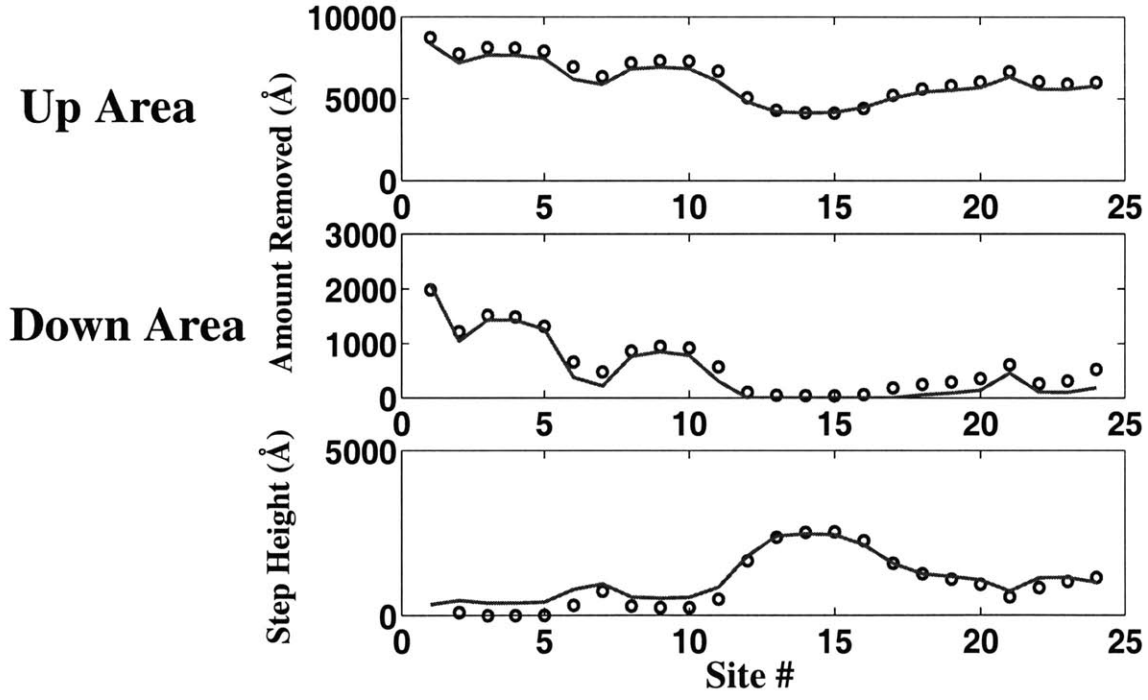


Figure 2.16: Model fit vs. experimental data for Phase 1 from Process E.

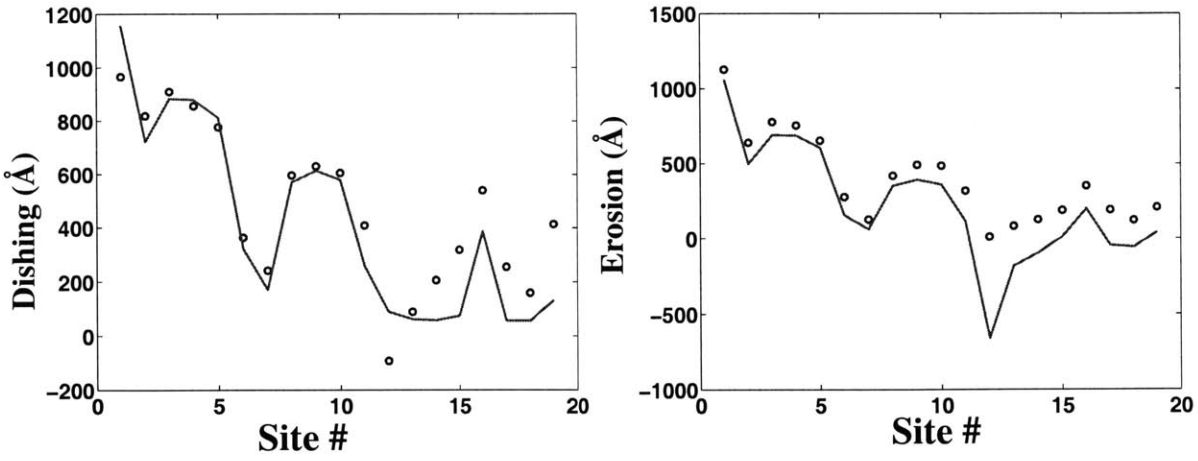


Figure 2.17: Model fit vs. experimental data for Phase 2 for Process E.

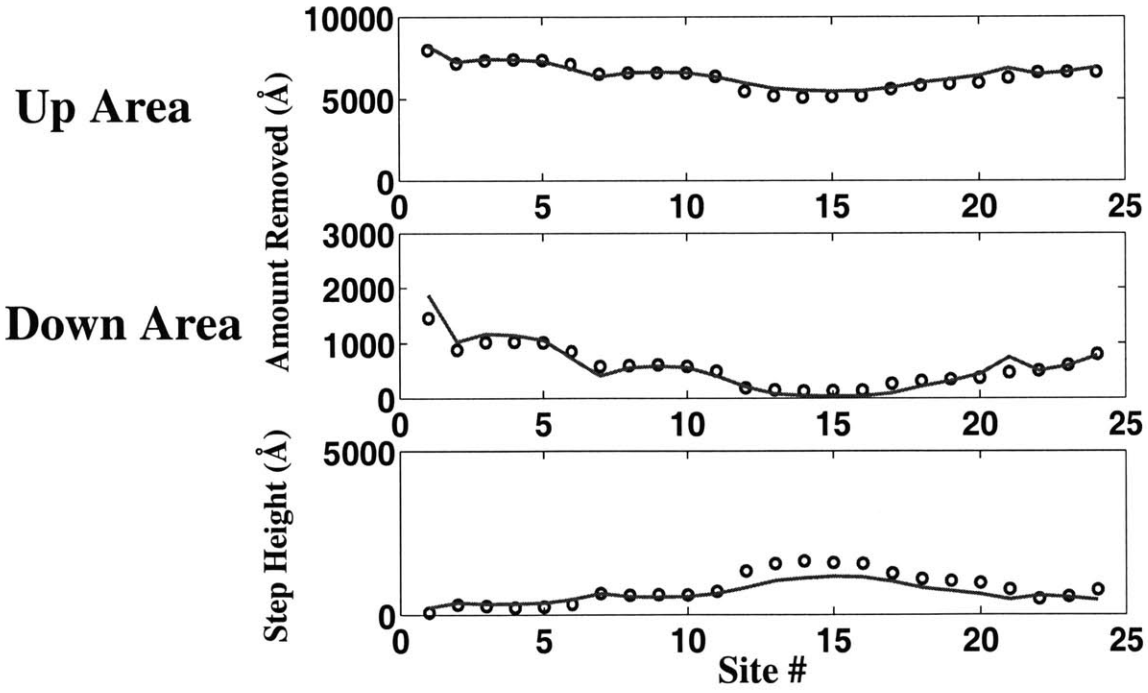


Figure 2.18: Model fit vs. experimental data for Phase 1 from Process H.

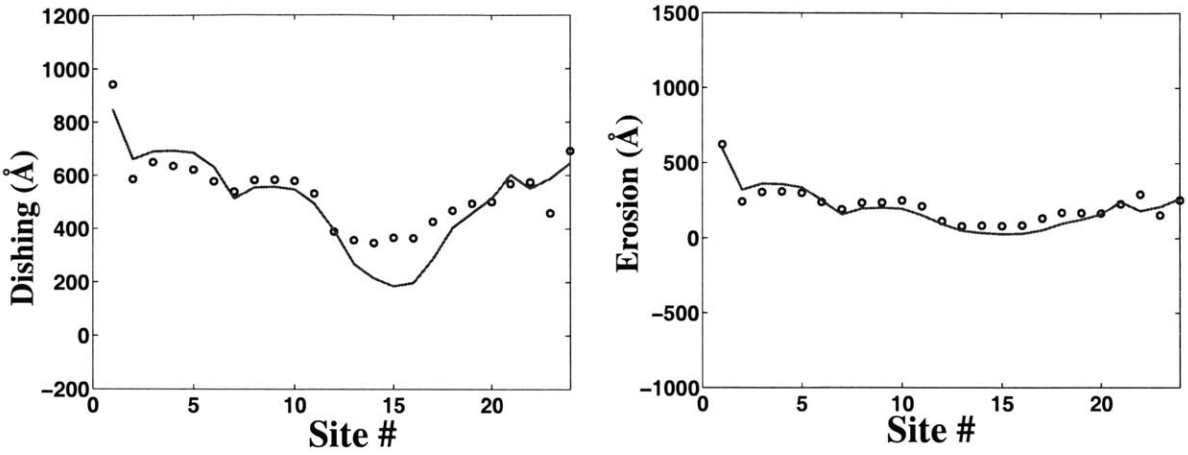


Figure 2.19: Model fit vs. experimental data for Phase 2 for Process H.

Of interest to note is the extremely long extracted planarization lengths for processes B, C, and G. Process G, in particular, has a much larger RMS fitting error than all of the other processes.

For Process B, the extracted planarization length for Phase 2 is 19.9 mm. However, a sensitivity analysis of the fitting error with planarization length reveals a fairly insensitive relationship of error to planarization length (see Figure 2.20). For other processes, the RMS error has a greater sensitivity to changes in planarization length (such as Process A, also shown in Figure 2.20). The insensitive response of RMS fitting error to planarization length tends to reduce the confidence in the result for nitride planarization length for Process B, since a wide range of planarization lengths will not appreciably change the fitting error.

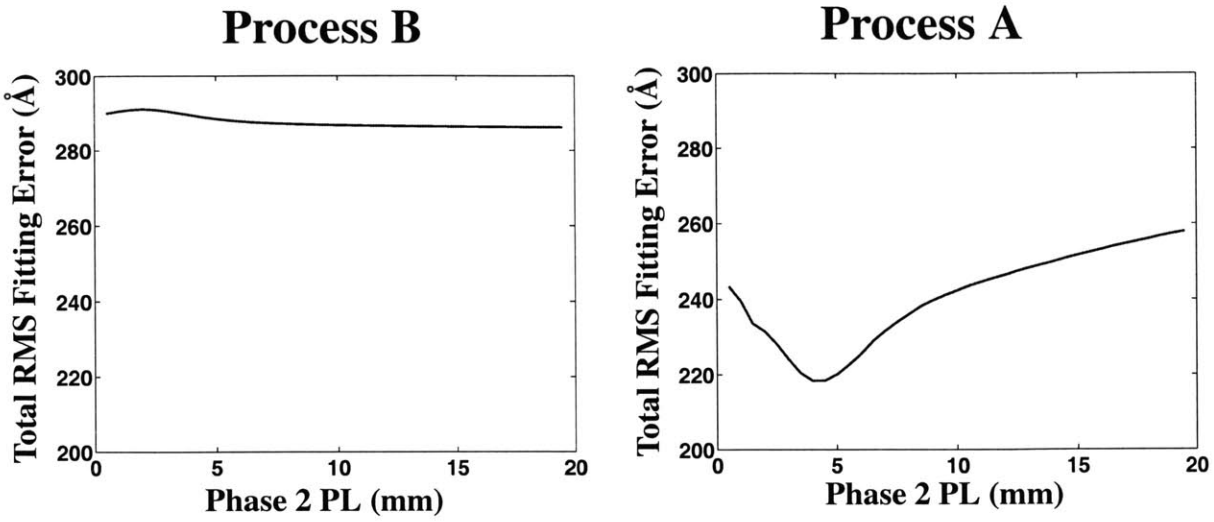


Figure 2.20: Sensitivity of RMS fitting error to Phase 2 planarization length.

The large error for process G can be explained by analyzing the data. Figure 2.21 plots erosion for Process G and for Process F. Expected behavior for erosion is to decrease with increasing effective pattern density. Process F illustrates this expected trend. However, the erosion data for Process G exhibits an unexpected trend for erosion with respect to pattern density, leading to large fitting errors. We conjecture that a very large wafer-level removal nonuniformity exists in this wafer, skewing the removal rates across the chip based on spatial location rather than based on pattern density.

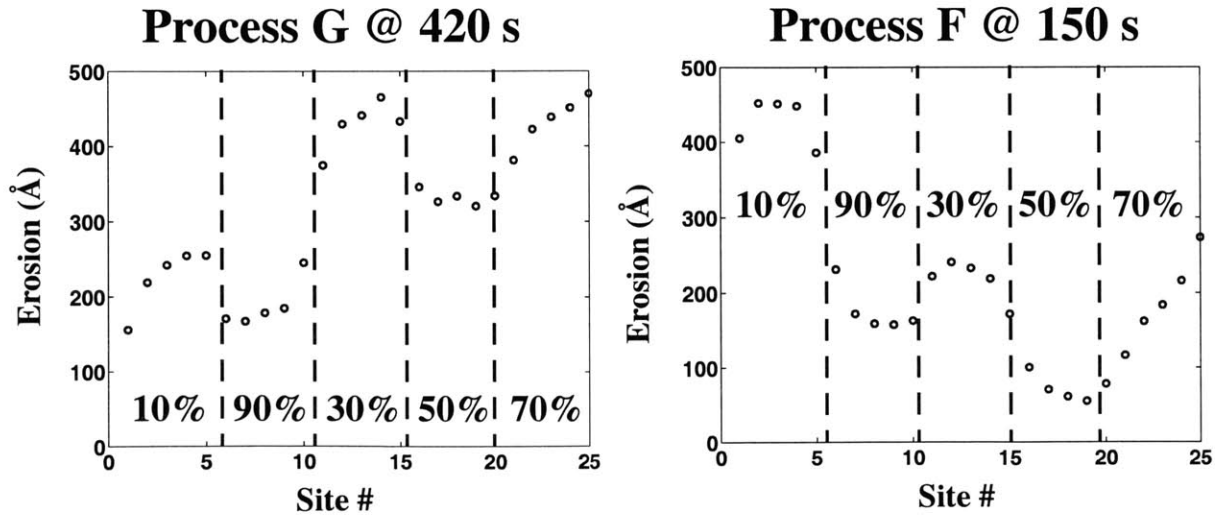
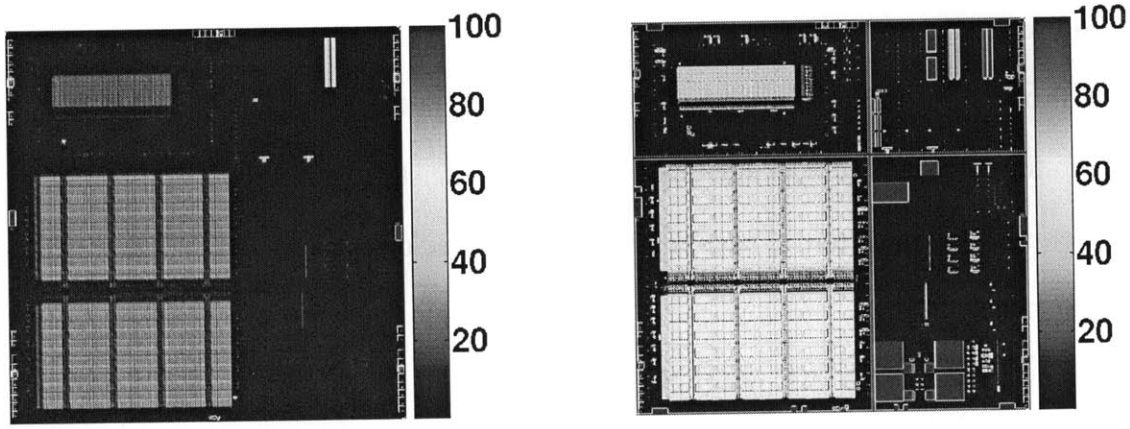


Figure 2.21: Comparison of erosion data for Processes G and F.

2.12 Modeling Applications

The key idea behind the modeling methodology proposed in this chapter is that it can be used for prediction of nitride erosion and oxide dishing on arbitrary layouts for a given characterized CMP process. In order to implement the model on an arbitrary layout, density information from the layout is required. Specifically, the local density map of the active area nitride layer and the local density map of the overburden oxide layer need to be generated. Example local density maps are shown in Figure 2.22. The nitride density map can be computed using the active area layer on a layout; the oxide density map is computed by biasing the nitride layer with the oxide deposition bias, subtracting out the reverse etchback layer (if one is used), and computing the density map of the resulting layout layer. In this work, we discretize the chip into cells 40 μm by 40 μm in size, and compute the local layout pattern density within each cell.



Oxide Layer Local Density **Nitride Layer Local Density**

Figure 2.22: Local densities for STI oxide and nitride layers for case study layout.

2.12.1 Dishing and Erosion Prediction

Trench oxide dishing and active area nitride erosion can be predicted using model parameters obtained through the characterization methodology. The example layout of Figure 2.22 is used as a test case, with model parameters as follows: 3 mm Phase 1 planarization length PL_{OX}, 2000 Å/min blanket oxide removal rate K, 20 second oxide time constant τ_{ox}, 2.3 mm

Phase 2 planarization length PL_{NIT}, selectivity s of 10, and pure nitride time constant τ_2 of 60 seconds. For this simulation, initial step height is 6600 Å and initial deposition is 9000 Å. Predicted dishing and erosion maps for the example layout are shown in Figure 2.23.

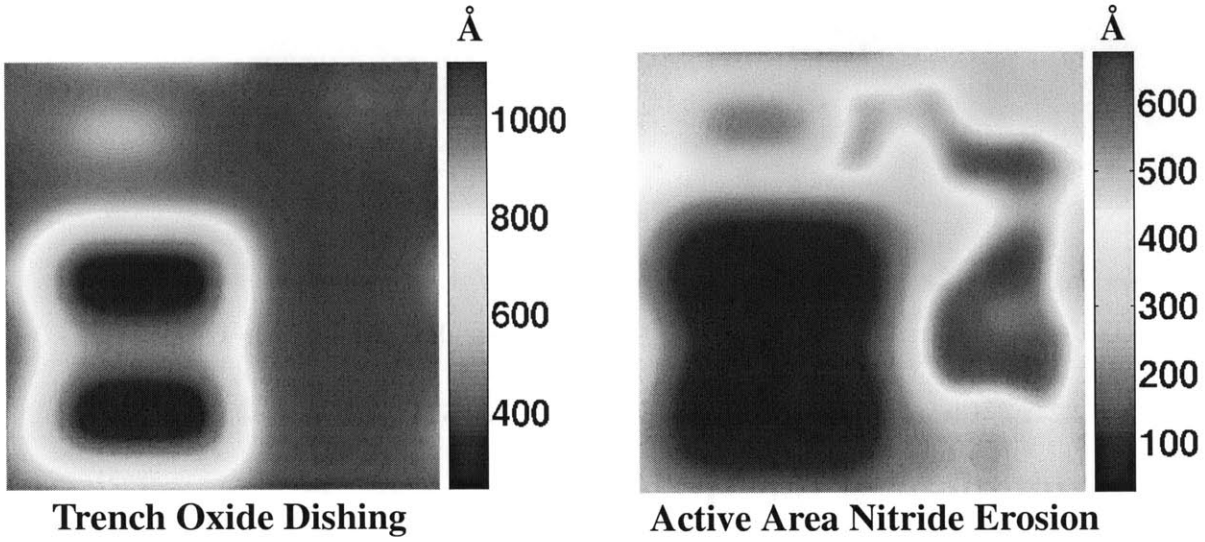
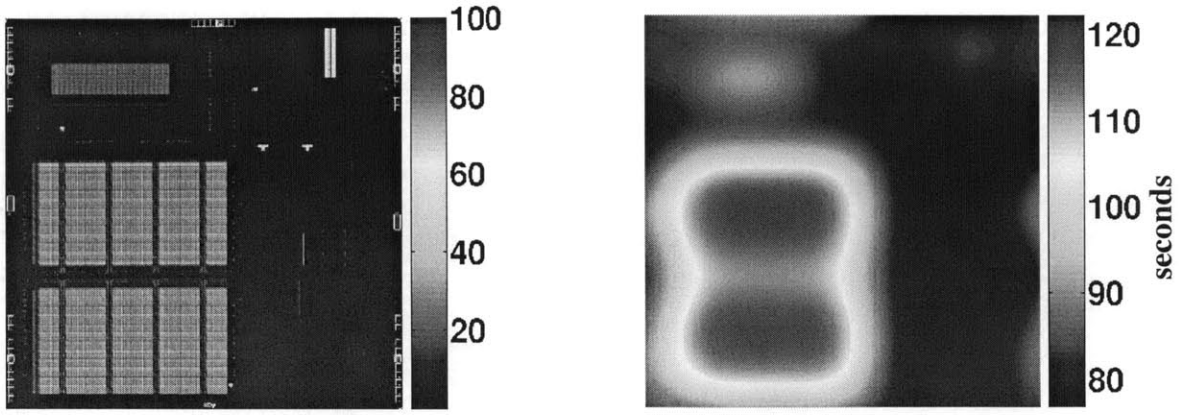


Figure 2.23: Dishing and erosion prediction for conventional STI CMP process.

The effect of pattern density on chip-scale dishing and erosion is clearly indicated. The regions on the chip that exhibit the lowest predicted dishing and erosion are the regions that have the highest nitride and oxide density. Similarly, low density regions on the chip exhibit the highest predicted dishing and erosion.

2.12.2 Clearing Time Prediction

Another useful piece of information that can be predicted using the STI CMP model is the nitride touch-down time. Unlike the dishing and erosion prediction, the nitride touch-down time prediction can be performed using only the local oxide density information, since oxide clearing is determined only by characteristics of the overburden oxide polish. An example prediction using a sample layout and process parameters is shown in Figure 2.24. We see that touch-down or oxide clearing begins at approximately 80 seconds, and completes at approximately 120 seconds. This information might be useful in conjunction with an endpoint detection system in an STI polish, to help set the length of the overpolish time to ensure full clearing of the die.



Oxide Layer Local Density **Nitride Touch-down Time**

Figure 2.24: Clearing time prediction for conventional STI CMP process.

2.13 Limitations of the Model

There are a few caveats that must be mentioned about the model developed in this chapter. First, the characterization test mask used for the work here has a minimum feature size of 10 μm. This is considerably larger than the minimum feature size of current technologies. Thus, while the model has been shown to work effectively for the large features greater than 10 μm, further study is required to verify and, if necessary, adapt the model for smaller features.

Second, the effect of line width and space has not been thoroughly studied in this work. Line width and line space has been demonstrated to have a significant effect on dishing and erosion of small features in copper CMP processes [81,82]. The effect of line width and line space on STI CMP dishing and erosion in structures with small feature sizes also needs further study; if these effects are found important in STI, model extensions similar to those proposed by Tugbawa [56] for copper CMP might be appropriate.

2.14 Non-Prestonian CMP Model

It has been conjectured that certain non-Prestonian behavior may result in improved planarity performance in a CMP process [76]. In this context, non-Prestonian behavior is defined

as a nonlinear relationship between removal rate and pressure. In order to accommodate this change, we can modify our STI CMP model: only the relationship between removal rate and pressure is changed from the Prestonian CMP system; the relationship between pressure and step height is the same. In this section, the ramifications of a non-Prestonian CMP system are explored for Phase 1 (oxide polish).

Typically, the most beneficial situation is where a low pressure results in a smaller removal rate than would normally occur in a Prestonian system, as illustrated in Figure 2.25a. The reason that this may improve planarity is that the down areas, which typically bear lower pressures than the up areas, will polish at a slower rate than in a Prestonian system, while the up areas polish at the same rate as a Prestonian system. This results in faster step height reduction.

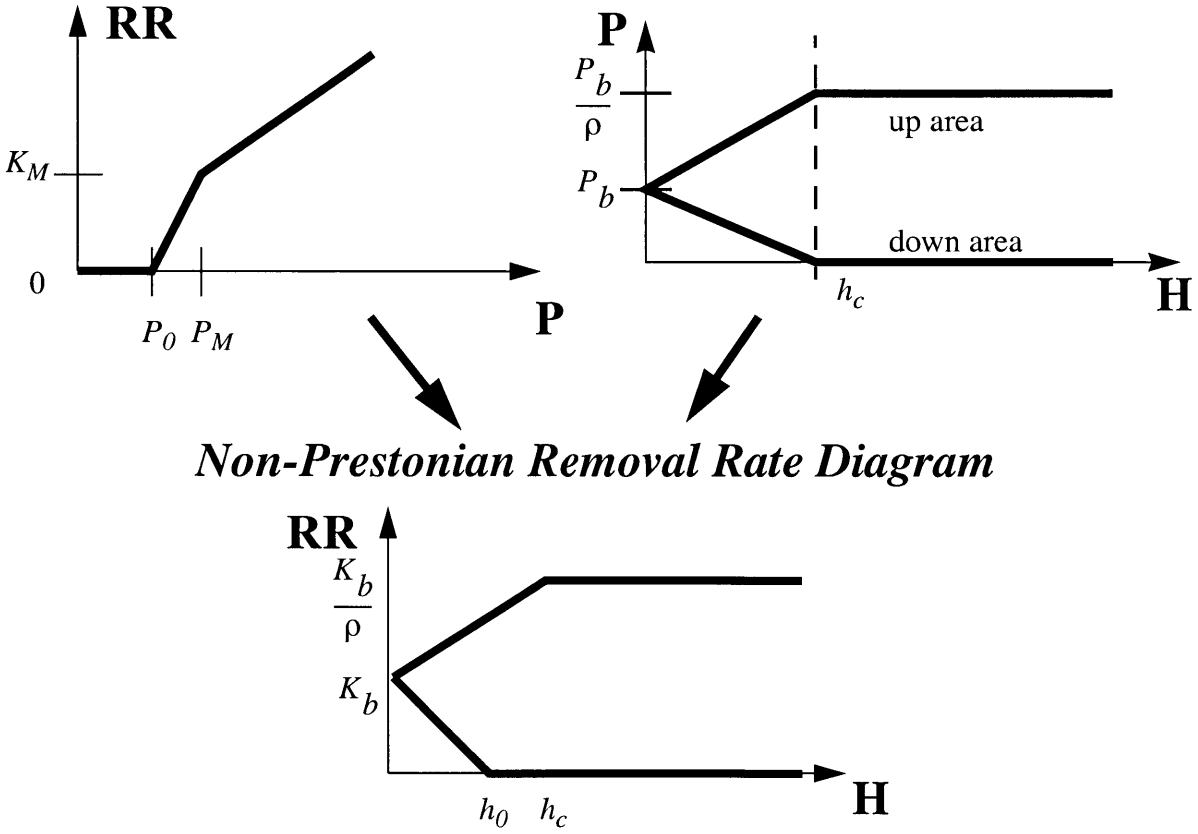


Figure 2.25: Derivation of a non-Prestonian removal rate diagram.

The removal rate versus pressure relationship shown in Figure 2.25 can be expressed as:

$$P \le P_0 \qquad RR = 0 \tag{2.57}$$

$$P_0 < P \le P_M \qquad RR = K\frac{(P - P_0)}{(P_M - P_0)} \tag{2.58}$$

$$P > P_M \qquad RR = \alpha(P - P_M) + K_M \tag{2.59}$$

2.14.1 Implications on AR equations

Since non-Prestonian behavior changes the removal rate diagram, the step height, removal rate, and amount removed equations all change. There may be additional subphases in the polishing process as well. The form of the revised model depends strongly on the relative values of P_0, P_M, P/ρ, and P_b in Figure 2.25.

The underlying benefit in this particular type of non-Prestonian system lies in two key ideas, based on the underlying principle that the down areas between features bear less pressure than the pressure borne by a blanket wafer, and up areas bear pressures higher than the blanket wafer pressure. The first idea is that step height reduction (i.e., planarization) proceeds faster if the up area removal rates are increased relative to the down area removal rates. Therefore, to improve the step height reduction relative to a Prestonian system, the removal rates at pressures higher than the blanket wafer pressure should either increase or remain the same as a Prestonian system, while the removal rates at pressures lower than the blanket wafer pressure should either decrease or remain the same as a Prestonian system. The second idea is that dishing can be reduced if the down area is at a pressure below the threshold pressure. This create low trench area removal during the CMP process.

In comparing a Prestonian case with this particular form of a non-Prestonian system, for valid comparison the systems are set up such that the blanket polish rates for the two systems are the same for the blanket polish pressure P_b. Using the arguments given in this section, the optimal situation occurs when P_b is less than P_M but greater than P_0. This is illustrated in Figure 2.26. P_b

is required to be greater than P_0 so that up area pressures are guaranteed to be non-zero.

Revised equations for step height, removal rate, and amount removed can be derived based on the values of P_0, P_b, P_M, K_M, ρ, and h_c. The exact form of the equations depends on the value of P_b/ρ relative to P_M. For this analysis, $P_b/\rho < P_M$ is assumed, and the step height, removal rate, and amount removed equations are derived as shown in Equations 2.60 through 2.74. Similar equations can be derived for the case of $P_b/\rho > P_M$.

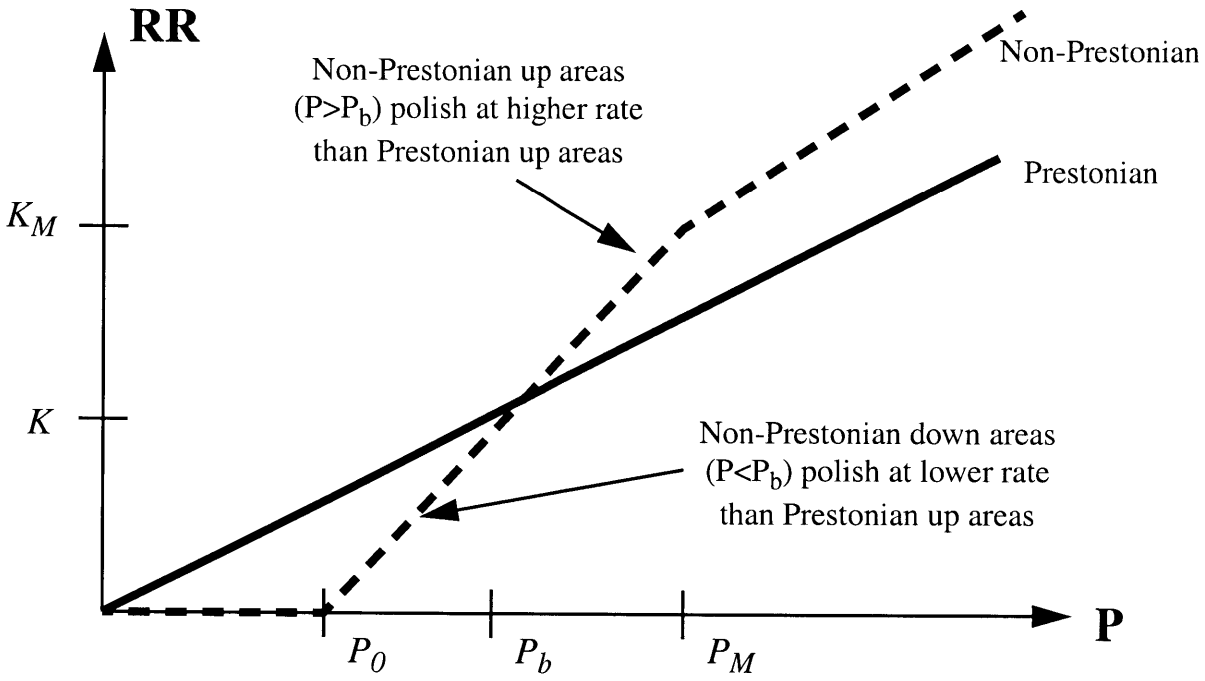


Figure 2.26: Prestonian vs. non-Prestonian relationship comparison, for $P_b < P_M$.

For $0 \le t < t_c$:

$$H = z_1 - K_1 t \tag{2.60}$$

$$RR_u = K_1 \tag{2.61}$$

$$RR_d = 0 \tag{2.62}$$

$$AR_u = K_1 t \tag{2.63}$$

$$AR_d = 0 \tag{2.64}$$

For $t_c \leq t < t_0$:

$$H = K_1 \tau_{ox1} e^{-\frac{t - t_c}{\tau_{ox1}}} - K_b \tau_{ox1} \tag{2.65}$$

$$RR_u = K_1 e^{-\frac{t - t_c}{\tau_{ox1}}} \tag{2.66}$$

$$RR_d = 0 \tag{2.67}$$

$$AR_u = K_1 t_c + K_1 \tau_{ox1} \left[1 - e^{-\frac{t - t_c}{\tau_{ox1}}} \right] \tag{2.68}$$

$$AR_d = 0 \tag{2.69}$$

For $t > t_0$:

$$H = h_0 e^{-\frac{t - t_0}{\tau_{ox2}}} \tag{2.70}$$

$$RR_u = K_b + \frac{h_0}{h_c}(K_1 - K_b) e^{-\frac{t - t_0}{\tau_{ox2}}} \tag{2.71}$$

$$RR_d = K_b - K_b e^{-\frac{t - t_0}{\tau_{ox2}}} \tag{2.72}$$

$$AR_u = K_1 t_c + (h_0 - h_c) + K_b (t - t_0) + \frac{h_0}{h_c}(K_1 - K_b)\tau_{ox2} \left[1 - e^{-\frac{t - t_0}{\tau_{ox2}}} \right] \tag{2.73}$$

$$AR_d = K_b (t - t_0) - K_b \tau_{ox2} \left[1 - e^{-\frac{t - t_0}{\tau_{ox2}}} \right] \tag{2.74}$$

where the following equations are used for parameters in Equation 2.60 to Equation 2.74:

$$t_c = \frac{z_1 - h_c}{K_1} \tag{2.75}$$

$$t_0 = t_c + \tau_{ox1} \ln\left|\frac{K_1 \tau_{ox1}}{h_0 + K_b \tau_{ox1}}\right| \tag{2.76}$$

$$\tau_{ox1} = \frac{h_c}{K_1 - K_b} \tag{2.77}$$

$$\tau_{ox2} = \frac{1}{\frac{1}{\tau_{ox1}} + \frac{K_b}{h_0}} \tag{2.78}$$

$$h_0 = \frac{h_c(P_b - P_0)}{P_b} \tag{2.79}$$

$$K_1 = K_M\left(\frac{\frac{P_b}{\rho} - P_0}{P_M - P_0}\right) \tag{2.80}$$

$$K_b = K_M\left(\frac{P_b - P_0}{P_M - P_0}\right) \tag{2.81}$$

As an example case, values of P_0=30 kPa, P_M=60 kPa, K_M=3000 Å/min, α=50, P_b=50 kPa, ρ=90%, h_c=3000 Å, and z_1=5000 Å are used. Comparing the resulting step height vs. time for the two situations shows a clear increase in step height reduction rate, as seen in Figure 2.27. These results indicates the benefits, in certain circumstances, of the non-Prestonian pressure. Similar modification of the model equations in these cases to study the effect on dishing and erosion is also possible. Model equations may be similarly derived for other possible non-Prestonian CMP systems. These equations will not be described in this work, since most of the

other systems do not provide a benefit in terms of planarization performance.

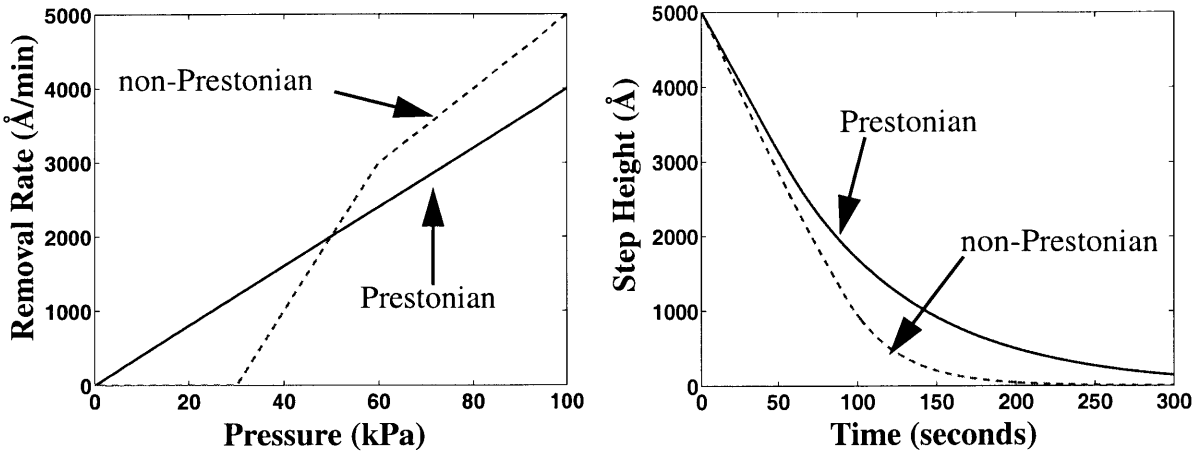


Figure 2.27: Comparison of Prestonian and non-Prestonian systems.

2.15 Numerical Solution for STI CMP

It is possible to numerically compute amount removed versus time and step height versus time results from arbitrary removal rate versus pressure relationships. The first step is the creation of the removal rate versus step height relationship. The two components for this are a removal rate vs. pressure relationship and a pressure vs. step height relationship. The removal rate vs. pressure relationship can be created from theoretical derivation or experimental data. The pressure vs. step height relationship is created from estimations of contact height, raised areas and steady-state removal rates of features at different densities. This relationship may contain values that are entirely theoretical, although certain aspects of it (such as contact height) may be estimated from experimental data.

These two functions can be combined to create a numerically derived removal rate versus step height relationship, as shown in Figure 2.28. The removal rate diagram may not be expressible in closed-form, but rather as a paired tabular set of values, one for up area removal rate and step height, and one for down area removal rate and step height, from which interpolations can be made.

