M. R. Oliver (Ed.)

Chemical-Mechanical Planarization of Semiconductor Materials

Springer

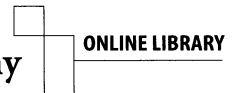Springer Series in
MATERIALS SCIENCE
69

Springer-Verlag Berlin Heidelberg GmbH

Physics and Astronomy

springeronline.com

Springer Series in
MATERIALS SCIENCE

Editors: R. Hull   R. M. Osgood, Jr.   J. Parisi   H. Warlimont

The Springer Series in Materials Science covers the complete spectrum of materials physics, including fundamental principles, physical properties, materials theory and design. Recognizing the increasing importance of materials science in future device technologies, the book titles in this series reflect the state-of-the-art in understanding and controlling the structure and properties of all important classes of materials.

Series homepage – springer.de

M. R. Oliver (Ed.)

Chemical–Mechanical Planarization of Semiconductor Materials

With 298 Figures

Springer

Dr. Michael R. Oliver
Rodel Fellow, Rodel Inc.
14625 NW Skyline Boulevard
Portland, OR 97231, USA
E-mail: MOliver@Rodel.com

Series Editors:

Professor Robert Hull
University of Virginia
Dept. of Materials Science and Engineering
Thornton Hall
Charlottesville, VA 22903-2442, USA

Professor R. M. Osgood, Jr.
Microelectronics Science Laboratory
Department of Electrical Engineering
Columbia University
Seeley W. Mudd Building
New York, NY 10027, USA

Professor Jürgen Parisi
Universität Oldenburg, Fachbereich Physik
Abt. Energie- und Halbleiterforschung
Carl-von-Ossietzky-Strasse 9–11
26129 Oldenburg, Germany

Professor Hans Warlimont
Institut für Festkörper-
und Werkstofforschung,
Helmholtzstrasse 20
01069 Dresden, Germany

springeronline.com

Preface

Chemical Mechanical Planarization (CMP) has emerged in the last two decades and grown rapidly as a basic technology widely used in semiconductor device fabrication. As a semiconductor processing step, it was developed at IBM in the mid 1980s. From this beginning the technology has been widely adopted throughout the semiconductor industry.

As basic CMP technology has been understood and accepted throughout the semiconductor industry, its uses in different parts of the semiconductor process have multiplied. This includes special steps for some special processing flows, such as for DRAM technology. In addition, the availability of CMP technology has enabled the implementation of new technologies, with the best example being copper interconnect technology. Copper could not be practically implemented into semiconductor process flows until the advent of CMP.

Unfortunately, the rapid acceptance and implementation of CMP technology in wafer fabrication has occurred without a corresponding rate of advance in the underlying science. Progress is being made in understanding the underlying CMP mechanisms, but, in general, it is slow and uneven. The most noteworthy exception to this trend is the science of metal CMP reactions, where the scientific understanding is actually driving much of the advance of the technology. There has been no corresponding progress in other CMP areas however.

In contrast to the science of CMP, the applied technologies of its components, including polishing tools, slurries and pads, has developed steadily for well over a decade. These gains have been large and have truly made CMP a stable, production-worthy semiconductor process. The goal of this book is to present and discuss the elements of CMP science and technology, as it relates to semiconductor processing technology. Industrial CMP has rapidly evolved, and will continue to evolve, but the fundamental approach to CMP has remained the same.

I would like to thank the chapter authors for their diligence and solid efforts in writing chapters that fit together well. I would also like to thank Peter Burke for his support in the early stages of the book. I am especially grateful to Claus Ascheron and Angela Lahee at Springer for their valuable guidance and to Peter Altenberg for preparing the text.

Portland,   *Mike Oliver*
October 2003

Contents

7 Fundamentals of CMP Slurry

8 CMP Cleaning

9 Patterned Wafer Effects

List of Contributors

D. Boning
Massachusetts Institute
of Technology
Room 39-567B
Cambridge, MA 02139, USA
boning@mtl.mit.edu

J. M. De Larios
Lam Research Corporation
4400 Cushing Parkway
Fremont, CA 94538, USA
john.delarios@lamrc.com

K. DeVriendt
IMEC
Kapeldreef 75
3002 Leuven, Belgium
Katia.Devriendt@imec.be

D. R. Evans
Sharp Laboratories
5700 NW Pacific Rim Boulevard
Camas, WA 98607, USA
devans@sharplabs.com

D. Hetherington
Sandia National Laboratories
MS 1084
P.O. Box 5800
Albuquerque, NM 87185-5800, USA
dhethe@sandia.gov

D. B. James
Rodel, Inc.
361 Bellevue Rd.
Newark, DE 19713, USA
djames@rodel.com

M. R. Oliver
Rodel, Inc.
14625 NW Skyline Blvd.
Portland, OR 97231, USA
moliver@rodel.com

K. M. Robinson
Maxxim Integrated Products
14320 SW Jenkins Rd.
Beaverton, OR 97005, USA
karl_robinson@or.mxim.com

D. J. Stein
Sandia National Laboratories
MS 1084
P.O. Box 5800
Albuquerque, NM 87185-5800, USA
dstein@sandia.gov

T. Tucker
Laredo Technologies
247 Many Lakes Dr.
Kalispell, MT 59901-8713, USA
ttucker@ix.netcom.com

1 Introduction

Michael R. Oliver

1.1 Original Motivation for CMP

Chemical Mechanical Planarization (CMP) was introduced initially into semi-conductor processing to planarize inter-level dielectrics. This technology enabled a greatly improved multi-level metallization integration approach. Once the technology was implemented into large scale manufacturing, both the hardware and the processes evolved by leaps and bounds. New processes were developed, with the second major application being tungsten polish.

As the equipment and consumables improved, the resultant improved process control enabled the use of CMP in more processes, such as copper and shallow trench isolation (STI). After less than twenty years, CMP is used up to 10 times in manufacturing semiconductor process flows.

With such large investments being put into CMP, the pressures for higher performance equipment and consumables have been great. In addition to evolutionary improvements in the more standard technology, using rotating tables, with pads and slurries, new approaches are continually being developed. For example, fixed abrasive pads, with the abrasives in the pad and not in the slurry, are being intensively exercised to provide improved performance at specific process steps, such as STI.

Other, more revolutionary, approaches such as electropolish for copper CMP are being developed. While these novel concepts are not yet widely accepted, some may very well gain a strong foothold in the next few years. Others will fall by the wayside. Certainly there will continue to be introductions of new technologies to replace more standard CMP processing.

1.2 CMP Technology and Its Technical Understanding

This subject of this book is CMP technology in semiconductor manufacturing. Most of the discussion is on applications, but there is some discussion of models of polishing mechanisms. These are mostly in Chap. 2 by Michael Oliver and Chap. 4 by David Stein. CMP began and grew as an approach to greatly enhance the capability of commercial semiconductor processes. Its initial application and subsequent enormous growth to date have not required a scientific understanding of individual polishing mechanisms, either

for silicon dioxide and other dielectrics, or for metals. The parent technologies, including glass polishing and silicon wafer polishing, provided the hardware, consumables (pads and slurries) and process with which to begin.

The fact that no quantitative models of individual CMP polishing events exist does not mean that efforts have not been made to develop them. Oxide modeling has progressed little beyond where Cook [1] left it in his 1990 review article. Several models have been proposed in literature, but no prediction capability or satisfactory mechanism studies have confirmed the models.

There has been more progress in metal polishing, but is far from complete. This subject forms the basis for Chap. 4 by David Stein.

The progress that has been made in the CMP arena to date has been made without a good understanding of the basic underlying mechanisms. As new demands continue to made on CMP, it is likely that the absence of scientific knowledge of the processes will become an increasingly larger limitation on progress. When that limitation begins to be recognized, this recognition will provide motivation for focused basic research to understand more clearly how CMP actually takes place. Success in that effort can then lead to future waves of progress in CMP.

While the scientific understanding of CMP mechanisms is limited, the development of CMP pads and slurries has relied upon in-depth knowledge of several key disciplines. David James, in Chap. 6, discusses many of the properties of pads and how they are determined by the chemistry of the polymers and other materials used in their fabrication. The manufacture of CMP pads relies heavily on polyurethane formation as a function of starting materials and the thermal history of the manufacturing process.

Similarly, slurry development requires understanding of abrasive particle properties and behavior in solution. Also, as discussed by Karl Robinson in Chap. 7, David Stein in Chap. 4 and David Evans in Chap. 3 and in the Appendix, electrochemistry is a key discipline to employ for slurry development, especially metal slurry development.

1.3 Applications of CMP to Semiconductor Processing

The general characteristics and behavior of CMP processing as practiced in semiconductor fabrication are reviewed by Michael Oliver in Chap. 2. The components of the CMP process, using silicon dioxide polishing as the model, are described. The planarization process and how it is affected by the relative wafer-pad velocity and downforce are described. In addition, the general effects of both pad and slurry properties on polishing performance are reviewed. This chapter was designed as a departure point for most of the other chapters in this book.

Metal CMP differs from the CMP of silicon dioxide and other dielectrics in several significant ways. David Evans discusses many of the unique aspects of metal polishing in Chap. 3. As semiconductor technology as a whole advances,

it is anticipated that metal CMP will be more widely used than dielectric CMP, and the issues he discusses will be the central ones to CMP process development.

1.4 Polishing Tools and Consumables of CMP Technology

The CMP process takes place when the wafer surface is moved across the pad, under pressure, in the presence of a slurry. The mechanical motion and down force are imparted to the wafer by the polishing machine, or tool. The pad surface provides the rough points, or asperities, which make contact with the wafer. The slurry provides the abrasive particles and the appropriate chemistry for the CMP process to proceed.

It is desirable for all points on the wafer surface to experience the same velocities and pressure, as well as the same pad surface properties and slurry properties throughout the polishing cycle. Because of the geometry of the wafer, and its motion on the polishing tool, which is usually rotational, this is not generally possible to achieve. As discussed in Chap. 5 on polishing tools, by Thomas Tucker, various techniques, such as wafer carrier rotation, are employed to average out the non-uniform wafer motion on the polishing pad. Wafer carrier rotation also minimizes effects associated with the direction of the wafer-pad velocity. Another mechanism is often employed to provide nearly uniform pressure across the wafer. A soft bottom pad underneath the pad which contacts the wafer is used to provide a much more uniform instantaneous pressure distribution at the wafer-pad interface than would be obtained without it.

David James reviews CMP pads and the issues associated with them in Chap. 6. The pads must be robust so as to have the same mechanical properties over the pad life of hundreds of wafers. This means that the pad material must be relative inert with respect to the slurry chemistries used.

Included in this discussion is the nature of the pad surface structure and its impact on CMP. Microstructure is often built into the bulk of the pad, such as by incorporating closed pores. This pore structure is exposed at the pad surface, and as the pad surface wears, old pores disappear and new pores from the bulk are exposed. In addition, there are other surface features that influence the CMP process. Grooves are often employed to facilitate slurry distribution.

The key surface feature that affects the CMP process are the small rough points [2] called asperities which are created on the pad surface by the process called conditioning. A pad surface is generally conditioned drawing diamond points held in matrix across the pad surface. This process creates points which stick up a few microns from the pad surface. As the polishing process wears these asperities down, they need to be regenerated.

Slurries have been in use for many different types of polishing for centuries before the advent of CMP. The material to be polished and the final smoothness requirements govern the constitution of the slurry used for any specific application. Since there exists in current CMP technology several different surfaces to be polished, it is to be expected that each CMP application requires its specific slurry formulations. Abrasive particles contained in the slurry can have problems of settling while being stored and agglomerating, among other difficulties. These are the subjects of colloid science, and the issues confronted in CMP slurry technology are similar to those in numerous other fields.

Since the widely varying, and continually tightening, requirements of different CMP steps require greatly differing slurries, slurry technology is quite complex and is becoming more so. Karl Robinson reviews the key slurry components and the issues associated with them in Chap. 7.

As noted, CMP polishing tools originally evolved from tools designed for polishing plain silicon wafers. Basically, the machines have one or more rotating carriers, each of which holds a wafer, usually face down. The rotating carriers are forced against the pad. The pad is usually rotated, so that the wafer-pad local velocity is on the order of 1 meter/second. There have been several new machine design approaches which have been introduced as CMP technology has grown, and some have survived. Also, numerous new ideas have emerged in carrier technology and conditioning technology. With new applications such as shallow trench isolation (STI) CMP and copper CMP, end point detection has been introduced into CMP technology. End point determination allows greatly improved control of the final wafer topography, especially when one film needs to be completely removed with minimum polish into the underlying film. All of these issues are analyzed by Thomas Tucker in Chap. 5.

1.5 Post CMP Cleaning

Post CMP cleaning has been a difficult problem right from the beginning of CMP implementation into semiconductor manufacturing. A substantial part of the difficulty is a consequence of the fact that CMP slurries have abrasive particles in them. The particles often are difficult to remove after the CMP process, especially if the wafer has begun to dry before the cleaning process begins. One result of this behavior has been the widespread implementation of a physical contact cleaning machine, which scrubs the wafer surface with a soft scrubbing brush. Indeed, this approach has been very widely adopted and has resulted in almost all new polishing machine designs incorporating a scrubbing tool within the CMP polishing machine so the wafers are "dry in, dry out."

Copper and low-k dielectric cleaning is a recent and demanding area. The cleaning of these surfaces and other issues of post CMP cleaning are addressed by John de Larios in Chap. 8.

1.6 Integration of CMP Into the Semiconductor Fabrication Process

CMP technology has been adopted by the semiconductor industry to provide new capabilities for the overall semiconductor fabrication process. In a representative process, which can have almost 200 steps, each step builds upon the previous steps, and must provide a well controlled structure for the following steps. The parameters that must be controlled for the overall process determine the process specifications for the CMP steps.

Initially, silicon dioxide CMP processes were timed processes, and the remaining specifications for silicon dioxide thickness were wide, reflecting the poor process control at the time. Even with these wide specification ranges, CMP provided a capability that was superior to the alternatives. As equipment, consumables and the overall process understanding improved, tighter specifications could be achieved.

Significantly, the better controlled CMP technology enabled the use of CMP to produce a high performance STI CMP process. As with most other individual steps in a semiconductor process, tighter specifications on individual process steps enable an overall semiconductor process capability. The pressure for improved device performance has directly led to a very strong need for improved CMP process performance.

Each semiconductor manufacturer has its own semiconductor fabrication sequence. As a result, there are multiple approaches for any step, including CMP steps. The issues associated with alternatives for a given step are common to most manufacturers, but the selection of a given approach is often determined by specific needs of each individual manufacturer.

For four of the most widely used CMP steps, the integration issues are analyzed in Chap. 10. Karl Robinson discusses silicon dioxide and tungsten integration concerns, Katia DeVriendt reviews STI integration and David Evans discusses copper CMP integration issues.

1.7 Pattern Dependency Issues

When a patterned wafer is polished with CMP, a perfectly planar final surface does not generally result. This less than ideal surface can create problems. The characterization of how the initial surface with a given topography changes during CMP is a critical issue for many semiconductor processes. Changes may have to be made in the CMP step and other steps to deal with

the height variations after the CMP step. The understanding and addressing of the issues associated with pattern density effects has been a significant issue ever since CMP was adopted into semiconductor manufacturing.

The evolution of the film surface height profile during polishing is a function of several factors, including machine design and the specific composition of the slurry for the specific polish step. However, more significant factors are the structure of the pad (including both the top pad and the sub pad) and the nature of the topography in the area within a fraction of a centimeter about the area on the wafer which is being evaluated. This well known, and very complex, subject is analyzed in Chap. 9 by Duane Boning and Dale Hetherington.

1.8 Other Issues

There are subjects associated with CMP that are not addressed here. Abrasive free slurries and their applications have some unique issues that are not discussed in either the chapter on pads or the chapter on slurries. Also, there is a continuing level of new technological approaches which could either replace CMP or change it substantially. Most of these approaches have vanished relatively quickly, but sooner or later, one or more novel approach may catch on and significantly alter one or more of the CMP technologies now in use.

Any of these potential alternatives will have to be superior to the CMP approach it is replacing, either in performance or cost, or both. CMP technology, as is the rest of semiconductor technology, is evolving, and any change to a new technology track will have to provide an advantage in the long term, i.e., for several technology generations. A new approach will have to address the concerns discussed in this book, including the issues of cleaning, integration, and pattern dependency.

The same requirements hold, of course, for all smaller, evolutionary changes. We can expect that the evolution of CMP will continue for several years to come, since incremental improvements in pads, slurries and equipment are announced at a high rate, especially compared with other areas of semiconductor technology. In addition the rate of filing of new patents in the area of CMP technology has increased in recent years and shows no signs of abating.

There is still much excitement in CMP technology awaiting us.

References

1. L.M. Cook, J. Non-Crystalline Solids, vol. 120, 152, 1990.
2. A.S. Lawing, Proceedings 2002 CMP-MIC Conference, 310, IMIC, Tampa, 2002.

2 CMP Technology

Michael R. Oliver

2.1 Background and Motivation for CMP

Chemical Mechanical Polishing, also often referred to as Chemical Mechanical Planarization (CMP), was initially used as an enabling technology to fabricate high performance multiple level metal structures. Specifically, after the first level of metal was fabricated, and a nearly conformal silicon dioxide interlevel dielectric (ILD) layer was deposited, the second level metal has several fabrication problems, including deposition, resist patterning and etching. These difficulties are caused by the steps in the topography over which this layer must be processed [1]. Other technologies, especially spin-on glass (SOG), reduce many of the problems of multi-level metal integration approaches, however SOG introduces additional difficulties of its own, and has been primarily used for two and three level metal structures [2].

From a technology point of view, the initial work to develop CMP for semiconductor fabrication was done at IBM [3], where they used the expertise of their own silicon wafer fabrication technology. This expertise included an understanding of the hardware: machines, pads, and slurries. The scientific understanding of CMP was largely based on that of glass polishing [4], but that theory itself was not quantitative. In 1990, Cook presented an excellent summary of the understanding of the mechanisms of glass polishing up to that date [5]. He emphasized the poor quantitative agreement of existing models with experimental results.

Once the technology and the required equipment were available, the application of CMP quickly spread beyond polishing inter-level dielectric (ILD) layers. For example, CMP began to be used instead of reactive ion etching (RIE) to remove tungsten which was deposited to fill the via openings between metal layers [6]. Another metal CMP application, the fabrication of inlaid trenches filled with metal, also called damascene, was proposed. This polishing technology has essentially been an enabling technology for the introduction of copper interconnects into standard semiconductor processing. Until the availability of CMP, copper was not used for interconnects even though it has a lower resistivity than aluminum for the reason that it could not be easily be etched by RIE [7].

Other applications for CMP have also emerged. A very significant one is the use of CMP as part of the shallow trench isolation (STI) process [8, 9].

Shallow trench isolation is an integration approach that allows transistors to be packed at a higher density by reducing the isolation spacing between adjacent transistors. Another use for CMP is polishing polysilicon via plugs and capacitor structures in memory devices [10].

The purpose of this volume is to describe the major applications of CMP in the current semiconductor technology. Broadly speaking, CMP technology can be divided into two areas, dielectric and polysilicon CMP and metal CMP. Oxide CMP, which is the polishing of silicon dioxide, will be used for this chapter as the vehicle to discuss the elements of CMP. This chapter will also cover the technology and application of other dielectric polishing applications.

The first section describes the elements of the CMP process. Development and refinements of the basic approach will follow. The last part of the chapter will address dielectrics other than silicon dioxide.

2.2 Description of the CMP Process

In the current standard approach, Chemical Mechanical Polishing takes place where the surface of the wafer to be polished is forced against a polishing pad. The polishing pad is covered with a liquid slurry which contains abrasive particles. The wafer is moved relative to the slurry-covered pad, and the rate at which material is removed from the wafer is often described by the heuristic equation called Preston's Law [11]:

$$RR = K_p * P * V \tag{2.1}$$

with

$$RR - \text{removal rate}$$
$$K_p - \text{a constant, Preston's coefficient}$$
$$P - \text{local pressure on wafer surface}$$
$$V - \text{relative velocity of the point on the surface of wafer}$$
$$\text{vs. the pad.}$$

This relationship is empirical, a system where material was removed by grinding. Numerous dielectric and metal CMP models have been, and are continuing to be, proposed in the literature, and for most, Preston's Law is only an approximation. However, for much of the data obtained in practice, especially silicon dioxide CMP, Prestons Law provides a reasonably good fit to the data.

2.3 Polishing Equipment

The first polishing machines on which semiconductor CMP processes were developed were rotary polishing tables. As the machine technology has ad-
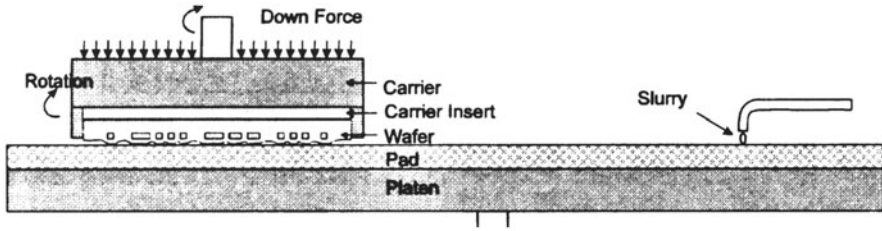
Fig. 2.1. Drawing of basic rotary CMP machine, showing wafer, carrier and platen (table). From US Patent 4,944,836. The retaining ring holds the wafer under the carrier insert (pad) (see text)

vanced, machine designs have evolved and other basic designs have been employed as well. However, most of the machines currently being sold as well as those in use are rotary tools.

A representative rotary polishing machine is diagrammed in Fig. 2.1, which is from [3], one of the early IBM patents. In such a machine, the polishing pad is circular and the wafer is placed in a carrier face down and is forced against the pad while the pad table, or platen, is rotated on its own axis.

The forces applied through the carrier on to the wafer are generally in the range of 1–10 psi, with oxide polishing usually in the higher end of the range and metal polishing in the lower end of the range. In practice, the table diameter is in the 20–26" range for commercial CMP machines which typically polish one wafer at a time. Figure 2.1 shows just the simplest configuration for a single table, single head (wafer carrier) system. In Chap. 5, Thomas Tucker reviews with details the many options for rotary designs as well as other designs. A key element of any polishing machine is to have well controlled pressures applied uniformly over the wafer as well as having controlled table and carrier rotation rates.

There are several other features in Fig. 2.1 that are to be noted. One is that slurry is dispensed from a tube in front of the wafer, so that as the table rotates, it is pulled under the wafer. Also, though not easily visible on this scale, the retaining ring around the edge of the wafer keeps the wafer in the carrier. The bottom of the retaining ring is recessed, usually about 0.008", from the plane of the bottom of the wafer.

The conditioner is a mechanism that moves a hard abrading surface, often a matrix with embedded diamond points, across the pad surface to roughen it. This is critical to CMP as an inadequately roughened pad surface results in a very low polish rate [13].

The slurry that flows onto the pad covers the roughened pad surface which moves under the wafer. The grooves on the pad allow more slurry to be brought under the retaining ring to the wafer face. As is discussed in Chap. 6, many pad structures also have small hollow spherical pores that are
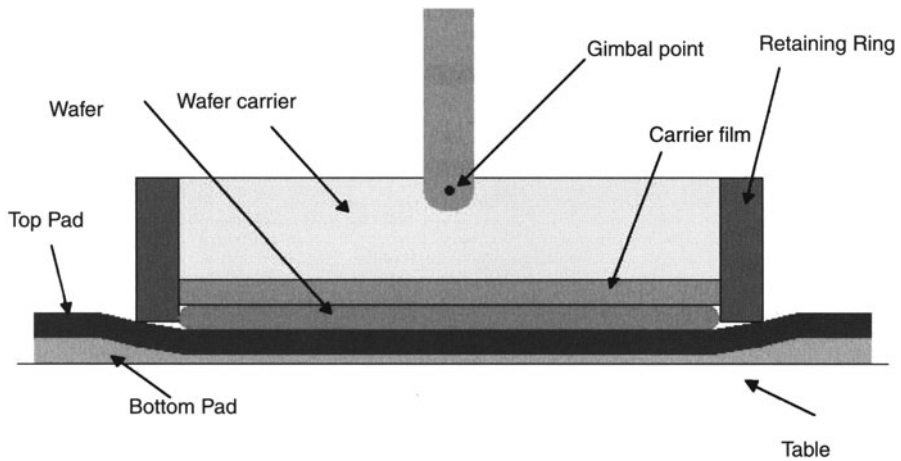
Fig. 2.2. Typical wafer carrier cross section, not to scale (see text)

exposed to the surface. These also contain slurry that is brought to the face of the wafer as the table rotates.

A second view, showing a simplified cross-section of a representative carrier, is shown in Fig. 2.2. This also shows a two-layer polishing pad, as well as the carrier film behind the wafer. Key features include the application of the down force from the carrier arm to the carrier at the gimbal point. The body of the carrier rotates about the gimbal point. Note that the gimbal point is above the wafer. The bottom pad layer and the carrier film are relatively compressible. Both films are generally about 0.050" thick and each compresses about 2–4% at pressures in the 5–7 psi range. The reason that both of these relatively compressible films are used is to maintain, over the entire wafer, a nearly uniform pressure at the wafer-polishing pad interface within the variations of the pad thicknesses, wafer thickness and the dimensional control of the table flatness relative to the carrier. It is worth noting that as machine and process tolerances become tighter, the sub-pad and carrier films can be thinner, since they will not have to compensate for as much mechanical variation.

Because the gimbal position (for most gimbal carrier designs) is about 1" above the wafer-pad interface, when the table rotates the friction at the wafer-pad interface causes a moment about the gimbal point, which increases the downward pressure at the leading edge of the wafer. Since the total constant downward force is applied at the gimbal point, a locally higher pressure at the leading edge of the wafer will also create a reduced pressure at the trailing edge. The exact instantaneous local pressures across the wafer will depend on properties of many of the elements in the system. One of the purposes of carrier rotation is to average out the leading and trailing edge effects [14]. This rotation averages the locally high removal rates at the leading edge

Fig. 2.3. Wafer polished for 60 seconds on Strasbaugh 6DS with no carrier rotation. Pre-polish wafer thickness was 10,000 Å and the *dark line* is the post-polish 7500 Å contour line. The *contour line* spacing is 250 Å. The leading edge is at the top of the wafer. Courtesy of David Evans, private communication

and the correspondingly low rates at the trailing edge, and can substantially reduce the non-uniform polish rate observed with a stationary carrier.

An example of such a polish rate variation is shown in Fig. 2.3. There for typical conditions except for no carrier rotation, the leading edge of the wafer has a higher polish rate than the trailing edge. There also is an effect of the outer side (here the right side) of the wafer polishes more quickly than the inside as it has a higher linear velocity and there is no velocity averaging by table rotation. At the trailing edge, there is less than 500 Å/min polish rate, and at the leading edge the rate is greater than 4250 Å/min.

Polishing conditions: ILD1300 silica based slurry

IC1400 perforated pad

Down force – 9 psi

Table rotation rate – 40 rpm.

However, with carrier rotation and in the absence of a wafer flat, or any other significant departure from rotational symmetry, the polish rate, and the total amount removed, will have close to radial symmetry. This behavior is widely observed on machines with rotating carriers.

A compressible carrier film between the carrier and the wafer is required to help provide a nearly uniform force on the back of the wafer with the variations in wafer thickness and top polish pad thickness. This is especially important as the pad wears with use. As discussed in Chap. 5, a trough is formed in the pad in the wafer path through abrasion during polishing. This trough can be quite deep, up to 25 µm or more lower than the edges of the pad. The trough formation during extended polishing is compensated for

and minimized in most current polishing systems by varying the conditioning conditions, primarily dwell time, as a function of radial position on the pad [15]. In general, though, some radial variation of pad thickness usually exists. Also, there are thickness non-uniformities due to manufacture of the polish pad as well as a lack of true planarity of the polishing table.

The compressible bottom pad, which is usually an impregnated felt or foam, is another component introduced to maintain a nearly constant pressure on the bottom side of the hard urethane polish pad (see Chap. 6). Because the top pad is less stiff than the wafer, the two layer stack of the urethane polish pad and the softer bottom pad determine key polishing features when polishing wafers with topography, i.e., wafers with device structures. (see Planarization section below and Chap. 10).

In summary, the purpose of the machine and pad elements together, is to provide as uniform polish conditions (pressure, velocity) as possible at all points on the wafer. The system is also designed to provide removal rate averaging through table and carrier rotation to minimize total variation of the amount of silicon dioxide, or other polished film, remaining across the wafer.

2.4 Polish Process

Several key elements of CMP are worth emphasizing. CMP of silicon dioxide surfaces requires certain specific properties of the slurries and pads. The slurries require the use of certain metal oxides as the abrasive particles. The oxide most widely used is silica (silicon dioxide), which can be fabricated by various methods (see Chap. 7). However, other metal oxide particles, such as ceria and manganese dioxide, can also be used. The slurry liquid needs to be aqueous. For maximum polish rates with silica slurry, the pH of the slurry should be in or near the range of 10.5–11.2. In this regime, the surface of the silicon dioxide film is strongly hydroxylated with internal bonds broken by interaction with the alkaline liquid. However, if the pH is much greater than 11.5, the silicon dioxide film will break down entirely and simply begin to dissolve [17].

There is a wide range of silica particle size that is used for oxide CMP. Mean diameters range from about 25 nm for some colloidal silica slurries to about 300 nm for some fumed silica slurries.

The specific properties of the particles and the solutions are covered in detail in Chaps. 3 and 7. The use of other abrasive particles or other liquids generally results in little or no material removal, only some level of surface scratching.

There are several types of polishing pads used for silicon dioxide polishing. Softer pads, such as poromeric pads, are often used for local smoothing or scratch removal, also call buffing. But such pads have poor planarization properties [18]. Urethane based pads are harder and are capable of producing

polished surfaces with longer range planarization (see Chap. 7). Though hard pads other than urethane pads are available, almost all silicon dioxide CMP is done with urethane pads. Urethane pads usually contain spherical pores or voids with diameters in the 30–50 μm range. These pores comprise about one third of the total pad volume, and also the same proportion of the top surface area. The surface of the pads can also be manufactured to have grooves or perforations. This surface pad structure aids slurry transport across the wafer surface.

The most significant feature of the urethane pads that is key to the CMP process is the formation of asperities on top of the pad by the process of conditioning. As noted, a typical conditioner has diamond points embedded in a matrix. This matrix is pressed against the pad while it is moving, and the conditioner is rotated. The surface of the pad is roughened to a level depending on the equipment and operating point. For representative conditions, in the space between the pore openings the pad surface has a roughness, R_a, of 1–5 μm with a spatial frequency of the same dimensions. An example of such a newly conditioned surface is shown in Fig. 2.4.

Empirically, the correct abrasive and liquid for the slurry, as well as pads with specific properties and appropriately conditioned surfaces are all required for the CMP process to occur. These lead to the working model pictured in Fig. 2.5 of how a specific film removal event takes place during CMP.

In Fig. 2.5, the silicon dioxide film is polished when an abrasive particle is forced against the film by an asperity of the pad. The particle is, under the force of the asperity pushing against the film, dragged along by the asperity at the relative velocity of the pad with respect to the wafer. However, the interaction of the particle with the film is not clearly understood. There recently have been proposed several alternative models of asperity-abrasive film
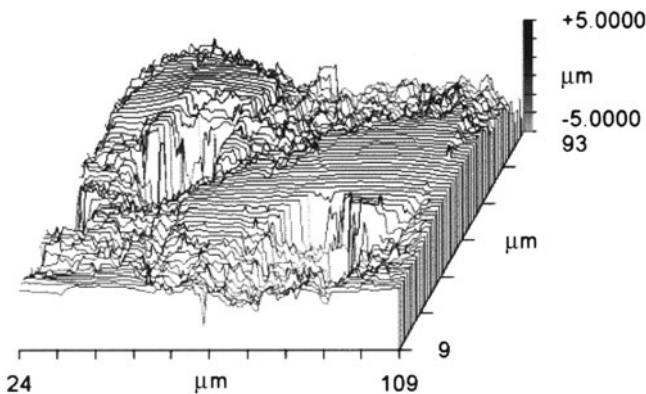
Fig. 2.4. Image of surface of conditioned IC1000 pad showing pores and conditioned surface over a $108 \times 144 (\mu m)^2$ area. Image taken with Zygo NewView 500. Courtesy of Robert Schmidt, Rodel
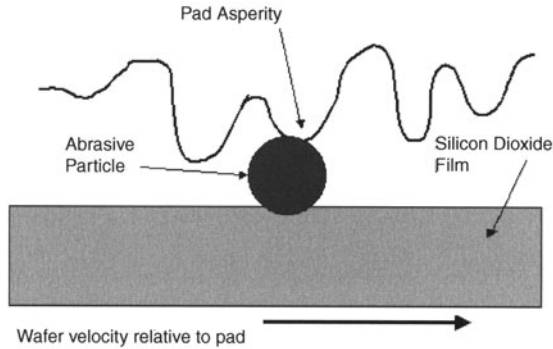
Fig. 2.5. Components (idealized) of film removal by CMP include the abrasive particle forced against the film by a pad asperity. The film to be polished moves relative the asperity with the abrasive

interaction that have led to overall relations between polish rate as various functions of pressure and velocity that do not follow Preston's Law [19, 20]. David Stein has compared several of these models to observed polishing data, especially in low pressure regimes [12], but none is an improvement over the simple Preston model. In an earlier work [5], Cook summarized the models for glass polishing to that date (1990), and much of this discussion is directly applicable to oxide CMP. Unfortunately, up to now there has been no clear quantitative, or semi-quantitive, interaction mechanism model proposed that is in reasonable agreement with the observed data.

The action of individual particles in polishing is repeated continually as the polishing process proceeds. As a result of carrier rotation, there is no preferred directionality for the paths of the polishing particles so that the sum of all the polishing events per unit time is the production of a average removal rate of the film.

2.5 Planarization

In contrast to glass polishing or silicon wafer polishing, for ILD polishing the goal of the CMP process is to planarize topography created by previous semiconductor processing steps. For other polishing steps such as STI CMP processes (see below) or metal CMP processes (see Chap. 3), CMP is used to remove an overburden of one material and stop on another material, leaving a planar surface.

In general, topographical features have different local polishing rates than do planar surfaces. Consider a polishing system where Preston's Law is a good approximation for the local polishing rate over a wide range of pressures, i.e.,

$$RR_0 = K_P P_0 V_0. \tag{2.2}$$

For the case where the velocity, V_0, and the term, K_P, are held constant, the local pressure determines the polish rate. This Prestonian relationship is generally valid for silica slurries polishing silicon dioxide. For the situation where the entire wafer has the same pattern density, then the local pressure on the top of each feature can be determined, as pictured in Fig. 2.6, as the average pressure applied to the pad divided by the pattern density, since the force per unit area applied to the pad is applied over the area of the pattern. For the case where the pattern density is some fraction of the total area, ϱ_1, then the down force is applied to this reduced area, and the polishing rate of each of the features will be increased to

$$RR_1 = K_P P_1 V_0 \qquad \text{here} \quad P_1 = P_0/\varrho_1. \tag{2.3}$$

Or here for the reduced density, ϱ_1,

$$RR_1 = K_P(P_0/\varrho_1)V_0. \tag{2.4}$$

For this sparse region where the feature density is uniform and at a density ϱ_1, the features will polish at a rate determined by the local Preston's Law. For example, if $F_1 = 0.25$, or 25%, then the features will polish at $(1/0.25)$, or 4, times the rate of the planar surface.

The variation in local polish rate with feature density does not require a simple Preston's Law relationship between pressure and polish rate. If the polish rate on a planar surface can be described as

$$RR = f_A(P), \tag{2.5}$$

then for uniform features of density ϱ_K, the removal rate is described by

$$RR_K = f_A(P/\varrho_K). \tag{2.6}$$
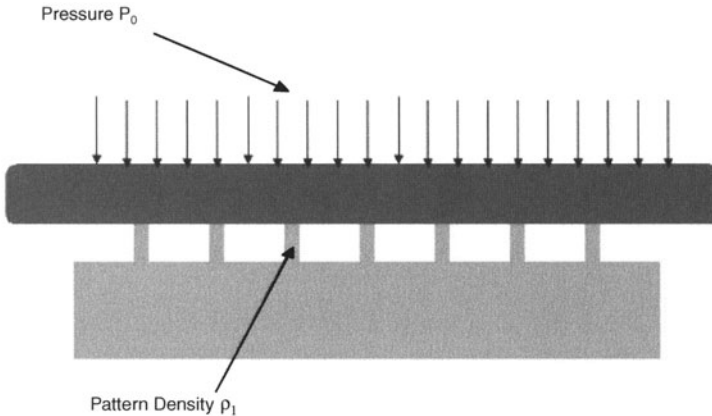
Fig. 2.6. Pressure P_0 is applied to pad and transmitted on to a wafer with a feature density ϱ_1, resulting in a higher local pressure on tops of features (see text)

Such a pressure dependence can and does occur in metal CMP and also in non-silica abrasive silicon dioxide polishing [21].

   The effect of local polish rate dependence for patterned features has been studied by many workers. In Fig. 2.7 is pictured some data showing increased polish rate with decreasing density [22]. When the features are eliminated, the polish rate then reduces to the rate for a planar surface. The time to reach this transition to planarity decreases with decreasing density since the local polish rate increases with decreasing pattern density.
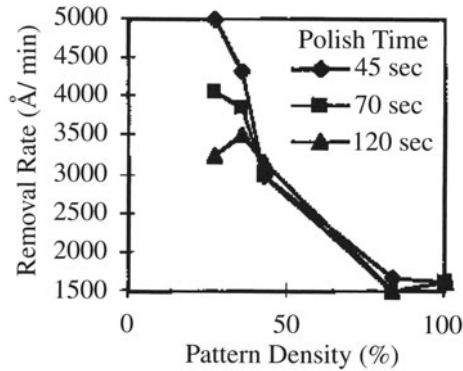
Fig. 2.7. Polish rate with a silica slurry as a function of silicon dioxide structure density. From [20]

   In most CMP systems, the polishing pad is somewhat flexible. Also, in practice, there are variable pattern densities within a die and across the wafer. Thus the pattern density in the vicinity of a given point affects the local polish rate. It has been shown ([21] and Chap. 9) that a weighting function of the local pattern density out to a certain distance can effectively determine the local polish rate. The weighting function is a decreasing function with distance, so that features close to the local area of interest have the greatest effect on the polish rate. The range over which pattern features can affect one another is a function of the polishing system, primarily the thickness and modulus of the polishing pad. Though the range can be modeled to be a fixed length, with no influence beyond that length, the actual interaction decreases gradually. These and related issues are covered in depth in Chap. 9.

   Areas with different local densities that are sufficiently separated will have independent polishing behaviors. Those areas with the lowest pattern densities will polish the most quickly, and those with the highest the most slowly. Once a given independent area is planarized, it will polish at the planar rate. This is pictured in Fig. 2.8 from [21], where low, medium and high density areas are pictured during stages of simultaneous polishing.

   As seen in Fig. 2.8, once the entire wafer has been planarized, different parts of the die and wafer will have different remaining amounts of the original
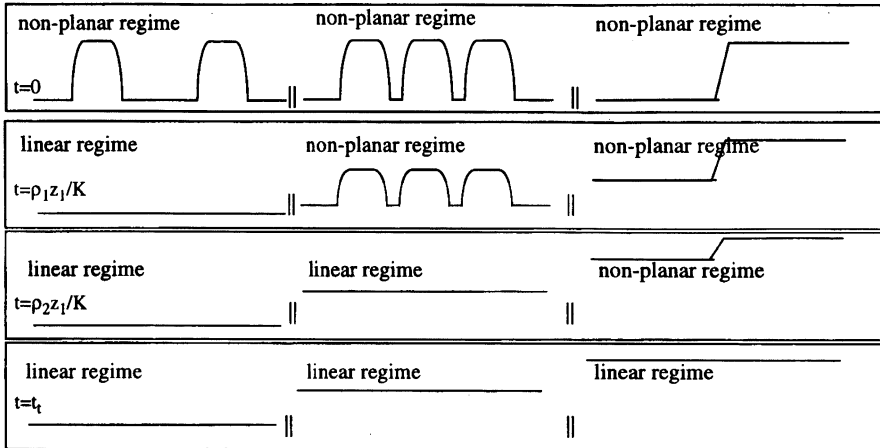
Fig. 2.8. Different final planarization thicknesses remain depending upon initial pattern density. See text. From [21]

film remaining. For adjacent areas the transition between these two areas will occur over a distance generally referred to as the planarization length. This is pictured in Fig. 2.9. The planarization length is a function of the interaction distance of the polishing system, and the amount of the polished film that has been removed during the polish step. Once local planarization is achieved as shown in Fig. 2.8, the planarization length will slowly grow as polishing continues. These lengths are generally hundreds of microns, and this subject also is discussed further in Chap. 9.

There are several consequences of the different clearing times and the resulting longer range thickness variations that exist once the topography has been removed. The first is that different wafers with different patterns will, in general, require different polish times to remove all of the topography. A very sparse metal pattern covered with a deposited ILD silicon dioxide layer
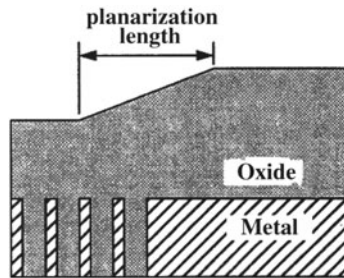
Fig. 2.9. Planarization length, the transition length between post-CMP high and low regions

will have its topography removed more quickly than will a wafer with a dense pattern. This well-known phenomenon often requires different polish recipes for the same process step for wafers with different patterns.

Another problem associated with variable feature clearing time is that, if one area of a chip has a sparse features such as the first pattern in Fig. 2.8 and another part has a dense pattern as does the third pattern, within die non-uniformity (WIDNU) will be large. This problem is a serious one. A frequently used approach to address this is to insert dummy structures so that all areas of the die will have similar feature densities. Dummy structures are isolated features that are designed to affect the CMP step and not have an electrical interaction [25]. In multi-level structures, the post-CMP surface of one level is the substrate upon which the metal and dielectric film of the next level are deposited. Since true planarity is not achieved at each level, the magnitude of the non-planarity can grow with multiple levels and is a key consideration in the integration process (see Chap. 10).

In addition to these long range planarization effects, there is a shorter range phenomenon that occurs when the height of the topography is reduced to the range of 400 nm or less. Ideally, no polishing at the bottom of a step in topography should occur until the step is removed. In practice, however, the bottom of the step begins to be polished before the step is removed [22] and the step height is not reduced at the ideal rate. As a result, in order to remove the step an extra amount of the film below the bottom of the original step must be removed. This additional amount of silicon dioxide that is deposited and then removed is another factor that must be accounted for in the overall integration considerations (see Chap. 10). Representative curves from [22] for various pattern densities are shown in Fig. 2.10. The ideal curves for each of the densities are compared to the data. It is, of course, desirable that the
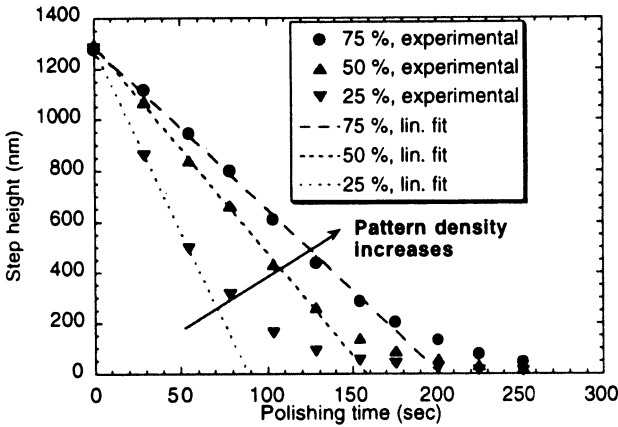
Fig. 2.10. Step height decrease vs. polish time for different density structures. Note departure from linear decrease below about 300 nm. From [22]

departure from the ideal be as small as possible. Slurryless, or fixed abrasive, technology, where the abrasive particles are embedded in the polishing pad, shows great potential for approaching such ideal planarization characteristics.

2.6 Polish Process Variables

The polishing removal rate, at least for silicon dioxide polishing on planar surfaces, is often well described by Preston's Law (2.1). This says that, for a given system, the removal rate is linearly proportional to local pressure and velocity between the pad and polished film. All of the other variables of the system are incorporated into the constant, K_p. These variables include the properties of slurry and pad, as well as the temperature. In addition, the properties of the material being polished are significant.

There is a very large range of possible system operating points, but in semiconductor fabrication, many of the components of the polishing systems are nearly standard across the industry, so variation of polishing performance within a narrow specific range is of most interest. However, as in other areas of technology, substantial changes in operating conditions are continually evaluated, and upon occasion, offer a significant advantage for some performance parameters, and then the new operating conditions (or equipment) are then adopted by a group of users.

Before discussing the variations of system parameters embedded in K_p, the ranges of pressure and velocity will be covered.

2.6.1 Pressure and Velocity Variation

For a given pad, slurry and polish film at ambient temperature, K_p, can be considered constant, and we can consider the pressure and velocity variations. For representative conditions, a typical slurry for polishing silicon dioxide contains 13 wt% solids of silica in a basic solution. For a standardly conditioned urethane polishing pad, a typical polish rate behavior as a function of average down force for a fixed table rotation frequency is shown in Fig. 2.11a. A corresponding curve for polish rate as a function of table rotation frequency for fixed average down force is shown in Fig. 11b. The film being polished is silicon dioxide deposited by plasma enhanced chemical vapor deposition (PECVD), which is a standard semiconductor deposition process. It can be seen that Preston's Law is in reasonable agreement with the data over the range tested, with some departure at very high table speeds.