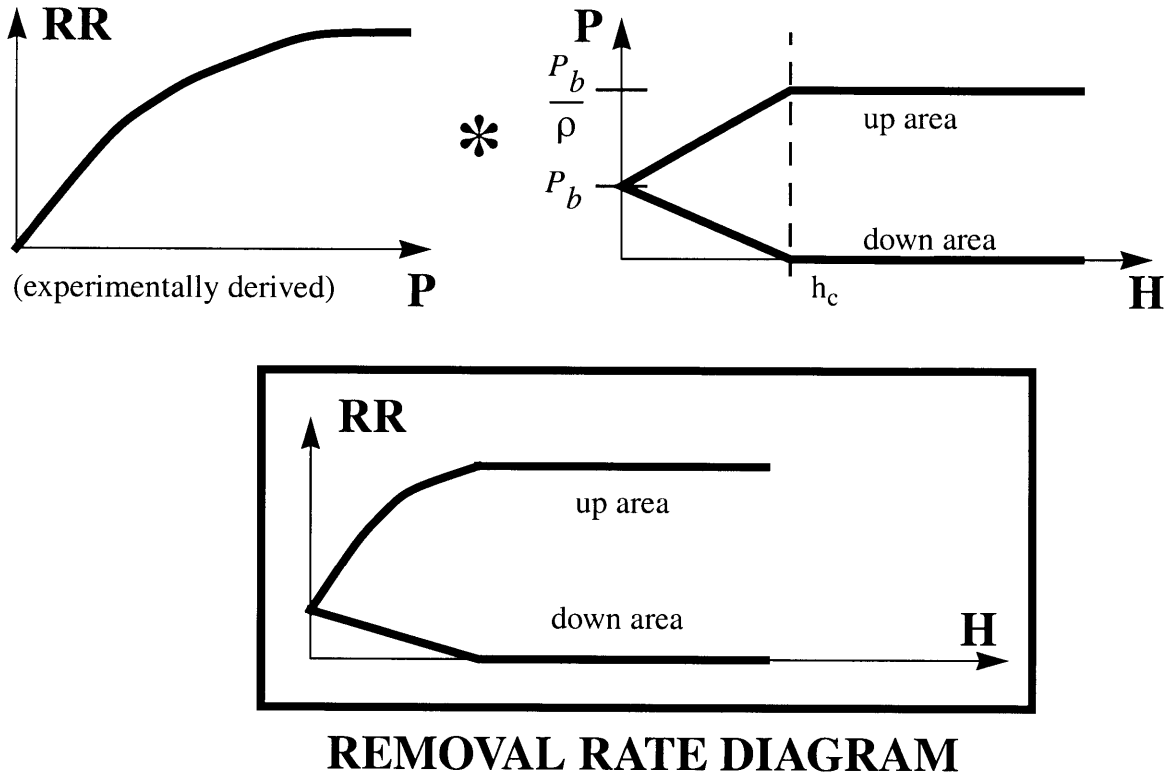


Figure 2.28: Numerically derived removal rate versus step height relationship.

2.15.1 Deriving Step Height, Removal Rate, and Amount Removed

To derive numerical solutions for step height, removal rate, and amount removed based on the arbitrary removal rate versus step height relationship, it is necessary to use numerical techniques rather than closed form algebraic solutions. First we re-examine Equation 2.6:

$$RR_{up} - RR_{down} = -\left(\frac{dH}{dt}\right) \qquad (2.6)$$

Given a numerically derived removal rate diagram, the left side of this equation can be fully determined by numerical subtraction of the up and down area rates in the tabulated removal rate function. Equation 2.6 can be rearranged (and then integrated on both sides) to give:

$$dt = \frac{1}{RR_{up}(H) - RR_{down}(H)} dH \tag{2.82}$$

$$t = \int dt = \int \frac{1}{RR_{up}(H) - RR_{down}(H)} dH \tag{2.83}$$

It is possible to numerically integrate the right side of Equation 2.83, creating a numerical solution for t as a mapping of H, where t as a mapping of H means that it is possible to produce (t, H) pairs. Our notation for t as a mapping of H will be $t^m(H)$. It is then possible to interpolate the (H, t) pairs onto a regular discretization in t to produce $H^m(t)$.

Finally, by numerically combining $H^m(t)$ with $RR_u{}^m(H)$ and $RR_d{}^m(H)$, it is possible to obtain $RR_u{}^m(t)$ and $RR_d{}^m(t)$. Numerical integration of the removal rates with time results in solutions for amount removed in up and down areas as mappings of time. A sample output of the numerical solution, using both Prestonian and non-Prestonian removal rate vs. step height curves, is illustrated in Figure 2.29.

The disadvantage of numerical solution method is that it cannot be usefully utilized for CMP process characterization, since the initial removal rate diagram is based on estimations of contact height and pressures. The calculation of the solution requires considerably more computation time than the solution involving closed form analytic equations, and so parameter extraction by way of an iterative least squares fit would be difficult.The numerical method is most useful as a tool for exploring theoretical variations the amount removed, removal rate, and step heights vs. time curves when there are modifications to the removal rate vs. pressure diagram.

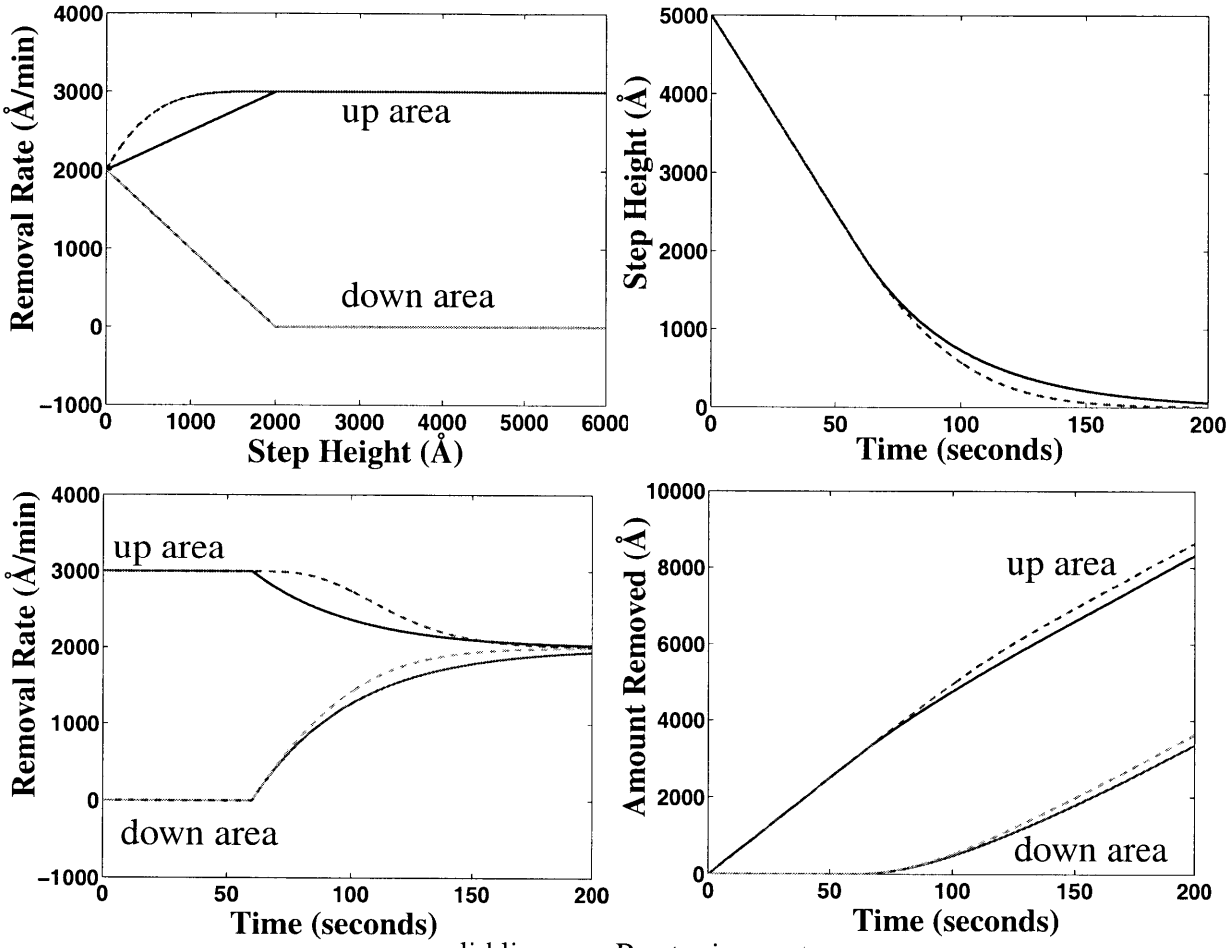


Figure 2.29: Numerically derived step height, removal rate, and amount removed vs. time.

2.16 Summary

This chapter has described a compact, closed-form STI CMP model that can be used to predict both oxide overburden and nitride overpolish phases of the STI CMP process. Based on the pattern density and pad compressibility effects, the closed-form model equations are derived resulting in a small number of descriptive model parameters.

A comprehensive modeling and characterization methodology has been presented. Validation of the modeling methodology has been demonstrated using experimental data from test wafers. The usefulness of the modeling technique has been illustrated through applications of

erosion and dishing prediction, as well as nitride touch-down time estimation.

Non-Prestonian CMP models have been briefly discussed. Given closed-form expressions for the removal rate vs. pressure diagrams, the ability to derive closed-form solutions for step height, removal rate, and amount removed as functions of time has been illustrated. Useful regimes of the non-Prestonian CMP removal rate diagrams have been described.

A numerical method for predicting step height, removal rate, and amount removed vs. time from arbitrary removal rate vs. pressure relationships has been described. This technique is not suitable for inclusion within an STI CMP process characterization methodology, but is useful as a tool for exploring the implications of theoretical removal rate vs. pressure relationships on STI CMP behavior.

Chapter 3
Nonconventional CMP Consumables

New CMP consumables are being researched with hopes of improving planarity and uniformity performance of the CMP process. These new consumables do not exhibit the polishing behavior that is typically expected from CMP processes using conventional CMP pads and slurries. Our goal in this chapter is to adapt our STI CMP model in order to accurately predict the new polish behavior shown by different consumable sets. This chapter generalizes the STI CMP model, and examines test cases based on several nonconventional consumables.

First, motivation for the use of nonconventional CMP consumables is discussed. The STI CMP model developed in Chapter 2 is then extended to enable capturing of nonconventional STI CMP behavior. The characterization and prediction methodology is revised to incorporate the nonconventional STI CMP model. The methodology is then verified using experimental data from three consumable sets: a fixed abrasive CMP pad, a high selectivity silica based slurry, and a high-selectivity ceria based slurry. An example of the dishing, erosion, and clearing time prediction is performed using an SRAM layout.

3.1 Motivation

The goal of improving the STI process is rooted in the dishing and erosion problems that occur in conventional STI processes. While dishing and erosion problems can often be approached via layout design optimization (see Chapter 4), recent research has also sought the development of new consumable sets that improve upon the dishing or erosion problems that plague conventional STI processes.

To that end, improvement is sought via two avenues. The first is reduction in erosion through the use of high material selectivity slurries. By creating a slurry that has a very low nitride

removal rate, while also maintaining an oxide removal rate comparable to conventional CMP processes, the erosion of the nitride overpolish phase will be less since the nitride polish rate is less.

The second method for reduction of erosion and dishing problems is the development of consumables with high selectivity to wafer topography. Selectivity to topography refers to a process where appreciable removal rates are exhibited only when local step heights exist during the CMP process; once the local step height is removed, the removal rates drop significantly. This is useful because it reduces post-CMP thickness variation of a single material CMP process. This benefits the STI CMP process because the oxide overburden phase does not exhibit pattern-sensitive variation in oxide clearing, thus reducing the erosion and dishing that is due to oxide overburden density mismatches.

With the development of new consumable sets, the modeling techniques for STI CMP developed in Chapter 2 may not be sufficient to properly predict post-CMP film thicknesses, dishing, and erosion. Specifically, the assumed density dependencies in the CMP models of Chapter 2 may not be applicable. Therefore, a new modeling technique is needed to deal with modeling of nonconventional consumable sets.

The modeling technique proposed in this chapter will be purely phenomenological; that is, the goal is to simply capture observed CMP behavior on patterns, and create models that predict polishing based on the observed behavior. The model parameter dependencies on density that can be investigated using the modeling techniques developed in this chapter may lend insight into the development of a more physical based model to explain nonconventional CMP process behavior.

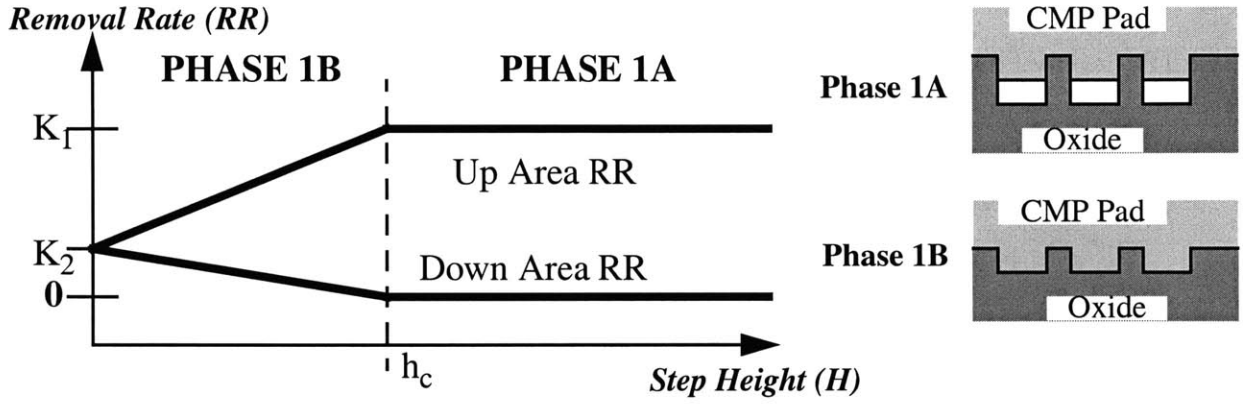
3.2 Generalizing the STI CMP model

In the pattern density and density-step-height based CMP models, a particular dependency of pattern removal rate on pattern density has been assumed. To generalize the model, this dependency will be broken, thus allowing for the modeling of CMP processes where the pattern

removal rate has an unknown dependency on density. In a similar fashion, any previous density dependencies of model parameters (e.g., contact height, steady-state dishing, steady-state removal rate, steady-state time constant, patterned nitride removal rate) will be broken to create a general model.

3.2.1 Rethinking Phase 1

Breaking density dependencies requires reformulating all of the model equations in the STI CMP model so that density is not explicitly included. This first requires creating a new removal rate vs. step height relationship. Recreating Figure 2.9 with no density dependency results in Figure 3.1, where K_1 indicates up area (patterned) removal rate, and K_2 is the blanket removal rate.



Phase 1A indicates polish before the CMP pad contacts the down areas.
Phase 1B indicates polish after down area has been initially contacted.

Figure 3.1: Generalized removal rate diagram for STI CMP polish, Phase 1.

New equations can then be derived from Figure 3.1 for removal rate as a function of step height:

Phase 1A:

$$RR_u = K_1 \tag{3.1}$$

$$RR_d = 0 \tag{3.2}$$

$$RR_u = \left(\frac{K_1 - K_2}{h_c}\right)H + K_2 \tag{3.3}$$

Phase 1B:

$$RR_d = -\left(\frac{K_2}{h_c}\right)H + K_2 \tag{3.4}$$

Forming the new differential equation based on Equation 2.6:

$$\frac{dH}{dt} = -\left(\frac{K_1}{h_c}\right)H \tag{3.5}$$

As in the conventional STI CMP model, this can be solved by assuming a boundary condition. There are two possible boundary conditions, depending on whether the initial step height is greater or less than the contact height (i.e., whether Phase 1 includes both Phase 1A and Phase 1B, or only Phase 1B). Solving for the two cases leads to:

Polish Situation

$h_c \le z_1$

$$H = h_c e^{-\left(\frac{t - t_c}{\tau_{ox}}\right)} \tag{3.6}$$

$h_c > z_1$

$$H = z_1 e^{-\left(\frac{t}{\tau_{ox}}\right)} \tag{3.7}$$

$$\tau_{ox} = \frac{h_c}{K_1} \tag{3.8}$$

Note that these are the same as the conventional STI CMP model, except that the definition of τ_{ox} has been changed to reflect the departure from the assumed density dependency. The step height equations can be used to derive removal rate and amount removed equations, similar to the conventional STI CMP model. The final relationship given in Equations 3.9 through 3.23 will have no explicit density dependencies in them; instead, K_1 and K_2 can implicitly accommodate a range of removal rate vs. pattern density relationships.

Polish Situation Includes Phase 1A and Phase 1B ($h_c < z_1$)

Phase 1A Equations ($t < t_c$)

$$H = z_1 - K_1 t \tag{3.9}$$

$$RR_u = K_1 \tag{3.1}$$

$$RR_d = 0 \tag{3.2}$$

$$AR_u = K_1 t \tag{3.10}$$

$$AR_d = 0 \tag{3.11}$$

Phase 1B Equations ($t_c < t < t_n$)

$$H = h_c e^{-\left(\frac{t - t_c}{\tau_{ox}}\right)} \tag{3.12}$$

$$RR_u = (K_1 - K_2) e^{-\left(\frac{t - t_c}{\tau_{ox}}\right)} + K_2 \tag{3.13}$$

$$RR_d = -K_2 e^{-\left(\frac{t - t_c}{\tau_{ox}}\right)} + K_2 \tag{3.14}$$

$$AR_u = K_1 t_c + K_2 (t - t_c) + \tau_{ox} (K_1 - K_2) \left[1 - e^{-\frac{(t - t_c)}{\tau_{ox}}} \right] \tag{3.15}$$

$$AR_d = K_2 (t - t_c) - \tau_{ox} K_2 \left[1 - e^{-\frac{(t - t_c)}{\tau_{ox}}} \right] \tag{3.16}$$

$$\tau_{ox} = \frac{h_c}{K_1} \tag{3.8}$$

$$t_c = \frac{(z_1 - h_c)}{K_1} \tag{3.17}$$

Polish Situation Includes Only Phase 1B ($h_c > z_1$)

$$H = z_1 e^{-\left(\frac{t}{\tau_{ox}}\right)} \tag{3.18}$$

$$RR_u = \frac{z_1}{h_c}(K_1 - K_2)e^{-\frac{t}{\tau_{ox}}} + K_2 \tag{3.19}$$

$$RR_d = -\frac{K_2 z_1}{h_c}e^{-\frac{t}{\tau_{ox}}} + K_2 \tag{3.20}$$

$$AR_u = K_2 t + z_1\left(1 - \frac{K_2}{K_1}\right)\left[1 - e^{-\frac{t}{\tau_{ox}}}\right] \tag{3.21}$$

$$AR_d = K_2 t - z_1\left(\frac{K_2}{K_1}\right)\left[1 - e^{-\frac{t}{\tau_{ox}}}\right] \tag{3.22}$$

where τ_{ox} is given by Equation 3.8. The generalized STI model can be fully solved if K_1, K_2, and h_c are known. These equations are location specific; that is, each location on the die will have a specific (K_1, K_2, h_c) value associated with it. These parameters have as yet unknown density dependencies that will be formulated as a step of the characterization methodology.

3.2.2 Rethinking Phase 2

To generalize Phase 2 (nitride overpolish) of an STI CMP process, model parameter dependencies on density are also broken, again resulting in revised equations that have no explicit density dependencies. The resulting removal rate vs. step height relationship is illustrated in Figure 3.2. There are several parameters that are included in this figure: K_{n1} is the patterned nitride removal rate, K_{ox2} is the blanket oxide removal rate, K_{n2} is the blanket nitride removal rate, K_{ss} and D_{ss} are the removal rate and step height (dishing), respectively, at steady-state, and d_{max} is the maximum amount of dishing that can occur in this phase.

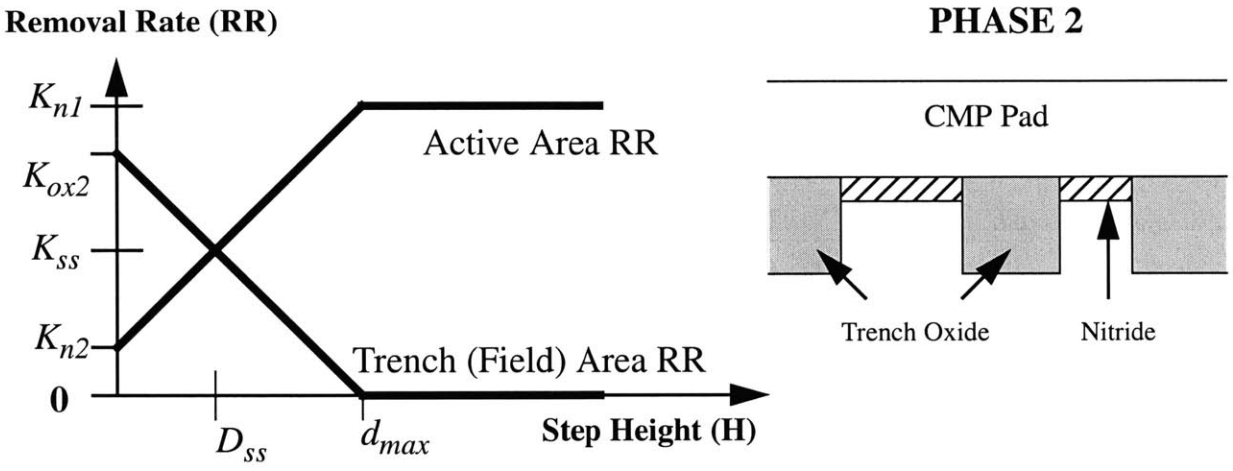


Figure 3.2: Removal rate diagram for STI CMP polish, Phase 2.

The removal rate equations from Figure 3.2 are:

$$RR_a = \left(\frac{K_{n1} - K_{n2}}{d_{max}}\right) H + K_{n2} \qquad (3.23)$$

$$RR_f = -\left(\frac{K_{ox2}}{d_{max}}\right) H + K_{ox2} \qquad (3.24)$$

Solving for when RR_a=RR_f leads to equations for D_{ss} and K_{ss}:

$$D_{ss} = \frac{d_{max}}{1 + \frac{K_{n1}}{(K_{ox2} - K_{n2})}} \tag{3.25}$$

$$K_{ss} = \frac{K_{ox2}}{1 + \frac{(K_{ox2} - K_{n2})}{K_{n1}}} \tag{3.26}$$

Using the concept that the difference in removal rates equals the rate of change in step height:

$$\frac{dH}{dt} = -(RR_a - RR_f) \tag{3.27}$$

Substituting Equation 3.23 and Equation 3.24 into Equation 3.27, and rearranging:

$$\frac{dH}{dt} + \left(\frac{K_{n1} + K_{ox2} - K_{n2}}{d_{max}}\right) H - (K_{ox2} - K_{n2}) = 0 \tag{3.28}$$

The general solution to Equation 3.28 is:

$$H = Ae^{-\frac{t}{\tau_{nit}}} + (K_{ox2} - K_{n2})\tau_{nit} \tag{3.29}$$

$$\tau_{nit} = \frac{d_{max}}{K_{n1} + K_{ox2} - K_{n2}} \tag{3.30}$$

Setting initial conditions that at time t_n, the step height is h_n, as was done for the conventional STI CMP model (see Section 2.5.2), and noting the second term of Equation 3.29 is D_{ss} gives the final equation for step height given in Equation 3.31. Having solved for H, it is now possible to derive expressions for removal rates and amount removed. Dishing and erosion expressions can then be obtained, forming the final model given by Equations 3.31 through 3.37.

Final STI CMP Model Equations for Phase 2

$$H = (h_n - D_{ss})e^{-\frac{(t-t_n)}{\tau_{nit}}} + D_{ss} \tag{3.31}$$

$$RR_a = \left(\frac{K_{n1} - K_{n2}}{d_{max}}\right)(h_n - D_{ss})e^{-\frac{(t-t_n)}{\tau_{nit}}} + K_{ss} \tag{3.32}$$

$$RR_f = -\left(\frac{K_{ox2}}{d_{max}}\right)(h_n - D_{ss})e^{-\frac{(t-t_n)}{\tau_{nit}}} + K_{ss} \tag{3.33}$$

$$AR_a = z_0 + K_{ss}(t-t_n) + \left(1 - \frac{K_{ss}}{K_{n1}}\right)(h_n - D_{ss})\left(1 - e^{-\frac{(t-t_n)}{\tau_{nit}}}\right) \tag{3.34}$$

$$AR_f = z_0 - z_1 + h_n + K_{ss}(t-t_n) - \left(\frac{K_{ss}}{K_{n1}}\right)(h_n - D_{ss})\left(1 - e^{-\frac{(t-t_n)}{\tau_{nit}}}\right) \tag{3.35}$$

$$D = (h_n - D_{ss})e^{-\frac{(t-t_n)}{\tau_{nit}}} + D_{ss} \tag{3.36}$$

$$E = K_{ss}(t-t_n) + \left(1 - \frac{K_{ss}}{K_{n1}}\right)(h_n - D_{ss})\left(1 - e^{-\frac{(t-t_n)}{\tau_{nit}}}\right) \tag{3.37}$$

where t_{nit} is given by Equation 3.30. The key model parameters for Phase 2 are D_{ss}, τ_{nit}, K_{ss}, and K_{n1}.

3.3 Model Parameters

Discussion of model parameters in this section refers to the parameters used to model single locations or densities. The full model parameter set also includes the planarization lengths for the oxide (PL_{OX}) and nitride overpolish (PL_{NIT}) phases of the STI CMP process, which are measures of the planarization capabilities of the CMP process on oxide and combined nitride/ oxide films. In the generalized density model, there are three primary model parameters for the oxide overburden phase (K_1, K_2, h_c) and four primary model parameters for the nitride overpolish phase (D_{ss}, K_{ss}, τ_{nit}, K_{n1}). In addition, there are other parameters in the model that are derived from the primary model parameters that have physical meanings, such as K_{ox2}, d_{max}, and K_{n2}.

3.3.1 Phase 1 Model Parameters

The patterned removal rate K_1 is the removal rate of a raised feature of this specific pattern density, before down area contact occurs. The blanket removal rate K_2 describes the average blanket removal rate for a particular die on the wafer. It should not be density dependent, but will vary by location due to within-die and within-wafer removal rate variability. The contact height h_c is the height at which down area contact (and appreciable down area removal) begins. It may also have dependencies on trench width. The oxide time constant τ_{ox} can be derived from K_1 and h_c via Equation 3.8, and describes how fast the step height will reduce once the down area is contacted by the pad.

In the conventional STI CMP model, K_1 has a linearly inverse dependence on density, with the proportionality constant equal to K_2, the blanket removal rate. The contact height h_c typically has been shown to decrease with increasing feature density, and in the conventional model it also has a linearly inverse density dependence. The oxide time constant in the conventional model is assumed to be independent of density; the product of the oxide time constant and the blanket removal rate is the proportionality constant for the contact height relationship.

3.3.2 Phase 2 Model Parameters

The parameters D_{ss} and K_{ss} describe the step height and removal rate, respectively, of the nitride overpolish phase once steady state has been achieved. The nitride time constant τ_{nit} describes how fast the nitride overpolish phase approaches steady-state. The parameter K_{n1} describes the nitride patterned removal rate. This is the removal rate of raised area nitride in the Phase 2 combined nitride/oxide system. The parameters K_{ox2} and K_{n2} (derivable from the four main model parameters) describe the oxide and nitride blanket removal rates. These should not be density dependent.

In the conventional STI CMP model, alternative model parameters are used that are independent of pattern density. Other than planarization length, the parameters were selectivity s and "pure nitride" time constant τ_2. The parameters used in the revised model have duals in the conventional CMP model; these have fixed density dependencies. D_{ss} in the conventional model has an approximately linear inverse dependence on density at high selectivities, and a weaker inverse density dependence at low selectivities (see Equation 2.32 and Equation 2.46). The parameters K_{ss} and τ_{nit} also followed this relationship. K_{n1} has an exact linear inverse dependence on density.

3.4 Modeling Methodology

The modeling methodology for the generalized STI CMP model is shown in Figure 3.3.

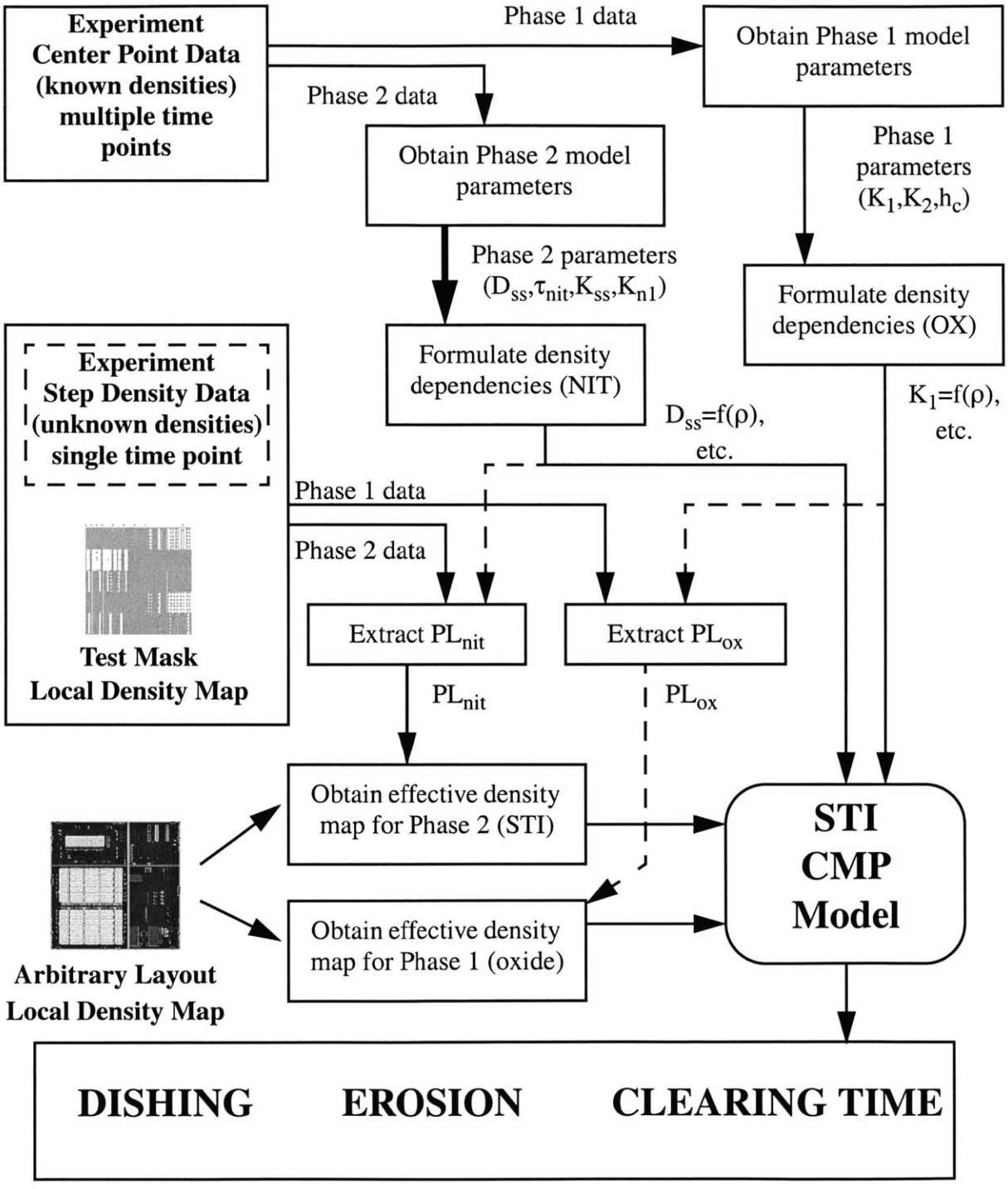


Figure 3.3: Characterization and modeling methodology for nonconventional CMP processes.

The methodology requires running wafers patterned with the characterization test mask on the CMP process of interest and gathering experimental data. Two regimes are focused on: center point data from structures with known densities, and step density data across the step density structures. Wafers are polished for multiple time steps, with pre- and post-CMP center point and step density data obtained for each time step.

The methodology begins by utilizing center point data at fixed structures with known densities. For each center point, a data vector of amount removed vs. time is obtained. This data vector is split into two vectors: before and after nitride clearing (this is obtained by checking whether the amount removed is less than or greater than the initial film thickness). Pre-clearing data is used to extract out Phase 1 parameters (K_1, K_2, h_c), and post-clearing data is used to extract out Phase 2 parameters (D_{ss}, τ_{nit}, K_{ss}, K_{n1}) for the density of that particular structure. The extraction procedure again uses an optimization loop to find optimal model parameters that minimize the RMS error of the model prediction to data.

If dishing and erosion data is available, then the extraction for Phase 2 can be simplified by first using dishing data to extract out D_{ss} and τ_{nit} (see Equation 3.36), and then using erosion data along with the extracted (D_{ss}, τ_{nit}) values to extract out K_{ss} and K_{n1}.

The extraction process is repeated for the center point data from various density structures across the test mask. The general idea is to build a set of model parameter-density pairs from which to extract out a model parameter density dependence. Although the effective pattern density at any point on the die may not be equal to the local pattern density, because center point data is used, the local density is a reasonable approximation to the effective pattern density.

A relationship between each model parameter and density is then conjectured, based on the gathered set of model parameter-density pairs. These new assumed relationships are then used to extract out the planarization length of Phase 1 and Phase 2 of the STI CMP process, again using an RMS error minimization routine.

3.5 Case Study I - Fixed Abrasive Pad

The first case study to demonstrate the generalized STI pattern dependent CMP model involves an experiment using a new type of pad design. Conventional CMP pads contain randomly shaped and sized asperities distributed across the surface of a polyurethane pad. These pads are used in conjunction with a chemical slurry that contains abrasive particles. A new type of pad has been reported by 3M where the abrasives are contained not in the slurry, but in the pad [71].

The fixed abrasive pad contains microreplicated structures in a fixed array. The structures contain abrasive particles (similar to those found in CMP slurries) and resin binder on a polyester backing [71]. The structure shape and dimension vary; one configuration is an array of cylinders 200 μm wide and 40 μm high, as illustrated in Figure 3.4.

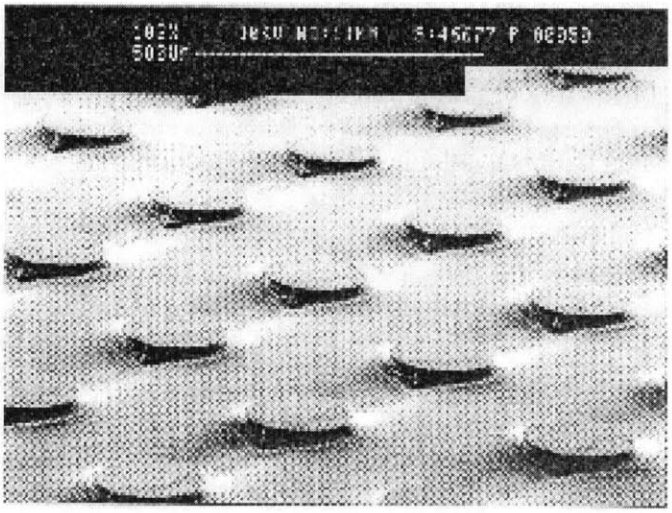


Figure 3.4: Illustration of the top surface of a fixed abrasive pad (from [72]).

The fixed abrasive pad is designed such that all particles required for the CMP process are embedded in the pad. Rather than a chemical slurry with abrasive particles, only deionized water (possibly with KOH added) is used during the polish. Polish rate on blanket wafers in a nominal process (using only deionized water) is extremely low (100 Å/min. [74]), but can be adjusted by modifying the pH (e.g., via addition of KOH) of the deionized water used during the polish [74].

In the case of two-material CMP, chemistry to modify film polish selectivity may also be included. Fixed abrasive pads are used in CMP processes because they exhibit strong selectivity to topography. It has been conjectured that the wafer topography activates the removal of abrasive from the fixed matrix by conditioning the pad surface and constantly exposing fresh pad minerals [73]. Work has been reported on implementing a viable fixed abrasive CMP process for copper [72] and STI [73,74].

3.5.1 Experiment Details

For this study, a fixed abrasive pad using ceria particles in the fixed matrix is used. It is not clear whether the nature of the nonconventional patterned removal-rate dependence on density is due to the ceria particles or to the embedded fixed pad structure. The goal in this study is to capture the combined effect of the pad and process.

A set of STI wafers with the test pattern (described in Section 3.5.2) is run to test the model. These wafers have an STI trench depth of 3100 Å, a pad oxide of 60 Å, and a nitride thickness of 1100 Å. The dielectric is an HDP silicon dioxide, deposited to 4300 Å in the trenches. The actual dielectric step height that results is 3000 Å, as measured by profilometry scans. Wafers are processed on an Obsidian Flatland 501 CMP tool. A 3M fixed abrasive pad (SWR 159 std) is used, consisting of a top pad of ceria particles in a cylinder matrix, with a subpad stack of 60 mils of polycarbonate and 90 mils of foam. Eight wafers are processed at a pH of 11.5 for times of 20, 40, 60, 90, 120, 150, 180, and 210 seconds.

3.5.2 Test Pattern

The test pattern used is similar to the one described in Section 2.10.2. The test mask is a 20 mm x 20 mm design, and contains 4 mm x 4 mm structures of fixed pitch and varying densities, along with 4 mm x 4 mm structures of fixed 50% local density and varying pitch. In addition, there are 2 mm x 2 mm structures of fine pitch lines for use in deposition characterization. An illustration of the test mask and measurement plan is shown in Figure 3.5.

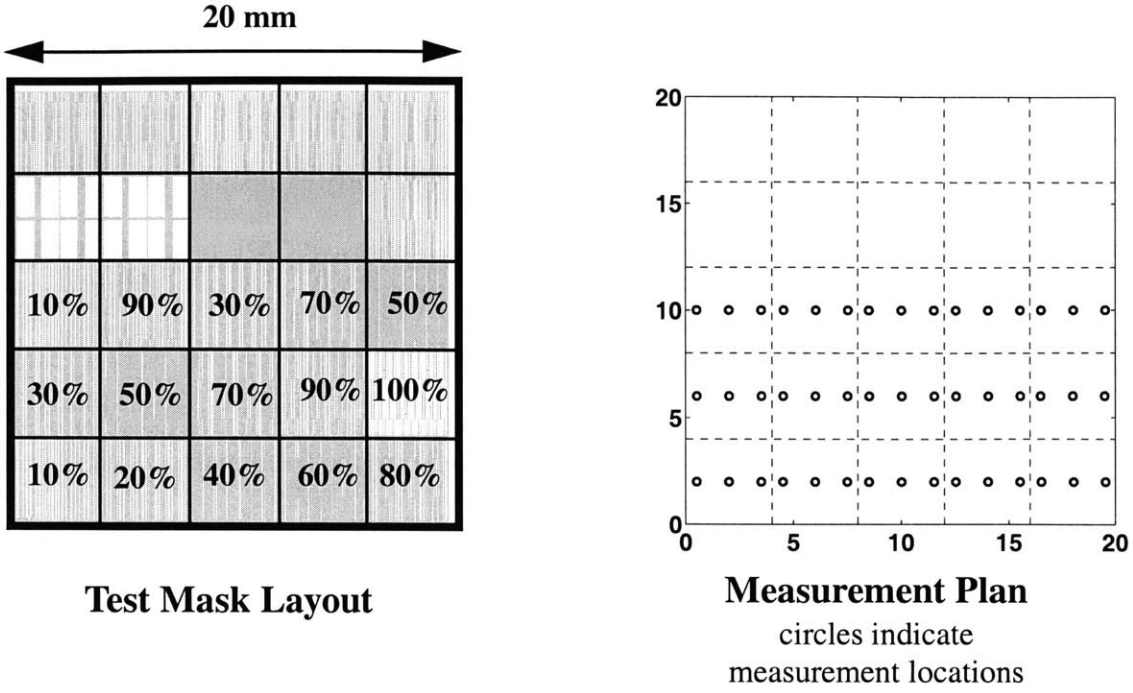


Figure 3.5: Test mask description and measurement locations used in fixed abrasive study.

3.5.3 Experimental Data

Wafers are measured after CMP on a KLA-Tencor UV1250-SE optical measurement tool. For the purposes of characterization, center point data measurements on the 10%-100% regions are used. Under the experimental conditions, the blanket removal rate after planarization is found to be low (less than 100 Å/min.). Although initial polish times were targeted for nitride overpolish, the low blanket oxide removal rate results in incomplete oxide clearing on the wafer, so that no Phase 2 data could be obtained. Thus, only characterization of Phase 1 is possible for this experiment.

3.5.4 Parameter Extraction

Experimental data is obtained from center points of the gradual density structures. Amount removed versus time data is used in the model parameter extraction procedure. The model parameters for Phase 1 (K_1, K_2, h_c) are extracted, and are shown in Table 3.1. Analysis of

the K_2 and h_c parameters reveals that for density structures at 60% and higher, these two model parameters have low sensitivity to fitting error below a certain step height. An example of the fitting error insensitivity to contact height variation is shown in Figure 3.6. The insensitivity is likely due to the nature of the experimental data used in the fitting process. We conjecture that structures of 60% or greater density did not polish sufficiently long enough for the contact height to be reached, which implies that the correct K_2 and h_c values may be lower than the extracted values in the table.

Table 3.1: Extracted Model Parameters for Fixed Abrasive Process

Drawn Density (%)	Extracted Model Parameters		
	K_1 (Å/min)	K_2 (Å/min)	h_c (Å)
10	12135	30	2331
20	8090	30	2993
30	4153	30	3158
40	1666	30	3262
50	1307	30	3150
60	348	150	2320
70	308	30	2320
80	134	70	2930
90	80	49	3120

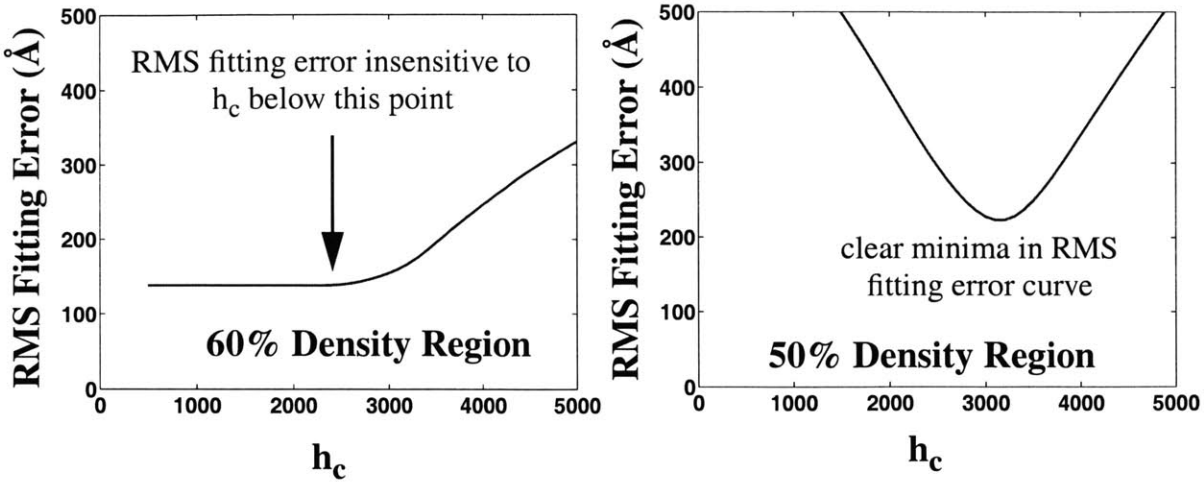


Figure 3.6: Fitting error sensitivity analysis for the h_c modeling parameter.

Predicted amount removed in up area versus measured data for the 10%, 40%, and 80% density structures is shown in Figure 3.7.The model prediction is computed from the extracted model parameters given in Table 3.1. The predictions show a reasonable fit to the measured data.

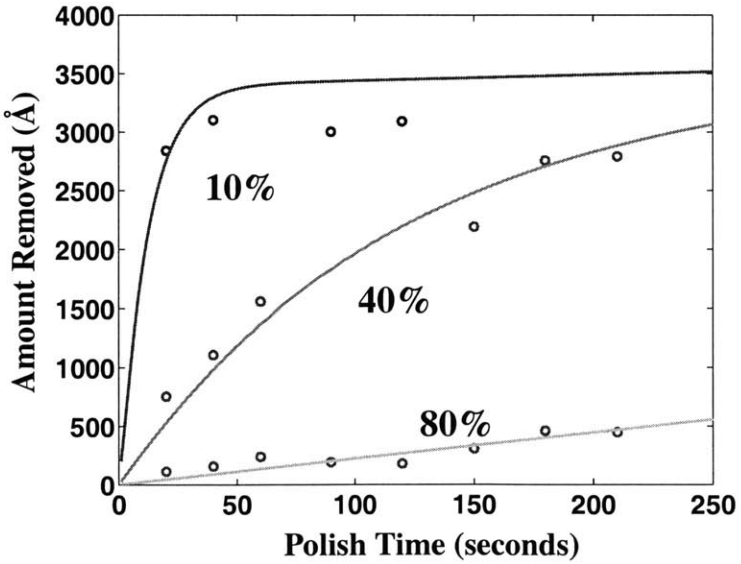


Figure 3.7: Predicted amount removed in up area vs. measured data.

3.5.5 Modeling Density Dependencies

Once model parameters have been extracted for specific known density structures, the functional dependence of the model parameters can be speculated on, and possible functional

forms can be constructed. To do this, knowledge of the physical meaning of the model parameters should be used in conjunction with the extracted parameter values.

The patterned removal rate K_1 conventionally exhibits a linearly inverse relationship with density. However, the extracted model parameters do not behave strictly in the classical inverse density relationship. Trying to fit an inverse density relationship results in considerable fitting errors, as shown in Figure 3.8. Therefore, a new function form should be used. We conjecture that the following form can be used to determine patterned removal rate:

$$K_1 = Ae^{\frac{-\rho}{\tau}} \tag{3.38}$$

where A and τ are fitting constants. A comparison of this function versus the classical density dependency is illustrated in Figure 3.8.

$$K_1 = \frac{K}{\rho}$$

K= 1193 Å/min.

$$K_1 = Ae^{\frac{-\rho}{\tau}}$$

A = 21900 Å/min. τ = 18%

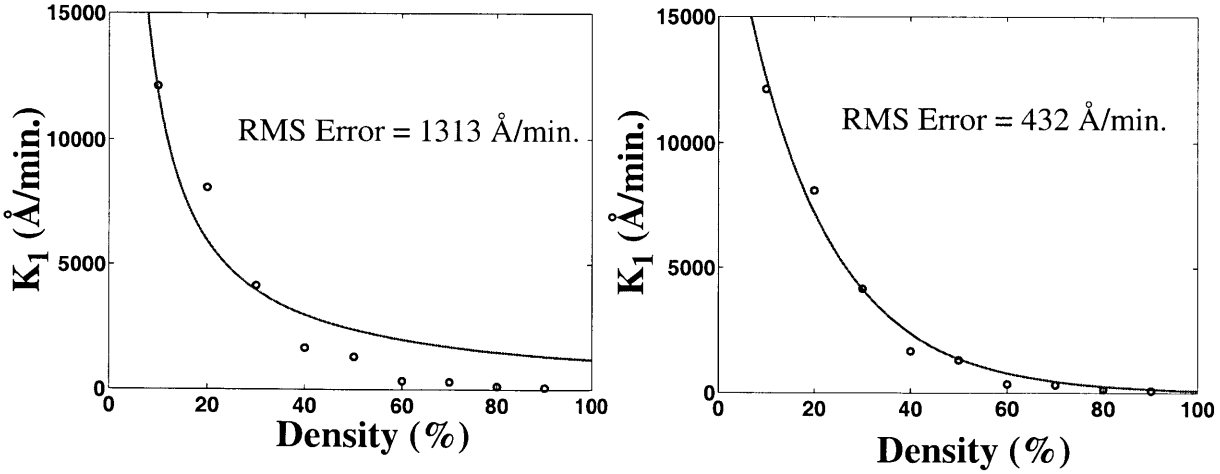


Figure 3.8: Comparison of conventional and revised K_1 density dependencies for fixed abrasive pad.

The blanket removal rate parameter K_2 should be independent of density, and from Table 3.1, an approximation of a constant K_2 seems reasonable. The value of K_2 used here is 30 Å/min.

The contact height parameter h_c conventionally decreases with increasing density. This does not agree with the extracted contact heights in Table 3.1. Therefore, a new functional dependency should be used. We conjecture that for the fixed abrasive pad, the contact height is independent of density, and has a fixed value of 2843 Å (the average of the h_c values in Table 3.1). A possible physical reasoning behind this functional form is that the microreplicated fixed abrasive pad surface enforces a particular fixed contact height on the CMP process.

3.5.6 Results

Once the model parameter dependencies on density are derived, the planarization length of the process can be extracted using RMS error minimization for model prediction to measured data. The extracted planarization length for this process is 1 mm. Normally, long planarization lengths are desired because the density averaging tends to improve the planarity. However, for the fixed abrasive process, the planarization capability is maintained by the low blanket removal rates. This prevents appreciable material removal from regions that reach local planarity the fastest (i.e., low density regions), while the local step heights in the regions of high density are removed. Therefore, a short planarization length for the fixed abrasive pad does not imply poor planarization capability, as it might for conventional CMP processes.

Using the planarization length of the CMP process and the derived model parameter functions of density, the post-CMP surface can be predicted and compared to the measurement data. The results of the experiment, as compared to the model prediction obtained using the calibration methodology described in Section 3.4, are shown in Figure 3.9. Two polish times are simulated (60 seconds and 210 seconds). The model predicts the data trends with a reasonable RMS error.

The model here appears to overpredict the amount removed for medium and high density

areas (see 210 second polish time data vs. model prediction). This is due to the fact that the functional form for the patterned removal rate K_1 overpredicts the patterned removal rate for medium to high densities (see Figure 3.8). Development of a more accurate functional form for patterned removal rate vs. density should reduce this overprediction.

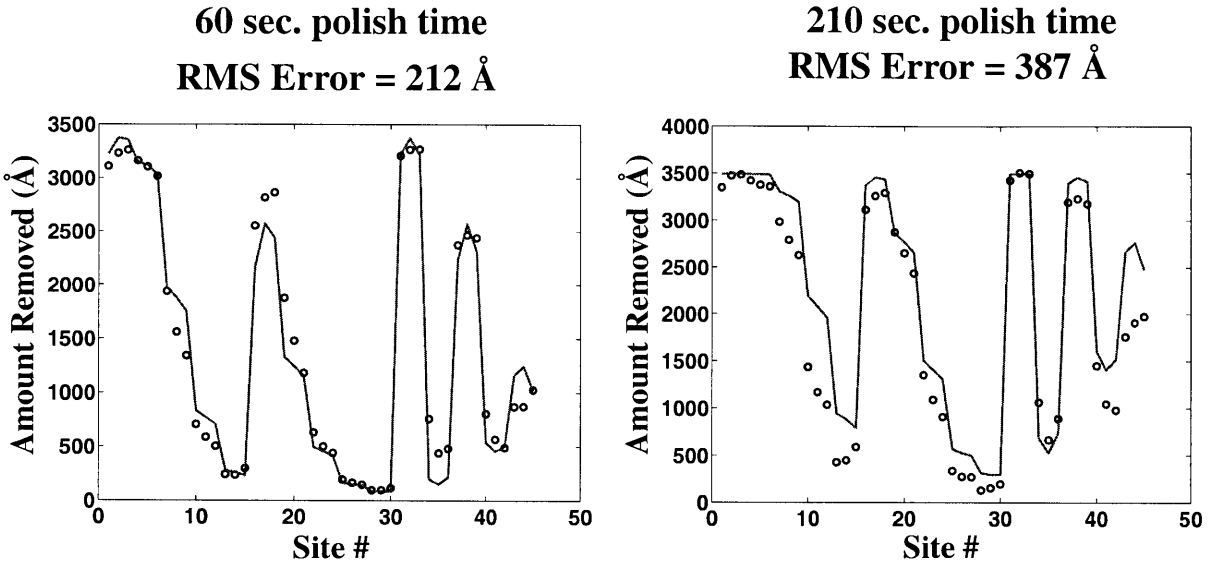


Figure 3.9: Experimental data vs. model prediction for up area data.

3.6 Case Study II - High Selectivity Silica Slurry

The second case study involves a high material-selectivity slurry. This slurry exhibits blanket oxide removal rates similar to standard oxide CMP slurries, with much lower nitride removal rates, for an overall selectivity gain from 3:1 (conventional) to 200:1 (high-selectivity) [75]. The slurry is silica-based, as in conventional oxide CMP slurries.

3.6.1 Experiment Details

A high selectivity silica slurry is used in this experiment in conjunction with a standard IC1400 stacked CMP pad. Polishing is performed on an IPEC 472 CMP tool. In this experiment, blanket wafers are polished at varying pressures to determine removal rates. This experiment is

performed to determine the nature of the removal-rate pressure relationship for these slurries. In addition, patterned wafers are also run to verify the characterization and modeling methodology on nonconventional consumable sets. These wafers are polished at five combinations of pressures and speeds, as shown in Table 3.2.

Table 3.2: Process Conditions for Patterned Wafers for HSS Silica Slurry

Process	Pressure (psi)	Speed (rpm)	Number of Wafers
1	3	30	8
2	3	90	8
3	9	30	8
4	9	90	8
5	6	60	8

3.6.2 Test Pattern

The test mask used to manufacture the patterned wafers in this experiment is the same one used for STI characterization, described in Section 2.10.2. Measurement locations are shown in Figure 3.10. The central cross section of points (through Y=10 mm) passes through five step density structures of 20%, 90%, 30%, 70%, and 50% density. The center points of these structures are used for density characterization for the model parameters, and the full data set is used for planarization length extraction.

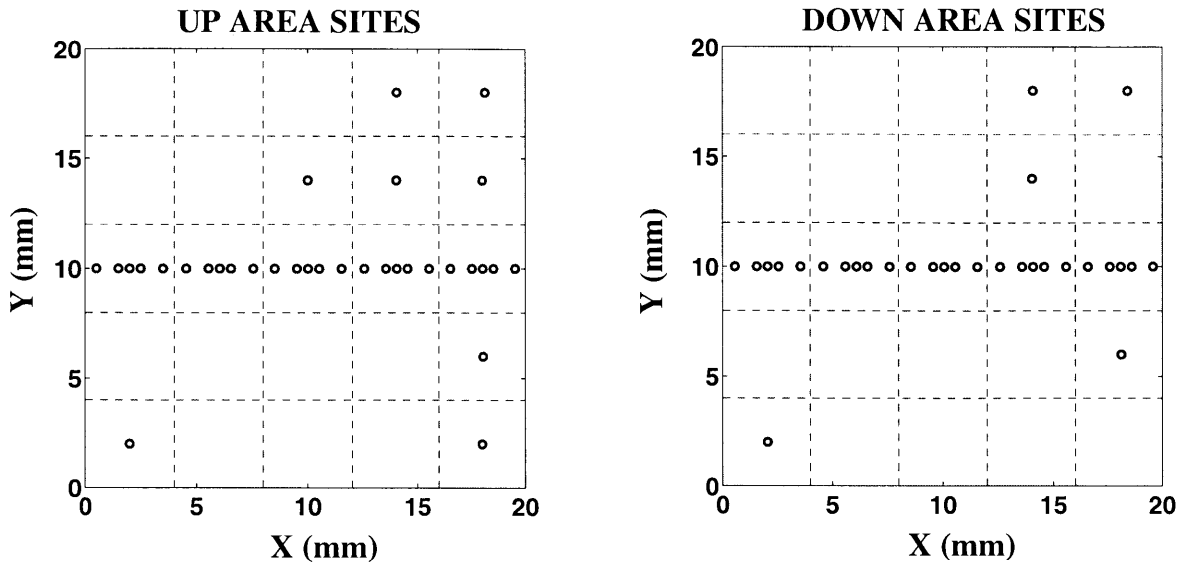


Figure 3.10: Measurement locations for patterned wafers for silica HSS experiment.

3.6.3 Experimental Data

A KLA-Tencor UV1250 SE system is used for film thickness measurement. For blanket wafers, a 49 point polar map measurement is used to determine an average amount removed for the given polish time. This is then used to determine average polish rate. For patterned wafers, the desired measurement locations are measured using the pattern recognition system.

3.6.4 Blanket Wafer Study

The results of the blanket wafer study with three different speed are shown in Figure 3.11. The blanket wafer results seem to indicate a fairly linear (Prestonian) relationship with pressure for the three different table speeds used. This implies that the non-Prestonian behavior of the slurry may not be due to a nonlinear response of removal rate to pressure.

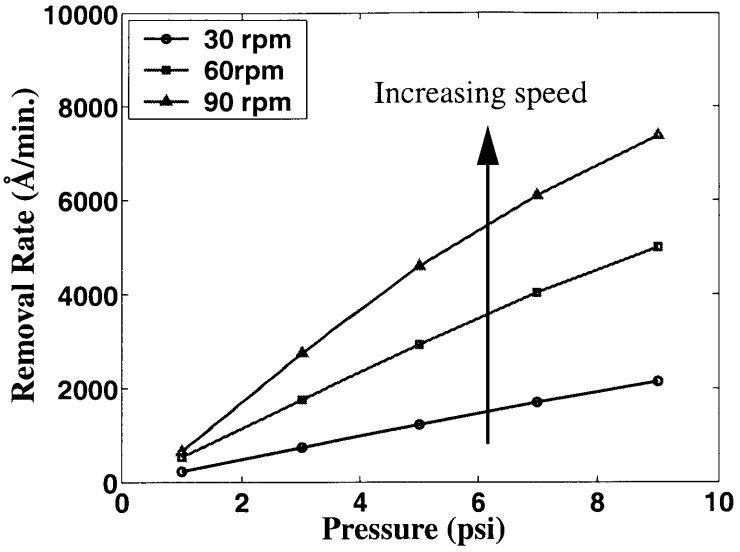


Figure 3.11: Blanket Wafer Removal Rate Data for Silica HSS.

3.6.5 Patterned Wafer Study - Parameter Extraction

For each CMP process in this experiment, the model parameter extraction routine is run on the center point data of the five density structures in the step density structure of the test mask. Model parameters for Phase 1 and Phase 2 are extracted and shown in Table 3.3. Certain density structures (e.g., the 20% density structure in Process 4) are marked "n/a" because they do not have sufficient valid amount removed, dishing, or erosion data to use for characterization.

Table 3.3: Parameter Extraction for Silica HSS, Process 1

Drawn Density (%)	Phase 1			Phase 2			
	K_1 (Å/min)	K_2 (Å/min)	h_c (Å)	D_{ss} (Å)	τ_{nit} (secs)	K_{ss} (Å/min)	K_{n1} (Å/min)
20	5176	912	11167	6813	315	50	52
30	2562	1052	6672	5251	224	50	53
50	1573	1273	4800	4354	214	50	54
70	1303	1210	4264	2912	91	50	54
90	1200	1200	2630	2057	44	50	50

Table 3.4: Parameter Extraction for Silica HSS, Process 2

Drawn Density (%)	Phase 1			Phase 2			
	K_1 (Å/min)	K_2 (Å/min)	h_c (Å)	D_{ss} (Å)	τ_{nit} (secs)	K_{ss} (Å/min)	K_{nl} (Å/min)
20	7021	2189	5126	5712	104	288	311
30	5007	2917	4712	6079	147	174	185
50	3812	3276	4392	4823	131	149	161
70	3243	3243	3789	3769	93	131	141
90	3045	3045	2803	2313	42	79	81

Table 3.5: Parameter Extraction for Silica HSS, Process 3

Drawn Density (%)	Phase 1			Phase 2			
	K_1 (Å/min)	K_2 (Å/min)	h_c (Å)	D_{ss} (Å)	τ_{nit} (secs)	K_{ss} (Å/min)	K_{nl} (Å/min)
20	7205	3197	7740	4058	56	335	375
30	5763	3309	6031	5228	106	456	604
50	4018	4018	3582	4143	98	64	65
70	3379	3379	3265	2971	13	50	50
90	2832	2832	2230	1757	8	50	51

Table 3.6: Parameter Extraction for Silica HSS, Process 4

Drawn Density (%)	Phase 1			Phase 2			
	K_1 (Å/min)	K_2 (Å/min)	h_c (Å)	D_{ss} (Å)	τ_{nit} (secs)	K_{ss} (Å/min)	K_{nl} (Å/min)
20	26229	6708	8629	n/a	n/a	n/a	n/a
30	18365	6780	6674	4552	37	659	659
50	11230	8317	3695	4263	45	364	364
70	9372	5263	3245	2843	15	375	375
90	7433	301	2052	n/a	n/a	n/a	n/a

Table 3.7: Parameter Extraction for Silica HSS, Process 5

Drawn Density (%)	Phase 1			Phase 2			
	K_1 (Å/min)	K_2 (Å/min)	h_c (Å)	D_{ss} (Å)	τ_{nit} (secs)	K_{ss} (Å/min)	K_{n1} (Å/min)
20	9666	4093	7928	7167	113	350	371
30	7464	4341	5886	5515	90	175	181
50	5287	4980	3953	4409	83	117	120
70	4530	4530	3391	2721	28	55	55
90	3951	3951	2368	1841	8	50	50

Once the model parameters are extracted for each process, then functional dependencies on density can be formulated. As an example we will study Process 5. There are seven total parameters to consider, three in Phase 1 and four in Phase 2. We shall analyze Phase 1 parameters first.

3.6.6 Patterned Wafer Study - Phase 1 Parameter Analysis

The extracted blanket removal rate K_2 is expected to be independent of density, and examination of the extracted model parameters does not disagree with this assumption. The parameter K_2 will be assumed constant, and equal to 4379 Å/min., which is the average of the extracted model parameters across the five density structures.

The patterned removal rate K_1 conventionally has an inverse density relationship. This does not agree with the extracted model parameters for Process 5. A revised relationship on density can be conjectured based on the extracted patterned removal rates. We propose the form:

$$K_1 = K_\rho \frac{(1-\rho)}{\rho} + K_c \qquad (3.39)$$

where K_ρ and K_c are fitting constants. This functional form exhibits a significantly better fit when compared to the extracted model parameters, as shown in Figure 3.12.

$$K_1 = \frac{K}{\rho}$$

K= 2181 Å/min.

$$K_1 = K_\rho \frac{(1-\rho)}{\rho} + K_c$$

K_ρ=1474 Å/min. K_c=3859 Å/min.

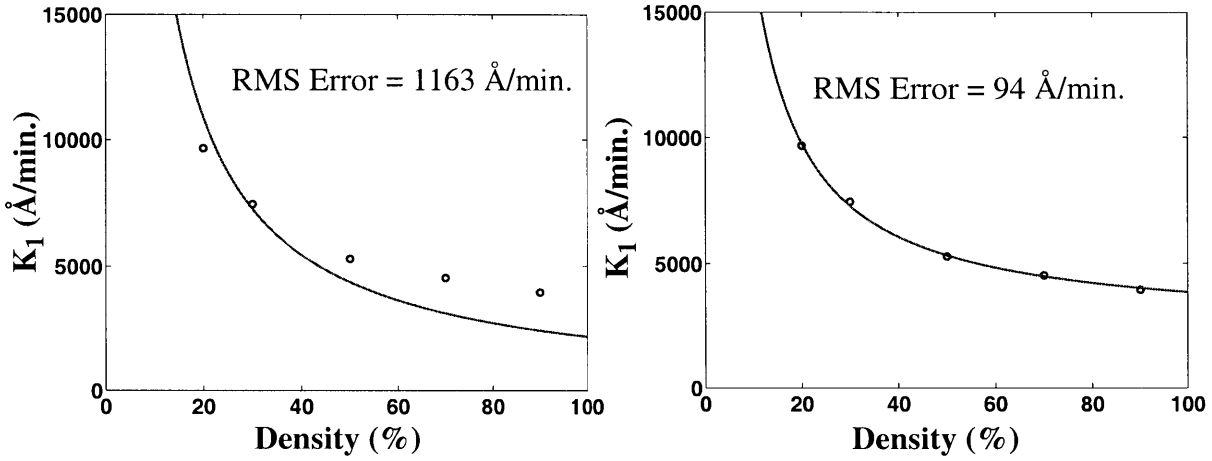


Figure 3.12: Comparison of conventional and revised K_1 density dependencies for HSS slurry.

The process of the formulation of model parameter density dependencies may be repeated for h_c. Analysis of the extracted contact height parameters leads to a proposed form that is the same as the one used for patterned removal rate:

$$h_c = A\frac{(1-\rho)}{\rho} + B \qquad (3.40)$$

with fitting constants of A and B. An interesting note is that in the conventional model, patterned removal rate and contact height were assumed to have the same functional dependence on density, and this is also the case for this particular CMP process. The proposed form produces an RMS fitting error (for optimized fitting constants A=1374 and B=2542) of 203 Å/min. compared to the extracted model parameters, an improvement over an inverse-density relationship which produces an RMS error of 603 Å/min.

3.6.7 Patterned Wafer Study - Phase 2 Parameter Analysis

The Phase 2 model parameters can be analyzed in the same fashion as the Phase 1 model

parameters. The following function forms are proposed for Phase 2 model parameters:

$$D_{ss} = A + B(\rho - C)^2 \tag{3.41}$$

$$\tau_{nit} = A + B(\rho - C)^2 \tag{3.42}$$

$$K_{ss} = A\frac{(1-\rho)}{\rho} + B \tag{3.43}$$

$$K_{n1} = A\frac{(1-\rho)}{\rho} + B \tag{3.44}$$

where A, B, and C are fitting constants for each function form. Optimized values for A, B, and C are given in Table 3.8.

Table 3.8: Optimized Fitting Constants for Silica HSS Phase 2 Model Parameters

	A	B	C
D_{ss}	1976	3754	1
τ_{nit}	9.7	163	1
K_{ss}	81	50	n/a
K_{n1}	85	50	n/a

3.6.8 Patterned Wafer Study - Planarization Length Extraction

Once the functional forms are developed and the function fitting constants are obtained, the planarization lengths for Phase 1 and Phase 2 can be obtained by taking step density measurements from a single time point and performing the planarization length extraction. The planarization length for Phase 1 is 3.1 mm, and the planarization length for Phase 2 is 3.7 mm.

3.6.9 Patterned Wafer Study - Prediction

The model is then used to predict the CMP behavior for this process for two polish times: one for Phase 1 regime (90 seconds) and one for Phase 2 regime (180 seconds). Model

comparisons to data are shown in Figure 3.13 and Figure 3.14. Model predictions show a reasonable fit to the measured data for both Phase 1 and Phase 2.

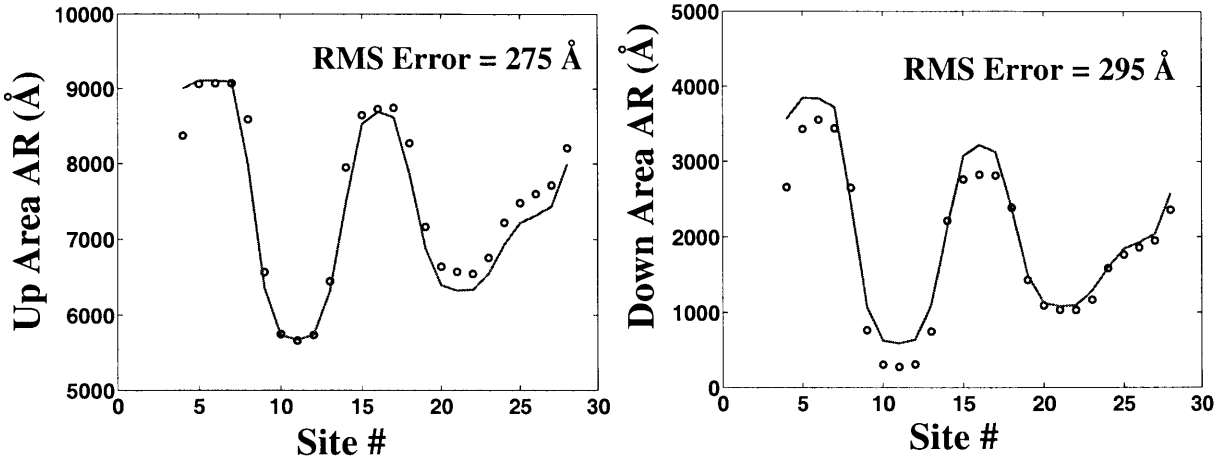


Figure 3.13: Model prediction vs. experimental data for Phase 1 (90 seconds).

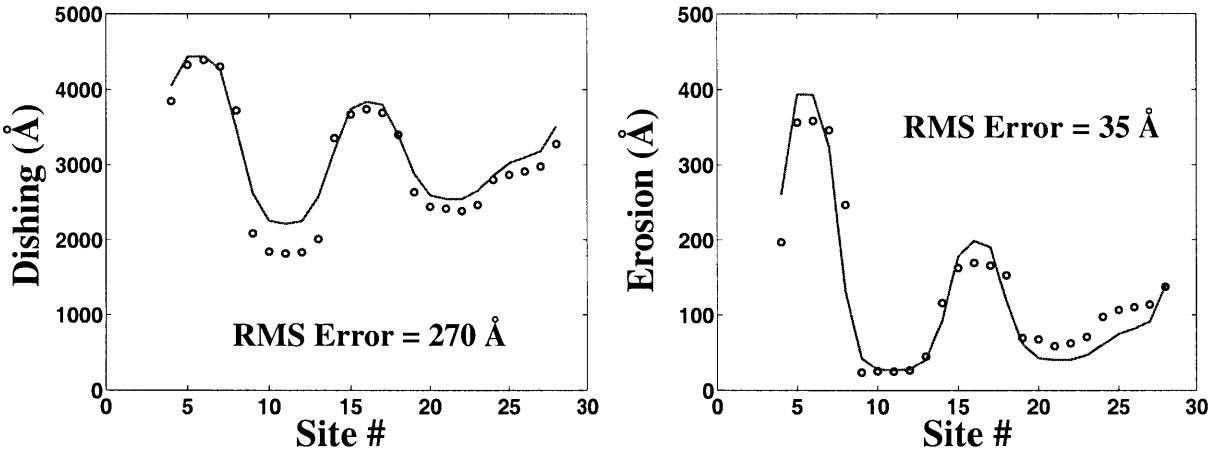


Figure 3.14: Model prediction vs. experimental data for Phase 2 (180 seconds).

The low blanket nitride removal rate of the high-selectivity silica-based slurry results in improved erosion (but worse dishing) relative to a conventional STI CMP process. This indicates that this process may allow for a greater overpolish window given a particular nitride loss budget.

3.7 Case Study III - High Selectivity Ceria Slurry

Conventional dielectric CMP slurries contain silica abrasive particles. Recent work has shown improvement in planarization capability as a result of using cerium-oxide particles in the slurry [76-80].

Nojo [76] analyzed the effects of surfactant concentration (from 2-5 wt %) on a nominal cerium oxide process. His metric is the relative removal rates of a nominal raised feature and down area on a test chip. He concludes that there is an optimal surfactant concentration such that raised/down area removal rate ratio is optimized, leading to better planarization capability of the process. Nojo also performed a blanket wafer study (results of which are shown in Figure 3.15) where he observes that there exists a threshold pressure below which removal rates are very low, and proposed that ceria slurries could thus be used for self-stopping and dishing-free oxide polishing [76].

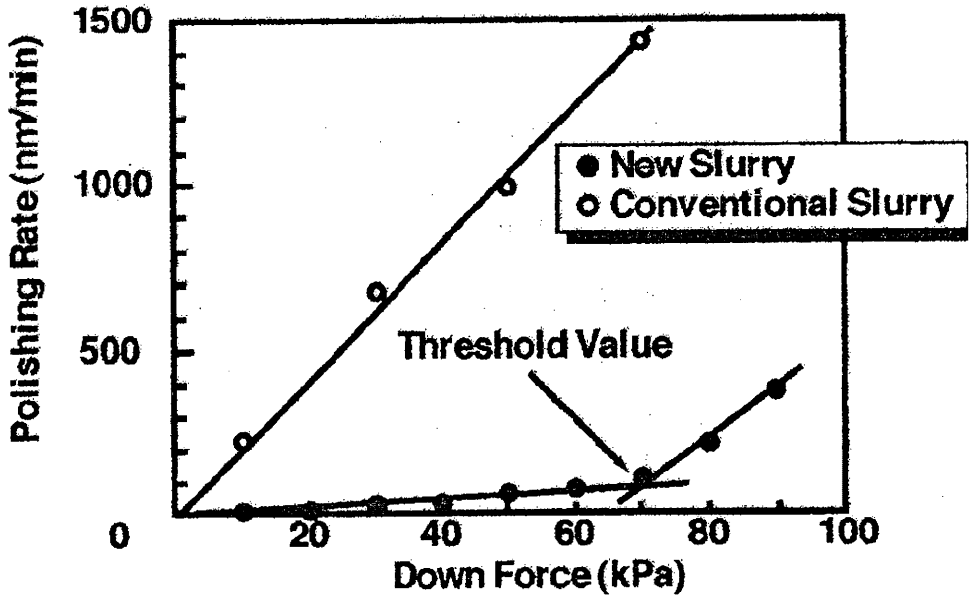


Figure 3.15: Removal rate vs. pressure for silica are ceria-based slurry (from [76]).

Evans [78] noted that ceria slurries offer high material selectivity for nitride/oxide polishing systems. Lee performed a patterned wafer study using silica and ceria-based slurries. He notes that for silica slurries, the planarization rate is dependent on feature density, while for ceria slurries, it is dependent on feature size and step height [79].

The mechanism behind the nonconventional behavior of the ceria slurry is not well understood. Nojo [76] argues that a protective layer is formed by the surfactant on the surface of the oxide that prevents material removal, and that only sufficient pressure (i.e., past the "threshold

pressure") will remove the protective layer and thus allow appreciable polish. Jouty [77], Lee [79], and Kwon [80] argue that the unusual material removal is due to abrasive particles that are trapped in the down area, preventing down area removal. Kwon argues that at sufficiently small step heights, the particles are freed and down area removal begins [80]. Zhao argues that the threshold pressure behavior illustrated in Figure 3.15 is due to the phenomena of rolling vs. sliding particles [9]. At low pressures, particles "roll" across the wafer surface, resulting in little material removal. At a certain pressure, particles begin to "slide" across the wafer, embedding themselves in the wafer surface, resulting in appreciable material removal.

Given the considerable disagreement over fundamental mechanisms, our purpose will be to use our generalized CMP model to capture pattern-dependencies in an empirical fashion.

3.7.1 Experiment Details

Ceria-based slurry is used with a standard IC400 stacked pad on an IPEC 472 rotary CMP tool. Seven blanket wafers were also polished at fixed table speed of 30 rpm at various pressures for 60 seconds to determine blanket wafer removal rate variation with pressure. Seven patterned wafers are also included in the experimental run to study the effect of ceria-based slurry on patterned wafers. These wafers are run at 7 psi and 30 rpm on the same pad and tool as the blanket wafers.

3.7.2 Test Pattern

The test mask used to manufacture the patterned wafers in this experiment is the same one used for STI characterization, described in Section 2.10.2. Measurement locations used in this experiment are the same as those used in the high material selectivity silica based experiment. The locations are specified in Section 3.6.2.

3.7.3 Experimental Data

The UV1250 SE system is used for film thickness measurement. For blanket wafers, a 49 point polar map measurement is used to determine an average amount removed for the given

polish time. This is then used to determine average polish rate. For patterned wafers, the desired measurement locations is measured using the pattern recognition system.

3.7.4 Blanket Wafer Study

The results from the blanket wafer experiment are shown in Figure 3.16. In this graph, removal rate is shown as a function of pressure. The relationship appears to be fairly Prestonian.

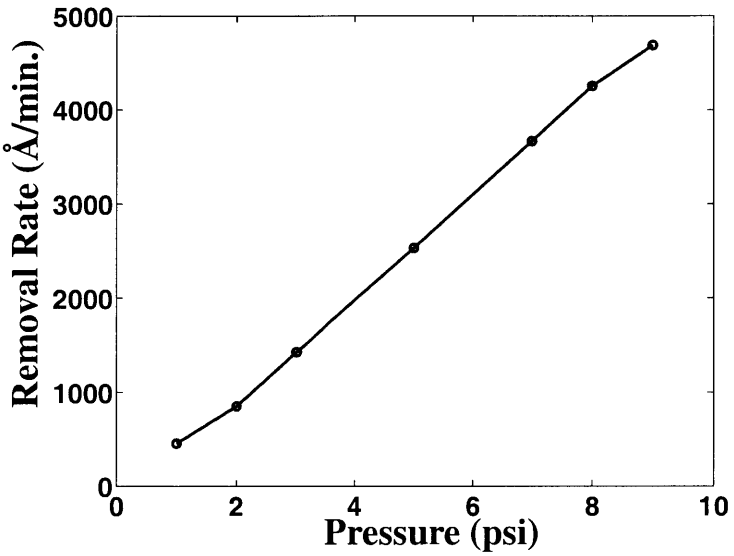


Figure 3.16: Blanket Wafer Removal Rate Data for Ceria HSS Slurry.

3.7.5 Patterned Wafer Study - Data Analysis

Before attempting to model the CMP behavior of this process, the data is first analyzed to observe unusual trends. Figure 3.17 is a plot of the up area amount removed vs. time for a variety of structures across the die. An interesting observation to note is that the data does not appear to extrapolate to zero amount removed at zero time. Rather, there appears to be a distinct time offset before appreciable removal begins. We speculate that this may be due to some fixed time at the beginning of the process that is necessary for the slurry chemistry to become active.

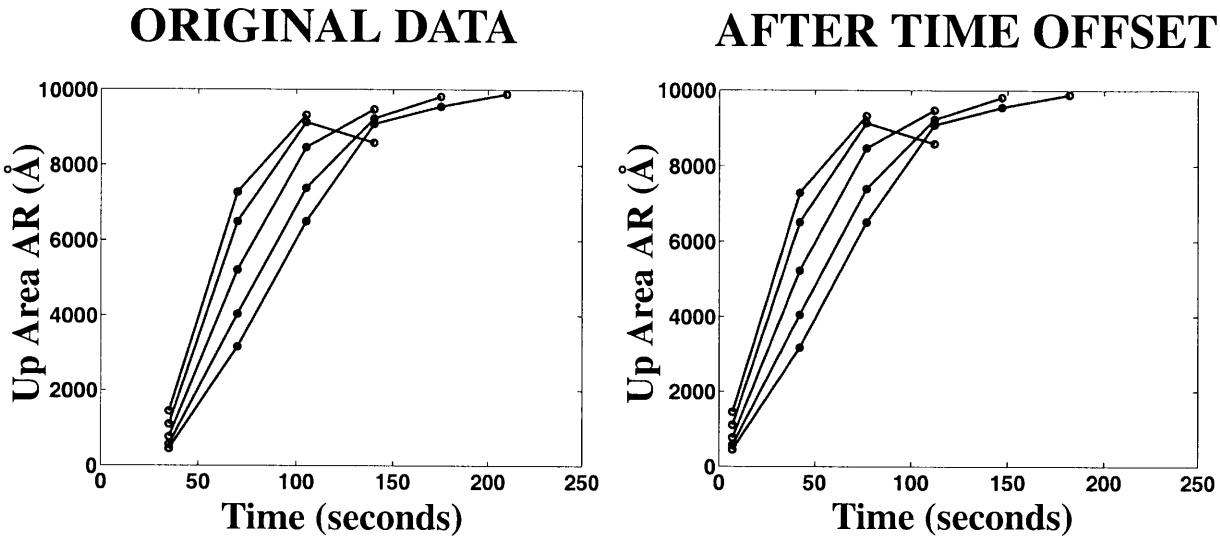


Figure 3.17: Examining the amount removed vs. time data for ceria HSS process.