In current practice, with pads and slurries like the above, average down forces on the wafer rarely exceed 10 psi. This is because the high total forces applied to the pad-wafer-slurry system result in the wafer not traveling smoothly over the pad surface, but sticking at points. This usually leads to wafer breakage or other forms of damage. As a result, with the current
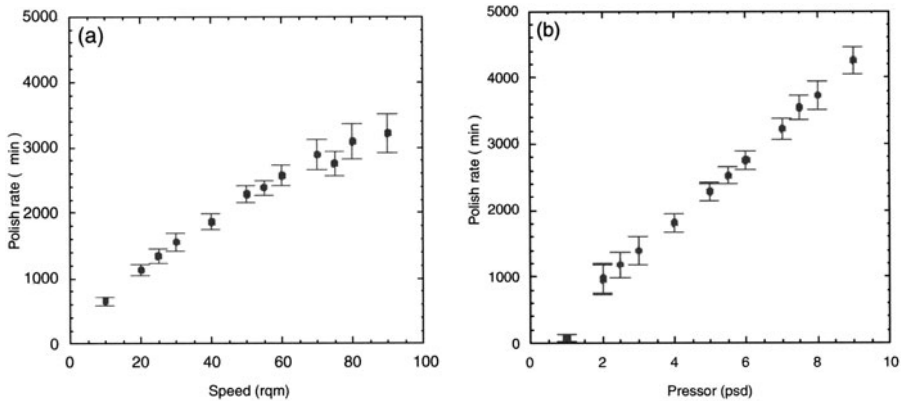
Fig. 2.11. (**a**) Polish rate vs. table rotation rate at 5 psi downforce on Westech 472 polisher. (**b**) Polish rate vs. downforce for 50 rpm table rotation rate on Westech 472 polisher. Both figures courtesy of David Stein

polishing machines, pads and slurries, semiconductor CMP processing is generally done below 10 psi. This, of course, may change over time with machine and pad design evolution. The table rotation rates shown in the two graphs of Fig. 2.11a and 2.11b are for a Westech 372M machine with an average radius position 16 centimeters from the center of the table. The magnitude of the instantaneous linear velocity of any point on the wafer is then given by

$$V = 2\pi r f, \tag{2.7}$$

where

V = magnitude of linear velocity, or speed, at that radius,

r = radius of the given point on the wafer with respect to the center of the table, and

f = rotation frequency of the table.

For this system at 60 rpm, or 1 rps, for the center of the wafer,

$$V = 1.01\,\mathrm{m/s}. \tag{2.8}$$

There is a trend with machine improvements to design machines to operate at higher linear velocities in order to produce higher polish rates. Representative polish speeds for newer equipment designs are up to twice this speed or more. This issue is addressed in Chap. 5. The carrier rotation rate also affects the average speed at a given point on the wafer, and this effect increases as a function of the position on the wafer relative to the center of the wafer. Rotation of the carrier serves to average the polish direction over the entire wafer, but at very high carrier rotation rates, it may change significantly the polish rate near the edge of the wafer. This effect may be used to improve the radial component of WIWNU.

2.6.2 System Factors

The significant factors incorporated into Preston's coefficient, K_p, include:

1. Film type and properties
2. Abrasive particles, type, size, and morphology and concentration
3. Slurry compositon and pH
4. Temperature
5. Pad constitution, both bulk and surface structure (including conditioning effects).

2.6.3 Film Type and Properties

The dielectric films that are of primary interest in CMP are silicon dioxide films, grown or deposited by different processes. Other dielectric films that are polished include silicon nitride and silicon oxynitride. Polycrystalline silicon (poly-Si) is also considered with these films. Metal films are covered in Chap. 3.

In semiconductor technology, silicon dioxide films are used for many different applications. The CMP removal rate is a function of the specific process and operating point by which the silicon dioxide film is formed. Among these different technological approaches are low pressure chemical vapor deposition (LPCVD) and plasma enhanced chemical vapor deposition (PECVD). For each of these approaches, the reactants and operating points (temperature, pressure, ionizing energy, etc.) can vary widely. For one specific use, a region of operation of PECVD called high density plasma (HDP) is employed. It has become the preferred deposition process for shallow trench isolation (STI) structures. The structure and properties of silicon dioxide vary with process and operating point of the deposition process, and these, in turn, influence the CMP removal rate [25, 26]. Specifically, the film density and number of open bonds appear to correlate with CMP removal rate. Thermally grown silicon dioxide is the densest type film used in semiconductor processing, denser than deposited films, and polishes more slowly. This is pictured in Fig. 2.12. In Fig. 2.12, the doped (BPSG) films polish more quickly than do the undoped (USG) films. The denser HDP films polish more slowly than do the APCVD films, with thermally grown silicon dioxide, the densest film, polishing the most slowly. Silicon dioxide films doped with phosphorus and sometimes boron are widely used for the first dielectric layer covering the active devices. In current semiconductor production, these films are now planarized with CMP.

These first dielectric layer films (this level is sometimes referred to as ILD0) can contain varying amounts of boron and phosphorus. The CMP removal rate is a strong function of both dopants. Two graphs of CMP results picturing this dependence are shown in Fig. 2.13. Over the range of dopants tested, the removal rates increase with concentration of both dopants [27, 29].

Remove rate of different oxide films
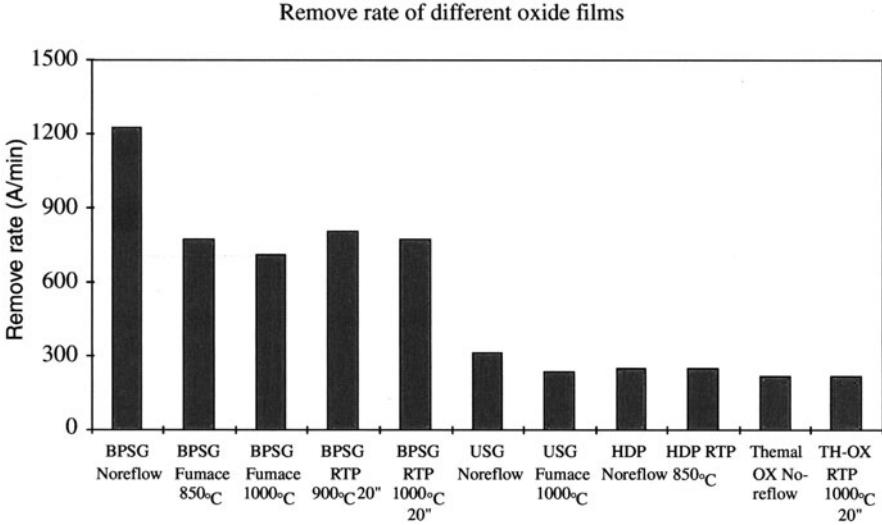
Fig. 2.12. Polish removal rates for different silicon dioxide films. Different deposition processes, dopant concentrations, and post-deposition anneals are compared, using thermally grown silicon dioxide as a reference. From [23]
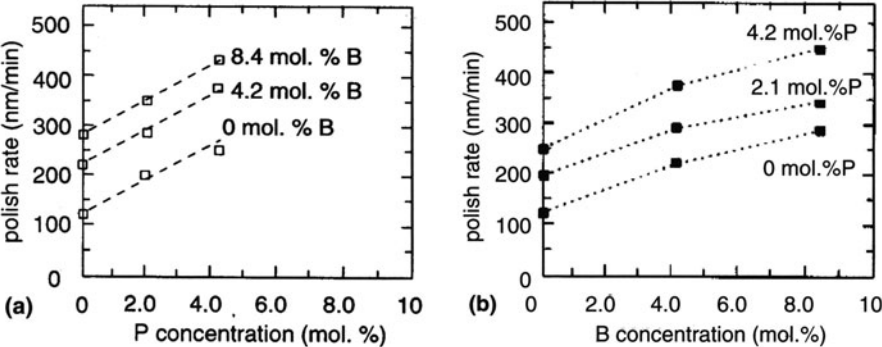
Fig. 2.13. (a) CMP rate of SiO_2 and BSG as a function of P concentration. (b) CMP rate of SiO_2 and PSG as a function of B concentration. From [25]

2.6.4 Abrasive Particles

The type, size, morphology and concentration of the abrasive particles in the slurry strongly influence the polish rate. If we consider initially only silica abrasive particles, there is a wide range of behaviors that are observed with changes in type, size, morphology and concentration. The primary types of silica abrasive used in CMP are fumed silica particles and colloidal silica particles. The fuming process [14] creates tightly bound aggregates of smaller primary particles. The aggregates are of irregular geometry. Colloidal par-

ticles are formed in solution and are, in general, nearly spherical, but the maximum particle size is usually smaller than can be achieved with the fuming process (see Chap. 7).

A typical silica abrasive slurry used in the industry is SS-12^{TM} supplied by Cabot. The concentration of abrasive as well as other properties is listed in Table 2.1. As noted, the pH is near 11. The abrasive particles are created by a fuming process, which is described in Chap. 7 and in [14]. Fumed silica particles employed in consist of an aggregate of tightly bound primary particles about 20 nm diameter with the mean aggregate size in the range of 100–300 nm.

Table 2.1. Properties of Cabot Semi-Sperse 12 (SS-12^{TM}) silica slurry. Courtesy of Cabot Corp

Property	Value
pH	10.9–11.2
Viscosity (cps)	≤ 15
Specific Gravity	1.071–1.078
Mean Aggregate Particle Size(nm)	130–180
% Solids	12.4–12.6

It was noted that abrasive particles are an essential component of the CMP system. Several researchers have shown that, in low concentrations and for other parameters held constant, that polish rate is linearly proportional to particle concentration in the slurry. At sufficiently high concentrations, the polish rate is sublinear with increasing particle concentration. For one type of particle, the colloidal silica used in 30N50pHN^{TM}, supplied by Rodel, Inc., the polish rate vs. particle concentration is shown in Fig. 2.14. As seen in the figure, the polish rate is linear with particle concentration up to about 20 weight %.

Particle size can also play a role, though in the range of particle sizes used in silica slurries, it does not appear to be a strong effect. For very small particles, the rate goes down for a given silica concentration as particle size is reduced. This is reviewed as well in Chap. 7.

2.6.5 Pad Conditioning

Pad conditioning is necessary to maintain the asperity structures on the surface of the polishing pad. The asperities on the pad surface force the abrasive particles against the wafer. The pad asperities need to be continually regenerated because they are worn down by the polishing process. Conditioning done concurrently with polishing is called *in-situ* conditioning while conditioning
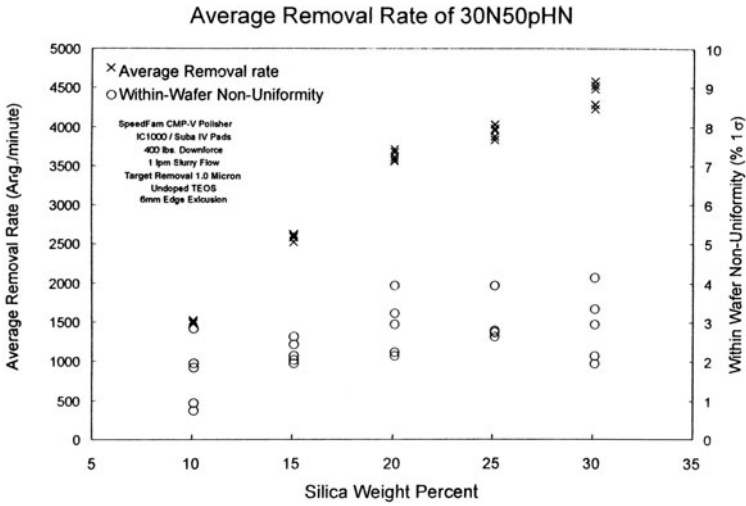
Fig. 2.14. Polish rate of silica slurry, 30N50pHN^{TM}, as a function of silica concentration in weight percent. Courtesy of Rodel, Inc.

done between wafer polishing cycles is called *ex-situ* conditioning. A representative graph of the reduction in polish removal rate when no conditioning is used to maintain the asperity profile is shown in Fig. 2.15, from [30]. The pad was conditioned normally between wafers (*ex-situ*) until this test was started. Then for this set of wafers no conditioning was done at all. Note that the polish rate decay is gradual and begins immediately when conditioning is not used.
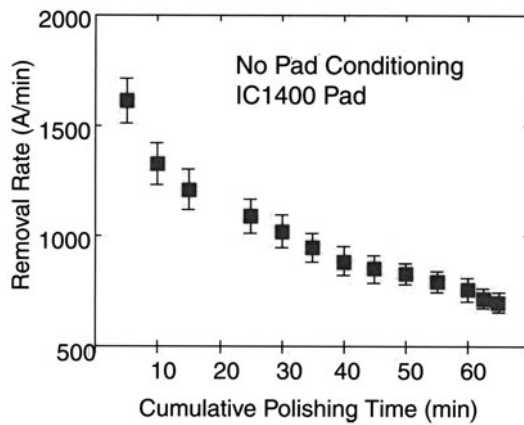
Fig. 2.15. Decrease of polishing rate in the absence of pad conditioning. From [27]

Because the removal rate decay begins immediately when conditioning stops, the latter part of any polishing step using an *ex-situ* conditioning process has a drop in the polishing rate during the polishing step itself [31]. The effective polish rate for a given step then is the average rate during the step and not the maximum rate. By using *in-situ* conditioning this problem is reduced, as the asperities are being generated simultaneously as they are being worn down. Some polishing cycles are 3 minutes long or more, and for these steps *in-situ* conditioning offers substantial throughput advantages.

The surface of a polishing pad is shown in Fig. 2.4. If one examines the land areas between the pore openings, the asperities created by conditioning this surface can be measured. This has been done for *ex-situ* conditioning where the polish rate over shourt intervals has been compared to the average asperity height measured on small coupons removed from the polishing pad [28].

The results for a standard IC1000$^{\text{TM}}$ pad taken over eight one minute intervals, with no intermediate conditioning, are shown in Fig. 2.16a and 2.16b. The average removal rate for each interval as well as the average asperity height is shown for areas between the pore openings, where the asperities created by conditioning this surface can be measured. This has been done for *ex-situ* conditioning where the polish rate over short intervals has been compared to the average asperity height measured on small coupons removed from the polishing pad [28].

The results for a standard IC1000$^{\text{TM}}$ pad taken over eight one minute intervals, with no intermediate conditioning, are shown in Fig. 2.16a and 2.16b. The average removal rate for each interval as well as the average asperity height is shown. As the asperity rate decreases, so does the polish rate. For representative polish conditions and rates, the asperity heights are in the 1–2 µm range. Note that the average asperity heights are much smaller than the average pore size of 30–50 µm.

2.6.6 Temperature and pH

Temperature and pH also affect the polish rate. As the pH increases in the regime near pH = 11, the polish rate increases. In practice, it is difficult to operate much above pH = 11.5, as the silica particles in the slurry begin to dissolve with time so that that the abrasive particles are not stable over time [15]. A representative curve of polish rate vs. pH is shown in Fig. 2.17. Also plotted is the polish rate vs. pH for CVD silicon nitride. From pH = 9.7 to pH = 10.7, the polish rate increases by about 20%. It is difficult to maintain silica in solution near and above pH = 11.5, so most silica slurries are made with pH near 11.

There is also an increase of silicon dioxide polish rate with ambient temperature near and above room temperature [26]. In [26], the authors attribute the increase to changes in the hard (IC1000$^{\text{TM}}$) pad mechanical properties. Fig. 2.18 shows data from [26] that demonstrates increased removal rate with
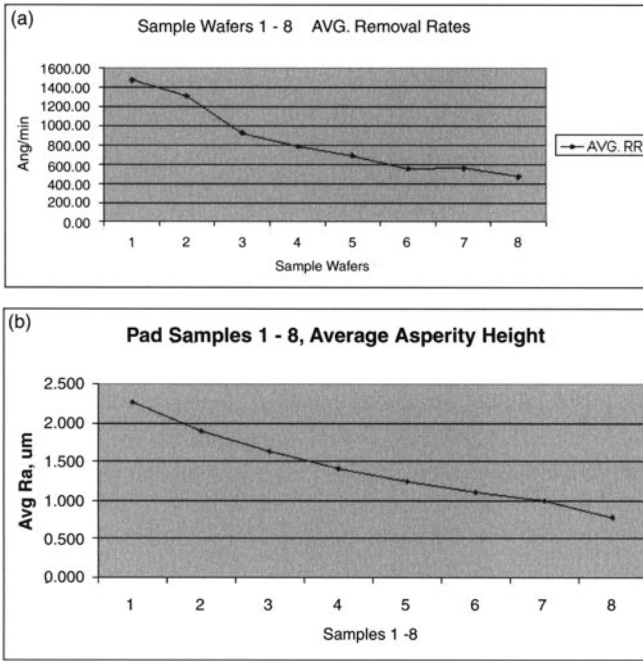
Fig. 2.16. (**a**) The average polish rate for 30 second intervals with no conditioning. (**b**) The average asperity height on the areas between the IC1000^{TM} pores at the end of the 30 second intervals shown in Fig. 2.16a. The measurement was made by a Zygo NewView 5000. From [26]

higher temperatures. In practice, most polishing machines have systems to heat or cool the platens in order to optimize a given polishing process.

2.7 Scales and Random Polishing Effects

The major variables of the silicon dioxide CMP process have been discussed on an elemental scale. These variables are the factors that locally affect polish rate, including properties of the silicon dioxide film being polished as well as the pad and slurry properties.

On the smallest scale, as pictured in Fig. 2.5, the silicon dioxide film removal occurs when an abrasive particle is forced against the silicon dioxide film by an asperity, and the relative velocity of the asperity (on the pad) and the film creates a path of film removal. Global film removal occurs as this action is repeated a very large number of times. The directions of the polishing paths are randomized by having the film (on the wafer) rotate relative to the pad during the CMP process.
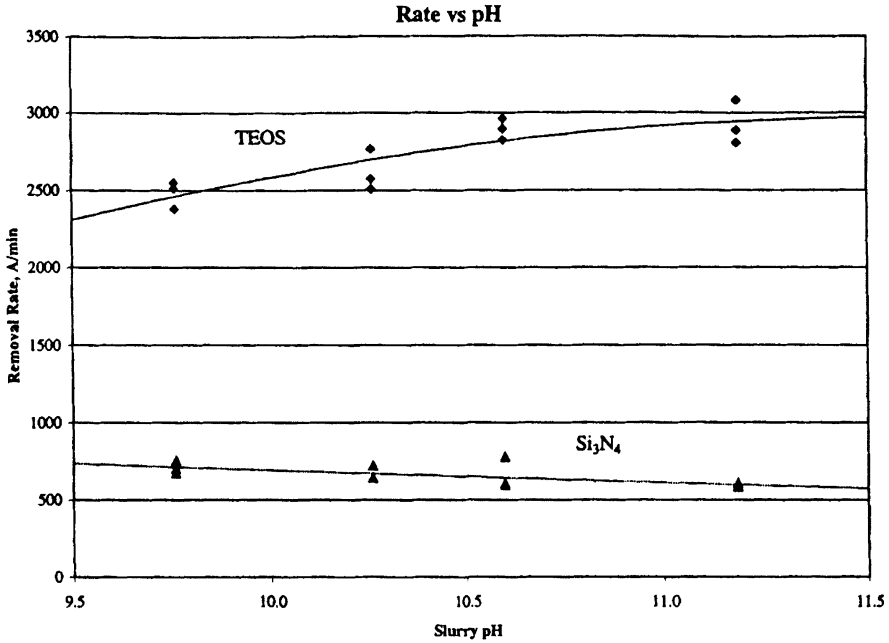
Fig. 2.17. Variation of CMP removal rate vs. pH for two films. Note that the TEOS (silicon dioxide) removal rate increases with increasing pH in this pH range

Local planarization is achieved when the higher points of the film surface topography are removed more quickly than the lower points. Global planarization issues associated with film pattern density variations have been discussed above. In addition to pattern density effects, there are several other issues that affect global planarization, especially on the wafer scale.

In Fig. 2.19, which is similar to Fig. 2.5, but on a somewhat larger scale. The dimensions of the features are near to scale. Metal lines are about $1\,\mu m$ high, and the range of asperity heights is 1–$3\,\mu m$. Here, the silica particles are pictured as irregular, but of course the shape and size distribution is determined by the manufacturing process (see Chap. 7). The asperity heights in a given process depend upon several parameters, with pad type and conditioner and conditioning process being the most influential.

If we look at the polishing system on a yet larger scale, features of the pad structure other than the asperity profile appear. In Fig. 2.20, the pores of an IC1000^{TM} pad are pictured and grooves are also shown. Both of these features enhance slurry flow between the wafer and the polishing pad (see Chap. 6). However, pads without grooves are sometimes used.

These local, random variations of the asperities and pore structure appear on a small scale. The statistical averages of these properties, such as the average asperity height or the specific gravity of the pad (as discussed in
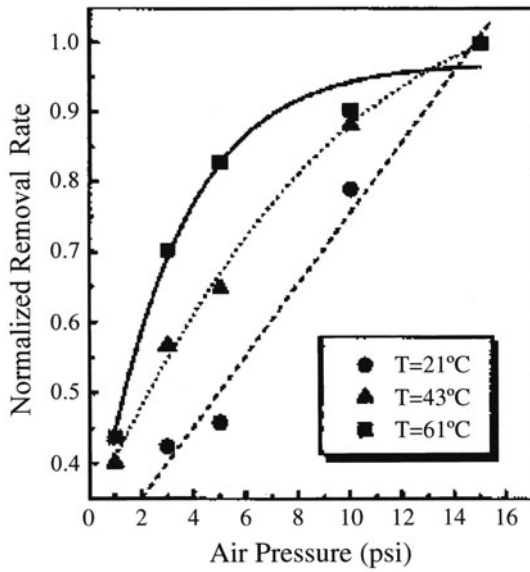
Fig. 2.18. The normalized oxide removal rate vs. air pressure used to provide the normal CMP load. Here, T is the slurry temperature, and the oxide removal rate for each slurry temperature is normalized to its highes value, respectively. In this experiment, an IC1000 pad was used. From [23]
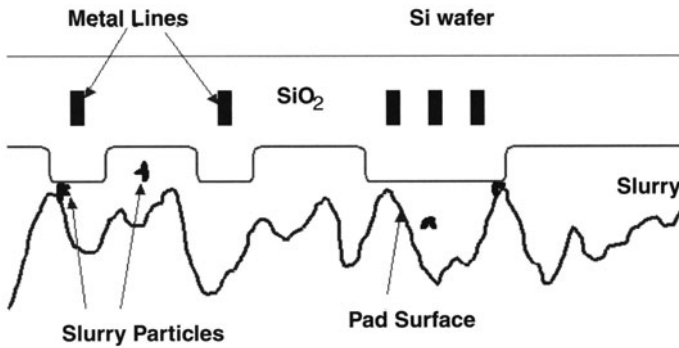
Fig. 2.19. Diagram of the elements of the CMP process showing a larger region than that of Fig. 2.5

Si wafer with SiO$_2$film
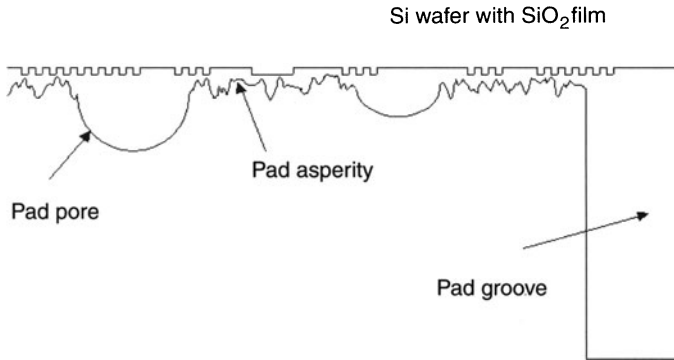
Fig. 2.20. Wafer with $\approx 2\,\mu$m features with asperities of comparable dimensions and pores of 30–50 μm diameters. Also the edge of a pad groove is shown

Chap. 6), do affect the observed polishing behavior. This is the result of the averaging process of a great many polishing events over large areas of the polishing pad in all directions.

2.7.1 Polish Rate and Other Variations Introduced by the System

CMP polishing systems have matured to produce consistent, well controlled processes at all scales, within-die, within-wafer, and wafer to wafer. The general approach has been to provide as uniform polishing conditions as possible at all levels. Because of wafer and table geometries, both the down force and velocity applied to the wafer are not uniform across the wafer nor over time. To minimize the effect on the resultant polished film of these variations, table and carrier rotation are employed. In addition compressible elements, such as carrier films and bottom pads, are used to provide a more uniform applied down force. Looking at the wafer and system as a whole, forces are provided between the back of the carrier film and the platen surface. However, the polishing process takes place at **the interface between the pad surface and the film being polished**.

In polishing machines, the local velocity over time at the pad surface-film interface is very well determined by the geometry of the machine. The local pressure, in contrast, is sensitive to any lateral dimensional variations in the layers of materials between the ideal carrier head surface and the ideal platen surface. Lateral thickness variations over the wafer and over all the wafer paths traversed across the rotating platen will create a time and pattern dependent variation of pressure at the pad surface-film interface at any specific point or thereafter. In addition, if the elastic constants of the compressible elements, the carrier film and bottom pad, change with position or slowly with time as many wafers are polished, these compression changes will also produce non-uniform applied pressure with respect to position and time.

The methods to improve the uniformity of applied pressure at the pad surface-film interface have focused on machine design improvements (see Chap. 5), and on replacing pads and carrier films when within-wafer non-uniformity becomes too large during production CMP. Two major design approaches that have been implemented on new machines and sometimes retrofitted on older machines are 1) the position dependent conditioning which is designed to keep the thickness of the polishing pad uniform and not allow a trough to form in the track of wafer travel, and 2) the fluid backed wafer carrier, where the local pressure applied across the back of the wafer is by a fluid, applied either directly to the back of the wafer or through a thin membrane. Both of these improvements are discussed in Chap. 5.

A second issue is the edge effect, which is a strong variation in the polish removal rate as a function of radial position near the edge of the wafer. This pressure variation and the observed polish rate as a function of radial position are pictured in Fig. 2.21a and 2.21b from [28]. This effect can be reduced by varying the bottom pad stiffness and thickness as discussed by Baker. Note that, while carrier rotation can average out, to a large degree, the leading edge to trailing edge variation shown in Fig. 2.3, it will not affect the edge effect since the magnitude of the effect only has radial dependence.
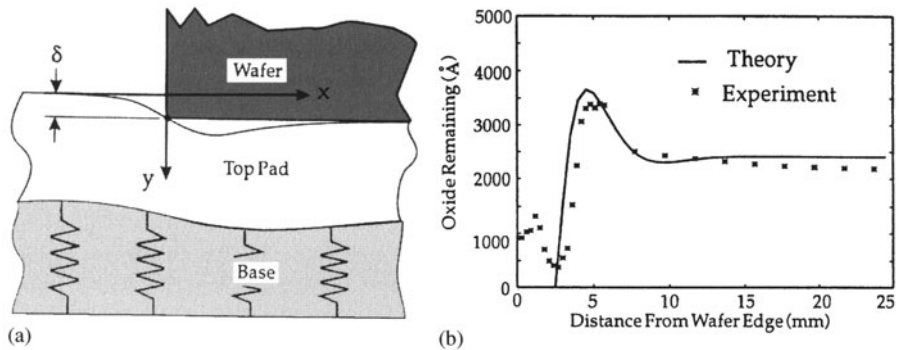
Fig. 2.21. (a) Model of pad structure that produces the edge effect. From [28]. (b) Model and data for remaining silicon dioxide at the edge of the wafer. From [28]

2.8 Random Effects

When we consider the entire wafer in the CMP system, there are several longer range mechanisms that can alter the polish rate across the wafer. One, discussed above, is the gimbal-induced force that causes the leading edge of the wafer to have a higher pressure on the pad than the trailing

edge, resulting in higher polish rate on the leading edge as shown in Fig. 2.3. This higher pressure on the pad can be reduced by lowering the effective gimbal point of the carrier, or by avoiding the gimbal effect altogether by providing pressure uniformly [32] to the back of the wafer by a fluid under pressure.

In the initial description of the pads, wafer and carrier films, the soft bottom pad and carrier film were required to make the applied pressure at the wafer-spolishing pad interface more uniform. The soft pads and carrier film improve the pressure uniformity, but since they act as springs they do not eliminate it. Referring to Fig. 2.2, dimensional variations in all the elements (wafer, pads, and carrier film) and elasticity variations in the bottom pad and carrier film all can lead to local pressure variations across the wafer at the wafer-polishing pad interface.

The effects of the variations in elasticity of the bottom pad and the thickness variations of both pads are reduced because of the averaging effects of table and carrier rotations. However, carrier film thickness and elasticity variations and wafer thickness variations are not reduced by this averaging and so lead to local polished film variations.

2.9 Slurries with Particles Other than Silica

Though silica-based slurries are primarily used for CMP of silicon dioxide and other dielectrics, other particles can also be used. As noted earlier, ceria, CeO_2, is known from glass polishing experience to be much more effective than silica , in terms of polishing rate, at polishing a glass film. This naturally has led to its evaluation as a CMP polishing abrasive.

The data for polishing with ceria abrasives [33] shows that indeed that, per abrasive particle, ceria polishes planar surfaces much more effectively than does silica. As shown in Fig. 2.22, the polish rate for planar silicon dioxide surfaces is higher with a slurry containing 0.5 wt% ceria than for silica based slurries containing 13 wt% silica. Ceria behaves differently than does silica for structured surfaces however. For a structured surface such as an ILD surface, silica based slurries polish the high areas at a rate that is higher than the planar rate. For ceria based slurries, this effect is much less pronounced, and sometimes the initial polish rate can be lower than the rate for planar surfaces [33]. However, this effect appears to be sensitive to the type of ceria used, as well as to certain additives for some slurries. Ceria slurries are also very sensitive to the nature of the silica film being polished. Diffferent depostion conditions can lead to a very different polishing characteristics [35, 36].

Another abrasive that has been tested for polishing silicon dioxide is manganese sesquioxide, Mn_2O_3 [34]. However it has not been extensively investigated, and though some initial data is interesting, the observed polish rate vs.
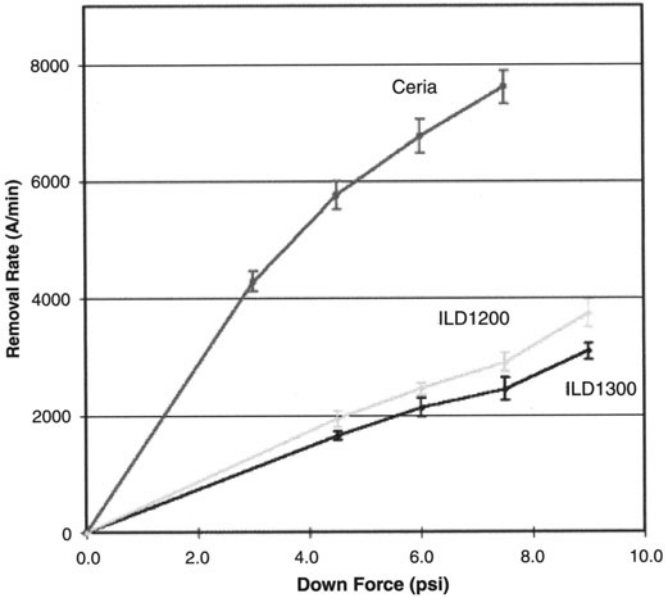
Fig. 2.22. Polish rates for two silica slurries with 13 wt% silica compared to a ceria slurry with 0.5 wt% ceria. The rotation rate is 40 rpm. From [33]

abrasive concentration is very non-linear, as shown in Fig. 2.23. This abrasive is used as a slurry in an alkaline pH range, but if the wafer is cleaned in an acidic solution, the residual particles are dissolved, thus simplifying the post-CMP cleaning of the wafers.
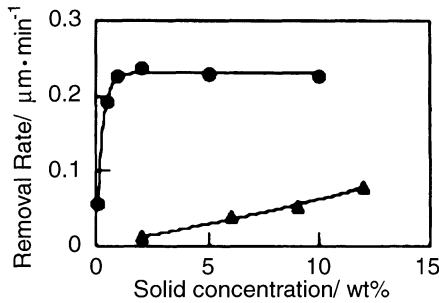
Fig. 2.23. Polish rate dependence of Mn_2O_3 slurry (●) compared to a silica slurry (▲). Note the very large polish rate as a function of concentration in the low concentration region. From [31] (©2003 IEEE)

2.10 Non-ILD Non-Metal CMP

2.10.1 Shallow Trench Isolation (STI)

The maturity of CMP technology aided the rapid introduction to semiconductor manufacturing of Shallow Trench Isolation (STI) technology [37]. As discussed in Chap. 10, this technology replaces LOCOS (LOCal Oxidation of Silicon) as an isolation approach. STI allows much tighter packing of transistors, thus increasing the number of transistors per unit area, with design rules being equal. As lithography dimensions, and their associated design rules, become smaller the relative advantage of STI vs. LOCOS becomes greater.

The standard approach to the CMP of an STI structure is shown in Fig. 2.24. The silicon substrate has a thin silicon dioxide buffer layer and a CVD silicon nitride layer on top of it. This stack is patterned and etched, with shallow trenches etched into the silicon substrate. The silicon nitride areas are where the active transistors will be placed. The structure is then filled with silicon dioxide, usually deposited with a technique called high density plasma (HDP). This deposition approach is able to fill very small trenches without leaving voids.

The CMP step is to remove all the silicon dioxide from the top of the silicon nitride, while polishing as little of the silicon nitride as possible. Then, in the overall process sequence the nitride is removed and the buffer silicon dioxide layer is etched off with hydrofluoric acid. The transistor gate oxide is then formed and polysilicon is then deposited as the gate electrode material.

CMP of the silicon dioxide-silicon nitride system has its unique considerations. When a standard silica based slurry is used, the silicon dioxide polish rate, for the same CMP conditions, is about three times greater than that of silicon nitride. This helps control the variation of the post-CMP thickness both within the die and within the wafer. Because of severe integration restrictions, as discussed in Chap. 10, a very tight control on the remaining silicon nitride thickness is critical. This is because the structure and shape of the edge of the transistor at the substrate-trench interface determine the electrical performance of the transistors. If there is much variability in the shape of the gate electrode edge, then the transistor performance and process yield will degrade.

The CMP step is to remove all the silicon dioxide from the top of the silicon nitride, while polishing as little of the silicon nitride as possible. Then, in the overall process sequence the nitride is removed and the buffer silicon dioxide layer is etched off with hydrofluoric acid. The transistor gate oxide is then formed and polysilicon is then deposited as the gate electrode material.

The behavior of CMP when polishing patterned wafers with variable density affects the STI process as it does the ILD process. Isolated raised features are polished quickly and dense areas are polished more slowly. This means
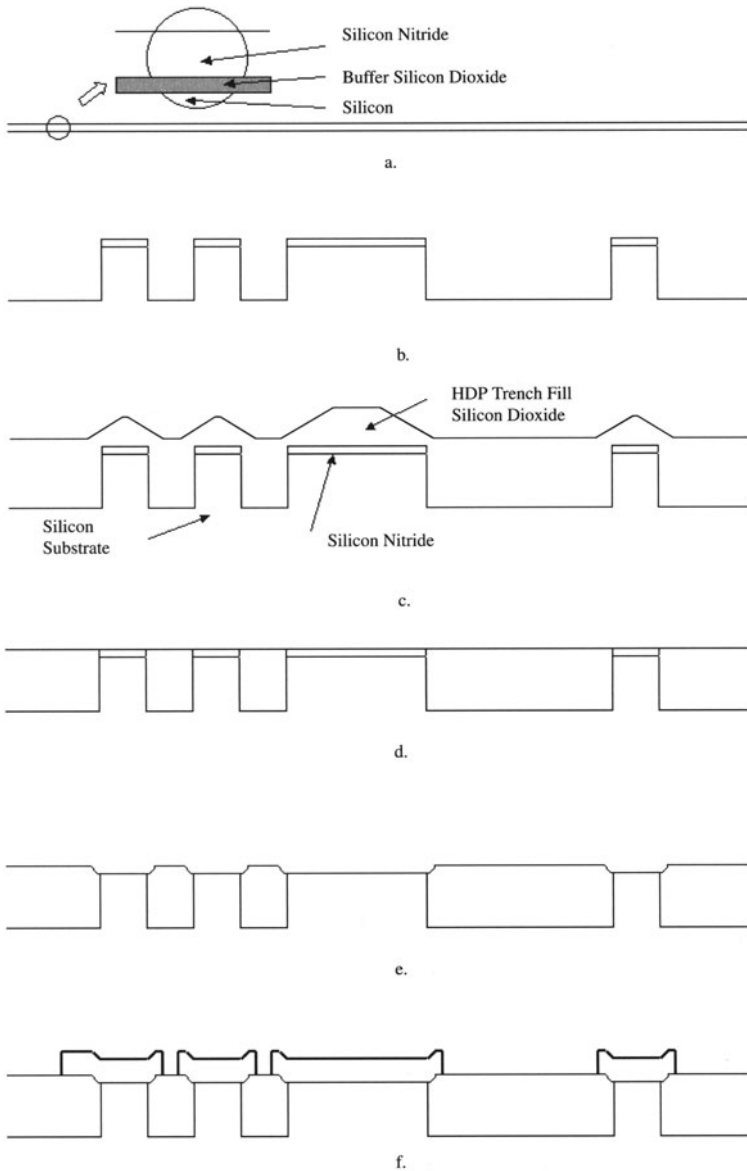
Fig. 2.24. Shallow Trench Isolation formation, including deposition and definition of poly Si gate. Starting with silicon nitride-buffer silicon dioxide films, the active areas are defined and the trenches are etched (**a,b**). After filling the trenches with silicon dioxide (here, HDP oxide), the oxide is planarized with CMP (**c,d**). Then the silicon nitride is removed, and the buffer silicon dioxide is also removed. Finally, the gate oxide is deposited and the poly-Si gates are formed (**e,f**)

that the nitride layer on an isolated feature is exposed and polished for longer times than is the nitride layer on a feature in a dense area.

Clearly, WIWNU and WIDNU of the remaining nitride and oxide thicknesses need to be controlled very tightly. The WIWNU issues are similar to those for ILD processing but the WIDNU issues are different. One major direction to minimize the WIDNU is to minimize the density variations so that, within a planarization length or so, all areas of the die have the same density pattern of silicon dioxide above the silicon nitride layer. In this way, the die itself will have less contribution to the variation in the time to clear the silicon dioxide from the silicon nitride.

Several approaches have been proposed to make the silicon dioxide density nearly uniform across the die. A reverse mask step is an approach that several groups have pursued [38]. This sequence is pictured in Fig. 2.25. Here, after the silicon dioxide is deposited, a reverse mask of the STI pattern, with features made slightly smaller than the STI mask itself, is patterned and the exposed silicon dioxide etched away. This leaves a fence of silicon dioxide around the edge of each feature but the total amount of oxide above the silicon nitride is sharply reduced.

A second approach is to make the density of STI structures as uniform as possible. This can be done by inserting dummy structures into the STI mask pattern. These structures are not to be anything more than isolated islands of conducting silicon. The dummy structure size and shape should approximate the active areas of most of the chip; usually this means minimize size transistors. This averaging by the use of dummy structures can only be approximate but it certainly can minimize the extremes of local pattern density variation [39, 40].

Another way to improve the performance of the CMP module is to make the slurry very selective. Slurry selectivity is the ratio of the polish rate of one film compared to another at the same CMP operating conditions. Here,
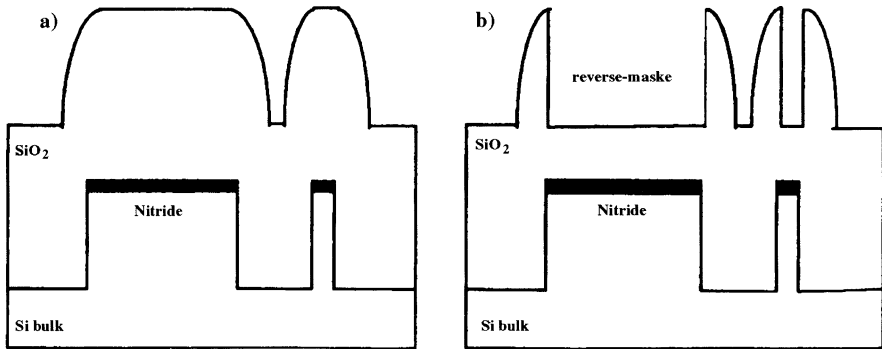
Fig. 2.25. The effect of reverse, or counter, mask etch back to reduce the amount of silicon dioxide over the silicon nitride prior to STI polish. From [34]

selectivity means rate selectivity, the ratio of the removal rates of silicon dioxide and silicon nitride. As noted above, for standard alkali-based silica slurries this selectivity is about 3:1. By using appropriate additives, very high selectivities, greater than 100:1, can be obtained [41]. These very high selectivity slurries can substantially reduce the non-uniformity of remaining silicon nitride across the die and wafer.

The use of high selectivity slurries is standard in metal CMP (see Chap. 3), and there are a number of unique effects that occur. Recess, or dishing, is one effect that occurs when the more slowly polishing layer is exposed. At that point the faster polishing layer continues to polish creating a dished area next to the slower polishing layer. For STI polishing this effect is very important and needs to be minimized [42].

Recess will affect the geometry of the gate electrode at the silicon-trench interface and needs to be controlled as tightly as the amount of nitride removal. Recess is influenced by the polishing operating conditions, most importantly the overpolish time required to clear the silicon dioxide from all the nitride structures [43, 44].

2.10.2 Polysilicon Polish

Silicon, including polysilicon, can be polished easily with basically the same types of polishers, and similar pads and slurries, that are used to polish silicon dioxide. In general, for the same polishing conditions, polysilicon (or polycrystalline silicon) deposited by LPCVD systems polishes more quickly than does silicon dioxide. However, with appropriate additives to the slurry, the rate selectivity between polysilicon and silicon dioxide can be made greater than 100:1 [45]. Using such slurries, several polysilicon CMP steps have been used. One is to polish polysilicon plugs, or vias, removing the layer of polysilicon on top of the ILD and leaving the plug filled with polysilicon [7, 46]. Another step that has been used is for polishing polysilicon to smooth it for subsequent processing [47].

In addition to these more widely used steps, there also have been integration sequences that have used polysilicon CMP in the formation of STI structures as well as interconnect structures.

2.10.3 Low K Dielectrics

With successive generations of semiconductor processed, the dimensions shrink but the materials also change. Copper is replacing aluminum because it has a lower resistivity. Also, dielectrics with lower permittivity, or dielectric constant (K), will replace silicon dioxide. Copper CMP and its associated issues are addressed in Chap. 3. The integration issues associated with different low K dielectrics are discussed in Chap. 10. In addition to the different low K dielectrics that are available, there also are several options for integration

approaches for the low K dielectrics. Some of these approaches employ thin barrier layers which are used as etch stops and CMP stopping layers.

One of the first generation lower K dielectrics to be integrated into a production semiconductor process was fluorinated silica glass, or FSG [48, 21]). The permittivity of undoped silica is about 3.9, and by doping with fluorine a reduced permittivity of 3.5–3.6 can be achieved. There are reliability issues which limit the fluorine concentration in the silica. The CMP removal rates for these glasses are at or slightly above those of undoped silica for the same CMP process conditions.

The dielectric materials that are used for permittivities lower than that of FSG behave, in general, very differently from silica. There are new materials being developed continually to provide improved performance of those already available. Types of materials include spin-on homogeneous organic based materials, such as SiLK^{TM} [49] and silsesquioxanes [50, 51]. A second group of materials being investigated closely include the CVD deposited carbon-doped silicas, with Black Diamond^{TM} [52] and Coral^{TM} being two commercially available materials. Another large group of materials are the porous materials where pores, or voids, are contained in the bulk of the material [53, 54]. Materials with voids have produced very low permittivities, some below 2.0, but are generally mechanically weak and are difficult to polish.

Most of the integration approaches for incorporating these low permittivity materials into the back end structure use, as noted, stop layers and the CMP process does not see the low K material. However, all materials have to be robust enough not to degrade under the pressure of the polishing process, and also must maintain good adhesion to their surrounding materials. Adhesion of many of these low permittivity materials is a significant issue.

At present, copper is being integrated into semiconductor processes, and the integration of low permittivity materials is lagging in its implementation. This lag is largely due to the many difficulties that have been encountered in integrating any low permittivity material beyond FSG into a multi-level metal process.

2.11 Conclusion

CMP of silicon dioxide was the initial application and is still the largest application of CMP in the semiconductor industry. A majority of the characterization of CMP processes has been on silicon dioxide processes. For this reason, the initial focus of the book has used these processes as the baseline. In addition, the characterization and evaluation of the consumables, pads and slurries, as well as polishing tools, has focused on silicon dioxide polishing.

Almost from the beginning of its application to semiconductor processing, CMP has been used for metal polishing as well as for dielectrics. It has also become an enabling technology for the introduction of copper as an interconnect metallization, since copper cannot be etched in a practical way in

semiconductor technology. The treatment of both the technology and modeling of metal polishing build upon, but are quite different from their silicon dioxide CMP counterparts.