This "slurry activation effect" can be modelled as a simple shift in the CMP time. The time offset can be estimated by linear extrapolation of the up area amount removed backwards to zero amount removed. For this case, the time offset t_{offset} is found to be 28 seconds. The model can be modified by replacing all instances of t with t', where:

$$t' = t - t_{offset} \tag{3.45}$$

3.7.6 Patterned Wafer Study - Parameter Extraction

After the time shift is introduced into the model, the model parameters for Phase 1 and Phase 2 can be extracted. The center points of five density structures (20%, 30%, 50%, 70%, 90%) are used in for the purposes of model parameter characterization. The lower density structures experience severe nitride erosion, resulting in insufficient valid measurement data for Phase 2 model parameter extraction. As a result, any functional dependencies may not accurately represent low density regions.

Table 3.9: Parameter Extraction for Ceria HSS

	Phase 1			Phase 2			
Density (%)	K_1	K_2	h_c	D_{ss}	τ_{nit}	K_{ss}	K_{n1}
20	13537	4244	4056	n/a	n/a	n/a	n/a
30	12131	4075	6005	n/a	n/a	n/a	n/a
50	7644	4585	4374	2021	31	1367	1902
70	5783	4193	4060	1873	33	903	903
90	4609	4609	4887	1532	32	465	465

3.7.7 Patterned Wafer Study - Phase 1 Parameter Analysis

Based on the parameters extracted from the data, we propose a model where K_2 and h_c are fixed constants, independent of density, with values of 4341 Å/min. and 4674 Å, respectively. In addition, we propose the following functional form for K_1:

$$K_1 = Ae^{\frac{-\rho}{B}} \tag{3.46}$$

with fitting constants A (9225) and B (0.59).

3.7.8 Patterned Wafer Study - Phase 2 Parameter Analysis

Based on the extracted parameters for Phase 2, τ_{nit} is constant (32 seconds), and the following forms for the other Phase 2 model parameters are proposed:

$$D_{ss} = Ae^{\frac{-\rho}{B}} \tag{3.47}$$

$$K_{ss} = Ae^{\frac{-\rho}{B}} \tag{3.48}$$

$$K_{n1} = Ae^{\frac{-\rho}{B}} \tag{3.49}$$

with fitting constants A and B, given in Table 3.10.

Table 3.10: Optimized Fitting Constants for Phase 2 Model Parameters

	A	B
D_{ss}	2862	1.51
K_{ss}	4784	0.40
K_{n1}	11573	0.28

3.7.9 Patterned Wafer Study - Planarization Length Extraction

Using the functional forms given in Equation 3.46 to Equation 3.49, the planarization lengths for Phase 1 and Phase 2 are extracted using a full dataset from a single time point (in Phase 1 or Phase 2) and performing planarization length extraction. In this case, planarization length for Phase 1 is 1.6 mm, and the planarization length for Phase 2 is 4.7 mm.

3.7.10 Patterned Wafer Study - Prediction

Prediction of CMP behavior vs. actual measured data for Phase 1 and Phase 2 is shown in Figure 3.18 and Figure 3.19. Phase 1 data is taken from a polish time of 70 seconds, while Phase 2 data is taken from a polish time of 140 seconds. The model shows a reasonable fit to measured data.

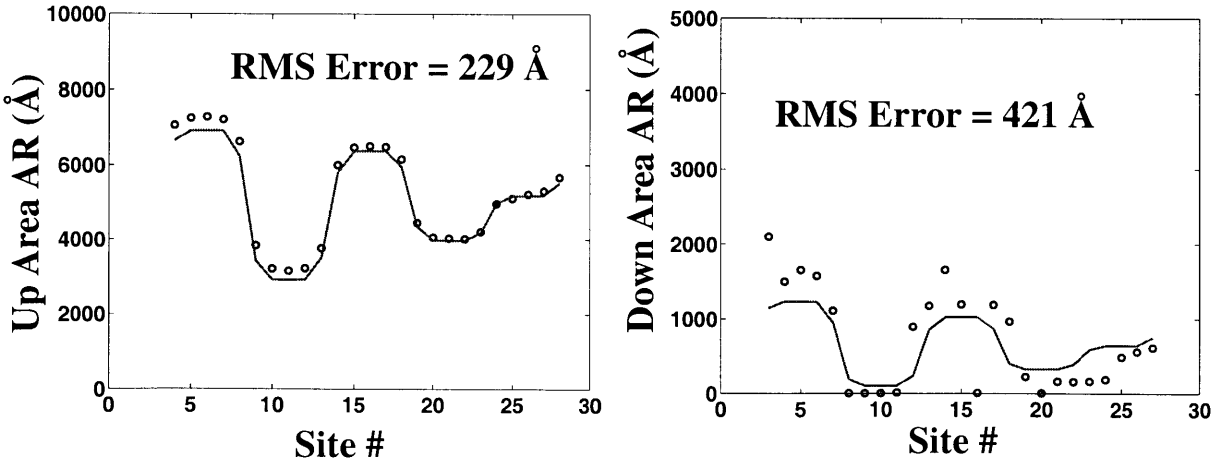


Figure 3.18: Model prediction vs. measured data for Phase 1 ceria HSS CMP process.

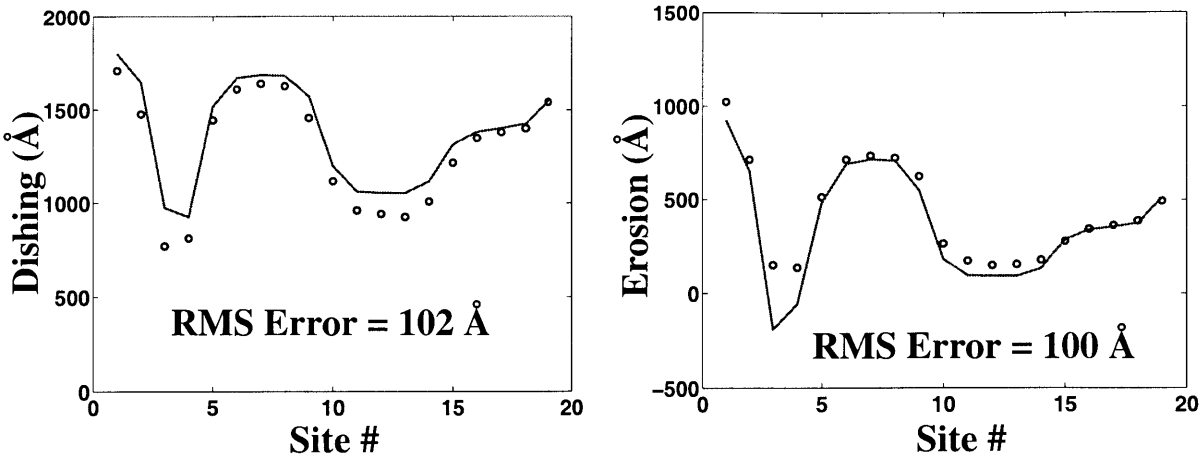


Figure 3.19: Model prediction vs. measured data for Phase 2 ceria HSS CMP process.

Based on the experimental results, it appears that the high-selectivity ceria slurry results in similar erosion and dishing results as a conventional STI CMP process. Further investigation into other pad and process condition combinations may provide improved erosion and dishing results.

3.8 Applications

The usefulness of the modeling and characterization methodology proposed in this chapter is the ability to predict CMP behavior on a characterized CMP process for arbitrary layouts. The three main quantities that can be predicted are chip-scale dishing and erosion for a particular polish time, and oxide clearing time prediction across a die. As an example, we perform a chip-scale simulation for the silica HSS CMP process.

3.8.1 Dishing and Erosion Prediction

Dishing and erosion is predicted for a second layout (the same SRAM chip case study layout used in Section 2.12). Local density maps for the oxide and nitride layers are used in addition to the revised model parameter functional forms as inputs to the predictor. The model parameters used are those characterized for silica HSS slurry process, described in Section 3.6. Simulation time is chosen to be 110% of the simulated clearing time of the region of the

overburden oxide layer with the highest density, or 93 seconds. Dishing and erosion predictions are shown in Figure 3.20.

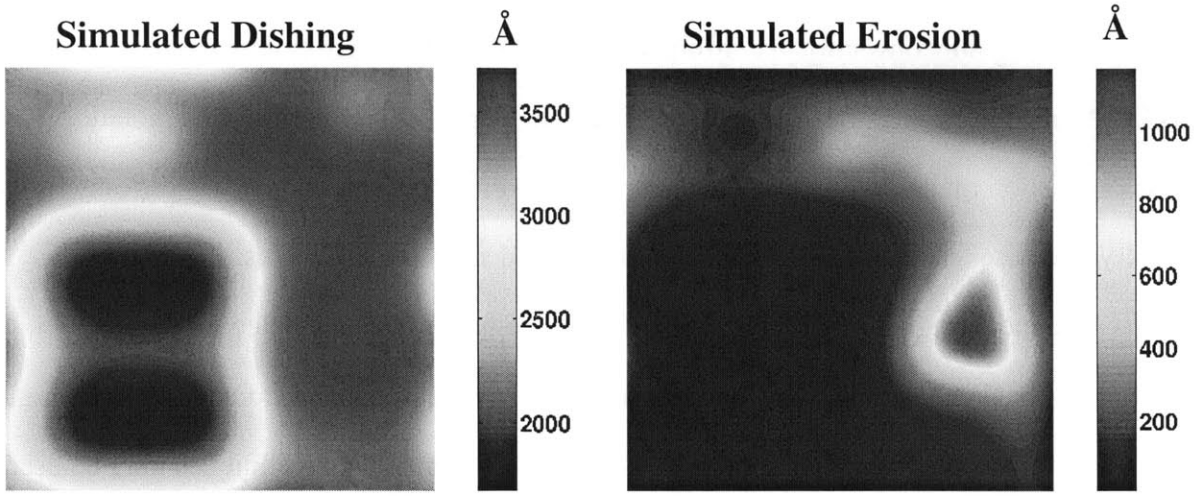


Figure 3.20: Dishing and erosion prediction for silica HSS slurry CMP process.

3.8.2 Clearing Time Prediction

The clearing time for the CMP process can also be predicted using the modeling methods described in this chapter. The prediction for the example CMP process is shown in Figure 3.21. Prediction of the clearing time is another possible decision metric for the feasibility of an nonconventional CMP process, since the minimum clearing time necessary to guarantee complete overburden oxide clearing ultimately determines the necessary polish time.

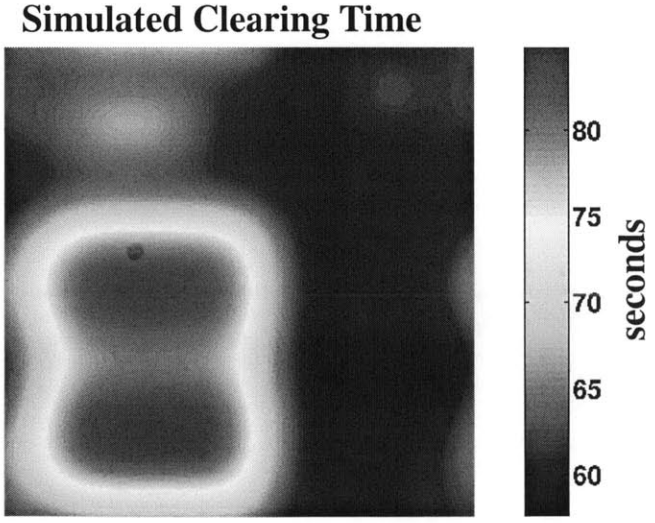


Figure 3.21: Clearing time prediction for silica HSS CMP process.

3.9 Summary

This chapter has described a modeling and characterization methodology for CMP processes using nonconventional consumables. The model used here is based on the conventional STI CMP model, but eliminates the assumed density dependencies used in that model. A characterization methodology is used where model parameters are first extracted from individual density structures, then modelled based on observed density dependencies.

The modeling and characterization methodology has been validated on three nonconventional consumable sets, and has shown to have reasonable accuracy in predicting the behavior of CMP processes using these consumables. The modeling techniques may be used for dishing, erosion, and clearing time prediction on arbitrary layouts.

The methods described here allow for immediate phenomenological capture of the effects of CMP processes using nonconventional consumables, while allowing for further investigation of model parameter density dependencies. Knowledge of density dependencies may provide insight towards physical sources of the unusual CMP behavior, and lead to development of more advanced models.

Chapter 4
Pattern Density Optimization Methods

A pattern density effect has shown to be a dominant factor in post-CMP film thickness [26, 27]. This chapter describes two methods of density modification, *dummy fill* and *reverse etchback*, that are used to correct pattern density mismatches and thus lower post-CMP film thickness variation. Previous work on dummy fill for dielectric CMP is reviewed in Section 4.1. Methods for density modification in layout design are described in Section 4.3. In Sections 4.4 and 4.8, two methodologies for chip-scale dishing and erosion reduction through inclusion of dummy fill and etchback structures are developed. These methodologies are verified on an example SRAM layout. Specific algorithms for dummy fill and reverse etchback structures are described in Section 4.6. The chapter is summarized in Section 4.10.

4.1 Previous Work on Density Optimization

Previous work on dummy fill for interlevel dielectric (ILD) CMP has been reported. Stine [45] noted the beneficial effects of metal fill on post-CMP thickness variation of ILD. Stine also analyzed the capacitance effects of fill structures, and proposed a method of selecting optimal fill by analyzing fill structures of various line widths and spaces and creating a metal fill design chart composed of contours showing structures with equivalent capacitance effect and contours showing structures with equal densities. Analysis of the metal fill design chart provides a selection method for dummy fill structure dimensions.

Kahng [46] analyzes the dummy fill problem and proposes a fill synthesis algorithm based on a linear programming approach. Features are added to satisfy two criterion: first, that the layout densities satisfy specified upper and lower density bounds for square windows of all possible sizes; and second, that the maximum variation of density over all possible window

positions on the layout is minimized. Kahng also analyzes implications of the actual filling patterns, in terms of impact on coupling and capacitance.

Tian [47] analyzes the dummy fill problem for STI by deriving a density-based STI CMP model and proposing an iterative algorithm for insertion of dummy fill. The proposed method involves beginning with the original layout, and finding the minimal effective density point on that layout. A fill structure is inserted at that point, and the new layout is reanalyzed to find the new minimal density point. The process is repeated until a termination criterion is met. Tian uses a termination criterion of the maximum predicted height difference between the lowest and highest effective density points on the layout. Tian's work does not propose an algorithm for computation of dummy fill structure density, but rather assumes the existence of a dummy structure library which can be exhaustively searched for the optimal dummy structure at each analysis step in the algorithm. One disadvantage of such a method is that the execution time of the dummy fill routine will increase with the size of the dummy structure library.

4.2 Importance of Density Optimization

A continuing CMP research goal in CMP is to reduce pattern-dependent effects in the post-CMP film. To date, elimination of pattern effects by CMP process and consumable design has proven elusive. Thus, alternatives at the design (or layout) level, namely density optimization are also being pursued. The concept of density optimization stems from the underlying basis of density-based CMP models.

4.2.1 Density Optimization for ILD

In ILD CMP, the main goal is to reduce the post-CMP thickness variation. Examining these models reveals that, absent any special processing considerations, a high density region will result in thicker post-CMP oxide films than in a low density region. Analyzing the pure density model (described in Equation 2.1 in Chapter 2, taken from [33]), one can derive a simple relationship, assuming removal of local step height everywhere, for ILD CMP [30]:

$$\Delta z = \Delta\rho \cdot z_1 \tag{4.1}$$

where Δz is the post-CMP thickness variation, $\Delta\rho$ is the effective pattern density variation, and z_1 is the initial step height. From this equation we can see that final thickness variation scales directly with effective pattern density range. Thus, a good method for minimizing post-CMP thickness variation is to minimize the pre-CMP density range across the die.

4.2.2 Density Range Reduction By Changing CMP Process

One way of reducing effective pattern density range is by increasing the planarization length of the CMP process. There are various ways of doing this, including decreasing down force [36], and using stiffer CMP pads [36,37]. Increasing planarization length means that more of the local die area is averaged out to compute effective density (i.e., the CMP process is influenced by more of the surrounding topography across the wafer), and so the effective density range should decrease. In the extreme case, a planarization length equal to the die size will result in a very small pattern density range (the range will not be zero due to the weighting filter used in the effective density calculation, which weights local densities closer to the point of interest more than local densities at points farther away).

4.2.3 Density Optimization for STI

In STI CMP, the main goal is to reduce the pattern-dependent erosion and dishing of the post-CMP STI films. By identifying the key contributors to erosion and dishing, it is possible to develop design strategies at the layout level that aim to reduce these problems. There are two major contributors to erosion and dishing in STI CMP. The first is the underlying nitride density, which directly affects the amount of erosion at a given location on the die. The second, and more subtle, cause is the overburden oxide density range, which affects the amount of overpolish time necessary to ensure complete oxide clearing in all regions on the die.

In ILD CMP, the main concern is the minimization of the post-CMP oxide thickness

variation. This also offers a benefit in STI because small post-CMP overburden oxide thickness variation translates to more uniform clearing across the die and reduces the polish time required to clear the overburden oxide in the highest-density locations on the wafer (and thus, the amount of overpolish time that low density locations on the wafer experience). This reduces the amount of dishing and erosion in these low density overburden oxide locations.

However, even if the overburden oxide density variation is zero, typically a nominal overpolish time is built into the CMP process to ensure clearing. This is due to wafer-to-wafer as well as within-wafer process variability in removal rates. Thus, the nitride overpolish phase requires careful study to analyze nitride layer pattern effect on dishing and erosion.

Using STI CMP models (discussed in Chapter 2), it is possible to predict post-CMP dishing and erosion, and identify potential problematic locations. For a given polish time, the underlying nitride density directly affects the amount of nitride erosion and oxide dishing. As explained in Chapter 2, the higher the nitride density, the lower the erosion. This is illustrated in Figure 4.1 (from Equation 2.56), which shows dishing and erosion vs. effective pattern density for a nominal overpolish time. These figures illustrate the effect of only underlying nitride density by using a completely uniform oxide overburden density so that oxide clearing occurs at approximately the same time everywhere.

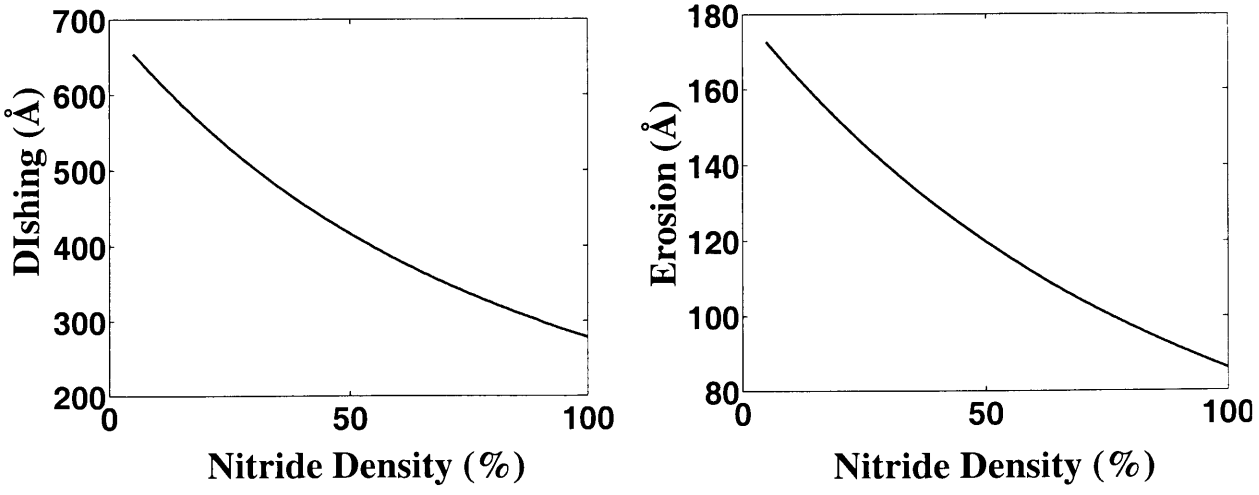


Figure 4.1: Dishing and erosion dependence on underlying STI nitride density.

Density modification in layout design can be used iteratively to reduce the erosion/dishing at these locations by raising the effective density in these regions.

In this thesis, two methods of density optimization for STI are discussed. The simpler method is to density-equalize the overburden oxide layer, similar to the approach used in ILD CMP, with the effect of reducing the difference in polish time required to clear the various locations on the wafer. A perfect density-equalized oxide overburden layer results in the oxide clearing the die at approximately the same time (subject to within-die/within-wafer removal rate variation). This methodology is discussed in Section 4.4. Minimizing the pattern density distribution of the overburden oxide can be performed without any modeling of dishing or erosion, although the planarization length of the oxide phase of the CMP process used will need to be characterized.

The second method involves analysis of both layers of the STI CMP process, the overburden oxide and the underlying nitride. This method involves enforcing a minimum nitride layer density using dummy fill, followed by minimization of the density range of the resulting overburden oxide layer using dummy fill and reverse etchback. Modification of the nitride and oxide layers seeks to reduce both major components of erosion and dishing. This methodology is explained in Section 4.8.

4.3 Techniques for Density Modification

This section discusses the design techniques that can be used to modify the density distribution of the active area and overburden oxide layers of a particular layout design. Two techniques are discussed: *dummy fill*, which is used to raise density, and *reverse etchback*, which is used to lower density.

4.3.1 Dummy Fill

Dummy fill [40] refers to the addition of features during the layout phase. These features are added solely to raise the density of specific regions on the layout (i.e., low density regions). This can be done on the front end (insertion of dummy active areas into a shallow trench isolation layout layer) or back end (insertion of dummy metal into metal layers). Conventional dummy fill utilizes a single feature (or *dummy feature*) arrayed over the original layout, with sections of the array removed if they interfere with existing features on the layout. Exact dummy feature dimensions can vary from layout to layout, and may depend on the design rule set used for a layout. Smart dummy fill refers to the implementation of an algorithm that analyzes the original layout pattern density distribution and customizes the dummy features to minimize the resulting effective pattern density distribution. Such analysis can be complicated since the modification of a single dummy feature on a layout may affect the effective densities of many surrounding points.

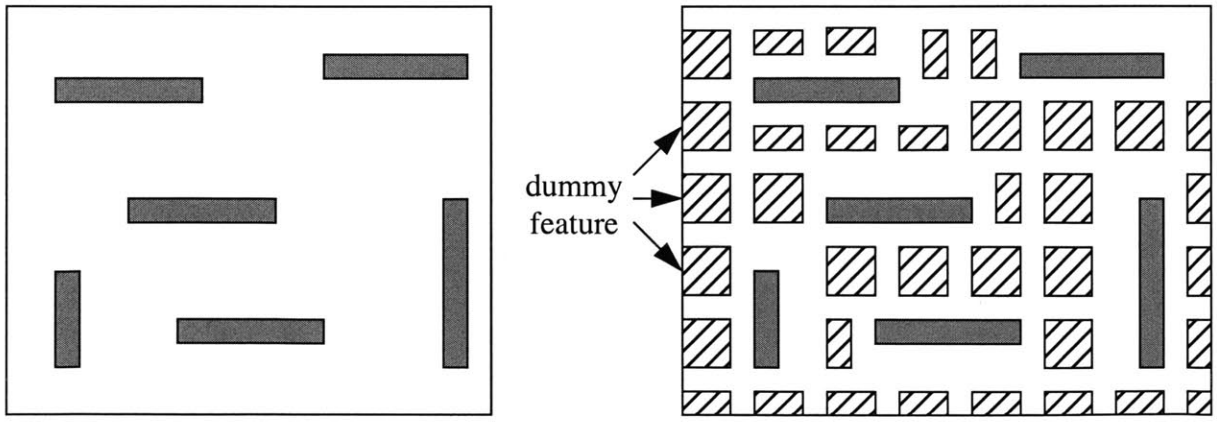


Figure 4.2: Illustration of dummy fill.

4.3.2 Reverse Etchback

Reverse etchback, first described by Davari [43], is a processing step performed after deposition and before CMP. It involves using a second lithography and etch to remove raised areas on the deposited film, thus lowering the film density. This procedure is normally used only for STI processes, where the underlying nitride layer serves as an effective etch stop. Etchback for other layers is difficult since a timed etch is required. The etchback mask is typically created by shrinking all of the features on a given layout by a fixed amount (called the *etchback bias*, which scales with technology feature size). For large features, this results in the removal of a large portion of the raised area material, and a resulting low density. Figure 4.3 illustrates the reverse etchback process. The etchback step is typically expensive since it requires an additional mask and processing steps, and most manufacturers would like to avoid this step if there are other successful methods that achieve the desired amount of planarization in films. However, reverse etchback has been stated to be the most effective in reducing post-CMP variation in heights between large and small active areas [44].

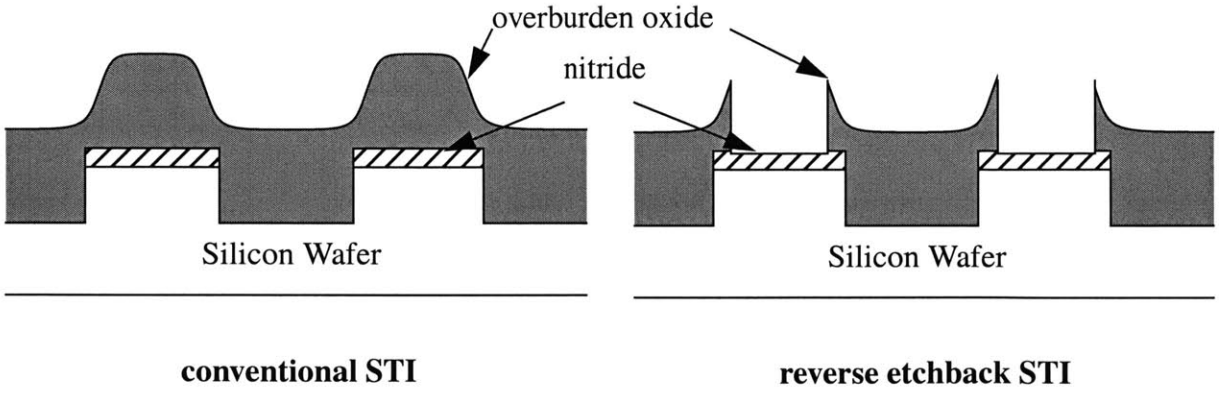


Figure 4.3: Illustration of the etchback process - cross section of STI device.

The resulting low density regions that arise from using conventional reverse etchback may not be desirable if there are other medium density regions on the layout. Therefore, it may be useful to use the process of *selective reverse etchback*. Selective reverse etchback refers to the customization of the etchback mask that results in less material removed than in a conventional

etchback process. This is accomplished by replacing large etchback features with arrays of selective etchback cells, and in some cases, complete removal of etchback features from the etchback mask. The goal is to raise pattern density in these regions that would normally have low post-etchback pattern density (i.e., large features on the original layout) such that the effective density range is minimized. This will result in longer polish times due to a higher average density, but better planarity due to lower pattern density range.

Creating selective etchback features to achieve a particular density is a fairly simple process. Without any etchback, a large feature will exhibit a local density of 100%. With complete etchback, the local density will be close to 0%. To create a selective etchback feature to achieve an eventual target density ρ, a selective etchback feature of (100-ρ)% is drawn. Since the etchback feature is what is *subtracted* from the original 100% local density region, the final density achieved will be the target density. This is illustrated in Figure 4.4.

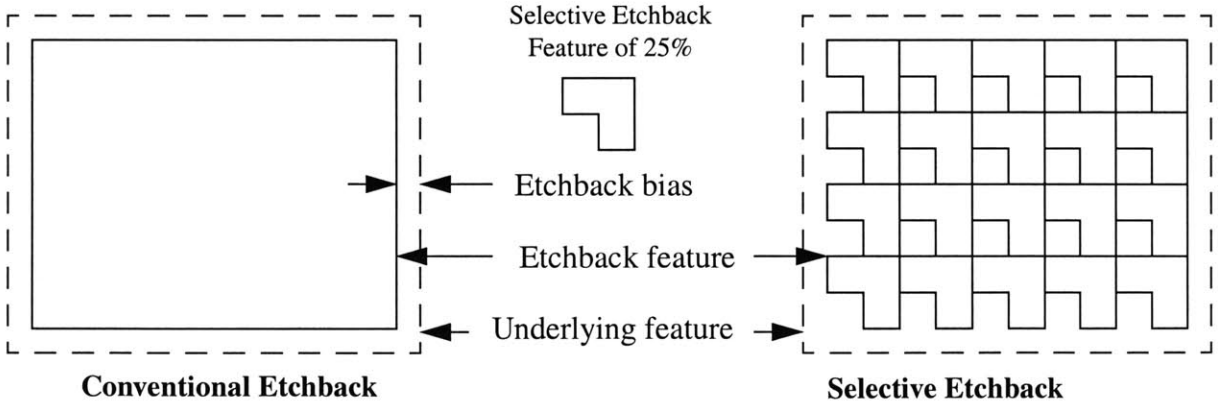


Figure 4.4: Illustration of selective etchback, top level (layout) view.

4.4 Density Optimization - Single Layer Analysis

The density optimization problem can be put into mathematical form by defining two matrices $\underline{\underline{\mathbf{L}}}$ and $\underline{\underline{\mathbf{D}}}$. $\underline{\underline{\mathbf{L}}}$ represents the original layout, discretized into small cells, each with a distinct local density value. $\underline{\underline{\mathbf{D}}}$ will be a matrix of the same dimensions as $\underline{\underline{\mathbf{L}}}$, and represents the incremental change in density that can be introduced using density modification techniques. The

density optimization problem can be stated as:

$$\{\underline{\underline{\mathbf{D}}} \mid \min\,[\text{range}(F(\underline{\underline{\mathbf{L}}}+\underline{\underline{\mathbf{D}}}))]\}$$

where F represents the function used to compute effective pattern density from a local pattern density matrix. In words, this can be stated as find the matrix $\underline{\underline{\mathbf{D}}}$ such that the effective pattern density function computed on ($\underline{\underline{\mathbf{L}}}+\underline{\underline{\mathbf{D}}}$) results in a matrix with the smallest range of density values.

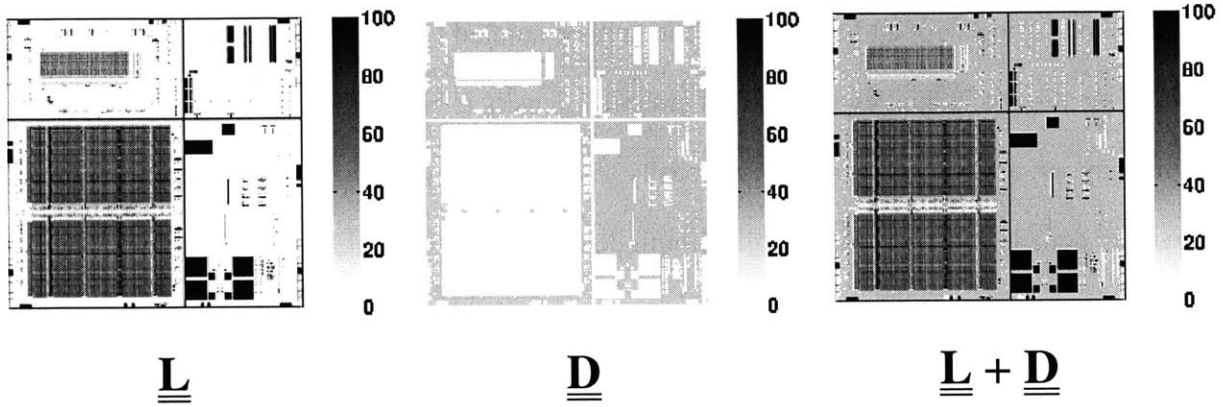


Figure 4.5: Examples of the layout, dummy fill, and combined density matrices.

The matrix $\underline{\underline{\mathbf{D}}}$ will typically have restrictions placed on it depending on the original layout density matrix and the design rules for a particular layout. For instance, consider the matrix cells in $\underline{\underline{\mathbf{D}}}$ that correspond to points on the original layout where a feature exists. These cells may not be assigned a value because it may be infeasible to implement density modification structures at these locations (due to the presence of existing features in the original layout). Thus, all of these matrix cells in $\underline{\underline{\mathbf{D}}}$ will have a value of zero. Similarly, design rules may restrict the number of cells in $\underline{\underline{\mathbf{D}}}$ that are allowed to be set to non-zero values.

The density optimization procedure is represented in flowchart fashion in Figure 4.6. It assumes that the CMP process to be used has already been characterized, and a planarization length has been obtained. The first step is density analysis of the original layout and computation of the effective density range. If this density range is sufficiently small, the layout requires no additional density modification structures. Assuming this is not the case, more analysis must be performed.

The second step is analysis of the density map and selection of the density modification feature characteristics. For conventional dummy fill, only one dummy feature is designed; for smart dummy fill, multiple dummy features are designed. There are many algorithms that can be used to select dummy feature characteristics; one simple method is to create a dummy feature that, when arrayed, has an effective density equal to the maximum density of the original layout. This method, along with additional methods, will be studied later in this chapter. One caveat that must be mentioned is that, even if the fill structure density is below the maximum density of the original layout, adding additional local density to a layout may result in a modified layout whose maximum density is larger than the maximum density of the original layout. Examples of this in the literature include a study by Liu [41], where an original layout with a maximum density of 60% resulted in maximum densities of 70% and 80% when modified with fill structures of 30% and 50%, respectively.

After the density modification feature matrix has been completed, the matrix is merged with the original layout local density matrix, and the effective pattern density across the resulting matrix is computed. If it is within the specifications of pattern density range, then the final density modification matrix is completed. If the pattern density range is still too large, the process is repeated, using analysis of the previous iteration to direct any changes. For example, if there still exists a particular low density region in the merged matrix of the previous iteration, then additional fill can be inserted into this region to raise the pattern density there.

After the proper density modification matrix is created, it is converted into actual layout features. This is done by creating layout to incorporate dummy fill features that achieve the particular density modification density matrix specifications. The density modification layout is then merged with the original layout to produce a revised layout. This resulting layout can then be analyzed for pattern density range to verify that the range has been reduced to the required amount.

This methodology is general and can be used on any particular layout. It requires only that the CMP process that will be used has been characterized (i.e., the planarization length of the process is known), in order that effective pattern densities can be computed.

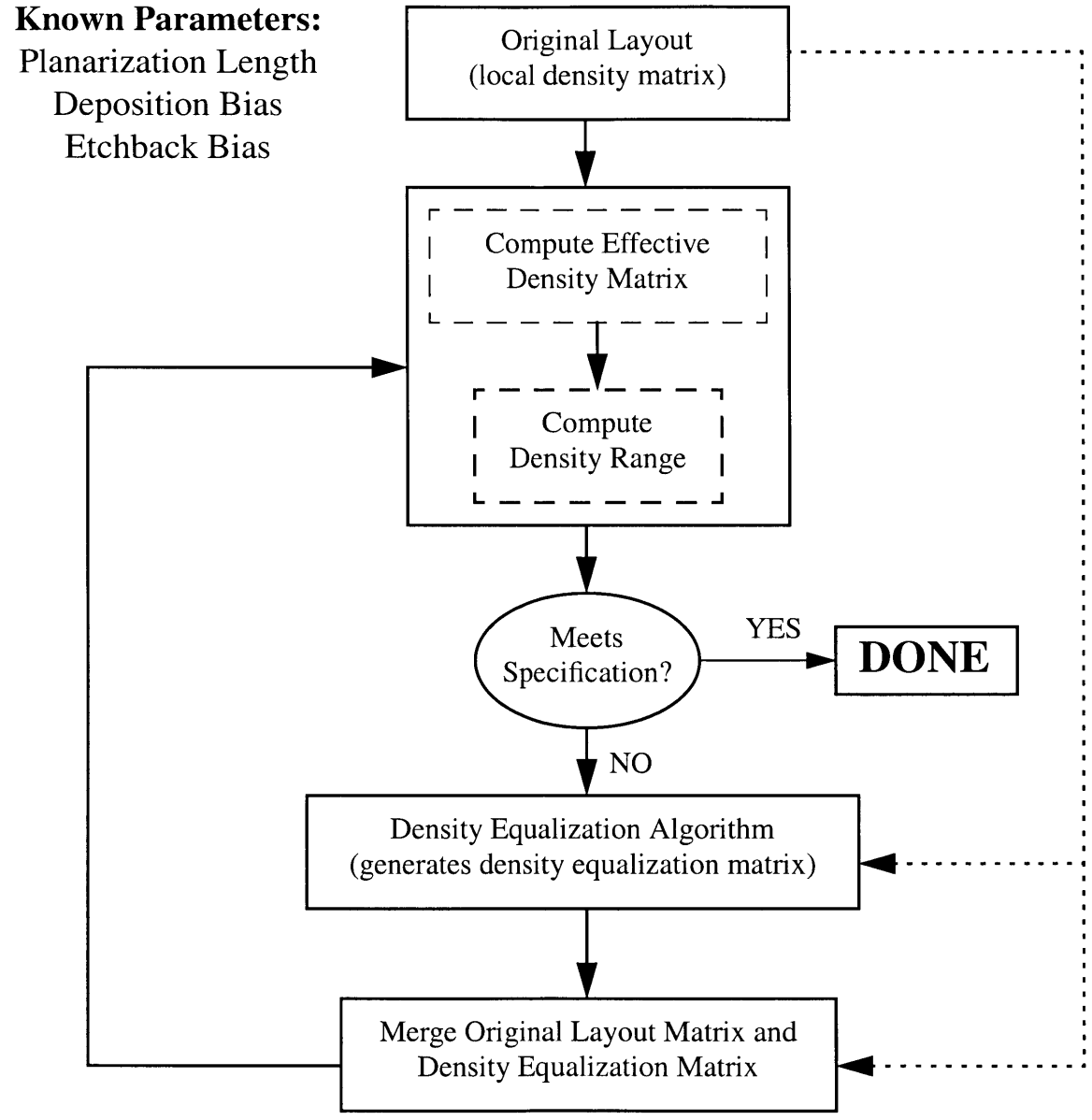


Figure 4.6: Flowchart for density optimization procedure (single layer).

4.5 Implementation

Implementation of density optimization involves converting the density modification matrix into dummy features and selective etchback cells. Creating selective etchback cells of particular density has been described previously; creating dummy cells involves deciding on a particular shape for a dummy feature and computing the dimensions of the feature to achieve the desired density.

There are many possibilities for feature shapes that can be used for dummy fill structures. Two such shapes are illustrated in Figure 4.7. Square features are used because they are simple to create in layout.

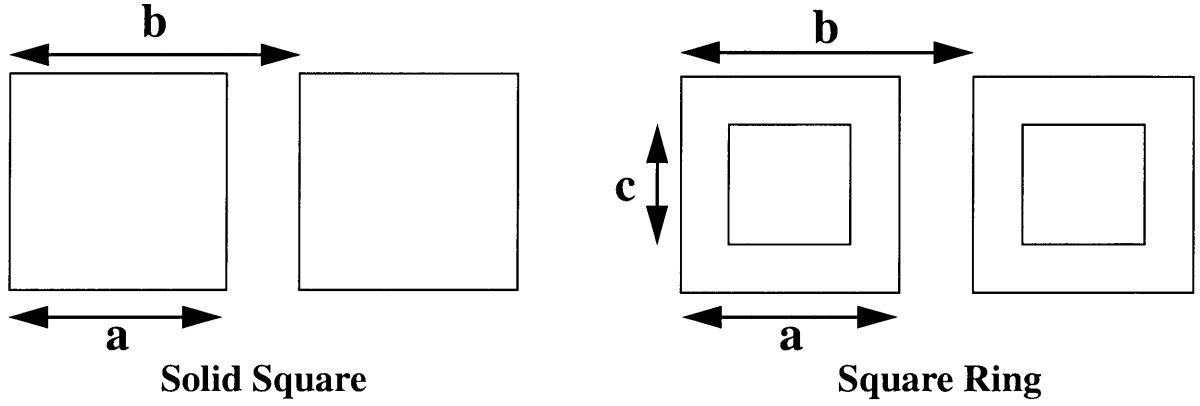


Figure 4.7: Example dummy structures.

4.5.1 Computing Dummy Feature Density

When computing the density of a dummy feature, it is important to consider the effects of deposition bias and etchback (if there is an etchback step). Deposition bias is described in Section 1.5.1 of Chapter 1. This effect can change the pattern density of a dummy structure, and so its effects must be included during the design of such structures. Of critical importance is the consideration of the effect on dummy features of any film (e.g., HDP) with a negative deposition bias. Dummy features smaller than the deposition bias in this case will cause negligible change in the film pattern density, since the deposition bias will effectively remove them.

Etchback bias refers to the amount that a feature is shrunk to produce its etchback feature.

This is illustrated in Figure 4.4. If no etchback is used, this is equivalent to having an etchback bias of half the feature size (i.e., the etchback feature has zero width and length).

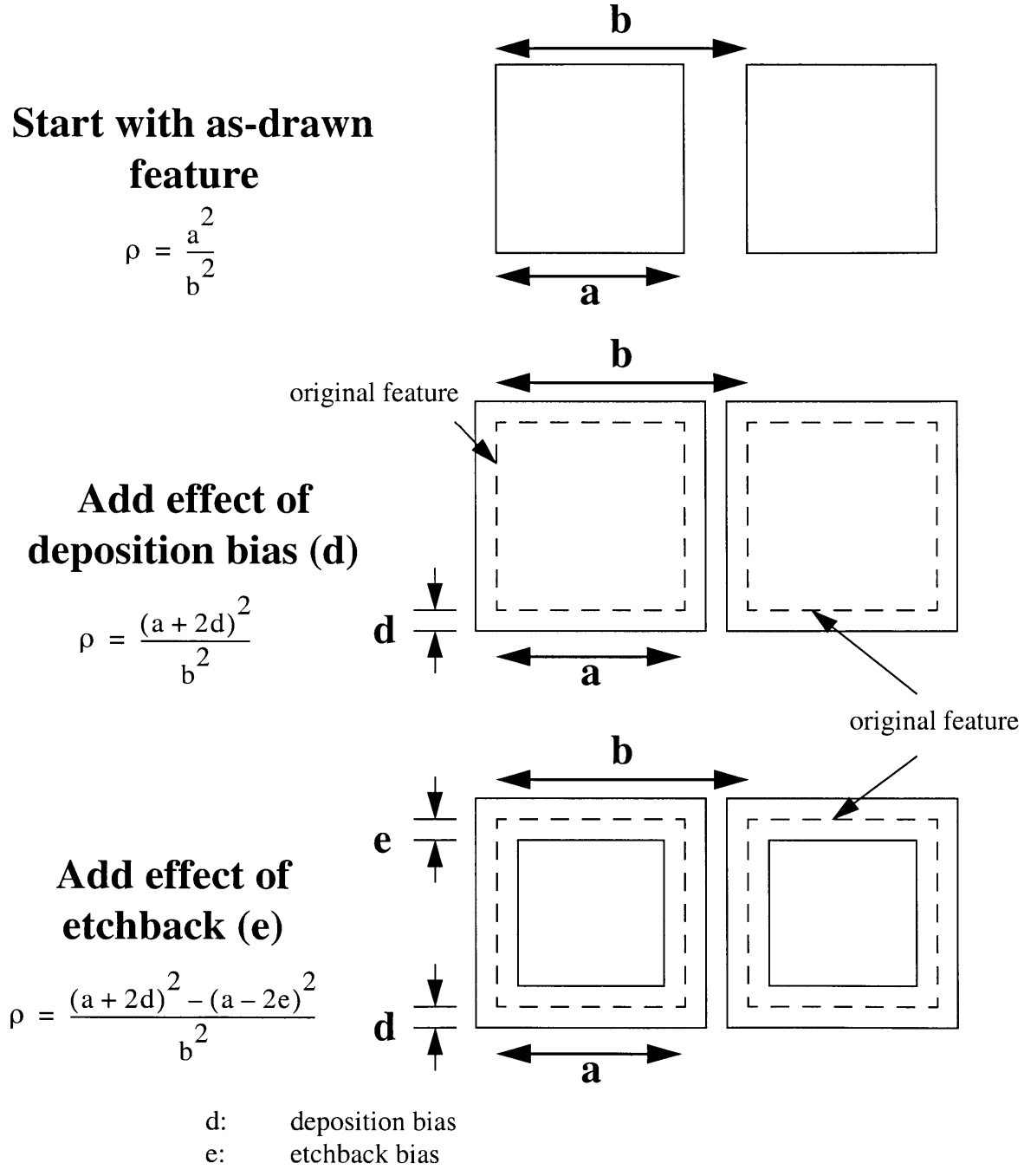


Figure 4.8: Illustration of density computation for dummy features.

The formulas for calculating the pattern density based on structure dimensions can be found using algebraic manipulation by considering the processing steps involved. A derivation for

solid squares is given in Figure 4.8. The equation for the local density of a square ring can be obtained in a similar fashion. The final formulas for local structure density are:

Solid Square:
$$\rho = \frac{(a+2d)^2 - (a-2e)^2}{b^2} \tag{4.2}$$

Square Ring:
$$\rho = \frac{(a+2d)^2 - (a-2e)^2 + (c+2e)^2 - (c-2d)^2}{b^2} \tag{4.3}$$

where a is the dimension of the feature, b is the grid spacing, c is the dimension of the square central hole of the square ring, d is the deposition bias, and e is the etchback bias. For a given STI process, d and e are known. The other parameters are then chosen by selecting fixed values for all but one variable (e.g., b for a solid square, or b and c for a square ring) and solving for the value of that variable.

The use of selective reverse etchback structures would modify the equations for density since a fixed etchback bias is no longer used to determine the etchback layer. Revised formulas can be computed based on the exact dimensions of the reverse etchback structures that are used.

Once dummy fill structures have been created, they may be arrayed to form additional layout layers which may be merged with the original layout to form a new layout. This new layout should exhibit significantly smaller pattern density range.

4.6 Algorithms

In this section, several alternative algorithms for selecting dummy feature densities are described. This choice of dummy fill algorithm can be difficult because it requires careful analysis of the initial layout, as well as constraints on location of dummy structures.

In general, the density optimization algorithm must take as input a local density matrix and a dummy fill allowed matrix $\underline{\underline{\mathbf{F}}}$ (which indicates the regions in which dummy features can be inserted), along with the planarization length (PL) of the CMP process of interest. The optimization procedure outputs a density modification matrix $\underline{\underline{\mathbf{D}}}$ that can be converted into dummy fill structures. All elements of $\underline{\underline{\mathbf{D}}}$ are restricted to non-negative values. Once the optimal $\underline{\underline{\mathbf{D}}}$ matrix is obtained, the original layout is revised. The analysis can be repeated using the revised layout to create a reverse etchback matrix $\underline{\underline{\mathbf{R}}}$, where the elements of $\underline{\underline{\mathbf{R}}}$ are restricted to be non-positive. Note that the $\underline{\underline{\mathbf{F}}}$ matrix for the dummy fill case is different from the $\underline{\underline{\mathbf{F}}}$ matrix for the reverse etchback case.

The dummy fill allowed matrix $\underline{\underline{\mathbf{F}}}$ is formatted as a matrix of the same dimensions as the local density matrix $\underline{\underline{\mathbf{L}}}$. $\underline{\underline{\mathbf{F}}}$ has a value of 1 where dummy features can be assigned and a 0 where dummy features are not allowed. For simplicity's sake, the matrix $\underline{\underline{\mathbf{F}}}$ used can be created by assigning a 1 to any element in the $\underline{\underline{\mathbf{F}}}$ matrix whose corresponding element in the local density matrix has a density value of zero.

Four algorithms will be discussed here. The first algorithm, called the maximum effective density algorithm, finds the maximum effective density value in the original layout and creates a fill layer by replicating a single fill structure of this maximum effective density in all valid dummy fill locations. The second algorithm, called the "offset from maximum effective density" algorithm, computes the effective densities at each location on the layout and creates a fill layer by inserting dummy structures with densities equal to the difference between the maximum effective density of the original layout and the effective density at each location on the original layout. The

third technique creates a fill layer by replicating a single fill structure, whose density is determined through density optimization analysis of the original layout. The final technique determines the fill structure by back-calculating the optimal fill structure from the computations used in the effective density calculation.

4.6.1 Maximum Effective Density Algorithm

The maximum effective density algorithm assigns each valid cell in the density modification matrix with the value of the maximum effective density of the original layout. The effective density map $\underline{\underline{\mathbf{E}}}$ is calculated from $\underline{\underline{\mathbf{L}}}$ and the planarization length, and the maximum value can then be found. Mathematically, this can be expressed as:

$$\underline{\underline{\mathbf{D}}} = \max(\underline{\underline{\mathbf{E}}}) \bullet \underline{\underline{\mathbf{F}}} \qquad (4.4)$$

where the dot notation (•) indicates elementwise multiplication. A small example (shown in Figure 4.9) helps illustrate this algorithm. Assume a small initial 4x4 $\underline{\underline{\mathbf{L}}}$ matrix. For simplicity, also assume that the effective density $\underline{\underline{\mathbf{E}}}$ matrix can be calculated by averaging the elements contained in a square of length N around it (N is equivalent to the planarization length in discretized units). For this example, let N have a value of 3.

4.6.2 Offset from Maximum Effective Density Algorithm

The "offset from maximum effective density" algorithm computes the effective density map $\underline{\underline{\mathbf{E}}}$, and then an offset map $\underline{\underline{\mathbf{O}}}$, which is a matrix of the differences between the maximum value of the $\underline{\underline{\mathbf{E}}}$ matrix and each element in $\underline{\underline{\mathbf{E}}}$. The final density modification matrix is computed by using fill allowed matrix $\underline{\underline{\mathbf{F}}}$ to mask out invalid density modification matrix cells from $\underline{\underline{\mathbf{O}}}$. Note that elements in the local density matrix with a density of zero can have a positive effective density due to the long range weighting scheme used in the effective density calculation.

$$\underline{\underline{\mathbf{D}}} = \text{mask}(\underline{\underline{\mathbf{F}}}, (\max(\underline{\underline{\mathbf{E}}}) - \underline{\underline{\mathbf{E}}})) \qquad (4.5)$$

where mask is the mask function. The mask function $\underline{\underline{M}}$=mask($\underline{\underline{A}}$, $\underline{\underline{B}}$) assigns elements in the output matrix $\underline{\underline{M}}$ the value of the corresponding element in the $\underline{\underline{B}}$ matrix if and only if the corresponding element in the $\underline{\underline{A}}$ matrix is 1. Otherwise, the element value in $\underline{\underline{M}}$ is set to zero. An example of the algorithm (using the same data set as in Section 4.6.1) is shown in Figure 4.10.

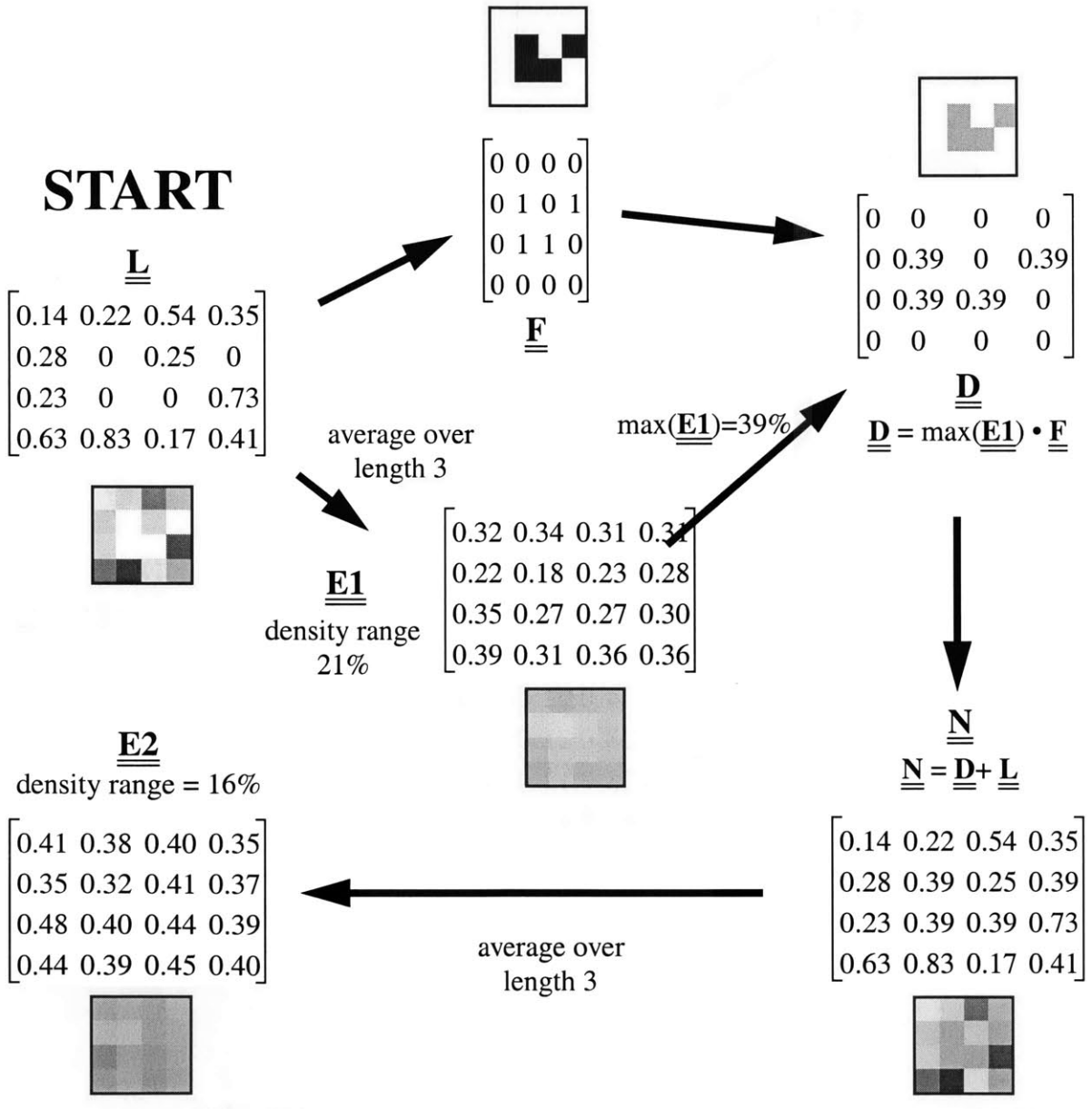


Figure 4.9: Example of maximum effective density algorithm.

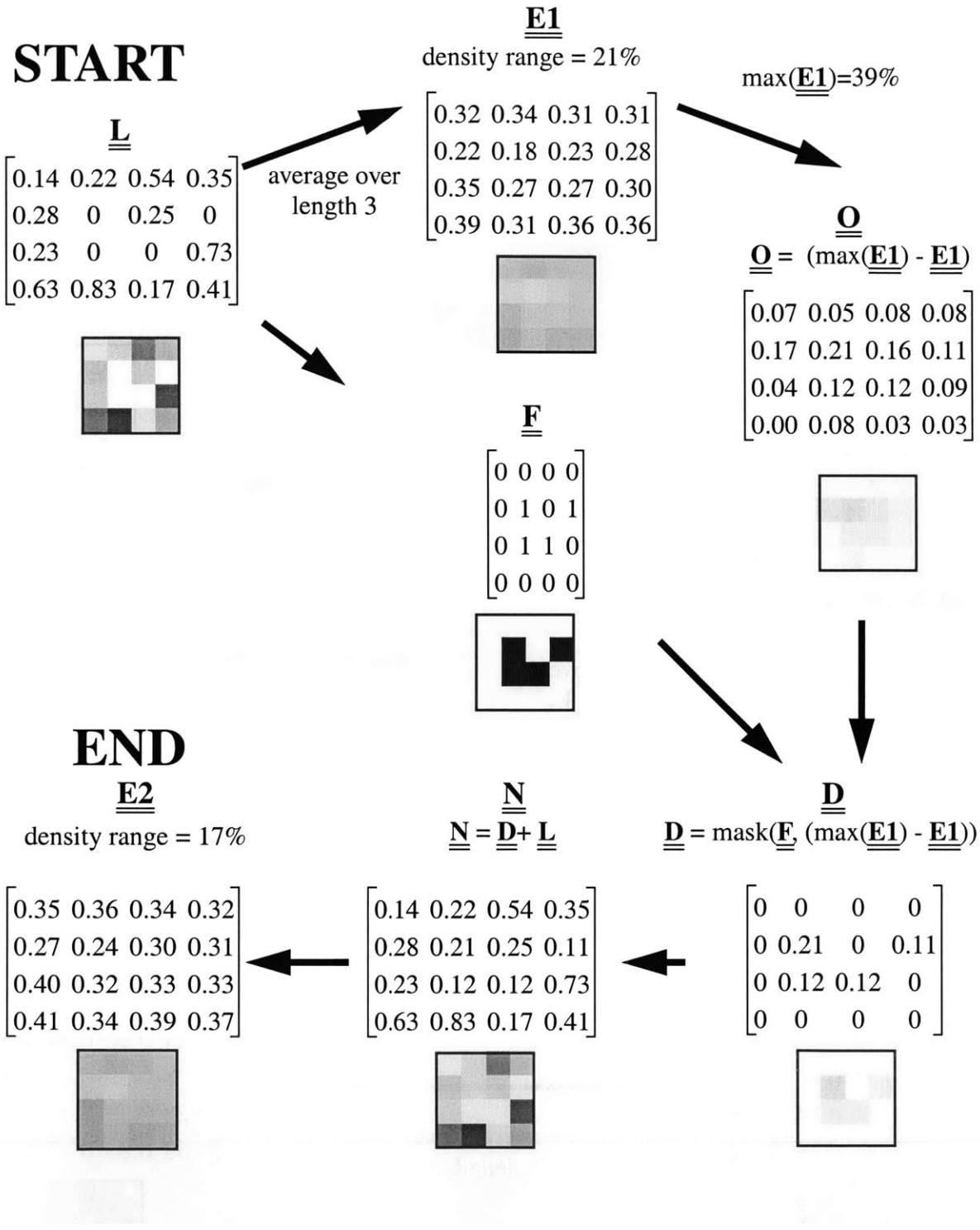


Figure 4.10: Example of offset from maximum effective density algorithm.

4.6.3 Optimal Single Value Algorithm

The optimal single value algorithm uses an optimization loop to compute the single optimal density value to use in the density modification matrix (i.e., an optimal value which all valid elements in $\underline{\underline{D}}$ are set to). It takes into account the fill allowed matrix $\underline{\underline{F}}$. The algorithm is illustrated in Figure 4.11.

Given:
$\underline{\underline{L}}$ layout density matrix
$\underline{\underline{F}}$ "density equalization allowed" matrix
PL planarization length of the CMP process

Find:
$\underline{\underline{D}}$ dummy fill matrix

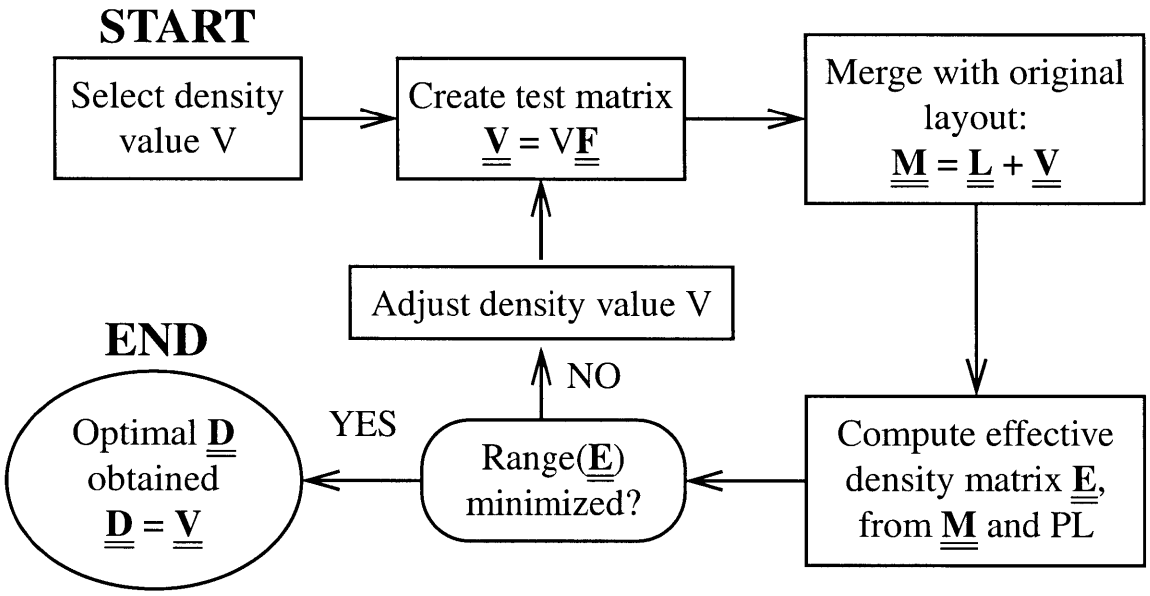


Figure 4.11: Optimal single value algorithm for density optimization.

4.6.4 FFT-Based Algorithm

The FFT-based algorithm uses FFT analysis to compute the optimal density values of the individual dummy fill structures. By considering the FFT analysis procedure by which effective density is calculated, it is possible to back-calculate the density matrix that results in a constant effective density matrix. Given a layout matrix $\underline{\underline{L}}$ and a weighting matrix function $\underline{\underline{W}}$ (which depends on planarization length and is the matrix implementation of the filter described in Figure 2.2), this can be expressed mathematically as:

$$\{ \underline{\underline{\mathbf{X}}} \mid \mathrm{IFFT}(\mathrm{FFT}(\underline{\underline{\mathbf{L}}} + \underline{\underline{\mathbf{X}}})) \bullet \mathrm{FFT}(\underline{\underline{\mathbf{W}}}) = \underline{\underline{\mathbf{C}}}\} \tag{4.6}$$

where $\underline{\underline{\mathbf{C}}}$ is a matrix, each of whose elements is the same value. If the element value of $\underline{\underline{\mathbf{C}}}$ is selected beforehand (e.g., the maximum effective density of the original layout $\underline{\underline{\mathbf{L}}}$), then Equation 4.6 can be rearranged to give:

$$\underline{\underline{\mathbf{X}}} = \mathrm{IFFT}(\mathrm{FFT}(\underline{\underline{\mathbf{C}}}) - \mathrm{FFT}(\underline{\underline{\mathbf{L}}}) \bullet \mathrm{FFT}(\underline{\underline{\mathbf{W}}})) \div \mathrm{FFT}(\underline{\underline{\mathbf{W}}})) \tag{4.7}$$

where • and ÷ mean matrix *element by element* multiplication and division. The density modification matrix $\underline{\underline{\mathbf{D}}}$ can be constructed by using the dummy fill allowed matrix $\underline{\underline{\mathbf{F}}}$ to mask off the values of $\underline{\underline{\mathbf{X}}}$ that are valid.

4.6.5 Comparison

We now compare results using these different dummy fill generation algorithms. The local density layout used is a version of a test chip with SRAM (similar to a layout later used for verification, described in Section 4.7).The original layout is 14 mm x 14 mm, and is discretized into 40 μm x 40 μm cells. This implies a local density matrix of 350x350 discretized units, for a total element count of 122,500 cells. There is total of 66,565 eligible dummy elements (i.e., elements in $\underline{\underline{\mathbf{F}}}$ with a value of 1).

Using the test layout, the optimal density matrix $\underline{\underline{\mathbf{D}}}$ was computed for each of the algorithms described here. A revised layout matrix $\underline{\underline{\mathbf{L}}}$' is then constructed, and a corresponding effective density matrix $\underline{\underline{\mathbf{E}}}$' is then computed. Density statistics of maximum, minimum, and range are obtained for each revised effective density matrix. The final results, in addition to the nominal density statistics, are shown in Table 4.1.

Table 4.1: Comparison of Density Optimization Algorithms

	Density (%)		
Analysis	Minimum	Maximum	Range
Nominal (Original Layout)	0.62	23.25	22.64
Maximum Effective Density	18.53	24.63	6.10
Offset from Maximum Effective Density	15.45	23.87	8.42
Optimal Single Value (27.16%)	20.96	25.05	4.09
FFT analysis	18.54	24.63	6.09

As can be seen, all of the algorithms show significant effective pattern density range reduction in comparison to the nominal layout. The optimal single value method appears to work best for this test case; however, it is unclear whether this is the best method for arbitrary layouts.

4.7 Case Study - SRAM chip

To verify the dummy fill algorithm, a case study is performed on an SRAM chip to produce dummy fill and selective reverse etchback structures. The layout is then modified to include these density modification structures, and the density distribution of the revised layout is computed and shows considerable improvement when compared to the original layout.

4.7.1 Test Mask Description

The STI layer of a test mask is used to verify the pattern density optimization methodology. This layout is 14 mm x 14 mm, and contains a 256K SRAM structure along with other test structures. An illustration of the chip is seen in Figure 4.13.

The STI process utilizes etchback with an original etchback layer that is produced by shrinking the original layout by 0.5 μm (i.e., an etchback bias of 0.5 μm). The film deposition bias is 0.3 μm. Localized pattern density information is created using a local discretization cell size of 40μm x 40μm. A histogram of the density distributions of the localized density information is

calculated and shown in Figure 4.13. The clustering of density at low density ranges reflects the fact that etchback is used; chips that do not include an etchback process will typically have density values spread further across the density spectrum, as shown in Figure 4.13.

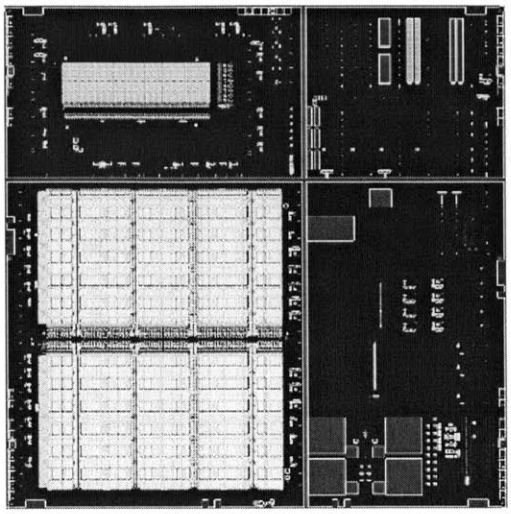


Figure 4.12: Case study test mask.

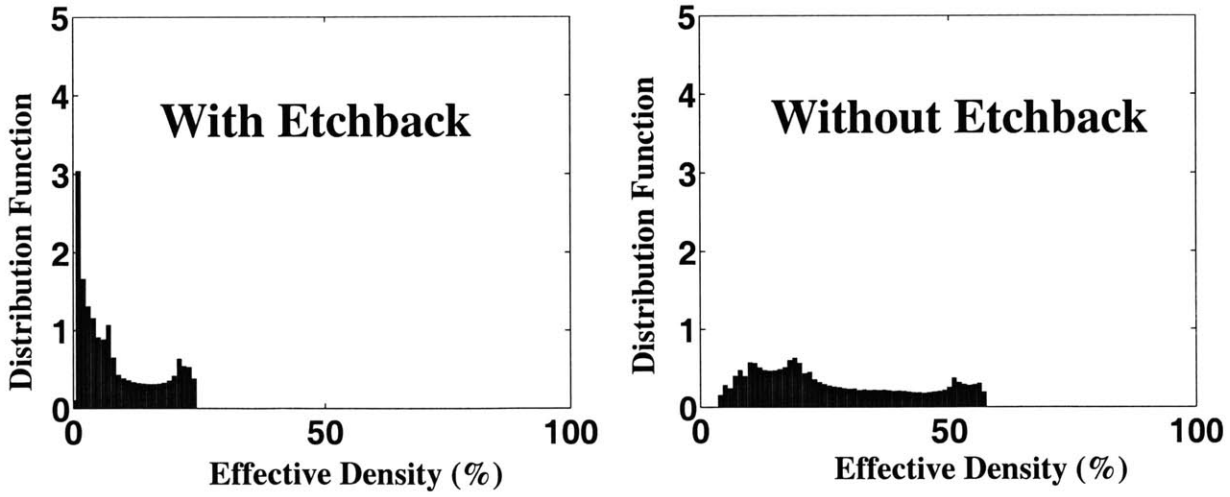


Figure 4.13: Case study density distribution chart.

4.7.2 Methodology Details

The density modification allowed matrix $\underline{\underline{F}}$ is restricted to areas that met design rule constraints. The dummy density is chosen by using the maximum effective pattern density present

in the original layout. The final density modification structure is implemented in layout. Selective etchback cells are also designed in areas where possible, and included in the final modified layout. Analysis on the final reticle is performed to verify effective density range reduction.

4.7.3 Results

Table 4.2 shows the simulated effects of dummy fill structures with varying densities. Resulting effective pattern density statistics of modified layouts are displayed as density variation, maximum, minimum, and range. Dummy fill structures of 10%, 30%, and 50% are shown.

The variation in effective pattern density is smallest for the 30% density structure, while using the 50% density structure results in a significantly smaller density range reduction. Table 4.3 shows the benefit of using selective etchback rather than conventional etchback.

Table 4.2: Density variation, dummy fill only

Test Case	Density Variation (%)	Max	Min	Mean
Original Layout (with conventional etchback)	30.36	30.46	0.10	9.20
Fill (10%), conventional etchback	22.49	30.46	7.97	13.86
Fill (50%), conventional etchback	28.18	47.83	19.65	35.65
Fill (30%), conventional etchback	**17.90**	**32.14**	**14.24**	**25.79**

Table 4.3: Density variation, comparison using selective etchback

Test Case	Density Variation (%)	Max	Min	Mean
Fill (30%), conventional etchback	17.90	32.14	14.24	25.79
Fill (30%), *selective etchback*	14.40	31.36	16.96	26.05

The final analysis showed that a modified layout using a density modification structure of density 30% would minimize the final effective pattern density range. Implementation of a 30% dummy fill structure, in addition to selective etchback structures of the same density, verified that the effective pattern density range had been significantly reduced.

Note that the final density variation results are larger than what was achieved on a similar layout in Section 4.6. This is due to the fact that the matrix of cells where dummy structures were allowed was further restricted due to design rule constraints, which eliminated additional fill structures which were present in the purely mathematical analysis.

4.7.4 STI Dishing and Erosion Implications

As mentioned, the single layer density analysis technique can be performed without full knowledge of characterization parameters of a CMP process (an estimation of CMP process planarization length may be used). However, it is more useful to see the actual benefit in terms of STI CMP dishing and erosion. A simulation of predicted dishing and erosion can be performed using models described in Chapter 2. For nominal parameters (PL_{OX} 3 mm, K 2000 Å/min, τ_{ox} 20 seconds, PL_{NIT} 6 mm, s 10, τ_2 100 seconds), the dishing and erosion for the original and revised (using selective etchback) layouts are shown in Table 4.4.

Table 4.4: Comparison of Predicted Dishing and Erosion

Test Case (polish time)	Predicted Dishing (Å)			Predicted Erosion (Å)		
	Mean	Min	Max	Mean	Min	Max
Original Layout (146 s)	752	291	1162	381	55	706
Revised Layout (148 s)	435	290	604	116	56	246

The revised layout reduced the average dishing by 42% and average erosion by 70%, verifying that density optimization can provide appreciable improvement over nominal layouts with respect to post-CMP STI dishing and erosion.

4.8 Density Optimization - (Dual Layer Analysis)

We have seen how the use of dummy fill and selective reverse etchback can optimize the density profile of the overburden oxide layer, improving the dishing and erosion that results in the STI CMP process. There is another method that involves analyzing the two layers of an STI

process in succession. The key idea is to reduce the erosion and dishing in the post-CMP films through modification of the density of both the oxide overburden layer and the underlying nitride layer. This process assumes the existence of an original layout for active areas on the die, knowledge of the deposition bias, and the existence of, or the capability to add, a reverse etchback layer. The flowchart procedure is illustrated in Figure 4.14.

The first step involves analysis of the underlying nitride layer. Section 4.2.3 described how erosion and dishing decreased with increasing nitride density. Therefore, to reduce nitride erosion, the first step is to enforce a minimum nitride density specification on the nitride layer. This can be accomplished by dummy filling the nitride area. This may be as simple as picking a single dummy fill structure and arraying it into open regions on the active area layer, or as complicated as customizing dummy fill based on effective nitride density.

The second step involves reduction of the range of the effective overburden oxide density, using a technique similar to the one described in Section 4.4. The main difference is that the input to this step is the modified layout including the dummy structures from the first step of this process. The layout is discretized into cells, and an algorithm is used to find the optimal local density modification (positive or negative) for each cell. The density modification is then used to raise or lower the density in each cell.

For modification to layouts for this second step, density can be raised only (via addition or modification of existing dummy structures) in the areas on the die where dummy fill structures are allowed on the original layout. Fill structures from the first step of this process can only be replaced with fill structures that have a higher density, since this will only raise the underlying nitride effective density (and thus, keep the minimum nitride density criterion valid). Density can be lowered using by creating or modifying a reverse etchback layer. Density can be lowered only in regions of the die that have underlying features, because the etchback process requires an underlying feature to serve as the etch stop.

Careful analysis must be performed to analyze the impact of density modification on the final density of the overburden oxide phase (affects nitride clearing time) and the overpolish phase (affects dishing/erosion). It is of critical importance to note that the effective density of the overburden oxide is not necessarily equal to the effective density of the underlying nitride. Also, certain processing steps that affect the overburden oxide density (e.g., deposition bias, etchback) will not affect the nitride density.

The final result is a revised active area layer and the addition or modification of an etchback layer. These modifications can be incorporated into the original layout layers to produce the final density-optimized layout. Implementation of dummy fill is performed using solid geometries (such as the solid squares shown in Section 4.5) assuming no etchback is used; implementation of reverse etchback is performed by creating selective etchback structures (as in Section 4.3) and arraying them in an etchback layer such that they exist only over active area features within the specified design tolerance for etchback features.

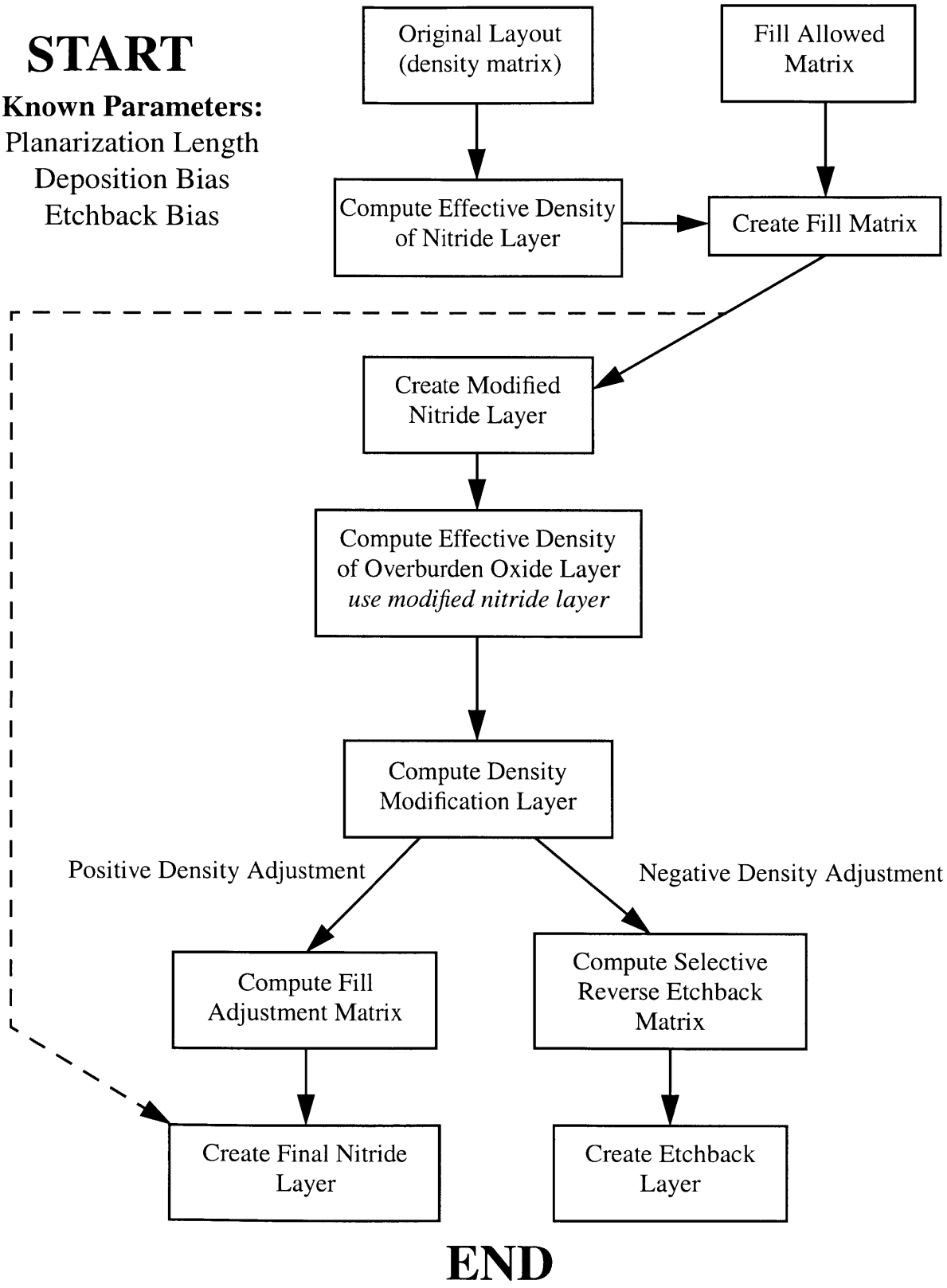
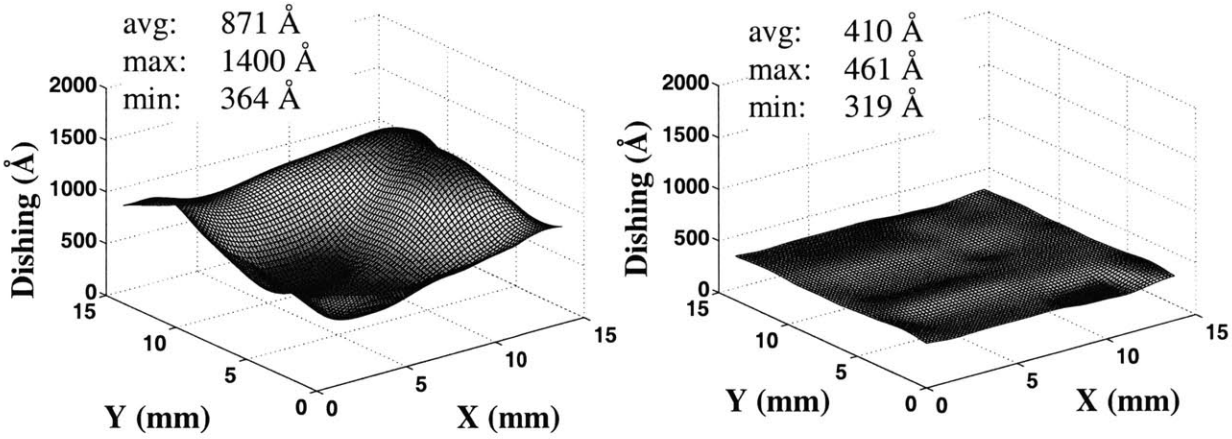


Figure 4.14: Flowchart for density optimization procedure (two layer).