For the above reasons, the book has been organized with the silicon dioxide CMP technology as the first technology chapter. The following chapters have addressed metal CMP technology and models, and then the hardware of the CMP process, pads, slurries and polishing tools. Finally, other key areas are addressed, including topography evolution and modeling, post-CMP cleaning, and overall process integration.

The application of CMP in semiconductor processing was introduced in the late 1980's, and is now a critical technological component in driving improved chip performance. Only in the past few years, though, has the scientific basis for the technology begun to receive much interest. At present, the number of publications focusing on a detailed understanding is small but is growing quickly. It is hoped that this book will provide a background that will enable the reader to be able to read the current literature and understand the science and technology of CMP as it evolves in the near future.

References

1. M. Fury, "The Early Days of CMP", *Solid State Technology*, 81, May 1997.
2. D. Pramanik, V. Jain and K.Y. Chang, Proceedings 1991 VMIC Conference, 27, 1991.
3. D. Moy, M. Schadt, C.K. Hu, F. Kaufman, A.K. Ray, N. Mazzeo, E. Baran and D.J. Pearson, Proceedings 1989 VMIC Conference, 26, 1989; See also U.S. Patent 4,944,836, July 31, 1990.
4. G. Nanz and L.E. Camilletti, IEEE Trans. Semiconductor Mfg., **8**, 382, 1995.
5. Cook, L. M., J. of Non-crystalline Solids, **120**, 152, 1990.
6. C.W. Kaanta, W.J. Cote, J.E. Cronin, K.L. Holland, P.I. Lee and T.M. Wright, IEDM Tech. Dig., 769, 1997.
7. C.W. Kaanta, W.J. Cote, J.E. Cronin, H.S. Landis, W. Hill and J. Ryan, Proceedings 1991 VMIC Conference, 144, 1991.
8. S.S. Cooperman, xxx, Journal Electrochem. Soc., **9**, 3180, 1994.
9. S.D. Hosali, A.R. Sethuraman, J-F. Wang, L.M. Cook and D.R. Evans, Proceedings 1997 CMP–MIC Conference, 52, IMIC, Tampa, 1997.
10. K. Itabashi, S. Tsuboi, H. Nakamura, K. Hashimoto, W. Futoh, K. Fukuda, I. Hanyu, S. Asai, T. Chijimatsu, E. Kawamura, T. Yao, H. Takagi, Y. Ohta, T. Karawawa, H. Iio, M. Onoda, F. Inoue, N. Nomura, Y. Satoh, M. Higashimoto, M. Matsumiya, T. Miyabo, T. Ikeda, T. Yamazaki, M. Miyajima, K. Watanabe, S. Kawamura and T. Taguchi, 1997 Symposium on VLSI Digest, 21, 1997.
11. F. Preston, J. Soc. Glass Technology, **11**, 214, 1927.
12. D.J. Stein and D.L. Hetherington, Electrochemical Soc. Proc., **99–37**, 217, 1999.
13. I. Ali, R. Roy, G. Shinn and C. Tipton, Proceedings 1997 CMP–MIC Conference, 311, IMIC, Tampa, 1997.

14. D.R. Evans and M.R. Oliver, Proc. MRS, San Francisco, 2001.
15. P. Truong and L. Blanchard, Proceedings 1998 CMP–MIC Conference, 351, IMIC, Tampa, 1998.
16. H. Muelenweg, F. Klaessig, W. Lortz, G. Varga and A. Gutsch, Proceedings 2000 CMP–MIC Conference, 325, 2000.
17. R.K. Iler, *The Chemistry of Silica*, 42, Wiley, New York, 1979.
18. Y.C. Shih, E.M. Shamble, J-F. Wang, A.R. Sethuraman, H-M. Wang, R.L. Lavoie and L.M. Cook, Proceedings 1997 CMP–MIC Conference, 237, IMIC, Tampa, 1997.
19. B. Zhao and F.G. Shi, Proceedings 1999 CMP–MIC Conference, 13, IMIC, Tampa, 1999.
20. J-Z. Zheng, V. Huang, M. Toh, C. Tay, F. Chen and B.B. Zhou, Proceedings 1997 CMP–MIC Conference, 315, IMIC, Tampa, 1997.
21. D.R. Evans, B.D. Ulrich and M.R. Oliver, Proceedings 1998 CMP–MIC Conference, 347, IMIC, Tampa, 1998.
22. M. Weling, S. Bothra, C. Drill and C. Gabriel, Proceedings 1997 CMP–MIC Conference, 65, IMIC, Tampa, 1997.
23. B. Stine, D. Ouma, R. Divecha, D. Boning, J. Chung, D.L. Hetherington, I. Ali, G. Shinn, J. Clark, O.S. Nakagawa and S-Y. Oh, Proceedings 1997 CMP–MIC Conference, 266, IMIC, Tampa, 1997.
24. J. Grillaert, M. Meuris, N. Heylen, K. DeVriendt, E. Vrancken and M. Heyns, Proceedings 1998 CMP–MIC Conference, 79, IMIC, Tampa, 1998.
25. C.N. Huang, H.B. Liu and J.T. Lin, Proceedings 1998 VMIC Conference, 632, IMIC, Tampa, 1998.
26. I.J. Malik, T. Mallon, B. Withers, R. Emami, D. Mordo and I. Goswami, Proceedings 1997 CMP–MIC Conference, 209, IMIC, Tampa, 1997.
27. M.A. Jaso, J.P. Gambino, K. Huckels, M. Ilg and G. Coleman, Proceedings 1997 CMP–MIC Conference, 15, IMIC, Tampa, 1997.
28. M.R. Oliver, R.E. Schmidt and M. Robinson, Electrochemical Soc. Proc., **2000–26**, 77, 2000.
29. C.W. Liu, B.T. Dai and C.F. Yeh, Thin Solid Films, **270**, 607, 1995.
30. D.L. Hetherington and K. Achuthan, Semicon West CMP Symposium Proceedings, San Francisco, July 1998.
31. A.R. Baker, Electrochemical Soc. Proc., **96–22**, 228, 1996.
32. T. Osterheld, S. Zuniga, S. Huey, P. McKeever, C. Garretson, B. Bonner, D. Bennett and R.R. Jin, Proc. MRS Symposium, **566**, 63, 2000.
33. S-I. Lee, C-I. Kim, H. Kim, J-H. Kim, C-W. Nam, S. Kim and C-T. Proceedings 1997 CMP–MIC Conference, 163, IMIC, Tampa, 1997.
34. S. Kishii, R. Suzuki, A. Ohishi and Y. Arimoto, IEDM Tech. Dig., 465, 1995.
35. M.R. Oliver, D.R. Evans, D.L. Hetherington, D.J. Sten, J.E. Stevens and S.D. Hosali, Proceedings 1999 CMP–MIC, 383, IMIC, Tampa, 1999.
36. D.J. Stein, D.L. Hetherington, M.R. Oliver, S.H. Hosali, D.R. Evans and B. Her, MRS, Paper M3.8, San Francisco, 2001.
37. J.T. Pan, D. Ouma, P. Li, D. Boning, F. Redeker, J. Chung and J. Whitby, Proceedings 1998 VMIC Conference, 467, IMIC, Tampa, 1998.
38. M. Jouty, M. Rivoire and T. Detzel, Proceedings 1999 CMP–MIC Conference, 329, IMIC, Tampa, 1999.
39. G.Y. Liu, R.F. Zhang, K. Hsu and L. Camilletti, Proceedings 1999 CMP–MIC Conference, 120, IMIC, Tampa, 1999.

40. B. Lee, D. Boning, D.L. Hetherington and D.J. Stein, Proceedings 2000 CMP–MIC Conference, 255, IMIC, Tampa, 2000.
41. T. Detzel, S. Hosali, A. Sethuraman, J-F. Wang and L. Cook, Proceedings 1997 CMP–MIC Conference, 202, IMIC, Tampa, 1997.
42. V.S.K. Lim, F. Chen, W.L. Goh, A. See, C.H. Loh, C. Lin, Q.H. Zhong and M. Xin, Proceedings 2000 CMP–MIC Conference, 177, IMIC, Tampa, 2000.
43. B. Withers, E. Zhoa, R. Jairath and S. Hosali, Proceedings 1998 CMP–MIC Conference, 319, IMIC, Tampa, 1998.
44. F. Chen, C. Tay, B.B. Hou, F.L. Chin and J-Z. Zheng, Proceedings 1998 VMIC Conference, 491, IMIC, Tampa, 1998.
45. S. Fang, R.B. Knamankar, B.B. Shinn, F. Abbasi and F. Zhang, Proceedings 1998 CMP–MIC Conference, 134, IMIC, Tampa, 1998.
46. G.H. Koh, xxx, Proceedings 1998 CMP–MIC Conference, 15, IMIC, Tampa, 1998.
47. M. Ravkin, J. Zhang, K. Mikhaylich, D. Hetherington and D. Stein, Proceedings 1999 CMP–MIC Conference, 297, IMIC, Tampa, 1999.
48. C.P. Chen, C.T. Lee, C.F. Lin, H.C. Yung and L. Fang, Proceedings 1996 CMP–MIC Conference, 82, IMIC, Tampa, 1996.
49. C.L. Borst, W.N. Gill and R.J. Gutmann, Proceedings 1999 CMP–MIC Conference, 409, IMIC, Tampa, 1999.
50. K. Barla, O. Demolliens, C. Gounelle, C. Lair, Y. Lafarges, V. Lasserre, S. Lis, E. Louis, C. Maddalon, Y. Morand, G. Passamard, F. Pires and C. Verove, Proceedings 1998 VMIC Conference, 25, IMIC, Tampa, 1998.
51. H.D. Jeong, H.S. Park, H.J. Shin, B.J. Kim, H.K. Kang and M.Y. Lee, Proceedings 1999 International Interconnect Technology Conference, 190, 1999.
52. M. Naik, S. Parikh, P. Li, J. Educato, D. Cheung, I. Hashim, P. Hey, S. Jenq, T. Pand, F. Redeker, V. Rana, B. Tand and D. Yost, Proceedings 1999 International Interconnect Technology Conference, 181, 1999.
53. C. Jin and J. Wetze, Proceedings 2000 International Interconnect Technology Conference, 99, 2000.
54. E.T. Ryan, H-M. Ho, W-L. We, P.S. Ho, D.W. Gidley and J. Drage, Proceedings 1999 International Interconnect Technology Conference, 187, 1999.

3 Metal Polishing Processes

D.R. Evans

3.1 Metal Polishing Processes

In a broad sense, metal polishing is an extremely ancient technology that has long been used to make beautiful objets d'art as well as utilitarian objects. However, within the context of modern microelectronics it is a much more narrowly defined technology that is invariably used to fabricate "damascene" structures. The origin of this terminology is obscure, however for integrated circuit fabrication damascene means microscopic, inlaid metal features that serve as "wiring" to connect individual electronic components, e.g., transistors, previously formed in an underlying semiconductor substrate (usually a single crystal silicon wafer). Of course, the overall objective is fabrication of functional devices such as microprocessors or memories. Naturally, to serve this purpose metal wiring must be inlaid in some insulating material (typically a silica based glass) with appropriate provision made for interconnections between wires and the substrate and between the wires themselves. Indeed, in the case of complex devices such as microprocessors or sophisticated logic circuits, several layers of both horizontally and vertically interconnected damascene structures must be fabricated to form a "multilayer interconnect" in order to accommodate required connections between tens of millions of components.

To preface subsequent discussion, one observes that fabrication of a simple, single layer damascene structure requires three basic steps. The first of these is etching a recessed image of the desired pattern into the substrate. This is illustrated in Fig. 3.1 for a small section of the substrate.

Within the context of CMP, the details of patterning and etching processes are irrelevant. However, one expects that at the present state of the art this will be done using deep UV photolithography and advanced dry etching [1]. Naturally, the second basic step is metal deposition. In principle, deposition of a single uniform and homogeneous metal layer would be ideal. In practice, more complicated processes involving deposition of two or more layers of different metals are generally required to obtain adequate adhesion between the metal layer and underlying substrate, to prevent diffusion of metal contamination to the underlying semiconductor, etc. In addition, metal deposition processes are generally not perfectly uniform and may have poor step coverage or pattern dependence. All of these issues affect subse-
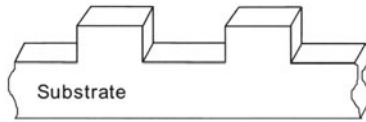
Fig. 3.1. Etched image of damascene structure
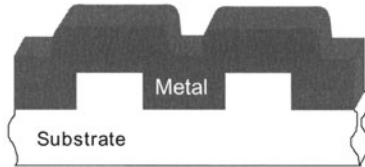
Fig. 3.2. Metal deposition
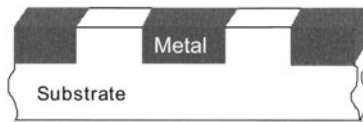
Fig. 3.3. Finished damascene structures

quent metal removal processes and will be discussed in detail subsequently. However, for conceptual purposes, deposition of a single uniform metal layer is illustrated in Fig. 3.2.

Again, precise details of the deposition process itself are unimportant. At present, physical vapor deposition (PVD), chemical vapor deposition (CVD), and electrochemical deposition (ECD) are all used by various manufacturers. The third and final basic step is removal of metal "overburden" lying outside of damascene structures. An ideal finished damascene structure is shown in Fig. 3.3.

Now, a perfect metal removal process would correspond to some form of "micro-machining" or "micro-milling" which achieves complete metal overburden removal simultaneously at all points of the substrate surface. Such a process could be visualized as due to the action of a perfectly stiff, smooth, and planar toolpiece that is in exact alignment with the substrate surface. Of course, CMP is quite different from this and is only a distant approximation to such a hypothetical process. For example, polishing pads are quite flexible and have surface structure that is commensurate with damascene feature dimensions. In addition, slurry particle sizes are often also of similar dimension. One consequence of these and other differences between CMP and micro-machining is a strong dependence of metal removal rate on feature pattern density, which, as is discussed elsewhere, can result in serious difficulties. In contrast, pad flexibility obviates the need for precise alignment of

pad and wafer surfaces, which would be quite difficult at best in any practical implementation.

Within the context of previous comments, it is useful to explore in more detail some of the basic phenomena specific to metal polishing processes. First of all, in any interconnect scheme, provision must be made for "off chip" connections. This requires the inclusion of "bond pads" in the circuit, usually around the periphery. Naturally, bond pads must be large enough to permit external wire connections and typically are square features having an edge length of about 100 μm. Figure 3.4 provides a schematic illustration of typical bond pad placement.

Here, the large gray center square represents the circuit "core", which consists of a complex array of device structures. The small peripheral squares represent bond pads connected to the circuit by wiring runs (gray line segments). Now, it often happens that the metal removal rate is much larger than the collateral dielectric removal rate. In this case, metal removal continues within a bond pad after all metal overburden has been removed in the surrounding field. Furthermore, dielectric is usually removed slowly from the field. Because, the pad is deformable, metal removal tends to be greater in the center of bond pads. Naturally, this is due to the local increase of pressure as is discussed in detail elsewhere within this volume. Indeed, if such "overpolishing" is sufficiently severe, metal can be completely removed in the pad centers as illustrated in Fig. 3.5 by an optical micrograph.

This phenomenon is known as *dishing* and is not just limited to bond pads, but occurs in any similar large feature. Associated removal of dielectric material in the field area is known as *field loss*. Again, due to pad deformation
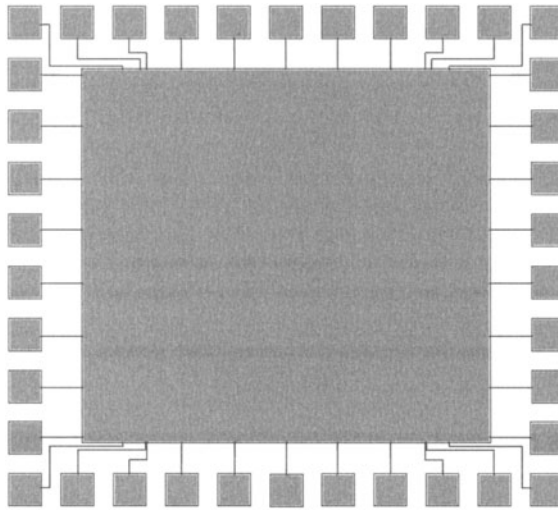
Fig. 3.4. Schematic of a "chip"
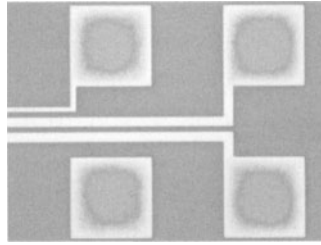
Fig. 3.5. Dishing of Bond Pads

and local pressure increase, field loss tends to be larger in proximity to bond pads or similar large features. Of course, corresponding phenomena also occur in dielectric polishing, however due to the fundamental differences in material properties and usage of metals and insulators it is of particular importance in metal polishing.

A closely related phenomenon is dielectric *erosion*. A simple way to visualize erosion is to consider a large metal feature, such as a bond pad, surrounded by a relatively large field area. If instead of a solid rectangle or other figure, this feature is made up of a dense array of much smaller metal features, then the feature as a whole tends to behave as if were a solid area of metal even though some dielectric is present between the features. The amount of dielectric removed corresponds to erosion. Clearly, this behavior is also a result of pad deformation and local pressure increase and can occur even if metal/dielectric selectivity is large. From this point of view, the metal and dielectric behave as a composite material. Hence, the effect of high metal/dielectric selectivity is an effective reduction of the composite removal rate in comparison to the removal rate of the pure metal itself. In practice, this results in a strong dependence of erosion on metal/dielectric feature density. Within this context, if pure metal corresponds to 100% feature density and pure dielectric to 0% feature density, then dishing can be identified with erosion in the limiting case of 100% density. Conversely, disregarding background field loss, erosion disappears in the case of 0% density. Of course, as well as feature density, erosion also depends on process and consumable parameters, relative feature sizes, etc [2]. Erosion and dishing are illustrated for various approximate feature densities in Fig. 3.6.

For any quantitative process characterization and control either within-a-wafer or wafer-to-wafer, measurements of dishing and erosion must be made on identical features within different dice or wafers. These features may be specially designed for this purpose or they may be part of a functioning circuit, however it cannot be overemphasized that such measurements must be consistently made to be useful.

One approach to minimization of both dishing and erosion is use of a slurry formulation that removes metal and dielectric materials at nearly the same rate, i.e., a non-selective slurry. Unfortunately, this can only increase field
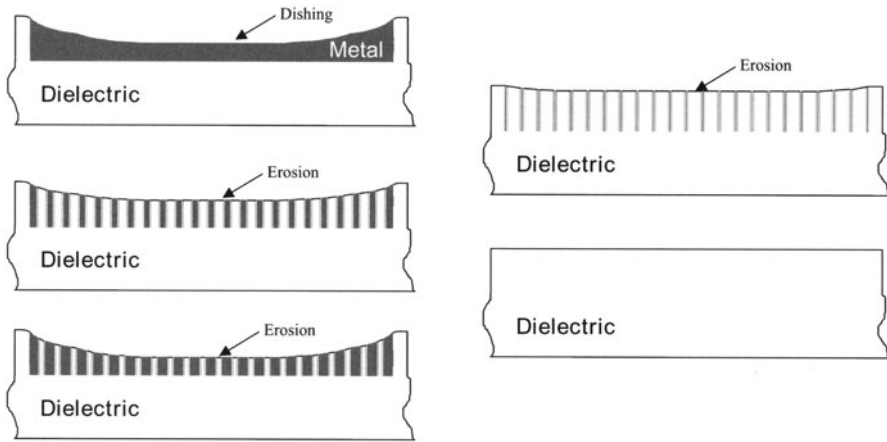
Fig. 3.6. Dishing and erosion for approximate feature densities of 100, 70, 40, 5, and 0%

loss and, thus, represents a fundamental trade-off between final planarity and material thickness uniformity. In addition, non-selective slurries tend to have low removal rates and, thus, a typical implementation involves a two-step polish. The first step employs a slurry having a high metal removal rate, which generally also implies selectivity toward removal of metal. The bulk of the metal overburden is removed in this step, which is followed by a second non-selective polish to remove the remaining metal in the field and a small amount of the underlying dielectric material. Obviously, the advantage to this approach is achievement of a very planar surface. However, as observed above, since some dielectric material is intentionally removed, finished metal and dielectric thickness and associated capacitance and resistance can be difficult to control.

An additional phenomenon often encountered in metal polishing is surface recession or just *recess*. Similar to dishing, recess can be understood as a degradation of final planarity arising after complete removal of metal overburden. However, unlike dishing, recess affects features regardless of size. While, the source of recess is not entirely clear in every case, it can be generally attributed to purely chemical attack or corrosion of the metal surface, either during overpolishing or perhaps during post polish cleaning and rinsing. As illustrated in Fig. 3.7, recess is characterized by a small, but sharp vertical step at the edges of metal features. In contrast, dishing and erosion are characterized by gradual changes in vertical dimension and relatively smooth topography.

Although undesirable, for a single layer damascene structure such departures from planarity may be manageable. However, as multilayer interconnect is built up, departures from planarity tend to combine causing severe difficulties in fabrication of upper layers. This may be partially addressed by
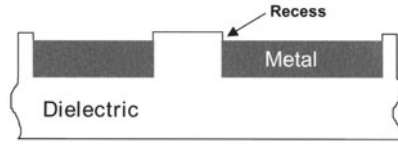
Fig. 3.7. Dishing of bond pads

relaxing design rules for upper layers, but nevertheless it is clear that there
will always be a strong motivation for achievement and control of planarity.

3.2 Evolution of Damascene Surface Morphology During Polishing

Irrespective of exact process details, for descriptive purposes an overall pol-
ishing process for fabrication of a metal damascene structure can be sep-
arated into several phenomenological stages. Within this context, Fig. 3.8
illustrates typical, initial metal morphology. The schematic on the left shows
metal/adhesion-barrier layers on top of a patterned dielectric layer. Of course,
the identity of these layers is immaterial for the present description, provided
that adhesion and other aspects of mechanical and electrical integrity re-
main adequate. The optical micrograph on the right shows a top down view
of a patterned wafer after initial metal/barrier deposition, but before any
polishing.

As is evident from the photograph, the unpolished metal layer is quite
grainy. (For convenience, this has not been represented in the schematic cross
section.) Although, the underlying pattern is not distinct, it is still visible and
in this case consists of four $100\,\mu\mathrm{m}$ bond pads and associated connections.
Of course, initial metal surface morphology is quite variable and is highly
material and process dependent.