4.9 Case Study

As a case study for the two-layer density optimization methodology, the same layout for the SRAM chip described in Section 4.7 is analyzed. For this simulation, a nominal set of model parameters is used (PL_{OX} 3 mm, K 2000 Å/min, τ_{ox} 20 seconds, PL_{NIT} 6 mm, s 10, τ_2 100 seconds). Predicted dishing and erosion is computed for the original layout, assuming no deposition bias and no etchback layer. Polish time is set to 110% of the polish time guaranteed to clear oxide everywhere on a given layout. Density optimization is then used to add dummy fill structures and a selective reverse etchback layer, and dishing and erosion is simulated for the revised layouts. The results are shown in Figure 4.15 and Figure 4.16. There is a clear reduction in dishing and erosion in the revised layout. In addition, the polish time is reduced from 207 seconds to 182 seconds due to the addition of the reverse etchback layer.



before density optimization **after density optimization**

Figure 4.15: Simulated trench oxide dishing, before and after density optimization.

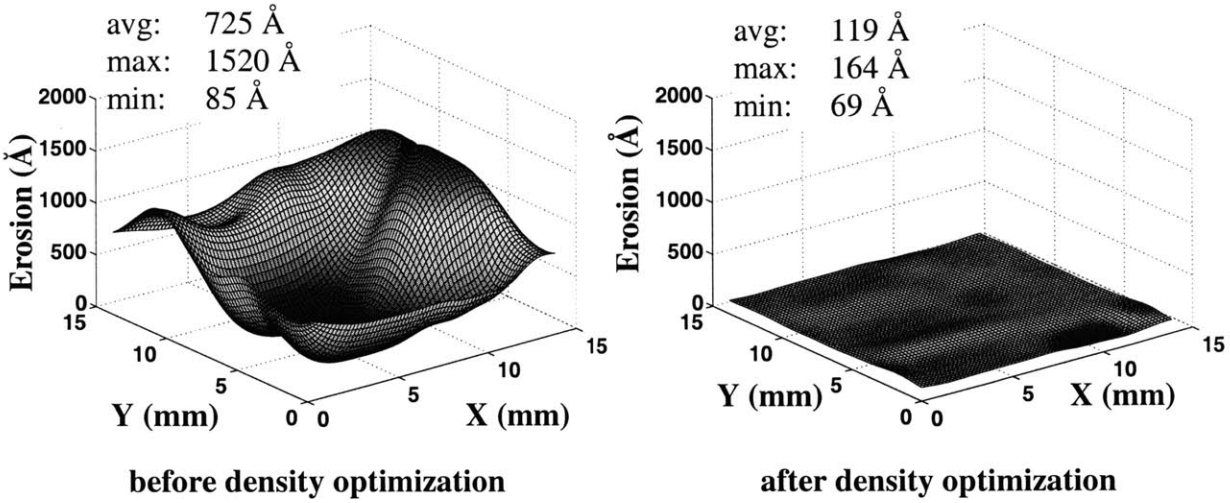


Figure 4.16: Simulated nitride erosion, before and after density optimization.

4.10 Summary

Pattern effects have been shown to strongly impact the CMP process. Nominal layouts can exhibit large effective pattern density ranges, which leads to large post-CMP film thickness variations. It is possible to reduce density variation through the introduction of density modification structures (dummy fill and selective etchback).

The causes of pattern-dependent dishing and erosion have been explained, and two methodologies for reducing these effects have been described. These methods use the addition of density modification structures into arbitrary layouts. One method requires only analysis of the oxide overburden layer (single layer analysis), and the second requires analysis of both the oxide overburden layer as well as the underlying active area layer (dual layer analysis).

An implementation of the single-layer analysis methodology in layout has been performed, with results verifying the validity of the methodology in reducing pattern density range. Potential algorithms for selection of structure densities have been detailed and demonstrated. A case study of the dual-layer analysis methodology has been performed on an actual layout. Simulated dishing and erosion results show considerable improvement over the

nominal layout.

It is critically important to analyze and create density modification structures on a layout by layout basis. This suggests that it is infeasible to create a generic dummy fill structure that can be used on arbitrary layouts, since the density profile statistics of arbitrary layouts can vary considerably from layout to layout. Creating a general (fixed-density) density modification structure for use in all arbitrary layouts will result in suboptimal effective pattern density range reduction since density distributions will vary from layout to layout.

Of equal critical importance when creating density modification structures is the consideration of other factors, including CMP planarization length, deposition bias, inclusion of an etchback step in the process flow (and if so, the value of the etchback bias), and layout design rule constraints. Each of these will affect evaluation of effective pattern density, or evaluation of the optimal density modification structures.

When performed properly, density optimization can be a powerful tool in reducing density-dependent concerns in CMP. This specifically affects STI CMP because it reduces the amount of overpolish time required, and so can reduce of eliminate problems due to trench oxide dishing and nitride erosion.

Chapter 5
Nanotopography

This chapter describes the existence of topography on raw silicon wafers (called *nanotopography*) and its effect on the CMP of films deposited on such wafers. The topography can result in excess thinning of surface films during subsequent CMP steps, which can be of serious concern in shallow trench isolation processing.

Wafer nanotopography is described and illustrated in Section 5.1. Implications of nanotopography on single material and dual material CMP systems are discussed in Sections 5.3 through 5.5. Section 5.6 reviews the previous work on nanotopography. Section 5.7 describes a wafer experiment using various combinations of wafer nanotopography signatures and CMP processes; the results of the experiment are given in Section 5.8. Sections 5.9 through 5.11 describe the development of an accurate nanotopography CMP modeling and characterization procedure which will allow for the proper diagnosis of potential problems due to wafer nanotopography in a given STI process. In addition, metrics for use in the diagnosis of a wafer's ability to yield are presented, and illustrated on a test wafer data set. Section 5.12 describes a simple estimator for the aggregate dishing and erosion effects resulting from wafer nanotopography and chip-scale layout pattern density effects. The chapter is summarized in Section 5.13.

5.1 What is Nanotopography?

Nanotopography describes nanometer-scale height variations that exist on lateral millimeter-scale wavelengths on unpatterned silicon wafers. The characteristics of the variation depend on the specific wafer manufacturing process used to generate a particular wafer. An illustration of wafer nanotopography is shown in Figure 5.1. The height variation is typically 100

nm with a lateral length scale between 1 and 20 mm.

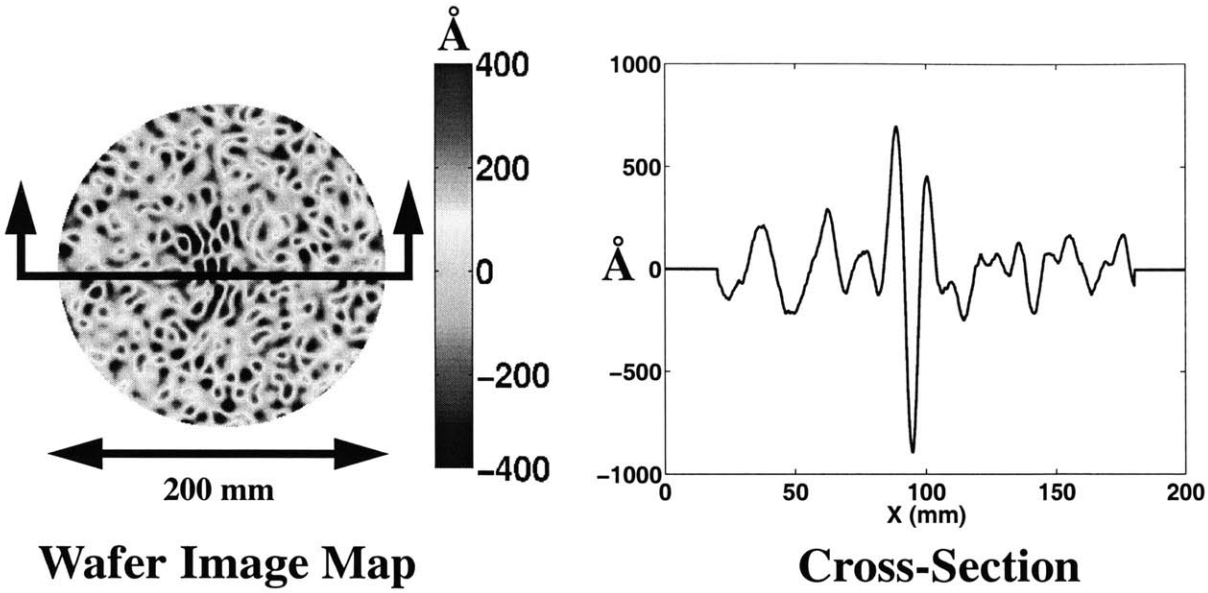


Figure 5.1: Illustration of wafer nanotopography as a wafer map and cross-section.

5.2 Nanotopography Characteristics

Nanotopography characteristics of a wafer depend on the wafer manufacturing process. Generally speaking, it is possible to characterize a wafer nanotopography type by two metrics: the lateral length scale of the variation, and the magnitude of the height variation.

To characterize the height magnitude, the simple metric of standard deviation of the height variation over the total wafer (or some specified region on the wafer) can be used. This metric can be easily computed from measured nanotopography data. Characterization of the length scale of the nanotopography features is more difficult. While it is possible to visually relate length scales of two particular nanotopography signatures, a more rigorous method for describing the length scales is desirable. One method that may be used is power spectral density analysis of the nanotopography data.

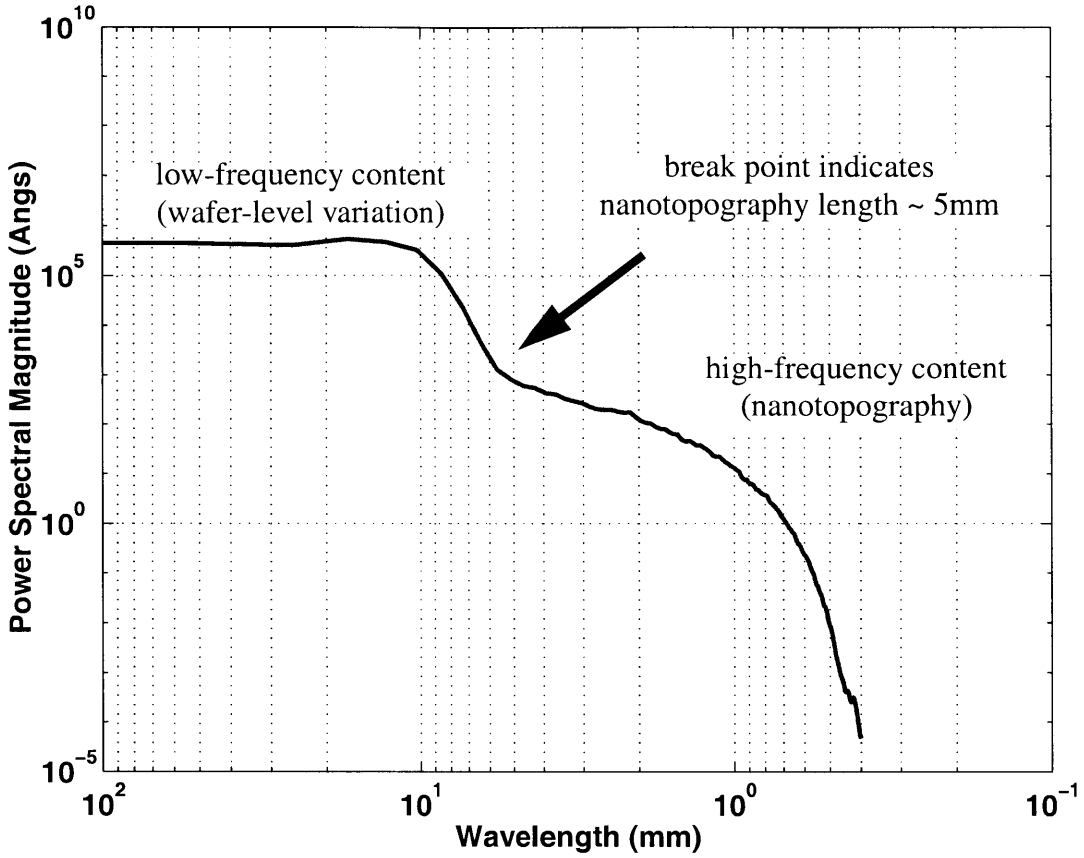


Figure 5.2: Characterizing nanotopography length using power spectral density analysis.

Given the nanotopography height variation data for a particular wafer (or a specified region on a wafer), the one-dimensional power spectral magnitudes of vertical and horizontal cross-sectional data within the dataset can be computed. The power spectral magnitudes can then be averaged and plotted as a function of wavelength. A typical plot of this is shown in Figure 5.2.

We conjecture that nanotopography data contains low frequency (long wavelength) content due to wafer-level CMP variation on the raw wafer, and higher frequency (short wavelength) content due to nanotopography. Analysis of the power spectral density plot for the break point between the short and long wavelength content gives an estimate for the length scale of the wafer nanotopography, or the *nanotopography length*.

5.3 Nanotopography Effects on CMP

Nanotopography is becoming a serious concern in IC fabrication [49]. Films deposited on wafers with certain nanotopography characteristics are conjectured to exhibit post-CMP film thinning, with the relative length scales of the nanotopography and the CMP process serving as a significant factor [58]. This thinning can cause yield concerns in STI processing, as well as possible lithography problems with respect to polysilicon critical dimension printability.

The thinning effect occurs due to the planarization capability of the CMP process over raised areas of material. Thinning is defined here as the preferential removal of material from the raised areas of pre-CMP films. This is illustrated in Figure 5.3.

Thinning of post-CMP oxide films is verified via experiments (described later in this chapter). Analysis of the post-CMP oxide thickness deviation shows a strong visual correlation to the initial wafer nanotopography, as shown in Figure 5.4.

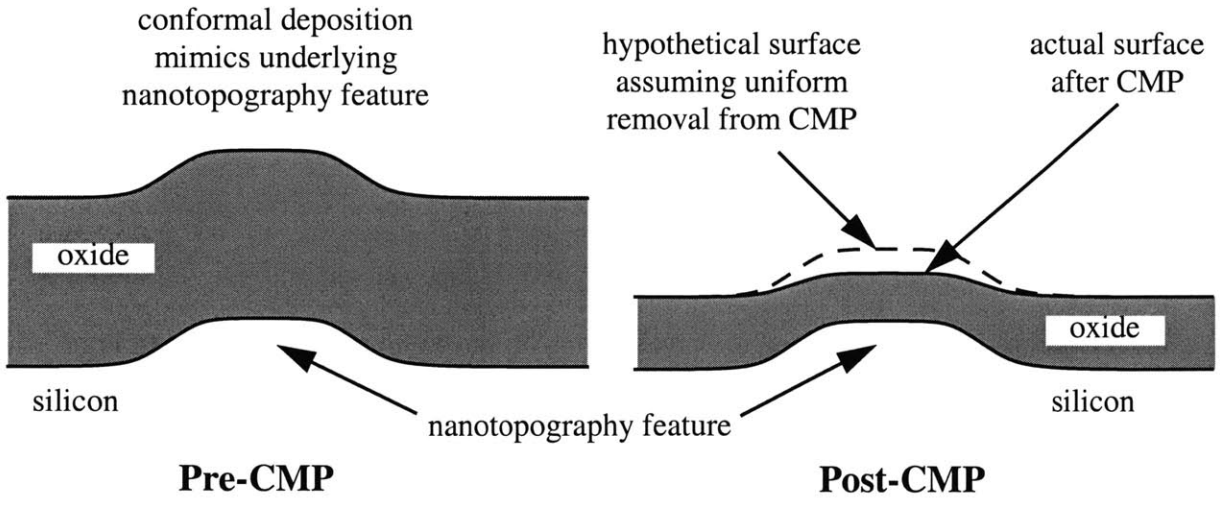


Figure 5.3: CMP of oxide film over nanotopography results in thinning of the surface film.

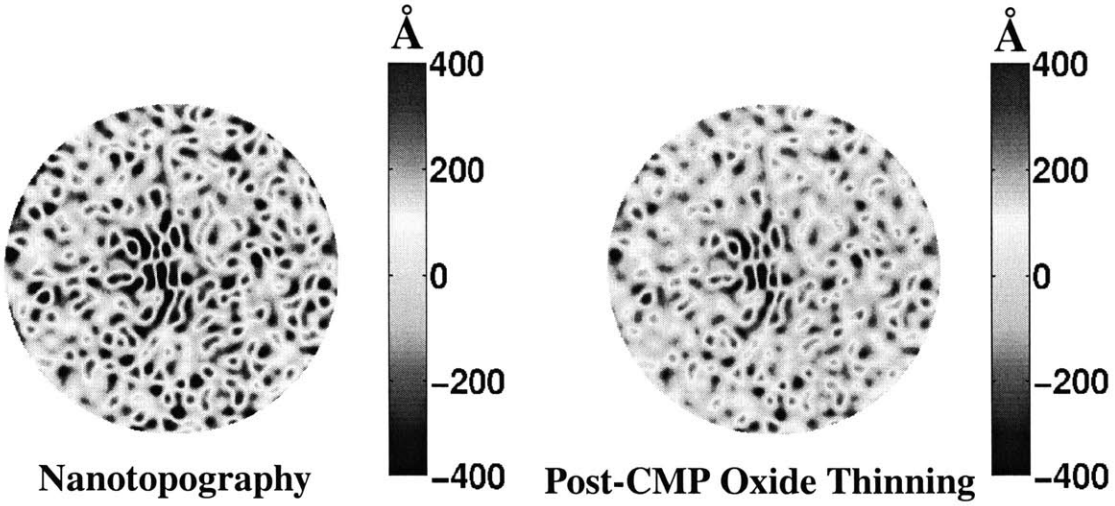


Figure 5.4: Experimental data for nanotopography and post-CMP oxide thickness deviation.

5.4 Effect of Relative Length Scales on Thinning

As mentioned, the thinning effect of nanotopography on post-CMP films occurs due to the planarization capability of the CMP process. The actual degree that a surface film will thin depends on the length scale of the nanotopography features relative to the length scale of the CMP process (planarization length). The thinning effect is directly related to the planarization efficiency of the CMP process; the better the planarization, the larger the thinning. Planarization efficiency is related to the planarization length of the CMP process; longer planarization length processes will result in more planarized post-CMP surfaces.

The planarization length defines the limit to the lateral range over which topography can be removed by the CMP process. Any feature smaller than the planarization length will be planarized given sufficient polish time; features larger than the planarization length will result in only a small amount of planarization. This is illustrated in Figure 5.5.

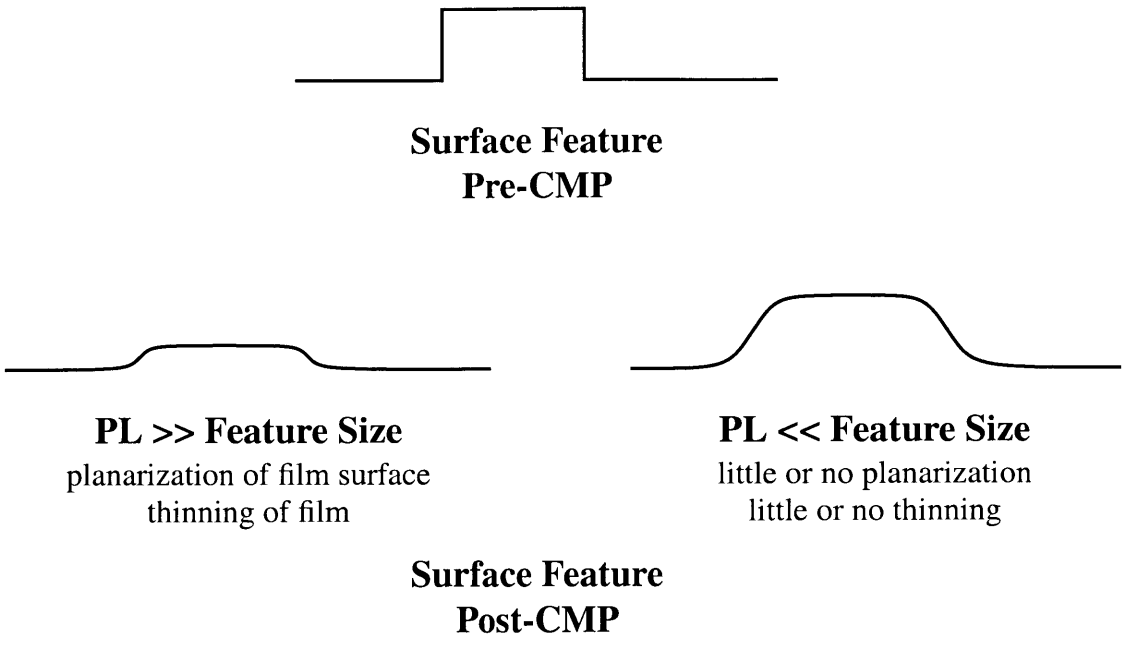


Figure 5.5: Implications of length scales in planarization and thinning effects.

5.5 Specific Concerns for Shallow Trench Isolation

In this thesis, it has already been described how, in STI CMP, pattern density variation can lead to yield concerns due to incomplete clearing of the overburden oxide, dishing of the trench oxide, and erosion of the active area nitride. As has been discussed, density variation problems can be approached in various ways (density modification via dummy fill and selective etchback, nonconventional consumable sets).

In this chapter, the film thinning effect of nanotopography on post-CMP film surfaces has been described. The thinning resulting from nanotopography interaction with CMP can potentially have serious yield implications in STI processing. Section 5.4 has described how the relative length scales of the CMP process and the underlying topography are important in determining thinning. In the case of STI, for certain nanotopography features, this can result in incomplete oxide clearing or excessive nitride erosion. An illustration of this is given in Figure 5.6. This details the case where identical STI devices are built upon underlying nanotopography

feature. In this case, there are two raised regions and one low region. Assuming that the two raised regions are separated by a distance less than the planarization length, the post-CMP polish will create yield problems: if the polish is timed to stop when nitride is first cleared, then the central low region will be uncleared; if the polish time is set to clear oxide in the low region, then the outer regions will experience nitride erosion.

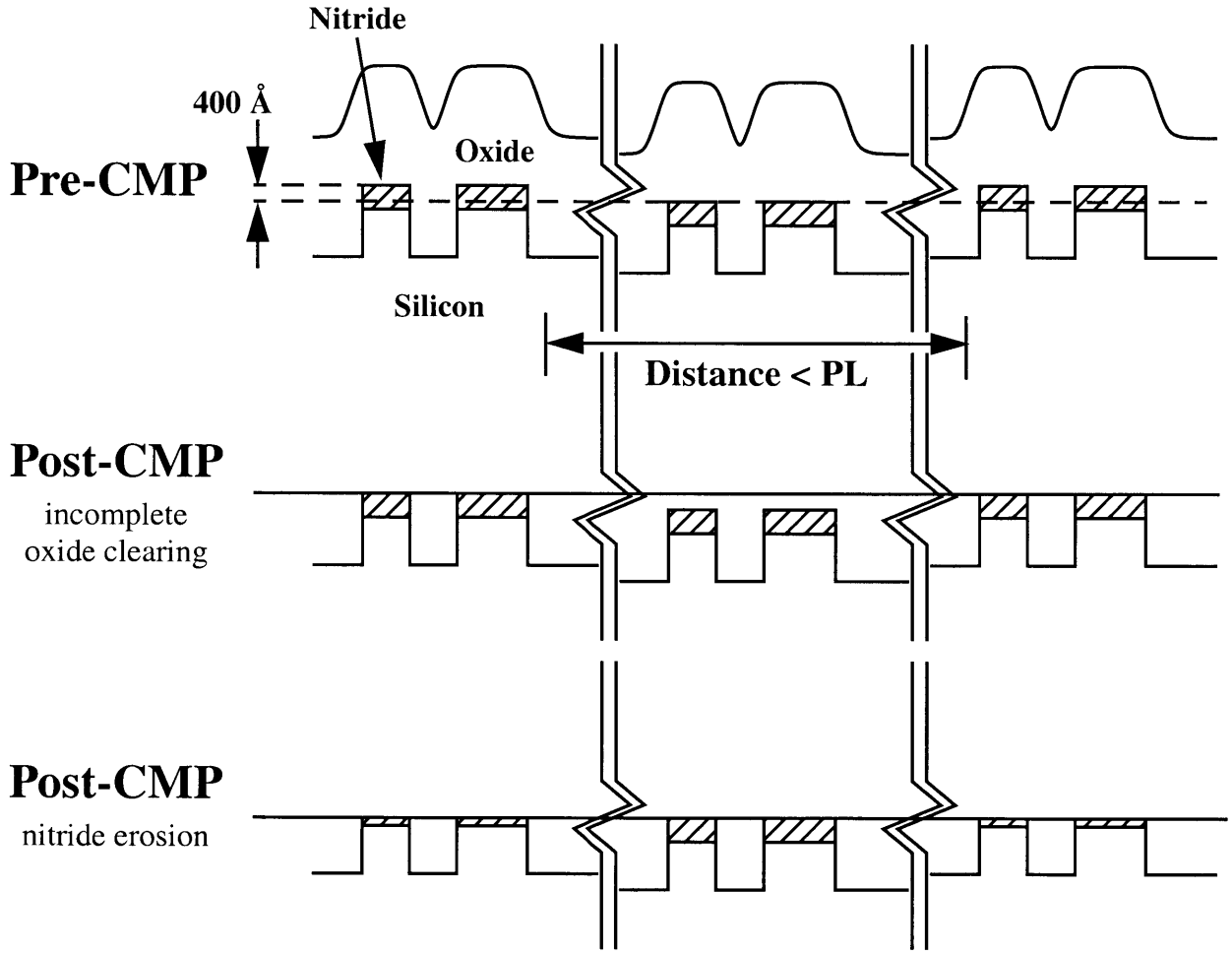


Figure 5.6: Nanotopography concerns for STI processing.

Note that this problem is purely due to nanotopography, and the yield problems from this effect are completely separate from the yield problems from density mismatches. Thus, the solution for density mismatch (e.g., density modification via dummy fill, selective etchback) will not help in this case. Additionally, certain consumable sets used to improve nitride erosion (such as using slurries with high nitride/oxide polish selectivity) will not help in this case, since Figure

5.6 shows that removal of the nitride in the outer two regions is required to clear the oxide in the central region. For a process with a low nitride removal rate, this will simply translate to additional polish time.

This analysis is worst-case since it neglects effects due to pad compressibility [39]. It may be possible to achieve the desired low area polish depending on the consumable set used. In this case, the low area polish is similar to the dishing phenomenon in dual material processes.

Considering the potential detrimental impact of nanotopography on device yields, it is useful to determine possible metrics for evaluation of the nanotopography effect on STI stacked films. Specifically, predictions of possible regions of concern due to potential device failure from incomplete oxide clearing or excessive nitride thinning would be practical. These metrics are described in Section 5.11.

5.6 Literature Review

Wafer nanotopography is a relatively recent topic but has become the subject of much attention. Metrics for wafer specifications of nanotopography have been developed based on maximal allowed peak-to-valley topography within specified analysis areas on the wafer [61] that reports defective areas on the wafer. Alternative metrics have also been proposed for wafer nanotopography characterization [62].

Tsujimura [64] has proposed modeling the effect of nanotopography on CMP process using boundary element methods to predict the post-CMP film thinning as a function of pad properties and CMP process conditions. He also proposes characterizing the goodness (high potential for device yield) of the wafer by specifying permissible peak-to-valley and wavelength combinations (PV-λ) given a specified post-CMP film thickness deviation tolerance, based on an original idea by Fukuda [65]. An example of this is illustrated in Figure 5.7.

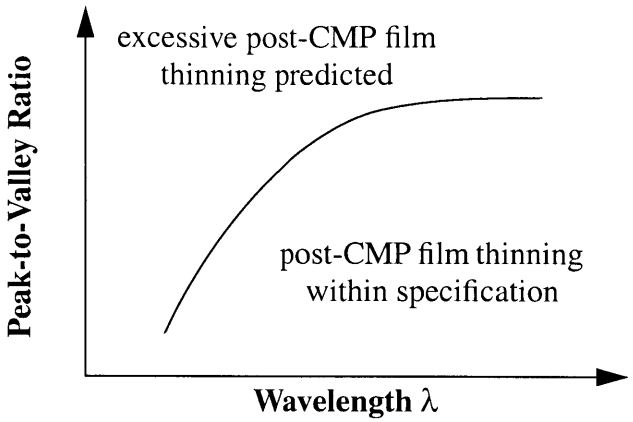


Figure 5.7: PV-λ curve for wafer nanotopography characterization (adapted from [64]).

Each PV-λ curve is specific for a fixed pad stiffness and Tsujimura's analysis, while valid for general wafer characterization, offers no specific identification of problem locations on a wafer, or the exact yield effect that results from problematic peak-to-valley/wavelength combinations.

In addition to attempts at modeling and formulating metrics, extensive experimentation and measurements have been performed in order to observe the effect of nanotopography on the CMP process. Fukuda, in association with JEIDA [65], ran extensive wafer experiments including a large variety of wafer types and CMP processes, and notes that nanotopography has an appreciable influence on post-CMP thickness for CMP processes using stiff pads. This analysis includes a filtering approach to remove wafer-level trends, but does not include any characterization of the CMP processes used (e.g., extraction of planarization length). Fukuda also stresses the necessity of quantitative nanotopography guidelines, and proposes one possible metric of peak-to-valley nanotopography height vs. nanotopography wavelength.

Schmolke investigates the relationship between nanotopography height and post-CMP oxide thickness [63]. He seeks to form a quantitative measure of the correlation between the two quantities using statistical methods. The basic premise is that some contribution to post-CMP oxide thickness deviation results purely from nanotopography, and some contribution results from

oxide deposition and random CMP variation [63]. The statistical analysis technique of auto-correlation and cross-correlation of initial nanotopography height to post-CMP thickness is used to separate the two components.

Auto-correlation is defined as:

$$ACF(\Delta x,\Delta y) = \langle NH(x+\Delta x,y+\Delta y)NH(x,y)\rangle - \langle NH(x+\Delta x,y+\Delta y)\rangle \langle NH(x,y)\rangle \quad (5.1)$$

Cross correlation is defined as:

$$CCF(\Delta x,\Delta y) = \langle OTD(x+\Delta x,y+\Delta y)NH(x,y)\rangle - \langle OTD(x+\Delta x,y+\Delta y)\rangle \langle NH(x,y)\rangle \quad (5.2)$$

where the angled brackets symbolize the average with respect to all positions (x, y) on the wafer. Schmolke considers a small experiment (ten wafers grouped into two nanotopography types) and analyzes the case for Δy of zero. He finds that the auto-correlation and cross-correlation functions show extrema at Δx of zero, and extracting the linear slope of a plot of CCF vs. ACF produces a "nanotopography contribution factor" α that is a measure of the contribution of nanotopography on the post-CMP thickness deviation. Schmolke's analysis does not separate out any wafer-scale CMP effects.

5.7 Wafer Experiment

An experiment has been performed to determine the effect of underlying wafer nanotopography on post-CMP oxide film thickness, and to specifically analyze the results of post-CMP oxide film thinning on various combinations of nanotopography signatures and CMP processes. This experiment uses eight sets of 200 mm epi wafers, each set containing wafers with four different nanotopography signatures. Data from a total of 120 wafers is reported here. Starting wafers have 1 μm of thermal oxide grown on them. Each set of wafers is run on a different CMP process, where process denotes a specific combination of tool, consumable set, and process conditions. Patterned characterization wafers are also run on the CMP processes to determine the planarization lengths for each process. Oxide thickness is measured before and after CMP processing, and the final thickness variation is compared to the initial wafer nanotopography.

5.7.1 Wafer Types

Four wafer types of varying nanotopography characteristics are used in this experiment. An image map illustration of the four wafer nanotopography types is shown in Figure 5.8. The wafer types vary both in lateral length scale, shape and orientation of nanotopography features, and height magnitude. Wafers were provided by MEMC and were specifically manufactured to produce particular nanotopography signatures. For identification purposes, the four wafer types are labelled DSP1, SSP1, SSP2, and SSP3. DSP1 wafers are double-sided polished wafers, while SSP1, SSP2, and SSP3 wafers are single-side polished wafers.

5.7.2 CMP Processes

Eight different CMP process splits are performed. The intent of the splits is to generate processes having a wide variety of planarization lengths. Different CMP tools and consumable sets are explored in order to generate these planarization lengths based on previously reported expectations (e.g, stiffer pads creating longer planarization length processes [36,37]).

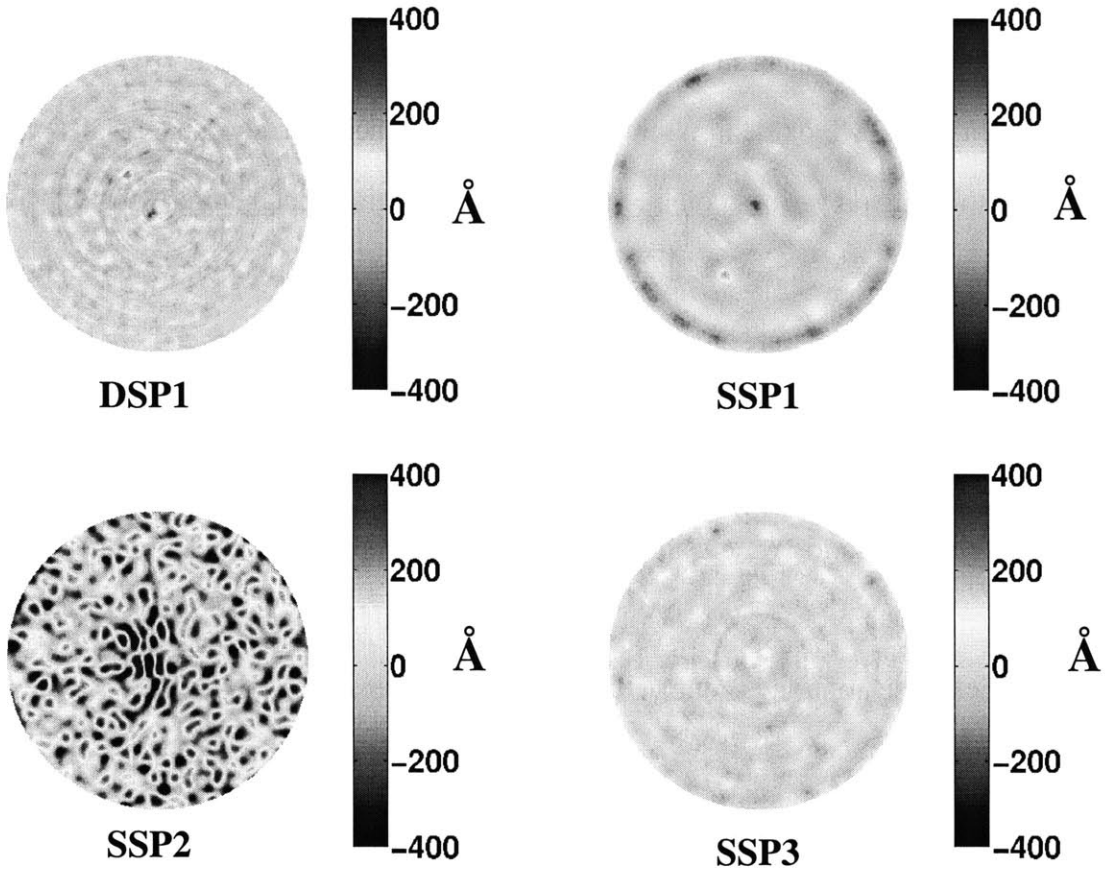


Figure 5.8: Wafer nanotopography signatures used in this experiment.

5.7.3 Metrology

Wafer nanotopography is measured using a Nanomapper^{TM} tool from ADE Phase-Shift. The technology used is a large area, direct height measuring, white light interferometry system [67]. Wafer data is measured and filtered with a 20 mm double gaussian filter to produce the final output data.

Pre-CMP and post-CMP oxide film thickness is measured using an AcuMap^{TM} film thickness measurement system from ADE Corporation. The AcuMap tool is a high speed film mapping metrology tool based on single point optical reflectometry that is capable of scanning the entire wafer at one time [68].

Wafer measurement are taken at nominal resolution, which is 0.905 mm x 0.905 mm for

the oxide thickness data, and 0.197 mm x 0.227 mm for the nanotopography data. A subset of wafers is also measured at higher resolution for post-CMP oxide thickness. The resolution in this case is 0.222 mm x 0.222 mm.