Polishing can be said to be in "early stage" from initial pad/wafer con-
tact until all original metal surface has been removed, including "down" areas
at the bottoms of features. Of course, this assumes that metal thickness is
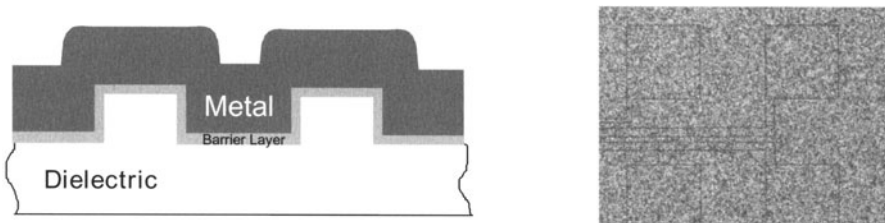
Fig. 3.8. Initial morphology
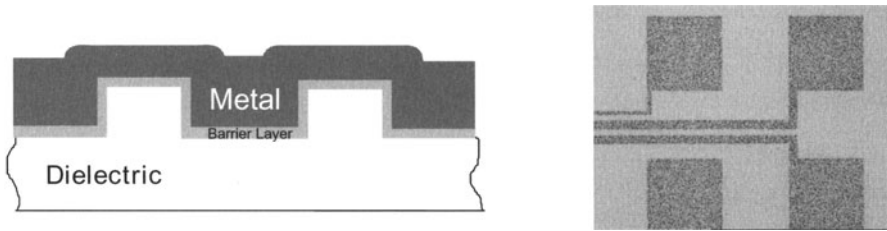
Fig. 3.9. Early stage polishing

greater than damascene feature depth. Indeed, this condition is generally required to achieve sufficient planarity. (A practical "rule of thumb" is that at a minimum, metal thickness must be one and a half to two times larger than damascene depth, although this becomes problematic in the case of very shallow features.) Thus, in early stage polishing the pad essentially contacts only high or "up" areas of the pattern. (In photolithographic terminology, for a positive photoresist process, these correspond to dark areas of the original pattern.) This situation is illustrated by Fig. 3.9. Clearly, the bond pad pattern has become quite distinct and it is obvious that all of the original surface roughness has been removed in up areas. However, graininess remains visible in down areas. Indeed, the surface between patterns is quite smooth and has a specular, mirror-like appearance. Since, surface roughness is removed rapidly once substantial contact is made with the pad, it is clear that the down areas have not been appreciably polished. Furthermore, there is no evidence of chemical etching in the down areas. Of course, this is just what is required for acceptable CMP. Clearly if chemical etching occurs to any significant degree, it will be very difficult to obtain a final, planarized metal surface.

Naturally, early stages of polishing are followed by an "intermediate stage", during which the wafer surface remains entirely covered by metal. In this case, the polish has not broken through to underlying barrier or dielectric layers, but it has passed the original level of down areas so that the surface is nearly planar, at least locally. This stage of polishing is illustrated by Fig. 3.10. Clearly, the bond pad pattern has nearly disappeared. (Indeed,
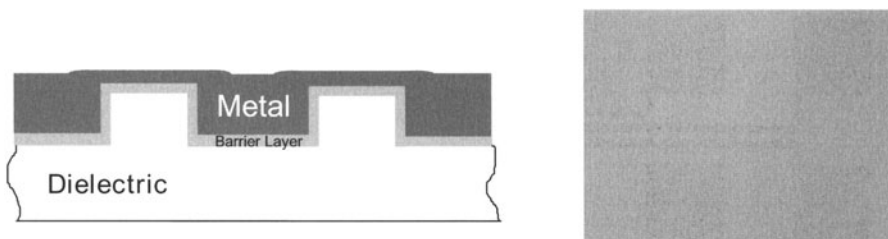
Fig. 3.10. Intermediate stage polishing
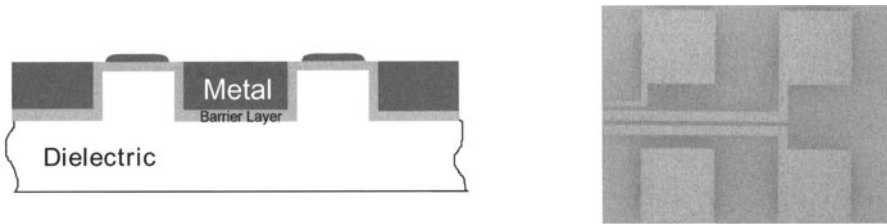
Fig. 3.11. Finish polishing

this particular micrograph was chosen because the patterning is still barely visible.) Microscopic examination of metal surfaces at this stage of polishing generally reveals an entirely featureless field, which is very difficult to bring to a sharp focus. Now, it is important to note that this does not mean that the surface has become perfectly planar. Long range non-planarity due to pattern density differences still remains. Moreover, shorter range non-planarity arising from corner rounding will also be present and is easily observed by profilometry. What is absent, having been removed by polishing, are sharp, vertical feature edges.

It comes as no surprise that "finish polishing" denotes the situation from breakthrough of the metal layer to underlying material at some point in the field until the field is entirely cleared of overburden with metal remaining only in damascene features. This situation is illustrated by Fig. 3.11. Here, the schematic cross section shows barrier metal exposed at feature edges. In normal polishing this is expected because of local pressure distribution in proximity to patterned features. This is also made evident in the micrograph by darkening at the edges of the bond pads. Of course, the more planar the polished surface becomes during the intermediate stage of polishing, the less important this effect will be. Indeed, if the metal surface could be rendered perfectly planar before all of the metal overburden is removed, finish polishing as such would not occur and the process would proceed directly from the intermediate stage as defined previously, to the process "endpoint". Obviously, such a situation is highly desirable and is one of the ultimate goals of any CMP process.

Of course, conclusion of finish polishing should define the ideal process endpoint. This designation is somewhat complicated by the barrier-adhesion layer since endpoint could be defined as complete removal of metal in the field or, alternatively, as complete removal of metal and barrier-adhesion layer. For the purposes of the present discussion, the former definition is adopted. In this case, the barrier-adhesion layer can be regarded as a "polish stop" as illustrated by Fig. 3.12. Here, it has been assumed that the particular metal polishing chemistry used is highly selective between metal and the barrier metal. Thus, the micrograph shows that, as desired, metal remains in the patterned features, but has been removed from the field to expose the barrier-
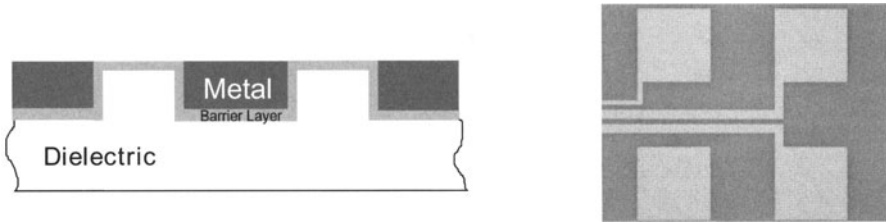
Fig. 3.12. Endpoint

adhesion layer. In this case, metal/barrier selectivity is sufficiently large so that the barrier layer is essentially intact everywhere in the field; however, there is just a subtle hint of thinning around feature edges. Obviously, an additional process step to remove the barrier metal is required. This can be accomplished by selective etching or by additional CMP, perhaps using a different chemistry or the same chemistry under different process conditions. In principle, if selective etching is used, the metal surface will be slightly raised above the field. Clearly, this is opposite of the situation described previously for surface recession. In practice, this situation is difficult to realize because barrier/adhesion layers are typically very thin and surface recession and/or dishing and erosion are of larger magnitude. Of course, if the barrier layer is removed using an additional CMP step, the metal surface should not be raised above the field.

Alternatively, if the primary polishing chemistry is not selective between the metal and the barrier–adhesion layer, then the barrier–adhesion layer can be removed in the field along with the metal at the end of finish polishing. In this case, the dielectric material itself can serve as a polish stop. Of course, the resulting morphology should be essentially the same as that generated using a separate barrier removal step and is shown in Fig. 3.13. Here, both metal and barrier–adhesion layers have been removed in the field and the interference color characteristic of the field dielectric thickness is visible in the area surrounding metal features. In the black and white micrograph this causes an enhancement in visual contrast relative to the previous micrograph for which the barrier layer was left intact.
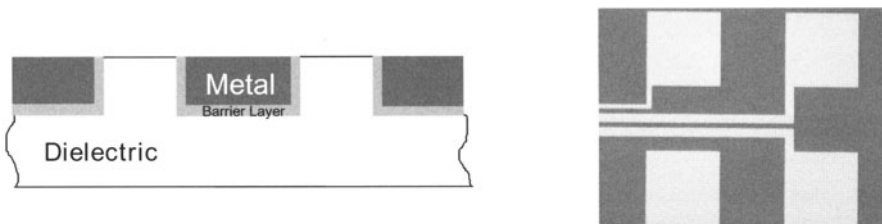
Fig. 3.13. Barrier removal
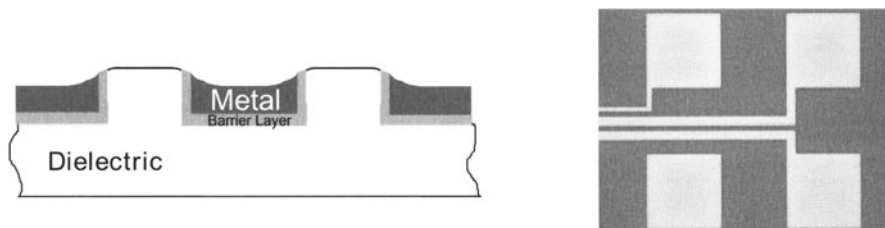
Fig. 3.14. Overpolishing

In practice, endpoint is not reached simultaneously everywhere on the wafer surface. Obviously, this is caused by a non-uniform, within-a-wafer removal rate. In addition to removal non-uniformity due to differences in feature density, non-uniformity can arise from many additional sources such as non-uniform backside pressure, metal thickness variation, slurry or polishing solution distribution over the wafer, etc. Thus, some overpolishing beyond endpoint is invariably required. Naturally, in any practical CMP process, the effective process stage will vary locally due to feature density as well as all other sources of non-uniformity. As a consequence, dishing, erosion, and recession will appear to some degree depending on the exact nature of the removal process. This is shown in Fig. 3.14 and represents the end state of a practical metal polishing process. However, if overpolishing is not too severe, as shown in the optical micrograph the metal pattern remains intact.

As a practical matter it is very difficult to optimize all stages of polishing using a single step process in which the same process conditions are used from initial removal through overpolishing. Typically, a multistep [3] process implemented on separate polishing stations is used. Naturally, a high removal rate with good planarity is desirable for the first step. Commonly, the first step is terminated during intermediate stage polishing with metal still covering the entire wafer surface. Alternatively, the first step may be terminated during finish polishing if due care is taken to prevent the onset of dishing and erosion. The second step polish (sometimes called "soft landing") is optimized to achieve and maintain a high degree of planarity. Generally, this requires that the primary metal removal rate is low. At the current stage of technological development there are a number of approaches to second step polishing. As mentioned previously, one might use the barrier layer as a stopping layer. In this case a third, separate barrier removal step would then be required. Of course, loss of planarity can occur and must be avoided during barrier removal. More commonly, the second step is optimized to remove metal and barrier layers at nearly the same rate, i.e., there is no selectivity between metal and barrier layers. In this case, the dielectric material itself can be used as a stopping layer. Even so, loss of planarity may still remain problematical. Again, as mentioned previously, a completely non-selective process that removes metal, barrier, and dielectric all at the same rate can be used to main-

tain planarity and can be implemented during the second step or as a separate third step. However, this comes at the cost of increased loss of dielectric in the field and more difficult control of final dielectric and metal thickness.

3.3 Specifics of Tungsten and Copper Polishing

The first full-scale application of metal CMP in semiconductor device manufacturing was for fabrication of tungsten plugs in association with standard etched aluminum alloy interconnect. The fundamental motivation was efficient filling of via holes between interconnect levels using a non-selective tungsten CVD process, which has much better conformality, i.e., step coverage, than does conventional PVD of aluminum alloy [4,5]. This interconnect scheme actually preceded the implementation of tungsten CMP. In the earliest implementation, the tungsten overburden was removed using an isotropic plasma etch process [6,7,8,9]. To avoid excessive etching in contact or via holes, this required that the tungsten deposition was sufficiently thick so that the effective tungsten layer thickness above via holes was equivalent to the thickness in the field. At a minimum, deposition thickness on the order of half of the dimension of the most critical via hole is necessary. Thus, in the mid to late 1980's, when tungsten plug technology first became widely implemented, a typical deposition thickness was 0.5–0.8 µm. Clearly, this required removal of a substantial amount of metal. Of course, subsequent, device and interconnect scaling has allowed tungsten deposition thickness to be reduced. Unfortunately, it was found that in a manufacturing environment plasma etch rate, uniformity, and selectivity were difficult to control and that "edge stringers", recessed plugs, keyhole defects, etc. were common. As an alternative, selective tungsten CVD was widely investigated to obviate the need for the "etchback" process [10, 11, 12, 13]. Unfortunately, despite its promise, selective tungsten CVD did not meet stringent requirements for defectivity and yield necessary for wide implementation into integrated circuit manufacturing.

The advent of dielectric CMP to planarize interconnect levels immediately opened the possibility that tungsten plugs could be fabricated using CMP rather than plasma etch back. Early results indicated that CMP was capable of superior performance with low defectivity and good circuit yield [14, 15, 16, 17]. Furthermore, implementation of tungsten CMP into a manufacturing environment was found to be reasonably cost effective and straightforward, even as a retrofit [18]. Of course, a prior dielectric planarization is required so that the substrate surface is substantially flat prior to tungsten plug fabrication. Obviously, this is essentially a single layer damascene structure. Although circuit manufacturers are reluctant to disclose process details, a few have been published. For example, in early work a typical down force used for tungsten CMP was 5 to 8 psi [19]. Likewise, proprietary slurry chemistries were typically used, however, these same workers indicate that tungsten to oxide selectivity can exceed 150:1. Of course, this selectivity

cannot be realized in densely patterned areas. (An absolute tungsten removal rate was not given.) As CMP process equipment and consumables have matured, the trend has been toward higher removal rates at lower down force. A typical production process might operate at a down force of 3.5 psi or less with a removal rate in excess of 400 nm/min [20]. Tungsten to dielectric selectivity has decreased in importance due to advanced hardware for endpoint detection and uniformity control.

Subsequently, tungsten plug technology has been extended to device contact holes as well interconnect via holes. Assuming a planarized pre-metal dielectric, fabrication of tungsten plugs in contact holes is illustrated in Fig. 3.15.

For simplicity, details of device isolation and self-aligned diffusion or silicidation schemes are not shown. Of course, the process sequence begins by patterning and etching of contact holes. As is standard practice, this is followed by deposition of a thin titanium nitride barrier layer followed by tungsten CVD. (The barrier layer is required to prevent a catastrophic interaction between the silicon substrate and tungsten hexafluoride during tungsten deposition.) Tungsten plugs are formed by removal of the tungsten overburden in the field by CMP.

Conventional aluminum alloy interconnect is subsequently fabricated by PVD and etching. Obviously, the planarized plug/dielectric surface eliminates stringent requirements for conformality of the aluminum alloy deposition. This is followed by interlevel dielectric deposition and planarization. As is illustrated in Fig. 3.16, at this point, via holes are opened and the overall process including tungsten CMP is repeated to fabricate a second layer of interconnect.
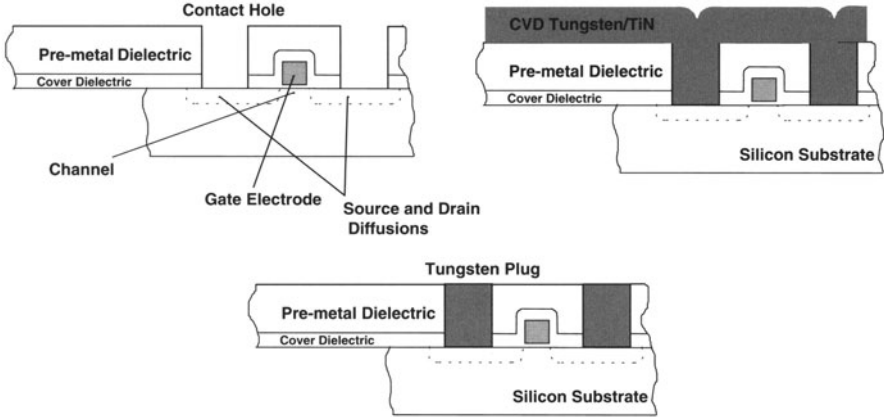
Fig. 3.15. Contact hole tungsten plugs
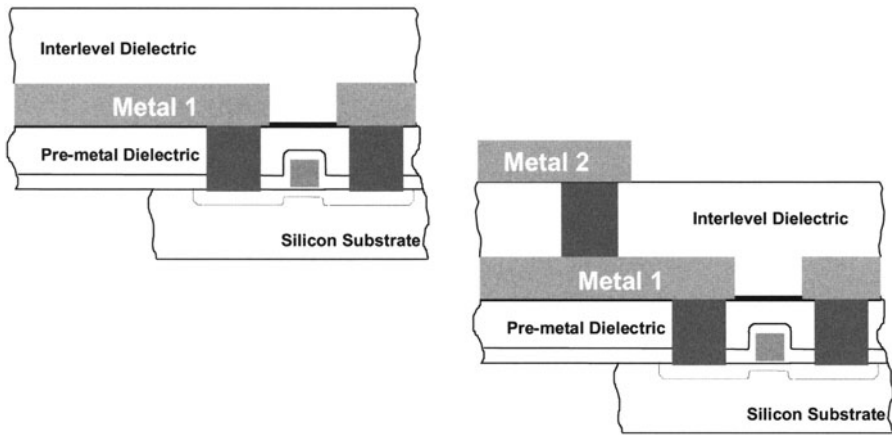
Fig. 3.16. Interconnect fabrication

Here, dark gray indicates contact and via hole tungsten plugs. In state-of-the-art interconnect schemes, plug and interconnect fabrication may be repeated six or more times.

Various methods have been used for endpoint detection in tungsten CMP. Historically, both spindle and platen motor drive currents have been used to provide an indication of polishing endpoint [21]. Unfortunately, although simple, this method is strongly affected by electrical noise and variation in consumables. Alternatively, observation of platen temperature during polishing using a suitable infrared sensor [22] has been used for endpoint detection [23, 24]. Although the temperature rise is relatively small, for a given slurry it is found to be proportional to the tungsten removal rate and reasonably repeatable [25]. Even so, at the present optical reflectometry represents the state-of-the-art for endpoint detection [26]. These issues are discussed in more detail elsewhere in this volume.

As shown in the Appendix, the Pourbaix diagram for tungsten strongly suggests that a neutral to acidic slurry chemistry should be used for CMP. (Clearly, tungsten is an acidic passivating metal with soluble tungstate anions produced at high pH.) Indeed, tungsten polishing was first implemented using a ferricyanide-phosphate chemistry at a pH from 5.0 to 6.5. Later, due to chemical hazards this was replaced by acidic ferric nitrate (pH $\cong 1.5$–2.5) as oxidizer. Alumina was generally used as abrasive. However, various studies [27] have shown that this chemistry is highly corrosive and contaminating both to equipment and substrates and has almost uniformly been replaced by less acidic media (pH $\cong 3.0$–4.0) with hydrogen peroxide, or persulfate or iodate ion as an oxidizer [28]. In addition, abrasives other than alumina have also been investigated [29].

A persistent difficulty with tungsten CMP is residual metal left on the dielectric surface after polishing. Such defects are colloquially called "puddles".
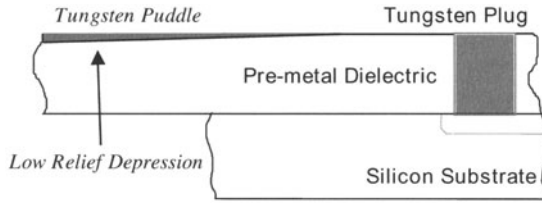
Fig. 3.17. Tungsten puddle

Obviously, a puddle can lead to undesirable electrical short circuits [30, 31] and, typically, as illustrated in Fig. 3.17, is caused by a low relief depression in the dielectric surface.

Moreover, although the depth of the depression may be slight, it is found that tungsten is often very difficult to remove and requires an unacceptable amount of overpolishing, which results in erosion of via hole arrays, dishing of large features if present, and a collateral large background field loss. Clearly, this exacerbates the situation for subsequent interconnect layers since low relief depressions are typically the result of previous metal dishing and dielectric erosion as well as other process steps. (Indeed, surface recession is a particularly troublesome cause of puddle formation due to associated sharp vertical steps.) These issues may be addressed either by improvement of dielectric polishing processes prior to plug formation or modification of the tungsten polishing process itself.

Tungsten is also quite chemically active as is indicated by half cell potentials summarized in the Appendix. As a consequence, formation of hydrous oxide layers is sometimes observed. This must be avoided or the hydrous oxide removed after CMP to prevent high via hole series resistance. Generally, adjustment of pH during polishing to higher values and/or treatment with an alkaline cleaning solution after CMP should serve to alleviate this problem.

In addition, it is undesirable to open "keyholes" or to roughen the metal surface during tungsten CMP. In the case of keyholes, aqueous media penetrates vertically into the via hole at the "seam" left by CVD resulting in oxidation and/or corrosion, thus degrading via hole resistance. Concomitantly, the surface of the metal may be significantly roughened. In either case, the problem is generally a consequence of direct chemical etching by the slurry. Ideally, the static chemical etch rate for any tungsten slurry should be zero. However, in practice because tungsten is quite chemically active, static etch rates of a few nanometers to a few tens of nanometers per minute can occur [32]. Clearly, keyhole formation itself is more an issue with the CVD process, however it can become an issue with CMP if the CVD process cannot be modified to eliminate the keyhole and completely fill via holes.

With the successful implementation of tungsten CMP, the advent of copper as an interconnect metallization using a damascene structure became practical [33]. Furthermore, although many attempts have been made to etch

copper conventionally, this has been uniformly unsuccessful for application to integrated circuit manufacturing [34]. Therefore, for the current state-of-the-art it would seem that CMP is an absolute necessity for fabrication of copper interconnect. Within this context, manufacturers have adopted various schemes for copper integration. The most conservative approach is the use of copper interconnect in upper wiring layers for power buses and long transmission lines while retaining conventional metallization in lower layers as a local interconnect. Since copper is a highly undesirable contaminant in the silicon substrate, this approach insures a large margin of separation. However, it has the disadvantage of requiring processing equipment for both conventional and copper damascene processing. Alternatively, damascene copper interconnect can be used for all circuit wiring. In this case, tungsten plugs are fabricated in device contact holes as usual; however, this process is followed by dielectric deposition rather than the usual aluminum alloy PVD. As described earlier, an image of the interconnect layer is then etched into the dielectric. This is followed by deposition of refractory barrier metal, typically a composite layer of tantalum and tantalum nitride. A thin seed layer of pure copper is then deposited on top of the refractory barrier metal. This seed layer is necessary for subsequent electrochemical deposition of copper. At present, both barrier and seed layers are generally deposited by PVD. The bulk of the copper layer is deposited by electroplating or electrochemical deposition, i.e., ECD. In a broad sense, ECD is a naturally more conformal deposition method than vapor coating methods. This can be simplistically understood by consideration of characteristic length scales, e.g., mean free paths, in liquid and gaseous media. Moreover, recent advances in copper ECD technology have resulted in "bottom up" filling processes which are in a sense "super-conformal" as indicated by the absence of seams and keyholes typical of conformal or sub-conformal processes. In any case, as shown in Fig. 3.18, copper deposition is followed by copper CMP resulting in a copper damascene structure.

It is clear that at this point that copper damascene and tungsten plug fabrication are conceptually quite similar. Of course, polishing chemistry and process conditions must be specific to each metal. Moreover, for subsequent fabrication of upper wiring layers, tungsten plugs are not necessary and the dual damascene method may be used. In a dual damascene process, images of both via holes and interconnect wiring are etched in dielectric materials before any metal is deposited. Although integration issues are complex (especially if a low dielectric constant insulator is used) and will not be discussed in detail here, it is evident that a choice must be made between "via first" and "trench first" process schemes [35]. In the case of via first processing, via holes are patterned on the planarized dielectric surface and partially etched through the dielectric layer. This is followed by patterning and etching of the wiring. In this case, it is critical that all photoresist is removed from the partially etched via holes since via etching must be completed at the same time that the wiring image is etched. In a trench first process scheme, the
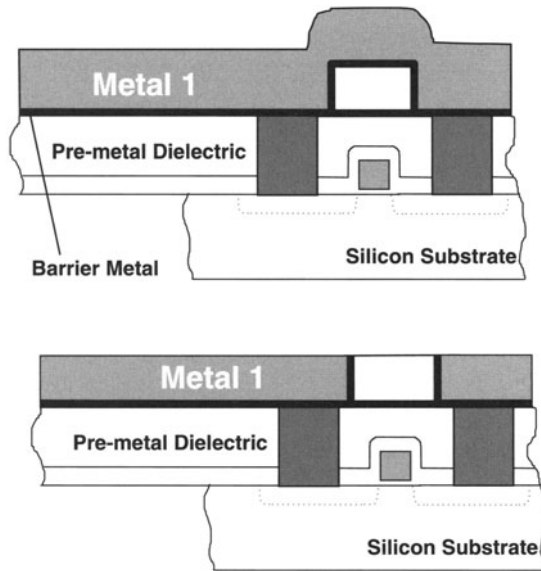
Fig. 3.18. Copper damascene fabrication

wiring is patterned and etched first. Hence, via holes must be patterned and etched on a non-planar surface. In either case, the result is an image of the interconnect wiring etched in the dielectric layer with via holes open to the underlying interconnect layer. Of course, patterning and etching is followed by metal deposition and CMP as shown in Fig. 3.19.

Here, both interlevel and intralevel dielectric layers are indicated since to maximize the performance of copper metallization low dielectric constant materials should be used as intralevel dielectric to minimize fringing capacitance.
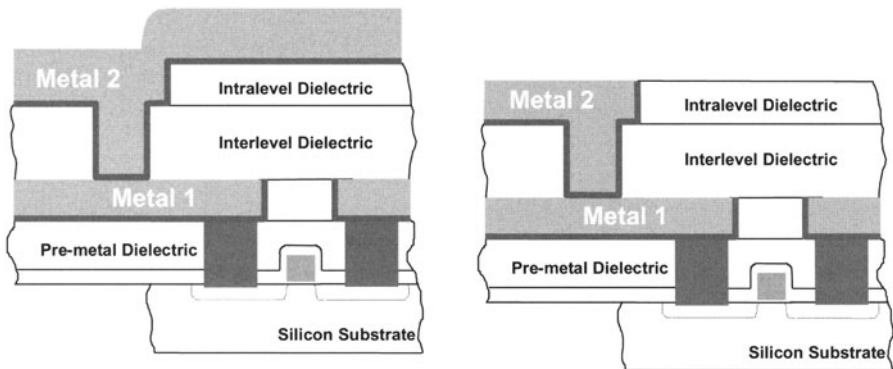
Fig. 3.19. Copper dual damascene fabrication

    The aqueous chemistry of copper is summarized in the copper Pourbaix diagram appearing in the Appendix. Clearly, in contrast to tungsten it is an alkaline passivating metal. In most early work, hydrogen peroxide [36] was used as an oxidizer for copper polishing. This chemistry typically gave removal rates in excess of 400 nm/min with reasonable uniformity. However, it proved very difficult to avoid dishing and especially recess due to static etching. Representative measurements of residual step heights on patterned copper wafers after CMP usually were about 100 nm at best and commonly could be much worse. (This is irrespective of the "beautiful" SEM's appearing in the literature and conference presentations.) This is easily observed in a typical series of profilometer traces as shown in Fig. 3.20. In addition, the relation of between the static etch rate of copper and its removal rate during a representative CMP process is shown in Fig. 3.21 as a function of hydrogen peroxide aliquot (relative volume of commercial 30% peroxide solution).
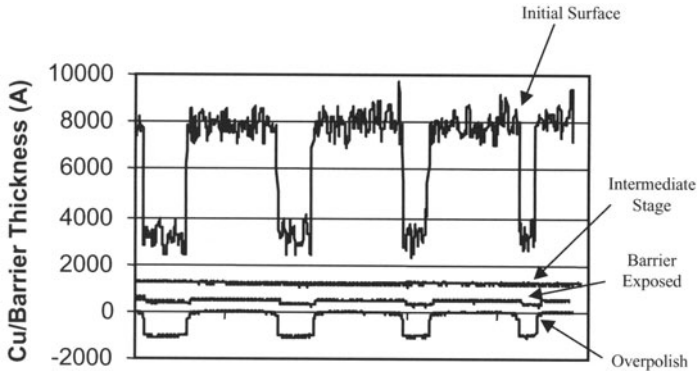
Fig. 3.20. Profilometer traces of damascene copper (60, 50, 40, and 30 μm wide lines)
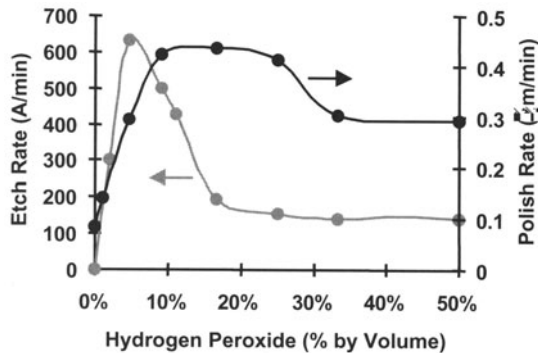
Fig. 3.21. Etch and polish rate of copper vs H_2O_2 aliquot

At the present, hydrogen peroxide is still used albeit at lower concentration and iodate ion is also under investigation for use as an oxidizer. While it would seem that alkaline chemistry could be used to polish copper, it turns out that this is not the case. A simple Pourbaix diagram does not indicate the degree to which a metal cationic species undergoes complexation. In particular, it is well known that copper forms a series of soluble complex cations with ammonia at high pH. The formation of these complexes is so thermodynamically favored that copper oxides and hydroxides are readily dissolved by ammonia. In addition, copper forms soluble hydroxo complexes with the hydroxide ion itself. Again, this tends to promote metal dissolution at high pH. Of course, in the acidic range, positively charged, soluble copper ions, i.e., cations, are favored and while attempts have been made to polish copper in dilute nitric acid, it is not likely that this is practical. Therefore, copper polishing chemistry is typically held close to a neutral pH. Fortunately, triazoles such as BTA are very effective corrosion inhibitors for copper. These can be added to the chemistry to suppress dissolution at almost any pH. This allows a much wider range of optimization for the slurry chemistry [37]. As in the case of tungsten, alumina is commonly used as an abrasive for copper polishing. However, for "soft landing" processes as described previously, colloidal or fumed silica may be used instead. The motivation for this is the reduction of "scratch" defects and the optimization of selectivity between dielectric, barrier metal and copper.

In contrast to tungsten, copper polishing is complicated by both the copper deposition process and the choice of barrier metal [38]. In early stages of development, MOCVD was the most attractive alternative for copper deposition because of its extremely good step coverage [39, 40, 41]. Unfortunately, copper layers deposited by MOCVD tend to incorporate fluorine and carbon contamination, which increases resistivity and degrades adhesion to barrier layers. Furthermore, after a period of experimentation, development of copper interconnect settled on tantalum and/or tantalum nitride as the most widely used choice for barrier layer material [42, 43]. It turns out that adhesion of copper to tantalum/tantalum nitride is particularly difficult to achieve. Indeed, adhesion of copper to tantalum/tantalum nitride can be achieved reliably only by copper deposition using advanced PVD. Of course, coverage and filling characteristics of PVD are not ideal for deep submicron structures. Therefore, it became obvious that a combination of PVD followed by a conformal deposition process is required for fabrication of damascene copper interconnect. One might have thought that this conformal deposition process would be MOCVD, however as observed previously, ECD was found not only to have excellent conformality, but superior purity as well [44, 45, 46]. Therefore, ECD has now become the standard copper deposition process.

However, ECD does present some issues that are relevant to copper CMP. It was found early on that ECD copper is extremely pure and undergoes substantial grain growth at room temperature for up to a few days after deposi-

tion [47, 48, 49]. After this time, the material is stable. Although specific information is sparse to nonexistent in the literature, polishing characteristics, in particular removal rate, can be expected to be affected by this recrystallization. Obviously, a simple solution to this is moderate heat treatment after ECD but before CMP. A more severe problem is related to enhancements in ECD associated with bottom up filling [50]. Bottom up filling is desirable to prevent void formation in small, high aspect features, however it leads to substantial thickness variations of the copper overburden. In particular, copper is thick over dense arrays of minimum dimension features and is thinner in the field or over larger features. This presents a severe problem for CMP. Clearly, unless overall planarization is very efficient and rapid, a thicker copper deposit over small features requires that large features and the field area are substantially overpolished. This can result in severe dishing and field loss.

In addition, specification of a particular barrier metal thickness inside of damascene features generally requires a much thicker deposit outside of features. This is an inherent limitation of PVD and can be understood by observing that the same material flux must cover a larger area, i.e., bottom and sidewalls, inside a feature than outside. Obviously, this effect becomes increasingly severe for high aspect features and is independent of step coverage modification due to the presence of ionization or some other scheme. The difficulty for CMP is that this thick barrier layer must be removed with minimum dishing or recess of copper. This suggests that CVD may be required for barrier deposition in the future. Moreover, since tantalum and tantalum nitride are difficult to deposit by CVD, alternative barrier materials may be necessary [51].

In general, copper polishing using conventional slurries and pads is nearly always subject to some residual non-planarity. Indeed, it has been found difficult to improve this except by non-selective polishing as a final process step. As observed previously, this makes dielectric and metal thickness much more variable and difficult to control. Recent development of copper polishing using fixed abrasive pads [52] or with no abrasive [53, 54] at all appears to offer a substantial improvement. In the case of fixed abrasive pads, alumina abrasive is incorporated into the pad material itself. Instead of slurry, a chemical solution is introduced to the wafer surface during polishing with this pad. Unfortunately, to date fixed abrasive pads have not been adopted for robust CMP manufacturing. Even so, planarization performance is often superior to conventional pads and slurries. This can be explained at least heuristically by the confinement of abrasive action to a relatively fixed plane.

Abrasive free polishing is similar conceptually, except that the fixed abrasive pad is replaced with a conventional pad [55]. At least in the case of copper, the conditioned pad structure itself acts as a mild abrasive. Thus, for an optimized chemistry, substantial copper removal rates of the order 500 nm/min are obtained without any conventional abrasive at all. In addition

to superior planarization performance, there is virtually a complete absence of any scratch defects.

For completeness it should be noted that endpoint detection issues in copper polishing are quite similar to those for tungsten polishing. Again, the state-of-the-art is represented by in-situ reflectometry [56].

3.4 Metal Polishing Chemistry

In broad terms, chemical mechanical polishing of metals is quite similar to that of the common dielectric materials. Indeed, within this general context, the same type of equipment used for dielectric polishing can also be used for metal polishing. Of course, in practice individual polishing machines are usually assigned to only one type of process, which, among other things, allows better process control and eliminates most cross-contamination. However, there is really no fundamental difference in hardware design. Furthermore, temporarily disregarding differences in chemistry, this is generally true for consumables such as polishing pads and carrier inserts as well. Indeed, carrier design, pad and insert properties, etc. affect both dielectric and metal removal in much the same way. This is to be expected since, as discussed elsewhere, microscopic interactions between abrasive particles, substrate and pad surfaces have many common features irrespective of materials. Thus, one finds that metal removal rates during CMP follow Preston's Law [57] at least approximately, that variations in local polish rates depend on feature density in much the same way both for metals and dielectrics, and that global planarization is similarly affected by pad flexibility and compressibility. Of course, these observations are of a very general nature and the "devil is in the details." Indeed, one should not uncritically attempt to predict the characteristics of metal CMP based on a prior knowledge of dielectric CMP. There are important differences. For example, the removal rate of silicon dioxide is highly dependent on pad conditioning for CMP using conventional alkali/silica slurries, e.g., Cabot SS-12, Rodel ILD 1200, etc. In contrast, conditioning is typically much less important (or at least of a different nature) for metal CMP. Additionally, high removal rate selectivity between overburden and underlying materials is encountered much more commonly in metal CMP processes. Although such a situation may be desirable since it allows the underlying material to act as a polish stop, it may also introduce difficulties in fabrication of an adequately planarized surface.

Metal polishing does differ from dielectric polishing in that it requires an oxidizing chemical environment. This is not necessary for CMP of dielectrics because these materials are generally semiconductor or metal oxides or nitrides and, as such, are already fully oxidized. In both metal and dielectric CMP, material removal is thought to occur by chemical formation of a "soft" modified surface layer, which is then removed by action of an abrasive during simultaneous dynamic contact of the polished surface, abrasive particles, and

polishing pad. Ideally, in both dielectric and metal CMP, formation of the modified layer, either by hydrolysis or oxidation, is self-limited. That is to say, a suitable chemical environment for CMP should not cause any significant chemical etching of the surface. Indeed, there should be no substantial material removal without the action of the abrasive and pad. Within the context of classical corrosion, this implies that during metal CMP the surface should be effectively passivated. This combination of surface passivation and abrasion is the essence of the Kaufman model, which was first proposed to describe tungsten polishing [58]. Other workers have extended this model [59]. However, more recent work by Stein [60] and others [61, 62] indicates that the Kaufmann model suffers from serious limitations. To be specific, a simplistic picture of passivation and abrasion of the metal surface cannot account for all of the material removed during CMP. Even so, the overall picture of metal polishing as the result of mechanical interaction of an abrasive and pad with a chemically modified metal surface remains generally accepted.

It is still most common for free solid abrasive (typically aluminum oxide or silicon dioxide for metal CMP) to be dispersed in an aqueous solution containing the appropriate oxidizing chemistry, either as a suspension or a true colloid, to form a slurry. Conventional metal polishing then follows. A more recent development is incorporation of abrasive properties into the polishing pad itself. In this case, during the polishing process slurry is replaced by a polishing solution containing only the chemistry. This situation can be realized either by impregnation of pad material with the same free solid abrasives customarily used in slurries (so-called "fixed abrasive" technology) or by use of either the inherent or a modified pad microstructure without incorporation of additional abrasive (so-called "abrasive free" technology). In any case, as material is removed from the metal surface, the modified surface layer must be rapidly regenerated by the chemistry. In this way, aggregate metal overburden is removed and a planarized surface is formed.

In all types of chemical mechanical polishing, the interaction of both abrasive particles and pad materials with the material to be removed is typically considered to be mechanical in nature. Under this assumption, chemical modification of a polished surface is controlled primarily by species dissolved in the liquid component of the slurry and subsequent removal of the modified layer is due to mechanical action of the pad and abrasive. Indeed, it is the chemical reactivity of the liquid that provides the fundamental distinction between chemical mechanical and merely mechanical polishing in which abrasive is suspended in an inert liquid. Of course, this is an oversimplification and it is not really possible to separate chemical and mechanical effects so neatly. For example, at high pH, surfaces such as silicon dioxide are modified by hydroxylation, but this is not the case at lower pH. However, it has been found that many metal oxide abrasives efficiently polish quartz glass under neutral or even acidic conditions. Furthermore, the morphology of the resulting surface is inconsistent with gross mechanical abrasion, i.e., surface fractures,

scratching, and other types of abrasive mechanical defects are essentially absent. Corresponding situations also exist for metal polishing. Therefore, as suggested by Cook [63], the abrasive itself may possess some sort of "chemical tooth" that participates directly in material removal. This further suggests that during contact between the abrasive and the polished surface, perhaps pressure induced oxidation/hydration followed by adhesion of contacting surfaces results in material removal. However, the detailed removal mechanism is not well understood even for a relatively simple case such as glass polishing. Moreover, for metal polishing this situation is further complicated by the possibility that solid abrasives, e.g., cerium or manganese dioxides, might even react directly, perhaps as oxidizing agents. Now, if in addition to oxidizers, for various purposes buffers, surfactants, and/or complexing agents are also added to the slurry or polishing solution, the removal chemistry may become very complex indeed. Therefore, since chemistry plays such a central role in metal polishing, it is useful to review some basic concepts.

3.5 Acid–Base Equilibria

By definition, aqueous solutions are formed when various substances are dissolved in pure water. Dissolution may involve only dispersion of electrically neutral molecular species, e.g., water soluble organic compounds such as sugars or alcohols or non-condensable gases such as O_2 and N_2, or it may involve simultaneous dissociation and dispersion of electrically charged ionic species, e.g., inorganic salts and organic or mineral acids. In any case, dissolved molecular or ionic species are identified as *solutes* with water acting as the *solvent*. Furthermore, as a useful informal convention or "rule of thumb"; a given volume of a classical aqueous solution should contain at least roughly ten times more solvent, i.e., water, than any solute. If the solution becomes more concentrated than this, then its character is likely to deviate drastically from that of a classical aqueous solution. (Indeed, substantial deviation from ideal behavior is generally observable even at much lower solute concentrations.)

Now, at reasonable temperatures, pure water undergoes partial dissociation or *autoprotolysis* to form equal amounts of aqueous hydrogen and hydroxide ions. Of course, these ionic species are electrically charged and are conventionally represented as H^+ and OH^-. Autoprotolysis is customarily represented as a chemical equilibrium:

$$H_2O \rightleftarrows H^+ + OH^- \qquad K_a = 10^{-14}.$$

Here, K_a is the autoprotolysis equilibrium constant, which at $300°K$ has a standard value of $10^{-14} M^2/L^2$. In passing, it should be noted that while conceptually useful, the primitive hydrogen ion, H^+, has no real physical existence in aqueous solutions. Alternatively, the hydronium ion, H_3O^+, which explicitly indicates solvent interactions, is often used instead of H^+ as a more

realistic description of the chemistry of aqueous media. Even so, there is really no conclusive evidence for the existence of H_3O^+ as a definite aqueous species. Indeed, it is a common consensus that the structural nature of aqueous hydrogen ions is even now only imperfectly understood. Therefore, within the present context, it is convenient to ignore such issues and continue with the primitive usage of H^+.

Since in pure water equal amounts of hydrogen and hydroxide ions are formed by dissociation, the concentrations of each species must be 10^{-7} molar (i.e., moles/liter) under equilibrium conditions. This motivates the standard definition of pH:

$$pH = -\log_{10}[H^+].$$

Here, $[H^+]$ denotes the molar concentration of hydrogen ions. Likewise, $[OH^-]$ denotes the hydroxide ion molar concentration and, although not used as commonly, pOH can be analogously defined, thus:

Clearly, due to the autoprotolysis equilibrium, pH and pOH cannot be independent, but must be related by the simple additive formula:

$$pH + pOH = 14.$$

Obviously, for pure water pH and pOH are equal and, thus both have a value of 7. This defines the condition of neutrality.

By definition, a classical or Brønsted acid is any chemical compound that dissociates to form equivalent concentrations of hydrogen ions and "spectator anions" when dissolved in water. Likewise, a Brønsted base dissociates in water to form equivalent concentrations of hydroxide ions and "spectator cations". Both Brønsted acids and bases can be classified as strong or weak. By definition, strong acids and bases are essentially dissociated completely when dissolved in water. Examples of strong acids are the common mineral acids such as hydrochloric, nitric, or sulfuric acids and examples of strong bases are the common caustics such as sodium or potassium hydroxide. In contrast, weak acids or bases do not dissociate completely. Therefore, in addition to spectator and hydrogen or hydroxide ions, a solution of a weak acid or base also contains a substantial concentration of the undissociated parent species. These concentrations are dependent on temperature and are determined by a material specific thermodynamic equilibrium constant. Acetic acid and aqueous ammonia, i.e., ammonium hydroxide, are common examples of a weak acid and a weak base.

3.6 Buffering

In metal CMP it is generally desirable, among other things, to control pH in a relatively narrow range to allow for better control of the removal chemistry. This is accomplished by *buffering*. On the acidic side of neutral (pH < 7), a buffer system is set up as a mixed solution of a weak acid and a salt of the

same weak acid, e.g., acetic acid and sodium acetate. Similarly, on the basic side (pH > 7), a buffer system is formed by a solution of a weak base and a corresponding salt, e.g., ammonium hydroxide and ammonium chloride. Strong acids and bases cannot be used as buffering agents.

To understand how a buffer controls pH, one must consider dissociation equilibria of weak acids or weak bases. In the case of a weak monoprotic acid, this equilbrium can be represented as follows:

$$\mathrm{HA} \rightleftarrows \mathrm{H}^+ + \mathrm{A}^-.$$

Here, A^- denotes the spectator anion formed when a molecule of the weak acid, HA, dissociates. In chemical terminology, A^- is identified as the *conjugate base* of the acid, HA. Naturally, hydrogen ions are also formed by dissociation of HA. Thus, the corresponding equilibrium expression has the form:

$$K_\mathrm{a} = \frac{[\mathrm{H}^+][\mathrm{A}^-]}{[\mathrm{HA}]}.$$

Here, [HA] is the concentration of remaining undissociated acid, $[\mathrm{A}^-]$ is the conjugate base concentration, and K_a is the acid dissociation equilbrium constant for HA. For aqueous solutions, these constants are extensively tabulated in standard reference literature. By definition, K_A is small, i.e., $\ll 1$, for a weak acid. Thus, for a solution of the pure weak acid the pH is given by the formula:

$$\mathrm{pH} = \frac{1}{2}\left(\mathrm{p}K_\mathrm{a} - \log_{10}[\mathrm{HA}]\right).$$

In close analogy to pH, $\mathrm{p}K_\mathrm{a}$ is just defined as $-\log_{10} K_\mathrm{a}$. Clearly, this expression is obtained by combining the reaction stoichometry with the equilibrium expression. Furthermore, since HA is only slightly dissociated, [HA] can be taken as just the nominal concentration determined from initial mixing conditions.