5.7.4 Post-Measurement Data Processing

The data sets are interpolated onto a common grid spacing of 0.2 mm x 0.2 mm, and then filtered with a 30 mm double-gaussian filter in order to remove wafer-level CMP effects. Finally, an alignment algorithm is used to correct for lateral (in both X and Y direction) and rotational offsets between the nanotopography and oxide thickness data sets. The algorithm is based on the assumption that the correct alignment produces the maximal correlation between the initial nanotopography data and the post-CMP oxide thickness deviation. It consists of a search over small lateral X-Y displacements to find the lateral correction, and a search over small rotational angles to determine the rotation correction. In most cases, the nanotopography and oxide thickness data are mean-centered.

5.8 Experiment Results

The total number of wafers used in this experiment is given in Table 5.1. In some cases, wafer measurement data are not included in the analysis due to the presence of wafer defects which may skew the analysis results. A total of 120 wafers are analyzed in this study. Planarization lengths of the CMP processes are extracted from measured data from CMP characterization wafers using the density-step height model [35]. Rather than study the effect of individual tool and process conditions on the CMP of nanotopography, we are interested in the relationship between planarization length and film thinning; the planarization lengths for the different process conditions are listed in Table 5.2.

Table 5.1: Wafer Counts

CMP Process ID	DSP1	SSP1	SSP2	SSP3	Total by Process
1	2	4	6	4	16
2	2	4	4	4	14
3	2	5	5	4	16
4	2	4	5	5	16
5	2	4	5	4	15
6	3	4	5	4	16
7	2	4	4	4	14
8	2	4	3	4	13
Total by Wafer Type	17	33	37	33	120

Table 5.2: CMP Process Planarization Lengths

CMP Process ID	Planarization Length (mm)
1	1.9
2	3.1
3	3.4
4	4.6
5	6.4
6	8.4
7	9.7
8	13.7

Initial wafer nanotopography height measurement data is compared to mean-centered oxide removed (MCOR) data. Measurement data is obtained at nominal resolution, with interpolation and alignment steps performed as described in Section 5.7.4. For the purposes of experimental data analysis, only the central 100 mm x 100 mm section of wafer data is used, in order to avoid any effects due to proximity to the wafer edge. In comparing the two datasets, two characteristics are studied: the similarity of shapes, and the degree of the height transmission from the

nanotopography to MCOR data. To that end, two metrics are used for characterizing the nanotopography effect: the first is a shape correlation coefficient, and the second is the ratio of the standard deviation of the MCOR data to the standard deviation of the nanotopography data. Both metrics are important to determine the impact of nanotopography on post-CMP film thickness. A graphical illustration of the two metrics is shown in Figure 5.9.

The correlation coefficient (c) is computed using the following formula:

$$c = \frac{\sum_i \sum_j (N_{ij} - \mu_N)(O_{ij} - \mu_O)}{\sigma_N \sigma_O} \tag{5.3}$$

where N is the nanotopography height, O is the MCOR, μ is the mean, and σ is the standard deviation. The correlation coefficient measures the aggregate effect of a point-by-point comparison of the offset from the respective means of the two datasets. The scale of correlation coefficients is from -1 to 1, with -1 indicating perfect anticorrelation, and 1 indicating perfect correlation. The standard deviation ratio (r) is computed using the following formula:

$$r = \frac{\sigma_O}{\sigma_N} \tag{5.4}$$

The ratio r gives a metric for the absolute magnitude of the height propagated from the nanotopography to the MCOR.

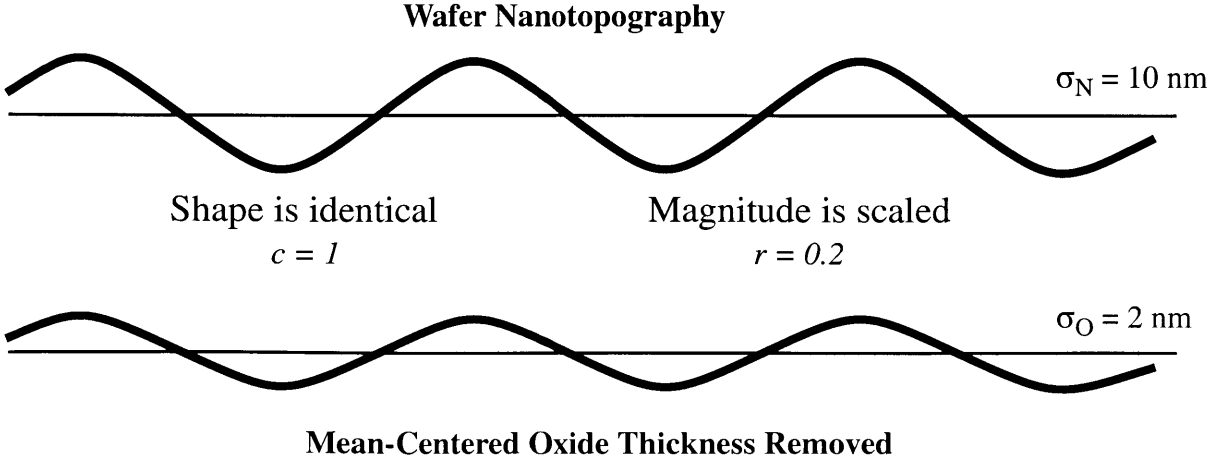


Figure 5.9: Illustration of metrics for nanotopography propagation.

Correlation coefficients and standard deviation ratios are computed for each wafer, and the results are then grouped by wafer type and by CMP process. Correlation coefficients and standard deviation ratios are shown in Table 5.3 and Table 5.4.

Table 5.3: Average Correlation Coefficients

CMP Process ID	DSP1	SSP1	SSP2	SSP3	Average by Process
1	0.09	0.10	0.47	0.12	0.24
2	0.17	0.12	0.43	0.15	0.23
3	0.33	0.17	0.68	0.29	0.38
4	0.10	0.12	0.47	0.12	0.22
5	0.63	0.48	0.84	0.63	0.66
6	0.36	0.51	0.89	0.46	0.59
7	0.73	0.81	0.95	0.82	0.84
8	0.62	0.57	0.92	0.61	0.67
Average by Wafer Type	0.38	0.35	0.69	0.39	

Table 5.4: Average Standard Deviation Ratio

CMP Process ID	DSP1	SSP1	SSP2	SSP3	Average by Process
1	0.36	0.33	0.07	0.34	0.24
2	0.76	0.78	0.19	0.79	0.61
3	0.45	0.42	0.15	0.36	0.33
4	1.35	1.27	0.28	1.12	0.93
5	0.52	0.37	0.34	0.38	0.38
6	1.27	1.25	0.64	1.14	1.03
7	0.88	1.01	0.76	0.86	0.88
8	1.00	1.00	0.7	1.02	0.95
Average by Wafer Type	0.85	0.79	0.37	0.76	

In general, the correlation coefficients and standard deviation ratios both appear to increase with increasing planarization length, and also vary significantly with wafer type. This agrees with our initial hypothesis that the amount of surface film thinning is dependent on both the wafer nanotopography length and the CMP process planarization length.

5.9 Modeling Approaches

In general, it is possible to abstract the CMP process as a (potentially nonlinear) transfer function, which converts initial wafer nanotopography data to final post-CMP film thickness. There are many possible forms that can be used for this transfer function, varying in implementation or computational complexity. A diagram of the abstract system is shown in Figure 5.10.

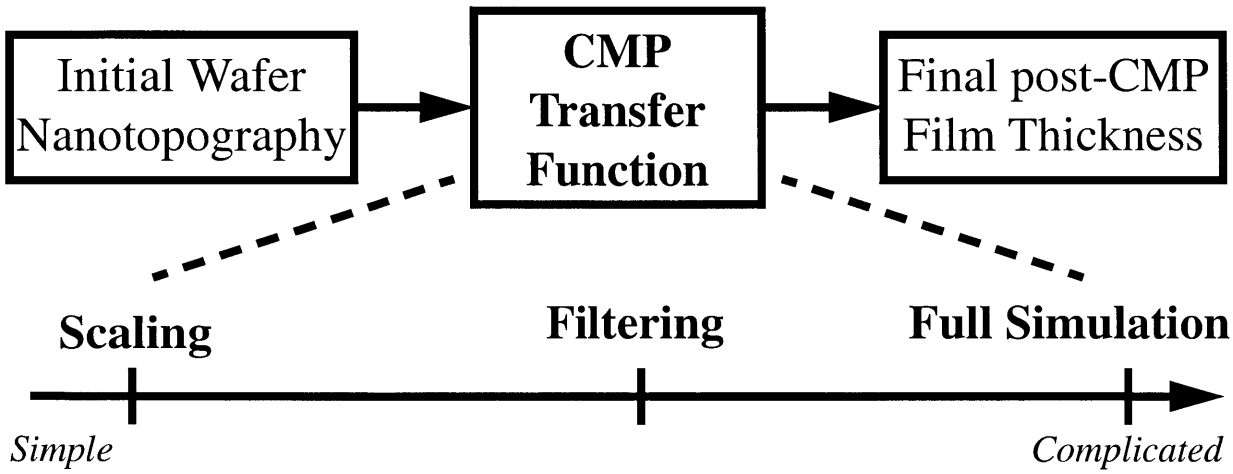


Figure 5.10: Forms for the CMP transfer function.

Three forms are included in this analysis: linear scaling, filtering, and full contact-mechanics based simulation. Each of the three will be described, and techniques for implementation of these models will be detailed. Two methods on opposite ends of the complexity spectrum will be examined initially, followed by a third method of moderate complexity. As a case study, each method will be performed on actual wafer nanotopography data, with the results compared to post-CMP film thickness data. The two components to be compared will be initial nanotopography and post-CMP oxide thickness deviation (OTD). This quantity is the same as mean-centered

oxide thickness removed (MCOR).

5.9.1 Scaling

In the first approach, the original nanotopography data is linearly scaled to produce the prediction of oxide thickness deviation. One prevailing idea of this work is that oxide on top of nanotopography features will preferentially be removed during the CMP process, and thus will thin compared to the average polish thickness. The degree of thinning will be assumed proportional to the original nanotopography. This approach is based on work by Schmolke [63] which is reviewed in Section 5.6. He proposes that the standard deviation of the post-CMP oxide thickness can be decomposed into a scaled portion due to nanotopography, and a "random" contribution from the oxide deposition and CMP process. This second component is due to oxide thickness variation uncorrelated to the initial nanotopography (OTD_{UNC}). This can be described as:

$$OTD = OTD_{UNC} + \alpha \cdot NH \tag{5.5}$$

where NH is the nanotopography height, and α is the scaling factor. The model is calibrated by obtaining the α parameter from experimental data.

$$\begin{aligned} \langle OTD \cdot NH \rangle &= \langle [OTD_{UNC} + \alpha \cdot NH] \cdot [NH] \rangle \\ &= \langle OTD \cdot NH \rangle + \alpha \langle NH \cdot NH \rangle \\ &= 0 + \alpha [\sigma_{NH}]^2 = \alpha [\sigma_{NH}]^2 \end{aligned}$$

$$\alpha = \frac{\langle OTD \cdot NH \rangle}{[\sigma_{NH}]^2} \tag{5.6}$$

where the angled brackets indicate the computation of the expectation (or average) of the quantity enclosed over spatial coordinates x and y. The final expression for α in Equation 5.6 can be computed on an individual wafer basis using pre-CMP nanotopography maps $NH(x, y)$ and post-CMP oxide thickness deviation maps $OTD(x, y)$. The functional dependence of α on wafer type

and CMP process is unknown. The resulting estimator for oxide thickness deviation is then:

$$OTD_{EST} = \alpha \cdot NH \tag{5.7}$$

5.9.2 Contact Wear Modeling

The second method for oxide thinning prediction is a physical simulation of the CMP process. This method involves using a contact-mechanics based model described by Chekina [51] and Yoshida [52]. This model formulates equations based on the physics of the contact between the wafer and pad (illustrated in Figure 5.11). The underlying physics behind the model is a fundamental relationship between local pressures on the pad and the resulting pad displacement, given in Equation 5.8 (taken from [51]).

$$w(x, y) = \frac{(1-\nu^2)}{\pi E} \int_{\omega} \frac{p(\xi, \eta)}{\sqrt{(x-\xi)^2 + (y-\eta)^2}} d\xi d\eta \tag{5.8}$$

where $w(x, y)$ is the displacement of the pad at point (x, y) on the wafer surface, $p(x, y)$ is the local pressure of the pad, ν is the Poisson's ratio of the pad, E is the Young's modulus of the pad, and ω is the simulation area. Local pressures can be used to derive removal rates via Preston's equation [34]. As the film surface evolves through time, the pad displacement changes as it contacts the evolving film surface, and local pressures and removal rates are modified. A typical implementation of the model involves discretizing the simulation into boundary elements, solving for pressures, computing the amount removed over a time step, updating the pad displacement, computing the new local pressures, and iterating for the duration of the polish time.

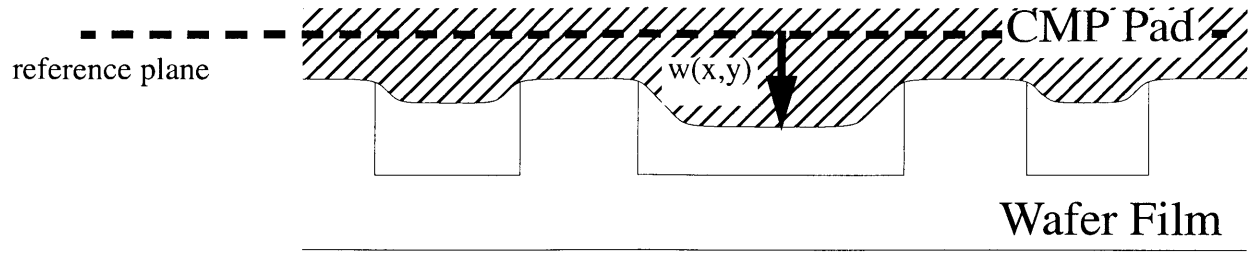


Figure 5.11: Wafer-pad contact diagram.

In the general case, wafer pad displacement and pad pressures are not fully known. Yoshida [52] offers a matrix manipulation algorithm that can be used to solve the matrix form of Equation 5.8. This method is accurate in computing pressures and displacement, but suffers from long computation times for large simulation domains. An alternative approach may be used in order to increase computation speed is to use Fast Fourier Transform (FFT) analysis. Equation 5.8 can be restated as a convolution between pressure and a kernel function, $k(x, y)$, representing the influence of local pressures on the pad displacement of neighboring regions on the pad.

$$w(x, y) = \frac{(1-\nu^2)}{\pi E}[p(x, y) \otimes k(x, y)] \tag{5.9}$$

$$k(x, y) = \frac{1}{\sqrt{x^2 + y^2}} \tag{5.10}$$

Equation 5.9 becomes a multiplication after a Fourier transform. Rearranging gives a solution for pressures that can be computed from displacement:

$$W = \frac{(1-\nu^2)}{\pi E}[P \cdot K] \tag{5.11}$$

$$P = \frac{\pi E}{(1-\nu^2)} \cdot \frac{W}{K} \tag{5.12}$$

where P, W, and K are the respective frequency domain representations for the $p(x, y)$, $w(x, y)$, and

$k(x, y)$ functions after a Fourier transform. The fundamental assumption behind using the FFT solution technique is that the pad contacts the wafer at every point on the wafer surface, and therefore the pad displacement $w(x, y)$ at every point is fully specified. Using the time-stepped algorithm approach, Equation 5.12 can then be used to compute local pressures and removal rates at each time interval. Oxide thickness deviation can be predicted by performing a contact-wear simulation using initial wafer nanotopography data to simulate the oxide thickness removed over the wafer during the CMP process, and then mean-centering the result. The key model calibration parameter is the Young's modulus E, which can be extracted from experimental data. E is actually an effective Young's modulus that may not be the same as the E of the polishing pad that can be obtained through material property testing.

5.9.3 Filtering

The third approach involves using a filter of the appropriate form to process the initial wafer nanotopography data to get an approximation of the final post-CMP film thickness. The filtering approach has been used to mimic the effects of a CMP process. Ouma [30] notes that an elliptical weighting filter can be effectively used to evaluate the pattern density "seen" by a CMP process as it polishes a wafer.

In this application, the original nanotopography data is convolved with a two-dimensional filter and scaled to obtain the final result. Fast Fourier transform (FFT) analysis is used to expedite the computational process. In the contact-wear simulation, FFTs are used to compute the pressures at each time interval in a time-stepped algorithm. In this approach, a single filter step is used to predict final oxide thickness deviation. The filter used is a double-Gaussian implemented in the frequency domain:

$$H(\omega) = 2e^{-k\left(\frac{\omega}{\omega_{cp}}\right)^2} - e^{-2k\left(\frac{\omega}{\omega_{cp}}\right)^2} \quad (5.13)$$

$$\omega_{cp} = \left(\frac{1}{1.331}\right)\frac{2G}{L_c} \quad (5.14)$$

where G is the discretization length, L_c is the filter cutoff length, and k is a constant equal to ln(2). The filter is computed in the frequency domain, multiplied by the transformed nanotopography data, and converted back to the spatial domain using inverse FFT. The resulting prediction is then scaled by a scaling factor in order to get the final prediction.

There are two key parameters used to specify this CMP transfer function: the cutoff length L_c that is used to define the filter shape, and the scaling factor s that is used to scale the filtered nanotopography data to produce the final result.

5.10 Simulation Results

For this analysis, wafers from three CMP processes (with planarization lengths of 3.4, 6.4, and 9.7 mm). Each of the four wafer types is also included in this study. Two wafers from most of the combinations of CMP processes and wafer types are included. To avoid edge effects which might complicate the analysis, only the central 100 mm x 100 mm region of the wafer is used in this analysis

5.10.1 Scaling

Scaling parameter α is computed for each data set using measured data and Equation 5.6. The results are shown in Table 5.5. The scaling parameter is fairly consistent within replicate wafers for the same wafer type and process combination. However, the extracted α value seems to vary from wafer type to wafer type.

Table 5.5: Computed Scaling Coefficients

Process	Planarization Length (mm)	Wafer Type	α	
			Wafer 1	Wafer 2
3	3.4	SSP1	0.07	0.08
3	3.4	SSP2	0.10	0.14
3	3.4	SSP3	0.13	0.11
5	6.4	DSP1	0.38	0.44
5	6.4	SSP1	0.19	0.17
5	6.4	SSP2	0.28	0.27
5	6.4	SSP3	0.27	0.25
7	9.7	DSP1	0.72	0.73
7	9.7	SSP1	0.84	0.85
7	9.7	SSP2	0.74	0.72
7	9.7	SSP3	0.71	0.69

5.10.2 Contact Wear

Using initial nanotopography data as an input to the contact wear model, the optimal value of E is determined by minimizing the root mean square (RMS) error between the final prediction oxide thickness deviation and the actual measured data. The polish time used in the simulation is set such that the average simulated film removal is approximately equal to the average measured film removal. Extracted values of E are given in Table 5.6. The values of E are fairly consistent at short and medium planarization lengths. At longer planarization lengths, E tends to vary from wafer type to wafer type, but is consistent within wafer type. In addition, E tends to increase with increasing planarization length, which fits in with expected relationships.

Table 5.6: Optimal Contact Wear Modeling Parameters

Process	Planarization Length (mm)	Wafer Type	E (MPa)	
			Wafer 1	Wafer 2
3	3.4	SSP1	20	25
3	3.4	SSP2	20	30
3	3.4	SSP3	30	25
5	6.4	DSP1	75	75
5	6.4	SSP1	60	50
5	6.4	SSP2	75	60
5	6.4	SSP3	60	60
7	9.7	DSP1	250	260
7	9.7	SSP1	900	800
7	9.7	SSP2	300	280
7	9.7	SSP3	375	360

5.10.3 Filtering

Optimal parameters for the filter based CMP model (filter cutoff length L_c and scaling parameter s) are obtained by varying the two model parameters simultaneously to minimize the RMS error between the predicted and measured oxide thickness deviation. The extracted parameters are shown in Table 5.7. Values of L_c are fairly consistent within processes. The long cutoff lengths obtained when extracting model parameters for Process 7 (the longest planarization length process analyzed in this study) in most cases are equal or close to the size of the simulation domain. This indicates that the best approximation for this process may be a simple scaling, since filtering with such long cutoff lengths results in only small modification to the original data. In such cases, the scaling parameter s is equal or close to the model parameter α obtained using the scaling method (see Table 5.5).

Table 5.7: Optimal Filter Modeling Parameters

Process	Planarization Length (mm)	Wafer Type	L_c(mm)		s	
			Wafer 1	Wafer 2	Wafer 1	Wafer 2
3	3.4	SSP1	8	9	0.45	0.42
3	3.4	SSP2	12	9	0.20	0.35
3	3.4	SSP3	7	100	0.48	0.10
5	6.4	DSP1	11	11	0.66	0.66
5	6.4	SSP1	18	16	0.43	0.44
5	6.4	SSP2	14	14	0.48	0.46
5	6.4	SSP3	12	13	0.56	0.51
7	9.7	DSP1	100	100	0.72	0.73
7	9.7	SSP1	92	100	0.85	0.87
7	9.7	SSP2	100	100	0.74	0.72
7	9.7	SSP3	100	61	0.71	0.71

5.10.4 Model Comparison

To compare the predictions of the three modeling techniques, the RMS error between the model predictions and measured data is computed. The results are summarized in Table 5.8. In all cases, the RMS error of the predictive models is less than the standard deviation of the measured oxide thickness deviation, which indicates that the methods used capture at least some degree of the relationship between nanotopography and oxide thickness deviation. There may be a 1-2 nm random post-CMP oxide thickness deviation (similar to OTD_{UNC} in Equation 5.5) that establishes a noise floor for all CMP nanotopography models. The RMS error for the filtering method is slightly smaller than the error from the scaling method, except for the longest planarization length process where it essentially equivalent. The contact wear modeling accuracy relative to that of the filtering method varies from process to process. For process 7, the contact

wear does slightly worse than the filtering in terms of RMS error. For the other processes in this study, the contact wear and filtering methods perform comparably. This suggests that filtering may be adequate in cases where the CMP of a single film is to be predicted. On the other hand, the contact wear model is useful because it can be applied to more complicated CMP situations, such as polishing of multiple material film stacks (such as those used in STI processing). A cross section comparison of the predictions of the three oxide thickness deviation models for one example wafer/process combination (Process 3, SSP2) is given in Figure 5.12.

Table 5.8: RMS Error Comparisons

					RMS Error (nm)					
Process	PL (mm)	Wafer Type	σ_{OTD}(nm)		Scaling		Contact Wear		Filtering	
			Wafer 1	Wafer 2	Wafer 1	Wafer 2	Wafer 1	Wafer 2	Wafer 1	Wafer 2
3	3.4	SSP1	1.20	1.70	1.18	1.67	1.15	1.65	1.13	1.65
3	3.4	SSP2	1.87	3.62	1.27	2.30	1.13	1.67	1.16	1.57
3	3.4	SSP3	1.16	1.08	1.08	1.00	1.01	0.94	1.00	1.01
5	6.4	DSP1	2.01	1.70	1.23	1.03	0.97	0.88	0.90	0.82
5	6.4	SSP1	1.05	1.04	0.81	0.85	0.74	0.79	0.77	0.81
5	6.4	SSP2	5.12	4.30	2.11	2.01	1.47	1.45	1.50	1.42
5	6.4	SSP3	1.38	1.36	0.94	0.96	0.84	0.87	0.83	0.90
7	9.7	DSP1	3.51	3.21	1.91	1.95	1.98	2.04	1.91	1.96
7	9.7	SSP1	3.42	3.38	1.98	1.39	2.04	1.44	1.98	1.40
7	9.7	SSP2	13.12	11.11	2.17	2.00	2.79	2.64	2.18	2.01
7	9.7	SSP3	2.82	2.90	1.49	1.45	1.58	1.55	1.49	1.43

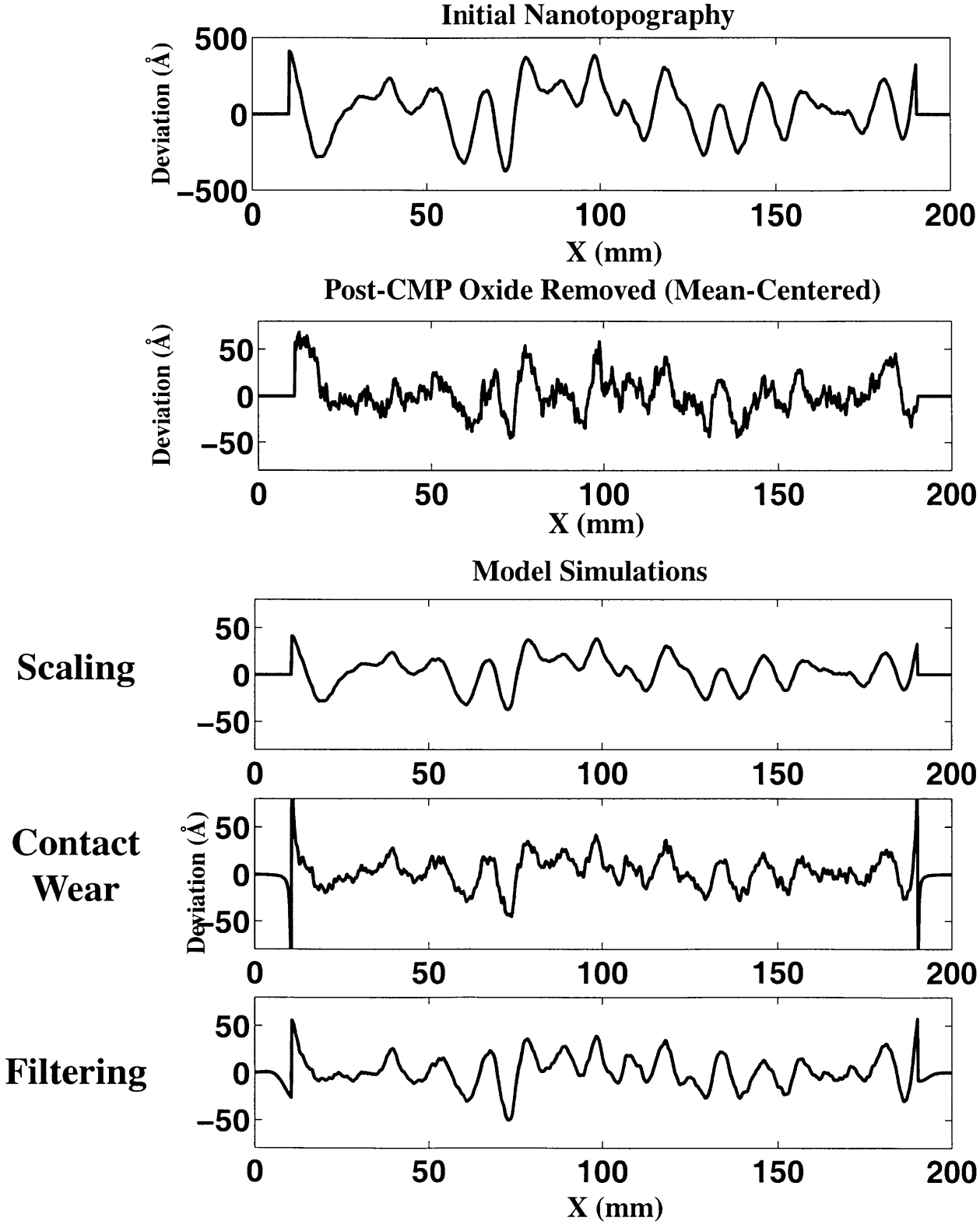


Figure 5.12: **Comparison of the predictions of the three CMP modeling methods.**

Another method of comparing the three modeling techniques is to compute the fractional variance of the oxide thickness deviation that is captured by each model:

$$R^2 = \frac{\sigma^2_{model}}{\sigma^2_{OTD}} = 1 - \frac{\sigma^2_{err}}{\sigma^2_{OTD}} \qquad (5.15)$$

The computed R^2 values are plotted for each of the modeling methods for the SSP2 case for three CMP processes in Figure 5.13. Two wafers are plotted for each CMP process. For the short and medium planarization length processes, the filtering and contact wear methods exhibit larger R^2 than the scaling method. At the longest planarization length process, the scaling exhibits a slightly higher R^2 and the filtering and contact wear models, although the difference between the three methods is small. The R^2 values increase as the planarization length increases, which implies that the modeling of the OTD becomes more accurate with longer planarization lengths.

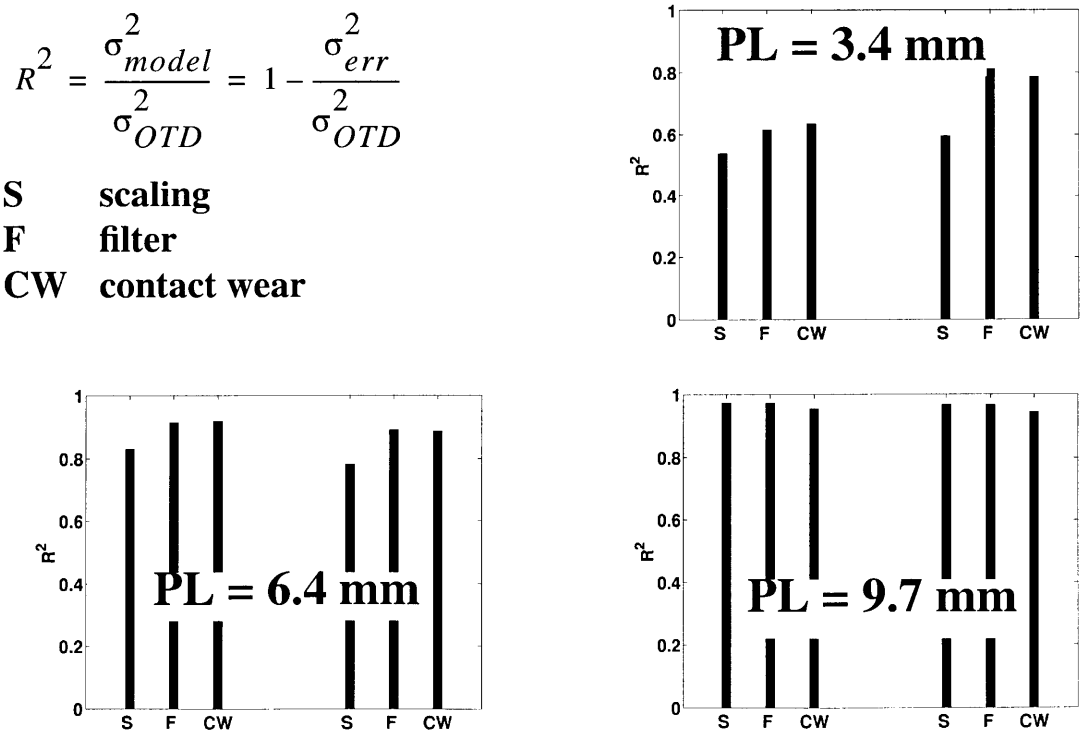


Figure 5.13: Model comparisons - fractional OTD variance.

5.11 Nanotopography Modeling for STI

The key motivation behind the development of accurate CMP modeling methods is to identify potential yield problems in STI CMP. Section 5.5 has illustrated the two major concerns of incomplete oxide clearing and excessive nitride erosion that occur due to nanotopography effect. The critical issue is whether an incoming wafer with a particular nanotopography signature will give rise to potential device failure, and to identify these failure locations on the wafer.

The modeling of an STI process requires modeling of a dual material system. The contact wear modeling methodology described in Section 5.9.2 can be used for modeling of an STI process on a wafer with underlying nanotopography. To model a dual material system, the contact wear model needs only a slight modification. The system is set up as an underlying nanotopography surface with two conformal films (silicon nitride, then silicon dioxide) deposited on top. The computation of pressure/displacement is exactly the same as for a single material CMP system, and a time-stepped algorithm is still used. The amount of remaining top surface film is tracked at each time interval along with the local pressures and pad displacements. In computing the local removal rates at each time step in the algorithm, if the top surface film still remains, then the nominal top surface film polish rate is used; otherwise, the underlying film (in this case, silicon nitride) polish rate is used. Boundary elements are advanced in time, and the algorithm iterates as usual until the final polish time is reached. This is illustrated in Figure 5.14.

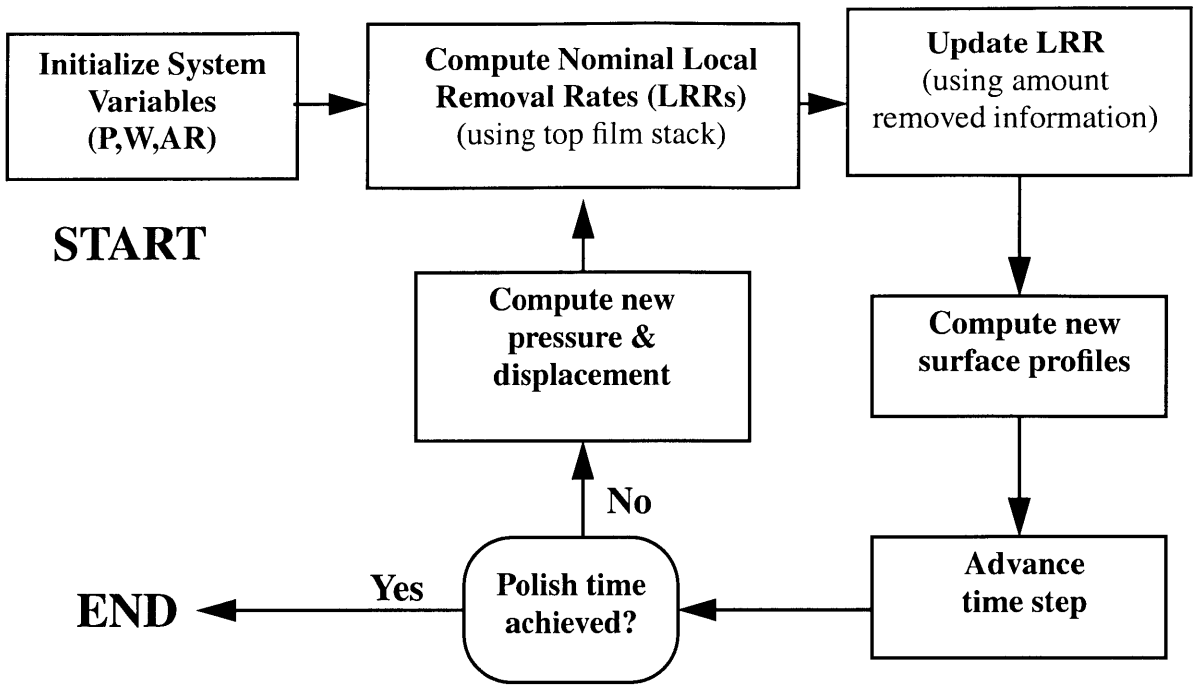


Figure 5.14: Revised contact-wear model algorithm for dual-material system.

Given a modeling technique for two stacked films on nanotopography, an STI process on an unpatterned wafer can be simulated to determine points of incomplete oxide clearing and excessive nitride erosion that is solely due to nanotopography. Predictive maps of oxide/nitride removed during CMP are used to determine potential failure points.

5.11.1 Incomplete Clearing

A simulated CMP polish is performed using initial nanotopography data from an SSP2 wafer. The simulation assumes that the initial wafer surface is deposited with 1000 Å of silicon nitride followed by 3000 Å of silicon dioxide. The dual-material stacked film contact-wear CMP model is used to simulate the CMP process. First, a polish is simulated where the polish time is set by simulating polish until the first point on the wafer clears (84 seconds), and polishing for a nominal overpolish time of 10 seconds. The total amount removed from the CMP process is determined, and areas of incomplete clearing are found by examining the amount removed wafer map for points with amounts of film removal less than the original top surface film deposition.

These points are identified in a third wafer map. This is the incomplete clearing map generation process, as illustrated in Figure 5.15.

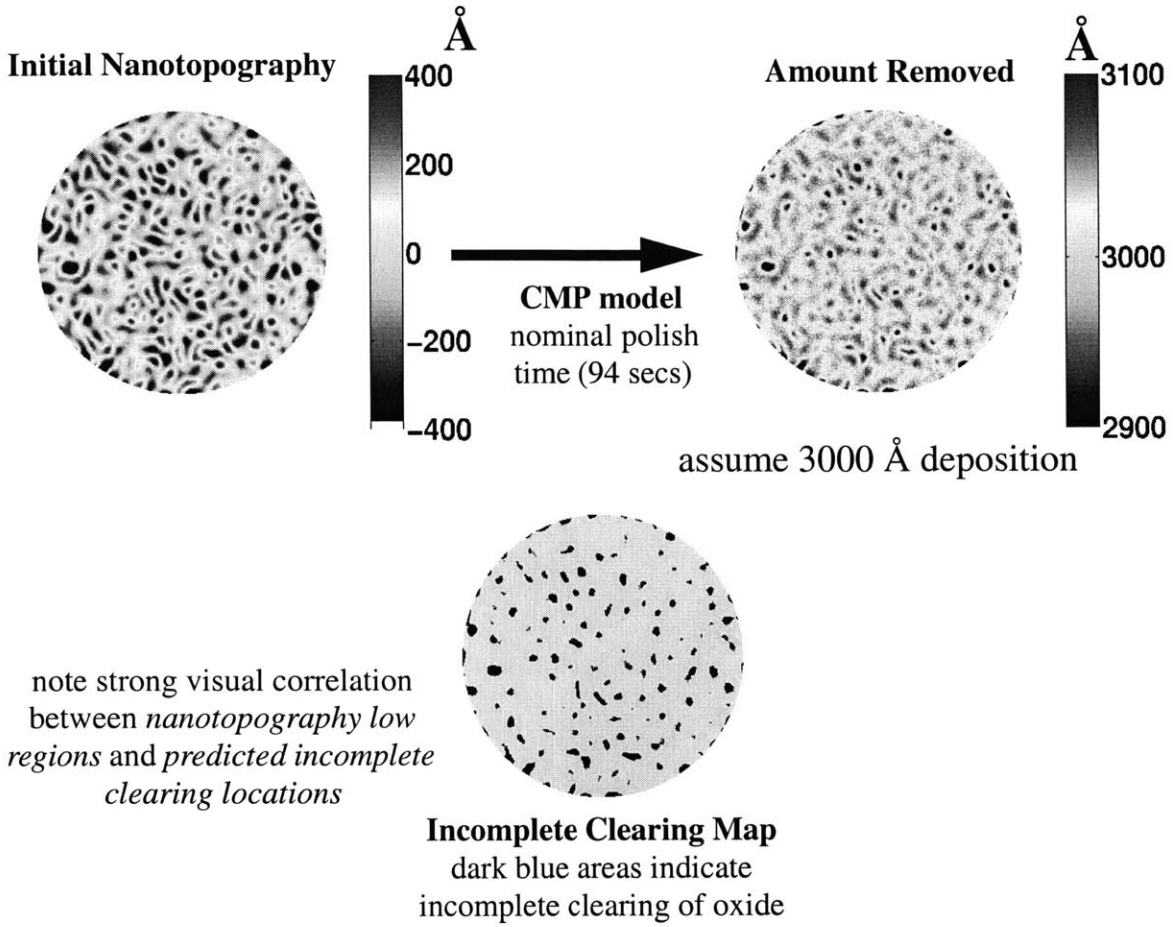


Figure 5.15: Using nanotopography CMP modeling to assess incomplete oxide clearing in STI.

In Figure 5.15, there is a strong visual correlation between initial nanotopography "low spots" and incomplete oxide clearing. This agrees with the original hypothesis in Section 5.5 that low nanotopography spots can result in incomplete clearing.

5.11.2 Nitride Erosion

The assurance of oxide clearing in STI CMP, as well as process variability in removal rates (both wafer-to-wafer and within-wafer), requires the inclusion of a certain amount of overpolish time built into the CMP process. This overpolish may result in excessive nitride erosion, both due to pattern-dependent erosion as well as effects from nanotopography. Incomplete oxide clearing

effects from nanotopography exacerbate this problem by requiring even more additional polish time due solely to nanotopography effects.

The same polish situation used in Section 5.11.1 is repeated, with the polish time set to just equal the time at which we have complete oxide clearing across the entire wafer. The resulting nitride loss due to the CMP process is computed. A nitride loss budget of 200 Å is used to determine points of potential device failure due to excessive nitride erosion, and a third map is generated showing these locations. This excessive nitride erosion map generation process is illustrated in Figure 5.16. In Figure 5.16, there is a strong visual correlation between initial nanotopography "high spots" and excessive nitride erosion. This agrees with the original hypothesis in Section 5.5 that high nanotopography spots can result in nitride erosion.

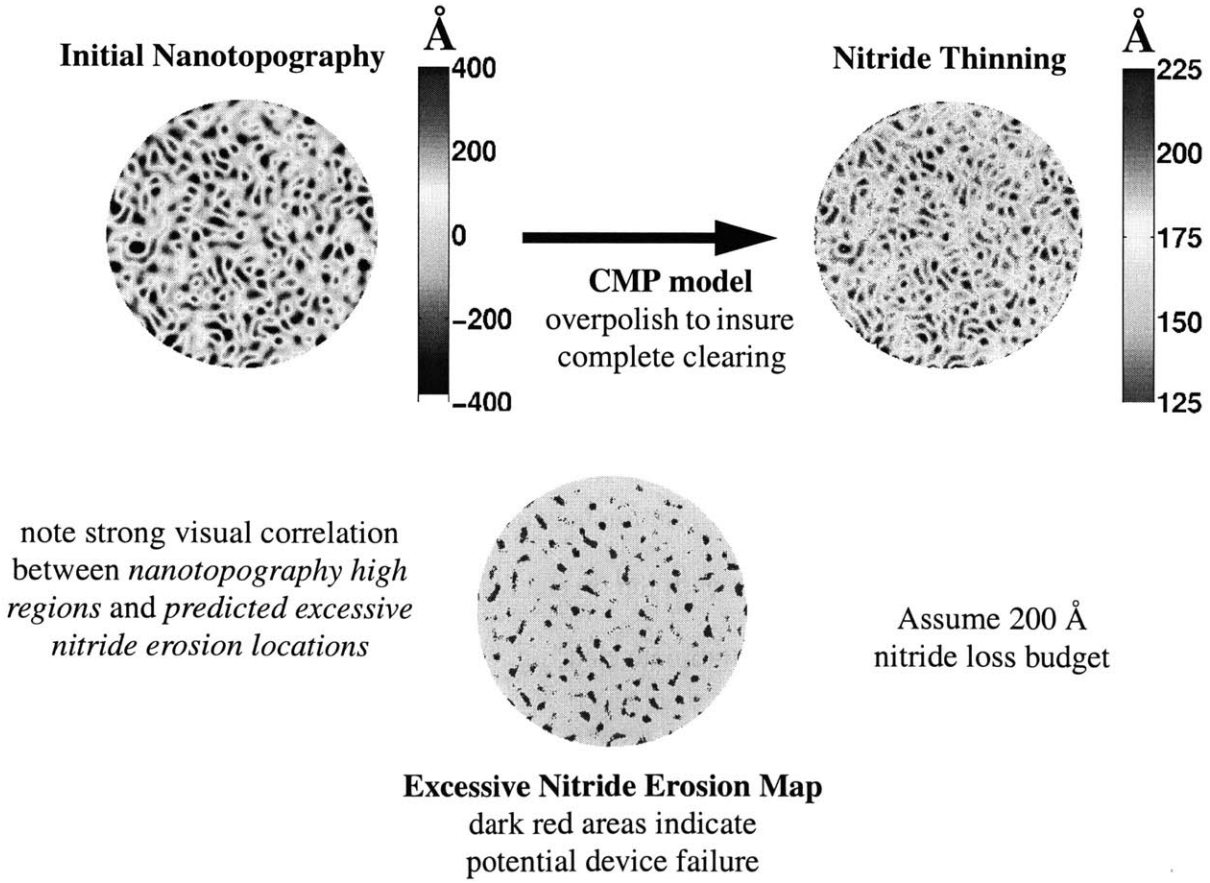


Figure 5.16: Using nanotopography CMP modeling to assess device failure in STI.

5.12 Combining Nanotopography and Pattern Dependent STI Effects

This section explores the implications of nanotopography on the chip-scale pattern dependent models developed in Chapter 2.

5.12.1 Analyzing Nanotopography Effects with Pattern-Dependent Models

Nanotopography implications on STI are a separate issue from the chip-scale pattern-dependencies discussed in Chapter 2. However, a comprehensive model of the STI CMP process should analyze both effects, as well as the effect of nanotopography on pattern-dependent dishing and erosion. As a simple estimator of the aggregate effect of nanotopography and layout pattern on the post-CMP erosion and dishing, an addition of the individual effects can be used. The erosion due to nanotopography can be calculated as a wafer map. The erosion due to chip-scale pattern dependency can also be computed, and then replicated into an array. This simulates a die grid on an actual wafer. These two quantities can then be added together to compute the aggregate erosion due to both wafer nanotopography and layout pattern dependencies. This is illustrated in Figure 5.17.

A more rigorous method would involve integrating the contact-wear wafer-scale CMP model with the density-based chip-scale CMP model. Since both models are based upon pressure distributions across the pad, it is likely that a simple addition of the two individually calculated effects will result in a certain amount of over-estimation. In addition, there may be interactions between the two effects that may result in an aggregate benefit or detriment to erosion and dishing. Further study is required.

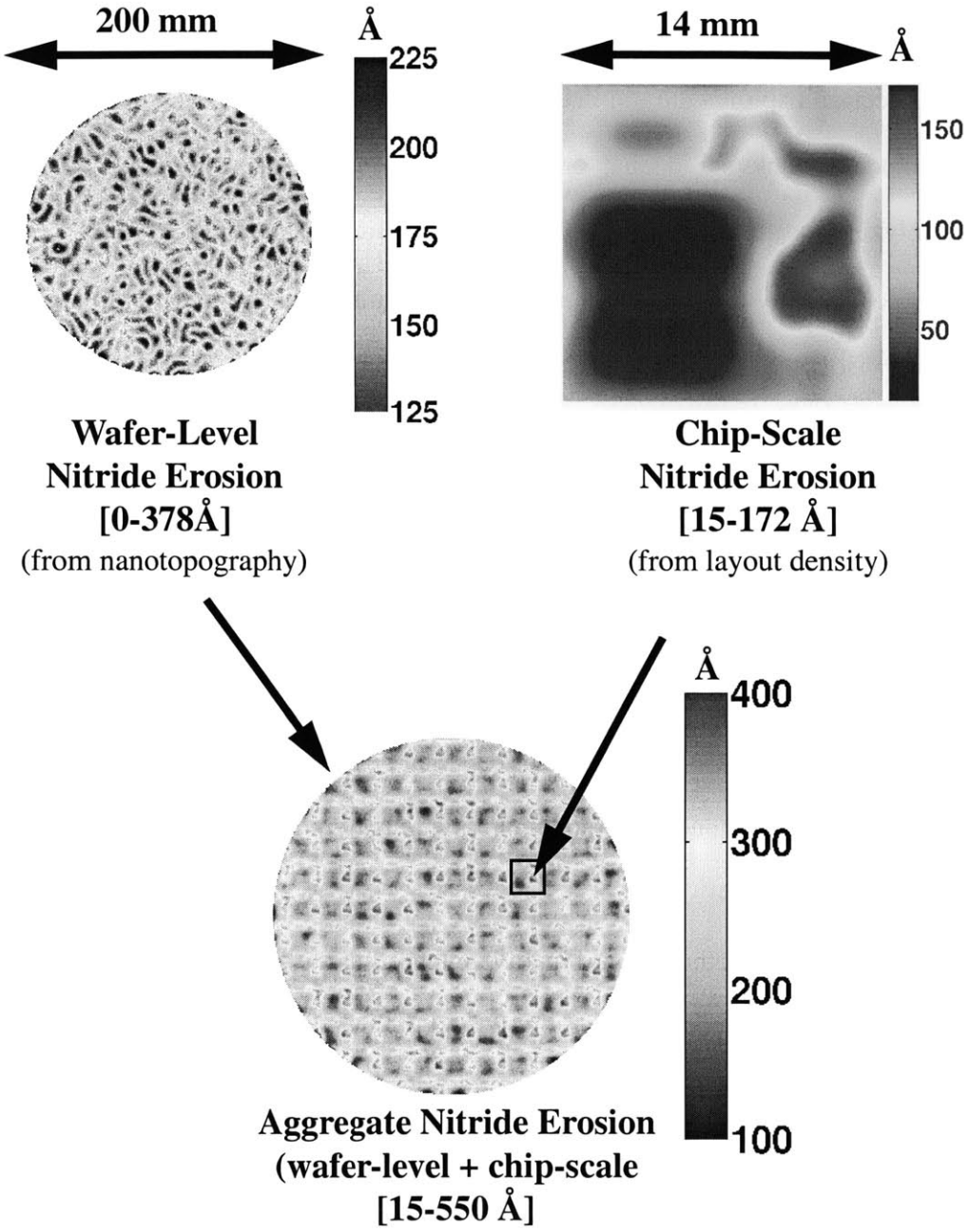


Figure 5.17: Adding effects of nanotopography and pattern density to determine total erosion.

5.12.2 Nanotopography Effect on Dishing and Erosion

Section 5.11.1 described how additional overpolish time is needed to insure that oxide is cleared everywhere on the wafer. This additional overpolish time has implications on the pattern-dependent dishing and erosion on a chip.

For the case study in Section 5.11.1, it is found that an additional overpolish time of 13 seconds is required to clear all oxide from the wafer. As a rough estimator of the additional dishing and erosion resulting from nanotopography, this overpolish time can be added to the nominal polish time in a chip-scale STI simulation. The results are shown in Table 5.9.

Table 5.9: Dishing and Erosion for Nominal and Nanotopography Cases

Test Case	Polish Time (secs)	Dishing (Å)			Erosion (Å)		
		Mean	Max	Min	Mean	Max	Min
Nominal	45	678	793	417	76	172	15
With Overpolish Time Due to Nanotopography	58	897	989	686	175	372	65

Average dishing increases by 32%, while average erosion increases by 130%, compared to the case without additional overpolish due to nanotopography. From the simulation results, the impact of nanotopography on chip-scale patterns as a result of additional overpolish time is significant.

5.13 Summary

Experiments using wafers with distinct nanotopography signatures in combination with CMP processes with distinct planarization lengths have been conducted. Results from this experiment verify the original hypothesis that substantial film thinning occurs when the CMP planarization length is larger than the length scale of the nanotopography, and that the post-CMP oxide thickness removed can be highly correlated with initial wafer nanotopography for CMP processes with long planarization lengths.

Three modeling techniques (scaling, filtering, contact wear) have been suggested as possible methods for analyzing the effect of nanotopography on post-CMP mean-centered oxide thickness deviation. Comparisons of the prediction errors of the three methods show that the contact wear and filter-based approaches perform slightly better than a simple scaling model

across a range of nanotopography types and CMP processes. Contact-wear models are preferred for use since they allow for straightforward prediction of dual material stacks such as those used in STI processing, while filter-based approaches do not offer a simple method of application for multiple film stacks.

As nitride erosion budgets in STI CMP decrease with future technology generations, the effect of nanotopography on nitride thinning will become more of a concern. Accurate modeling of the nanotopography effect on film thinning is essential for determining the impact of nanotopography on device yields.

Predictive mappings of post-CMP film thicknesses across the wafer can be used for the diagnosis of potential yield problems due to incomplete oxide clearing or excessive nitride erosion. Such device failure maps complement nanotopography wafer data and metrics based on that data, and enable evaluation, comparison, and development of improved wafers and STI CMP processes.

An integrated model that predicts the combined effects of wafer-level nanotopography and chip-scale layout pattern dependencies is a useful tool in diagnosis of the yield of a particular combination of wafer type, CMP process, and layout. A simple estimator that adds the contributions of the two effects can be used as a first-order approximation of an integrated model. Further development of a complete model that truly integrates the wafer-scale contact mechanics along with the chip-scale density-based model is needed.

Chapter 6
Conclusions

This work has examined key pattern related issues associated with the shallow trench isolation chemical mechanical polishing process. A shallow trench isolation modeling and characterization methodology has been developed in this thesis. The model has been verified with experimental data. Applications and extensions of the model have been described and demonstrated. This concluding chapter summarizes the main contributions of this research and applications of the concepts developed, and describes future work that builds on the research in this thesis.

6.1 Thesis Contributions

A major contribution in this work is the development and implementation of a characterization and modeling methodology for STI CMP. An STI CMP model has been developed, implemented, and verified using experimental data. A characterization method has been defined that can be used to extract a small set of characterization model parameters from a given CMP process that are applicable for that CMP process on any layout.

Layout design optimization techniques have been developed. Methods to analyze the pattern density of arbitrary layouts, diagnose potential density-dependent dishing and erosion problems, and modify the existing layout to reduce these problems have been developed and demonstrated.

A generalization of the STI CMP characterization and modeling methodology has been developed, and applied to the modeling of STI CMP with nonconventional consumable sets. Phenomenological models have been derived and verified using experimental data for three nonconventional consumable sets. These include fixed abrasive pads, high-selectivity silica based

slurry, and high-selectivity ceria based slurry CMP processes.

The impact of wafer nanotopography on STI CMP has been studied. Modeling methods for assessing the additional erosion and dishing due to the consideration of nanotopography have been developed and implemented, and verified on experimental data. Methods for characterization of wafer nanotopography have been described.

Together, the CMP modeling and layout improvement approaches contributed by this research aid in the understanding and improvement of current and future STI CMP processes.

6.2 Applications

A number of applications are enabled by the contributions from this thesis. The characterization of STI CMP processes allows for the prediction of nitride erosion and dishing on arbitrary layouts, which is a key benefit to chip designers. The availability of an accurate predictor reduces or eliminates the necessity of running costly characterization experiments for each layout and CMP process. The density optimization techniques provide methods for nitride erosion and dishing reduction through layout modification.

A typical chip designer can use all of the tools and methods described in this work to diagnose and improve STI erosion and dishing on arbitrary layouts. First, the STI CMP process to be used can be characterized using the test wafers and characterization methodology described in Chapter 2. Second, the dishing and erosion of an arbitrary layout can be predicted using the models developed in Chapter 2. Third, the feasibility and performance of alternative consumable sets can be explored using the techniques and models in Chapter 3, and a candidate nonconventional consumable can be evaluated using the calibrated erosion and dishing predictions using the models in this chapter. Fourth, erosion and dishing can be reduced using layout design optimization techniques described in Chapter 4. Finally, the impact of wafer nanotopography on STI dishing and erosion can be assessed using the models in Chapter 5. This allows the designer to minimize the effects of STI dishing and erosion through proper selection of

CMP process, layout design, and wafer.

6.3 Future Work

Several topics in STI CMP have not been investigated in the course of this research. Some are extensions of the research work, and some are completely new topics that relate to STI CMP.

One major advancement of the modeling techniques described in this thesis is the development and implementation of an integrated wafer and chip-scale model that takes into account the effects of wafer nanotopography and chip-scale pattern variation. Such a model needs to be two-tiered, due to the different discretization scales of the two problems. A proposed method is to discretize the wafer on a coarse scale, compute pressure variations across the wafer surface due to nanotopography, and feed these pressures into a chip-scale simulation. Each advancement of the chip-scale surface in time is then fed back into the wafer-level model, updating the wafer-scale surface profile. Updated wafer-scale pressures are then fed back into the chip-scale model, allowing for the combined effects of the two models to be computed simultaneously. This is illustrated in Figure 6.1.

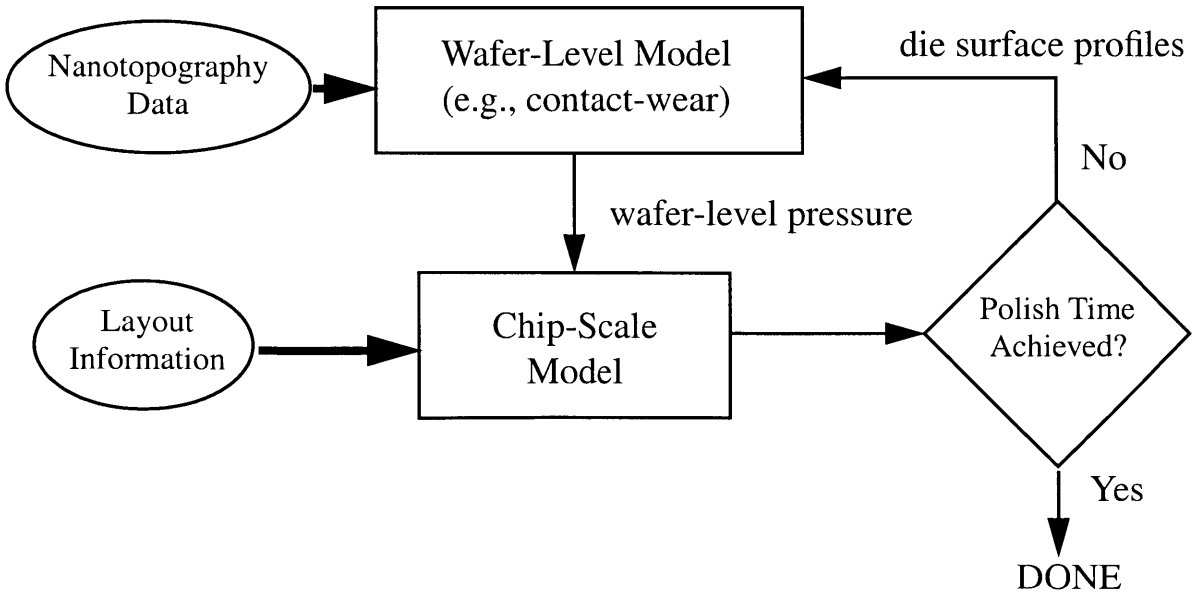


Figure 6.1: Block diagram of an proposed integrated STI CMP Model.

Further study into modeling of STI CMP for small feature STI structures is required. This work has focused on STI CMP of structures greater than 10 μm; realistic device active areas can be significantly smaller. In addition, the effect of active area width and trench oxide width on the STI CMP polish should also be studied. Model extensions to incorporate these additional effects may be required.

Another topic that has not been discussed in this thesis is the development of more physically based models. All of the modeling work in this thesis is semi-empirical or phenomenological, and connecting the models here with physics of the CMP process would be a major advancement. In particular, determining physical reasons for the empirically determined density dependencies for nonconventional consumable sets may provide insight into developing revised models for these consumables.

6.4 Final Remarks

The shallow trench isolation process is a critical part of the semiconductor manufacturing process flow, and the chemical mechanical polishing process is a key enabler for STI processing. This work provides a set of effective characterization and modeling methods that can be used as a diagnostic tool to examine the various issues associated with STI CMP, from chip-scale pattern dependent erosion and dishing prediction, to layout design optimization for STI, to the impact of wafer nanotopography on STI CMP.

Bibliography

[1] S. Sivaram, H. Bath, R. Legegett, A. Maury, K. Monning, and R. Tolles, "Planarizing Interlevel Dielectrics by Chemical Mechanical Polishing," *Solid State Tech.*, pp. 87-91, May 1992.

[2] C. Yu, P. C. Fazan, V. K. Mathews, and T. T. Doan, "Dishing effects in a chemical mechanical polishing planarization process for advanced trench isolation," *Applied Physics Letters*, vol. 61, no. 11, pp. 1344-1346, September 1992.

[3] F. B. Kaufman, D. B. Thomson, R.E. Broadie, M. A. Jaso, W. L. Guthrie, D. J. Pearson, and M. B. Small, "Chemical-mechanical polishing for fabricating patterned W metal features as chip interconnect," *J. Electrochem. Soc.*, vol. 138, no. 11, pp. 3460-3465, 1991.

[4] J. M. Steigerwald, R. Zirpoli, S. P. Muraka, and R. J. Gutmann, "Pattern geometry effects in the chemical-mechanical polishing of inlaid copper," *J. Electrochem. Soc.*, vol. 142, no.10, pp. 2841-2848, 1994.

[5] L. M. Cook, "Chemical processes in glass polishing," *J. Non-Crystalline Solids*, vol. 520, pp. 152-171, 1990.

[6] A. S. Lawing, "CMP Pad Conditioning and Pad Surface Characterization," *MRS Spring Meeting*, Paper I5.3, 2002.

[7] S. Runnels, M. Kim, J. Schleuter, C. Karlsrud, and M. Desai, "A modeling tool for chemical mechanical polishing design and evaluation," *IEEE Trans. on Semi. Manuf.*, vol. 11, no.3, pp. 501-510, August 1995.

[8] D. Dornfeld, "Mechanical Aspects of CMP," *Proc. VMIC Conf.*, pp. 105-112, June 2000.

[9] B. Zhao and F. Shi, "Chemical Mechanical Polishing in IC Processes: New Fundamental Insights," *Proc. CMP-MIC Conf.*, pp. 13-21, February 1999.

[10] G. Fu, A. Chandra, and S. Guha, "A Generalized Material Removal Model for the Chemical Mechanical Polishing (CMP) Process," *Proc. VMIC Conf.*, pp. 113-122, June 2000.

[11] Y.-T. Su, "Investigation of Removal Rate Properties of a Floating Polishing Process," *J. Electrochem. Soc.*, vol. 147, no. 6, pp. 2290-2296, 2000.

[12] D. Stein and D. Hetherington, "Review and Experimental Analysis of Oxide CMP Models," *Electrochemical Society Proceedings of the Third International Symposium on Chemical Mechanical Planarization in IC Device Manufacturing*, 99-37, pp. 217-233, 1999.

[13] R. Jairath, S. Chadda, E. Engdahl, W. Krussel, T. Mallon, K. Mishram, A. Pant, and B. Withers, "Performance of OnTrak System's Linear Planarization Technology (LPT) for Dielectric CMP Processes," *Proc. CMP-MIC Conf.*, pp. 194-201, February 1997.

[14] A. Yagishita, T. Saito, K. Nakajima, S. Inumiya, Y. Akasaka, Y. Ozawa, G. Minamihaba, H. Yano, K. Hieda, K. Suguro, T. Arikado, and K. Okumura, "High Performance Metal Gate MOSFETs Fabricated by CMP for 0.1 μm Regime," *IEDM Technical Digest*, pp. 785-788, 1998.

[15] E. B. Brown, Modern Optics, Van Nostrand-Reinhold, New York, 1965.

[16] S. Wolf and R. N. Tauber, Silicon Processing for the VLSI Era: Vol. 1 - Process Technology, Lattice Press, Sunset Beach CA, Chapter 13, 1986.

[17] International Technology Roadmap for Semiconductors, International SEMATECH, 2001.

[18] S. Sundararajan, D. Thakurta, D. Schwendeman, S. Mrarka, and W. Gill, "Two-Dimensional Wafer-Scale Chemical Mechanical Planarization Models Based on Lubrication Theory and Mass Transport," *J. Electrochem. Soc.*, vol. 146, no. 2, pp. 761-766, 1999.

[19] R. Subramanian, L. Zhang, and S. Babu, "Transport Phenomena in Chemical Mechanical Polishing," *J. Electrochem. Soc.*, vol. 146, no. 11, pp. 4263-4272, 1999.

[20] J. Tichy, J. Levert, L. Shan, and S. Danyluk, "Contact Mechanics and Lubrication Hydrodynamics of Chemical Mechanical Polishing," *J. Electrochem. Soc.*, vol. 146, no. 4, pp.1523-1528, 1999.

[21] J. Luo, D. Dornfeld, Z. Mao, and E. Hwang, "Integrated Model for Chemical-Mechanical Polishing Based on a Comprehensive Material Removal Model," *Proc. CMP-MIC Conf.*, pp. 25-32, March 2001.

[22] S. Mukherjee and J. Levert, "Planarization of Copper Damascene Interconnects by Spin-Etch Process: A Chemical Approach," *MRS Spring Meeting*, Paper E8.10, 2000.

[23] M. H. Tsai, S. W. Chou, C. L. Chang, C. H. Hsieha, M. W. Lin, C. M. Wu, W. S. Shue, D. C. Yu, and M. S. Liang, "CMP-Free and CMP-Less Approaches for Multilevel Cu/low-k BEOL Integration," *IEDM Tech. Digest*, pp. 80-83, 2001.

[24] S. D. Hosali, A. Sethuraman, J.-F. Wang, L. M. Cook, and D. Evans, "Planarization Process and Consumable Development for Shallow Trench Isolation," *Proc. CMP-MIC Conf.*, pp. 52-57, February 1997.

[25] S. Wolf, Silicon Processing for the VLSI Era: Vol. 2 - Process Integration, Lattice Press, Sunset Beach CA, Chapter 13, 1990.

[26] P. Burke, "Semi-empirical modeling of SiO_2 chemical-mechanical polishing planarization," *Proc. VMIC Conf.*, pp. 379-384, 1991.

[27] E. Chang, B. Stine, T. Maung, R. Divecha, D. Boning, J. Chung, K. Chang, G. Ray, D. Bradbury, S. Oh, and D. Bartelink., "Using a Statistical Metrology Framework to Identify Systematic and Random Sources of Die- and Wafer-level ILD Thickness Variation in CMP Processes," *IEDM Tech. Digest*, pp. 499-502, 1995.

[28] D. Ouma, B. Stine, R. Divecha, D. Boning, J. Chung, I. Ali, and M. Islamraja, "Using Variation Decomposition Analysis to Determine the Effects of Process on Wafer and Die-level Uniformity in CMP," *First International Symposium on Chemical Mechanical Planarization (CMP) in IC Device Manufacturing*, 190th Electrochemical Society Meeting, October 1996.

[29] D. Ouma, D. Boning, J. Chung, G. Shinn, L. Olsen, and J. Clark, "An Integrated Characterization and Modeling Methodology for CMP Dielectric Planarization," *International Interconnect Technology Conference*, pp. 67-69, 1998.

[30] D. Ouma, Modeling of Chemical Mechanical Polishing for Dielectric Planarization, *Ph.D. Thesis*, MIT, November 1998.

[31] J. Grillaert, N. Heylen, E. Vrancken, G. Badenes, R. Rooyackers, M. Meuris, and M. Heyns, "A novel approach for the elimination of the pattern density dependence of CMP for shallow trench isolation," *Proc. CMP-MIC Conf.*, pp. 313-318, February 1998.

[32] UV-12x0SE Thin Film Measurement Systems User's Guide, KLA-Tencor, 1997.

[33] B. Stine, D. Ouma, R. Divecha, D. Boning, J. Chung, D. Hetherington, I. Ali, G. Shinn, J. Clark, O. Nakagawa, and S.-Y. Oh, "A Closed-Form Analytic Model for ILD Thickness Variation in CMP Processes," *Proc. CMP-MIC Conf.*, pp. 266-273, February 1997.

[34] F. W. Preston, "The Theory and Design of Plate Glass Polishing Machines," *J. Soc. Glass Technol.*, vol. 11, p. 214, 1927.

[35] T. Smith, S. J. Fang, D. Boning, G. B. Shinn, and J. A. Stefani, "A CMP Model Combining Density and Time Dependencies," *Proc. CMP-MIC Conf.*, pp. 97-104, February 1999.

[36] R. Jin, J. David, B. Abbassi, L. Liauzu, T. Osterheld, and F. Redeker, "A Production-Proven Shallow Trench Isolation (STI) Solution Using Novel CMP Concepts," *Proc. CMP-MIC Conf.*, pp. 314-321, February 1999.

[37] D. Goetz, "The Effect of Subpad Construction on Pattern Density Effects for Slurry-Free CMP," *Proc. CMP-MIC Conf.*, pp. 234-241, February 1999.

[38] J. Grillaert, M. Meuris, N. Heylen, K. Devriendt, E. Vrancken, and M. Heyns, "Modelling step height reduction and local removal rates based on pad-substrate interactions," *Proc.CMP-MIC Conf.*, pp. 79-86, February 1998.

[39] J. Grillaert, M. Meuris, N. Heylen, K. Devriendt, E. Vrancken, and M. Heyns, "Modelling the Influence of Pad Bending on the Planarization Performance During CMP," *MRS Spring Meeting*, pp. 45-50, 1999.

[40] E. Nowak, "Pattern filled photo mask generation for integrated circuit manufacturing," *US Patent #5,597,668*. January 28, 1997.

[41] G. Liu, R. Zhang, and K. Hsu, "Chip-Level CMP Modeling and Smart Dummy Fill for HDP and Conformal CVD Films," *Proc. CMP-MIC Conf.*, pp.120-127, February 1999.

[42] S. Sun, "Active Layer Mask with Dummy Pattern," *U.S. Patent #5902752*. May 11, 1999.

[43] B. Davari, C. W. Koburger, R. Schulz, J. D. Warnock, T. Furukawa, M. Jost, Y. Taur, W. G. Schwittek, J. K. DeBrosse, M. L. Kerbaugh, and J. L. Mauer, "A New Planarization Technique, Using a Combination of RIE and Chemical Mechanical Polish (CMP)," *IEDM Tech. Digest*, pp. 61-64, 1989.

[44] M. Nandakumar, A. Chatterjee, S. Sridhar, K. Joyner, M. Rodder, and I.-C. Chen, "Shallow Trench Isolation for advanced ULSI CMOS Technologies," *IEDM Tech. Digest*, pp. 133-136, 1998.

[45] B. E. Stine, D. Boning, J. Chung, L. Camiletti, F. Kruppa, E. R. Equi, W. Loh, S. Prasad, M. Muthukrishnan, D. Towery, M. Berman, and A. Kapoor, "The Physical and Electrical Effects of Metal Fill Patterning Practices for Oxide Chemical Mechanical Polishing Processes," *IEEE Trans. on Electron Devices*, vol. 45, no. 3, pp. 665-679, March 1998.

[46] A. Kahng, G.Robins, A. Singh, and A. Zelikovsky, "Filling Algorithms for Layout Density Control," *IEEE Trans. on Computer Aided Design*, vol. 18, no. 4, pp. 445-462, 1999.

[47] R. Tian, X. Tang, and M. Wong, "Dummy-Feature Placement for Chemical-Mechanical Polishing Uniformity in a Shallow-Trench Isolation Process," *IEEE Trans. on Computer Aided Design*, vol. 21, no. 1, pp. 63-71, 2002.

[48] B. Lee, D. Boning, D. Hetherington, and D. Stein, "The Use of Smart Dummy Fill and Selective Reverse Etchback for Effective Pattern Density Equalization," *Proc. CMP-MIC Conf.*, pp. 255-258, February 2000.

[49] S. Xu, "Effects of Silicon Front Surface Topography on Silicon Oxide Chemical Mechanical Planarization," *Electrochem. and Sol. State Lett.*, vol. 1, no. 4, pp. 181-182, 1998.

[50] K. V. Ravi, "Wafer Flatness Requirements for Future Technologies," Future Fab International, no. 7, pp. 207.

[51] O. G. Chekina and L. M. Keer, "Wear-Contact Problems and Modeling of Chemical Mechanical Polishing," *J. Electrochem. Soc.*, vol. 145, no. 6, pp. 2100-2106, June 1998.

[52] T. Yoshida, "Three-Dimensional Chemical Mechanical Polishing Process Model by BEM," *Electrochemical Society Proceedings of the Third International Symposium on Chemical Mechanical Planarization in IC Device Manufacturing*, 99-37, pp. 593-604, 1999.

[53] J. Grillaert, M. Meuris, E. Vrancken, N. Heylen, K. Devriendt, and W. Heyns, "The Use of a Semi-Empirical CMP Model for the Optimization of the STI Module," *Proc. CMP-MIC Conf.*, pp. 105-112, February 1999.

[54] E. Tseng, M. Meng, and S. C. Peng, "Modeling and Discussion for STI CMP Process," *Proc. CMP-MIC Conf.*, pp. 113- 119, February 1999.

[55] H.-W. Chiou and L.-J. Chen, "One Step Effective Planarization of Shallow Trench Isolation," *Proc. IITC Conf.*, pp. 199-201, 1998.

[56] T. Tugbawa, T. Park, D. Boning, T. Pan, P. Li, S. Hymes, T. Brown, and L. Camilletti, "A Mathematical Model of Pattern Dependencies in Cu CMP Processes," *Proc. ECS Meeting*, pp. 605-615, October 1999.

[57] B. Lee, D. Boning, and L. Economikos, "A Fixed Abrasive CMP Model," *Proc. CMP-MIC Conf.*, pp. 395-402, March 2001.

[58] B. Lee, T. Gan, D. Boning, P. Hester, N. Poduje, and W. Baylies, "Nanotopography Effects on Chemical Mechanical Polishing for Shallow Trench Isolation," *Proc. ASMC Conf.*, pp. 425-432, September 2000.

[59] B. Lee, D. Boning, W. Baylies, N. Poduje, P. Hester, Y. Xia, J. Valley, C. Koliopoulus, D. Hetherington, H. Sun, and M. Lacy, "Wafer Nanotopography Effects on CMP: Experimental Validation of Modeling Methods," *MRS Spring Meeting*, Paper M4.9, 2001.

[60] D. Boning, B. Lee, W. Baylies, N. Poduje, and J. Valley, "Modeling and Mapping of Nanotopography Interactions with CMP," *MRS Spring Meeting*, Paper I1.5, 2002.

[61] *Guide for Reporting Wafer Nanotopography*, Semi M43-0301, SEMI Semiconductor Equipment and Materials International, March 2001.

[62] P. Wagner, "Alternative Nanotopography Metric," SEMI Standards Meetings, Advanced Wafer Geometry Task Force, *SEMICON/Europe*, April 2001.

[63] R. Schmolke, R. Deters, P. Thieme, R. Pech, H. Schwenk, and G. Diakourakis, "On the Impact of Nanotopography of Silicon Wafers on Post-Chemical Mechanical Polished Oxide Layers," *J. Electrochem. Soc.*, vol. 149, no.4, pp. G257-G265, 2002.

[64] M. Tsujimura, "Analysis of CMP planarization performance for STI process," *MRS Spring Meeting*, Paper M3.6, 2001.

[65] T. Fukuda, "JEIDA Nanotopography Report (II)," SEMI Standards Meetings, Advanced Wafer Geometry Task Force, *SEMICON/West*, July 2000.

[66] N. Tamura, H. Niwa, M. Hatanaka, M. Kase, and T. Fukuda, "The Influence of Wafer Nanotopology on Residual Film Thickness Variation after Chemical Mechanical Planarization," *Proc. ECS Meeting*, Abstract 495, May 2000.

[67] J. Valley, ADE Phase-Shift, 2000, *private communication.*

[68] Y. Xia, K. Daniell, and G. Gardopee, "Rapid CMP Process Characterization Through Fast Full Wafer Film Thickness Mapping," *Proc. CMP-MIC Conf.*, pp. 121-124, March 2001.

[69] J. T. Pan, D. Ouma, P. Li, D. Boning, F. Redeker, J. Chung, and J. Whitby, "Planarization and Integration of Shallow Trench Isolation," *Proc. VMIC Conf.*, pp. 467-472, June 1998.

[70] J. Gagliardi and T. Yo, "Total Planarization of the MIT 961 Mask Set Wafers Coated with HDP Oxide," *Proc. CMP-MIC Conf.*, pp. 373-378, March 2000.

[71] M. Fayolle, J. F. Lugand, F. Weimar, and W. Bruxvoort, "Evaluation of a New Slurry-Free CMP Technique for Oxide Planarization," *Proc. CMP-MIC Conf.*, pp. 128-133, February 1998.

[72] V. Koinkar, R. Golzarian, Q. Luo, M. VanHanehem, J. Shen, P. Burke, T. Fletcher, L. C. Hardy, J. Kollodge, J. Trice, T. Engfer, and E. Funkenbusch, "Chemical Mechanical Planarization of Copper Interconnects Using Fixed Abrasive Polishing Pad," *Proc. CMP-MIC Conf.*, pp. 58-65, February 2000.

[73] A. Römer, T. Donohue, J. Gagliardi, F. Weimar, P. Thieme, and M. Hollatz, "STI CMP Using Fixed Abrasive: Demands, Measurement Methods and Results," *Proc. CMP-MIC Conf.*, pp. 265-274, March 2000.

[74] L. Economikos, F.-F. Jamin, A. Ticknor, and A. Simpson, "Evaluation of Fixed Abrasive Pads for STI Planarization," *Proc. CMP-MIC Conf.*, pp. 553-562, March 2001.

[75] *Elexsol STS-2 specification sheet*, available at www.rodel.com

[76] H. Nojo, M. Kodera, and R. Nakata, "Slurry Engineering for Self-Stopping, Dishing Free SiO_2-CMP," *IEDM Technical Digest*, pp. 349-352, 1996.

[77] M. Jouty, M. Rivoire, and T. Detzel, "The Effect of Feature Size and Counter Mask on Oxide CMP Removal Rate in Shallow Trench Isolation," *Proc. CMP-MIC Conf.*, pp. 329-332, February 1999.

[78] D. R. Evans, and B. D. Ulrich, "Pattern Dependence and Planarization Using Silica or Ceria Slurries for Shallow Trench Isolation," *Proc. CMP-MIC Conf.*, pp. 347-350, February 1998.

[79] S.-I. Lee, C.-I. Kim, H. Kim, J.-H. Kim, C.-W. Nam, S. Kim, and C.-T. Kim, "The Effects of Slurries with Pattern Size and Step Height in Shallow Trench Isolation Chemical Mechanical Polishing," *Proc. CMP-MIC Conf.*, pp. 163-168, March 2000.

[80] B.-H. Kwon, J.-H. Lee, S. Kweon, S.-Y. Lee, B.-C. Kim, I.-Y. Yoon, S.-I. Lee, and J.-G. Lee, "The Polishing Characteristics of Ceria-based Slurry and Silica-based Slurru in SRAM Device," *Proc. CMP-MIC Conf.*, pp. 218-224, March 2001.

[81] J. M. Steigerwald, R. Zirpoli, S. P. Murarka, D. Price, and R. J. Gutmann, "Pattern Geometry Effects in the Chemical Mechanical Polishing of Inlaid Copper Structures," *J. Electrochem. Soc.*, vol. 141, no. 10, pp. 2842-2848, October 1994.

[82] T. Park, T. Tugbawa, D. Boning, J. Chung, R. Muralidhar, S. Hymes, Y. Gotkis, S. Amalgir, 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. VMIC Conf.*, pp. 437-442, 1998.