Now suppose that in addition to the pure acid, HA, a significant concentration of the conjugate base, A^-, is independently added to the solution. This can be done either by direct addition of a salt containing A^- or by adding a strong base to the solution to partially neutralize some of the weak acid. Again, an expression for pH can be obtained from the reaction stoichometry and the equilibrium expression:

$$\mathrm{pH} = \mathrm{p}K_\mathrm{a} + \log_{10}\left(\frac{[\mathrm{A}^-]}{[\mathrm{HA}]}\right).$$

The important feature to observe in this formula is that pH is determined by $\mathrm{p}K_\mathrm{a}$ and the ratio of A^- and HA concentrations. Clearly, if the concentrations of A^- and HA are exactly equal, then pH and $\mathrm{p}K_\mathrm{a}$ are also equal. More to the point, if additional acid or base is either added to or chemically generated within the solution, in order to maintain equilibrium A^- is

converted to HA by excess acid or HA is converted to A^- by excess base, respectively. Provided that the amount added is not too large, the concentration ratio remains relatively close to unity. Since the dependence on the concentration ratio is logarithmic, this means that pH is maintained within a relatively small range about pK_a. A buffer system will continue to maintain a stable pH until substantially all of either A^- or HA is consumed by excess acid or base, respectively. By definition, the amount of excess acid or base that can be absorbed by a buffer determines *buffer capacity*. Of course, an entirely analogous treatment can be formulated for a weak base. Therefore, a buffer system can be obtained for any given pH value by a judicious choice of a weak acid or base having an optimum value of the dissociation equilibrium constant.

In passing, one should recall that any weak acid (or base) has a corresponding conjugate base (or acid). Therefore, conjugate acid and base dissociation constants are not independent, but are formally related since the

Table 3.1. Selected Acid Equilibrium Constants

Acid/Base System	K_a	pK_a	Charge state*
Acetic acid/Acetate	1.76×10^{-5}	4.75	0
Ammonium/Ammonia	5.59×10^{-10}	9.25	+1
Benzoic acid/Benzoate	6.46×10^{-5}	4.19	0
Citric acid/Dihydrogen Citrate	7.24×10^{-4}	3.14	0
Dihydrogen Citrate/Hydrogen Citrate	1.68×10^{-5}	4.77	−1
Hydrogen Citrate/Citrate	4.07×10^{-7}	6.39	−2
Oxalic Acid/Hydrogen Oxalate	5.90×10^{-2}	1.23	0
Hydrogen Oxalate/Oxalate	6.40×10^{-5}	4.19	−1
Pthallic acid/Hydrogen Pthallate	1.3×10^{-3}	2.89	0
Hydrogen Pthallate/Pthallate	3.09×10^{-6}	5.51	−1
Tartaric/Hydrogen Tartrate	1.04×10^{-3}	2.98	0
Hydrogen Tartrate/Tartrate	4.55×10^{-5}	4.34	−1
Hydroxylammonium/Hydroxylamine	9.33×10^{-7}	6.03	+1
Ethylammonium/Ethylamine	1.56×10^{-11}	10.81	+1
Ethylenediammonium/Aminoethylammonium	1.94×10^{-11}	10.71	+1
Aminoethylammonium/Ethylenediamine	2.73×10^{-8}	7.56	+2

Note: The suffix "-ium" denotes cationic species, the suffix "-ate" denote anionic species, all names denote neutral molecules.

* Charge state is for the acidic species; the conjugate base will have one unit of charge less (e.g., acetate has a charge of 1, ammonia has a charge of 0, etc.).

product of the two is exactly the autoprotolysis constant, K_a. Hence, litera-ture compilations are often given in terms of acid constants only, since cor-responding base dissociation constants can be trivially obtained from them. Furthermore, for acids that can dissociate more than once, e.g., phosphoric or pthalic acid, independent dissociation equilibrium constants are associated with each dissociation reaction. Selected acid constants are given in Table 3.1.

3.7 Oxidation–Reduction Reactions

As observed at the outset, for metal CMP an oxidizing chemical environment is required. Now, in colloquial usage, the term oxidation means chemical combination of a given material with oxygen, generally from the atmosphere. This is usually thought of in terms of combustion of fuels or tarnishing or rusting of metals. However, in a chemical sense, oxidation has a very pre-cise meaning. To be specific, a chemical element or compound is oxidized when it undergoes a net loss of valence electrons. Therefore, when a metallic substance, e.g., copper, is oxidized, neutral metal atoms lose electrons and, thus, take on a net positive charge. This might seem no different than simple ionization and, indeed, simple aqueous metal ions having the full net charge are often formed by oxidation. However, oxidation processes generally occur within a complex chemical environment in which the metal may be closely associated with other chemical species and in this way the net charge may be "hidden". Therefore, to cover all situations, the term *oxidation state* is introduced. By definition, a pure elemental metal is said to have an oxida-tion state of zero. Upon oxidation, this changes to a positive value, which corresponds numerically to the hidden net charge irrespective of the actual charge of the relevant chemical species.

Naturally, an inverse process called *reduction* must exist which involves a net gain of valence electrons. To relate this terminology to colloquial us-age, a metal atom is oxidized when it loses valence electrons to atmospheric oxygen to form metal oxide, but since oxygen gains electrons from the metal, it is reduced. (Of course, in the presence of not just oxygen, but also water or carbon dioxide, hydroxides and/or carbonates may be formed as well.) Similarly, within the context of metal CMP, the oxidizer species contained in the slurry or polishing solution undergoes reduction when the metal sur-face is oxidized. In general, since free electrons do not appear as a stable species in an aqueous environment, oxidation and reduction processes must always occur simultaneously such that electron transfer is exactly balanced. Hence, the overall process including both oxidation and reduction is called an *oxidation-reduction* reaction. This is commonly abbreviated to just "redox reaction." Clearly, a general characteristic of an oxidation-reduction reaction is the transfer of valence electrons between chemical species to increase the oxidation state of one species and to decrease the oxidation state of the other. This does not happen in acid-base reactions, which involve only dissociation

of acidic or basic compounds in water (or other solvents) to form dissolved species.

3.8 Half Reactions

Now, for the description of oxidation-reduction processes, it is generally convenient to divide an overall reaction into two explicit "half-reactions"; one of which involves the oxidized reactant and the other of which involves the reduced reactant. Of course, free electrons must now formally appear as transient reactants or products in these half-reactions, even though they do not exist physically in an aqueous environment. By definition, a reduction half-reaction has free electrons as a reactant species and an oxidation half-reaction has free electrons as a product species. Clearly, the overall reaction is obtained by combining oxidation and reduction half-reactions such that the same number of electrons occur as reactants and products and, hence "cancel out." Furthermore, it is obvious that any oxidation half-reaction becomes a reduction half-reaction (and vice-versa) if designations of reactants and products are inverted. Therefore, as a matter of convention, oxidation-reduction chemistry can be summarized exclusively either in terms of formal reduction or formal oxidation half-reactions. Indeed, both conventions are commonly in use, however, in recent years reduction half-reactions have become the preferred description. Thus, a generalized reduction half reaction can be represented as follows:

$$X_{ox} + ne^- \rightarrow X_{red}.$$
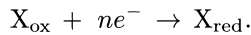
Here, X_{ox} and X_{red} denote oxidized and reduced species, respectively. Of course, the preceding half-reaction implies reduction of one equivalent (i.e., some consistent quantitative chemical unit of measure such as moles, molecules, or atoms) of X_{ox} to one equivalent of X_{red} and involves the transfer of n equivalents of electrons. Furthermore, X_{ox} and X_{red} may not correspond to just a single species, but rather may represent several species in aggregate. Indeed, in aqueous solutions solvent species, i.e., water and hydrogen or hydroxide ions, commonly appear as reactant and product species in half-reactions. Usually, they are not themselves oxidized or reduced, but are required to maintain overall charge and mass balance. As a consequence, any half-reaction that includes solvent species interacts with acid-base equilibria that might be present and, therefore, is generally dependent on pH.

3.9 Electrode Potentials

Association of a numerical *electrode potential* with any particular half-reaction serves to quantify oxidation-reduction chemistry. These values are generally determined experimentally and as in the case of acid dissociation

constants, extensive compilations of electrode potentials appear in the reference literature. Historically, oxidation-reduction chemistry was studied in connection with battery technology. In a battery, oxidation and reduction processes proceed on two independent electrodes immersed in an aqueous solution or *electrolyte*, which are physically separated by a membrane (or some other means). As a consequence of the separation, electrons transferred in the overall chemical reaction appear as electrical current flowing through an external circuit. The open circuit potential developed by a battery is a proportional measure of the thermodynamic free energy change of the overall reaction. Furthermore, the overall free energy change can be separated into independent contributions from oxidation and reduction processes. As a matter of convention, these are expressed as equivalent electrical potentials, i.e., electrode potentials. Of course, as with many thermodynamic quantities, these potentials must be evaluated relative to some fixed standard. By convention, this is the potential of the *standard hydrogen electrode* or SHE, which consists of an inert platinum electrode immersed in a unit normal acid solution, i.e., pH = 0, under one atmosphere of hydrogen gas at 300°K. Clearly, the reduction half-reaction associated with SHE can be written as follows:

$$H^+ + e^- \rightarrow \frac{1}{2}H_2.$$
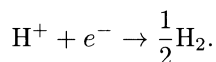
Here, one unit of aqueous hydrogen ion is reduced to one half unit of hydrogen gas. While it is not always realizable physically, the electrode potential appropriate to any half-reaction can be envisioned hypothetically as the open circuit potential generated by a battery made by combining SHE and an electrode on which the desired half-reaction is occurring. Therefore, this implies that the standard electrode potential of SHE is exactly zero. In addition, the *standard reduction potential* for any particular reduction half-reaction is defined as the corresponding electrode potential one expects to observe under conditions that all aqueous species are of unit concentration, i.e., unit molar, all gaseous species are at a pressure of one atmosphere, and the temperature is 300°K.

As a practical matter, standard reduction potentials serve to quantify the *oxidation strength* of a particular chemical species. In general, oxidizers appear as reactants in reduction half-reactions, such that the more positive the corresponding reduction potential, the stronger the oxidizer. For example, fluorine gas is the strongest oxidizer known and appears in the corresponding reduction half reaction:

$$\frac{1}{2}F_2 + H^+ + e^- \rightarrow HF.$$
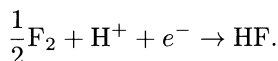
The corresponding standard reduction potential is 2.9178 V. Similarly, hydrogen peroxide, which is a commonly used oxidizer for metal polishing, appears in the half-reaction:

$$H_2O_2 + 2H^+ + 2e^- \rightarrow 2H_2O.$$

In this case, the standard reduction potential is 1.776 V. Obviously, hydrogen peroxide is not as strong an oxidant as fluorine, but it still has considerable oxidation strength.

Now, as mentioned previously, standard reduction potentials have been extensively tabulated. Furthermore, these are customarily arranged in numerical order (either ascending or descending) to define the *electrochemical series*. In principle, an oxidizer corresponding to a more positive reduction potential should be able to oxidize reduced species appearing as a product in a reduction half-reaction having a less positive reduction potential. In the preceding example, this means that fluorine gas should be able to oxidize water to hydrogen peroxide according to the overall reaction:

$$F_2 + 2H_2O \rightarrow 2HF + H_2O_2.$$
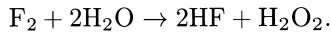
Of course, a relevant half-reaction for metal polishing is reduction of aqueous tungstate ion to metallic tungsten:

$$WO_4^{2-} + 8H^+ + 6e^- \rightarrow W + 4H_2O.$$
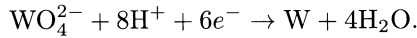
The corresponding standard reduction potential is found to be about 0.05 V. Clearly, hydrogen peroxide can be expected to oxidize metallic tungsten according to the formula:

$$W + 3H_2O_2 \rightarrow WO_4^{2-} + 2H_2O + 2H^+.$$
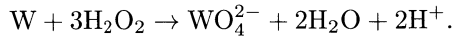
This overall reaction is actually observed since an aqueous solution of hydrogen peroxide is known to dissolve metallic tungsten at an appreciable rate.

Of course, standard conditions (unit concentrations, etc.) are only rarely, if ever realized in practice. Thus, it is desirable to obtain electrode potentials for actual conditions if at all possible. In principle, if E^0 and E are defined respectively as the standard reduction potential and the corresponding reduction potential adjusted to non-standard conditions, then this can be accomplished by means of the *Nernst equation*, which for a generic reduction half-reaction, takes the form:

$$E = E^0 + \frac{RT}{nF} \ln \left(\frac{A_{ox}}{A_{red}} \right).$$

Here, R is the perfect gas constant, T is absolute temperature, F is Faraday's constant, and n is the stoichometric coefficient of electrons in the reduction half-reaction. By definition, A_{ox} and A_{red} denote "aggregate activities" of oxidized and reduced species, i.e., X_{ox} and X_{red}, respectively. In practice, it is convenient to convert natural logarithms to base 10 logarithms:

$$E = E^0 + \frac{\varphi}{n} \log_{10} \left(\frac{A_{ox}}{A_{red}} \right).$$

In addition, *effective thermal potential*, φ, is defined as $(\ln 10)R/F$ and has a nominal value at $300°$K of 59.159366 mV (or mV per equivalent).

As is usual, A_{ox} and A_{red} are obtained by taking the product of individual chemical activities each raised to the power of its corresponding stoichometric coefficient. Thus, for the tungstate/tungsten half-reaction appearing previously, one can write:

$$E = E^0 + \frac{\varphi}{6} \log_{10} \left(\frac{A_{\mathrm{WO_4^{2-}}} A_{\mathrm{H^+}}^8}{A_{\mathrm{W}} A_{\mathrm{H_2O}}^4} \right).$$

It is evident that in this case the value of n must be 6. Now, activities of dissolved species correspond closely to and for most purposes can be estimated as molar concentrations. Furthermore, since chemical effects of water and/or solid materials such as metallic tungsten, are constant as long as any amount remains present, the corresponding activities are conventionally defined to be unity. Therefore, it follows that,

$$E = E^0 + \frac{\varphi}{6} \log_{10} \left([\mathrm{WO_4^{2-}}][\mathrm{H^+}]^8 \right) = E^0 + \frac{\varphi}{6} \log_{10} [\mathrm{WO_4^{2-}}] - \frac{4\varphi}{3} \mathrm{pH}.$$

Of course, as the electrode potential, E, becomes more positive, oxidation of metallic tungsten to tungstate ion becomes more difficult. Clearly, this implies that in the absence of competing effects, the presence of aqueous tungstate ion tends to suppress oxidation of metallic tungsten. In contrast, when the tungsten overburden is cleared from the substrate surface the tungstate concentration in the slurry must fall. It is evident from preceding considerations that this tends to promote tungsten dissolution in patterned features. At least in some cases, this may account for anomalously large rates of metal loss sometimes observed for particular patterned features. Furthermore, such issues are not limited to CMP, but may arise in other aqueous processes such as cleaning after CMP or other process steps. In addition, such effects are not limited to tungsten and can be observed for other types of metal polishing, most notably copper.

If for convenience, one assumes that all species except hydrogen ions are of unit activity, i.e., unit concentration, one atmosphere gas pressure, condensed phase, etc., then Nernst's equation can be expressed generally in terms of pH only:

$$E = E^0 + \frac{\varphi}{n} \log_{10} [\mathrm{H^+}]^{n_{\mathrm{H}}} = E^0 - \frac{n_{\mathrm{H}} \varphi}{n} \mathrm{pH}.$$

Here, n_{H} is the stoichiometric coefficient of hydrogen ion in the reduction half-reaction. By convention, n_{H} is taken to be positive if $\mathrm{H^+}$ appears as a reactant and negative if $\mathrm{H^+}$ appears as a product. Clearly, this general expression is consistent with the particular result obtained for the tungstate/tungsten half-reaction.

In passing, it should be mentioned that in the absence of complexing agents, metals that form surface oxide or hydroxide layers in alkaline media enter solution as soluble *aquo* cations. (An aquo cation is formed when

a simple "bare" metal ion complexes with some number (usually six) water molecules.) However, some metals form surface oxide layers in acidic media. In this case, instead of aquo cations soluble *oxo anions* may be formed in alkaline media. (An oxo anion is formed by a central metal atom strongly bonded to some number (usually four) oxygen atoms and is characteristic of high metallic oxidation states.)

3.10 Complexation

In the preceding discussion, only metal, metal containing compounds (oxides or hydroxides) and ions, and solvent species, such as hydrogen or hydroxide ions and water molecules, have been considered as chemically active. All other species have been treated as inactive spectators. Of course, this is a gross oversimplification. In particular, very many metal containing complexes are known to exist and can form in aqueous media. By definition, a metal *complex* is a chemical species that has a metal atom in a central position to which various *ligand* species are chemically bonded. The metal atom is usually in one of its common oxidation states, although complexes are known for almost any oxidation state (including zero). Since, the central metal atom is electron deficient due to oxidation, typical ligand species are electron donating molecules or ions. The number of ligands is called the *coordination number* and is typically four or six, but different coordination numbers are possible depending on the particular metal. Typically, the ligands form a regular geometrical arrangement around the central metal atom. For a coordination number of 6, this is a regular octahedron. For a coordination number of 4, both tetrahedral and square planar arrangements can occur.

To illustrate this further, one can consider the nature of the hypothetical aqueous species, M^{2+}, introduced previously. At first glance, it might appear that this just represents a metal atom that has lost two valence electrons. Indeed, simple atomic ions can and do exist in gases and plasmas, but as mentioned previously in aqueous solution they commonly form aquo cations, which in the present example consists of molecular water ligands coordinated with a central M atom in an oxidation state of two. Assuming the most common coordination number of 6, aqueous M^{2+} corresponds to the structure shown in Fig. 3.22.
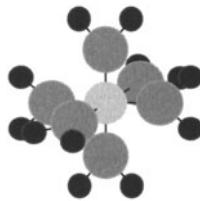
Fig. 3.22. Aquo cation corresponding to M^{2+}
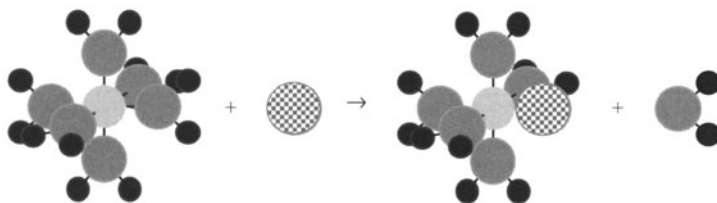
Fig. 3.23. Ligand exchange with an aquo cation

Here, the metal atom is represented by the central light gray sphere, and oxygen and hydrogen atoms by dark gray and black spheres, respectively. Since water molecules are uncharged, the cationic aquo complex, $[M(H_2O)_6]^{2+}$, has an overall charge of positive two.

In general, water is a relatively weak ligand. Indeed, there are many other aqueous species that complex much more strongly than water. Therefore, if such a species is present, it may displace water from the simple aquo complex to form a different, more stable complex. This reaction is illustrated pictorially in Fig. 3.23.

Here, the checkered sphere represents an alternative complexing species, perhaps a halide or pseudohalide ion, which displaces a water molecule from the aquo complex. Clearly, if the ligand is a sufficiently strong complexing agent, all of the water molecules may be displaced to form a complex containing only the central metal and the alternative complexing species. If, as above, the coordination number is 6, then complete ligand exchange is represented in Fig. 3.24.

Clearly, if the ligands are anionic species such as halide ions, the complex is no longer cationic, but becomes anionic with an overall charge of negative four. Even so, the central metal atom remains in its original oxidation state, which in this case is 2. Of course, if the ligand species is not so strong a complexing agent, then not all of the water may be displaced. This results in a series of complex species related by the appropriate equilibria. Furthermore, these equilibria are usually pH dependent. Indeed, this situation becomes especially complicated if one of the ligand species is the hydroxide ion, and *hydroxo* complexes are formed (as often happens in alkaline solutions). In this case, the equilibrium between aquo and hydroxo complexes can just as well be considered as an acid-base equilibrium as a complexation equilibrium.
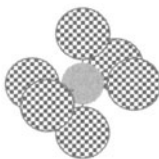
Fig. 3.24. Complex formed by complete ligand exchange

Some examples of common complexing agents are, as mentioned, aqueous anionic species such as halide ions, i.e., fluoride, chloride, bromide, and iodide, pseudohalide ions, e.g., cyanide, thiocyanate, etc., and various carboxylate anions. Furthermore, there are a number of neutral species such as ammonia and amines, which also can form strong complexes. Furthermore, there is a particularly interesting group of complexing species called *chelating agents*. (The resulting complexes are often called *chelates*, which is derived from the Greek word for "claw".) Typically, these are large organic molecules or ions that coordinate more than once with a metal atom. One of the most common of these is EDTA (ethylenediaminetetraacetic acid), though many others are known. This particular compound contains two amine and four carboxyl functional groups in every molecule. Each one of these functionalities can complex with the central metal atom, hence one single molecule of EDTA can satisfy sixfold coordination entirely by itself.

In general, complexation results in dissolved metallic species that are thermodynamically more stable than simple aquo ions. Therefore, in an oxidative environment this tends to promote dissolution of the metal and reduce the effect of any passive surface layer. For metal CMP, complexation can be generally expected to increase the removal rate, though precise effects are difficult to predict. Complexing agents can be added intentionally to the slurry or polishing solution as might be the case with phthalates or EDTA, or they may be formed as byproducts of oxidation-reduction, e.g., halide ions. Furthermore, amines or carboxylic acids can act as both buffering and complexing agents.

Again, for completeness one should observe that aquo and hydroxo complexes are formed in water if the oxidation state of the metal is relatively low, typically 4 or less. If the oxidation state is higher, e.g., 6 as is the usual case for tungsten, then the metal atom is so highly electron deficient that oxo anions are formed instead. Now, it can happen, usually for an oxidation state of 5, that an oxo cation is formed, which has characteristics of both oxo and aquo or hydroxo species. Not surprisingly, the chemistry becomes quite complicated; however, such situations are generally not important in metal CMP as currently practiced and, thus will be ignored.

In a typical oxo anion the central metal atom is very strongly bound to three or four oxygen atoms. In the most common case of four, these invariably form a tetrahedral arrangement. As is indicated on the right hand side of Fig. 3.25, hydrogen atoms can also be bound to oxygen atoms; usually

Fig. 3.25. Oxo anion

one or two, but this bonding is weak and these hydrogens are highly acidic. (In principle, an electrically neutral acid containing the oxo anion can be formed, however, these are often unstable.) Examples of common oxo anions are chromate and permanganate. (Obviously, for CMP, tungstate is of particular importance.) Although oxo anions have some similarities to typical complex species, they are usually not considered to be true complexes because the "oxide ligands" are so strongly bound to the central metal atom that they do not undergo ligand exchange reactions to any appreciable degree. Clearly, this means that common complexing agents cannot displace the oxygen atoms to form bonds with the central metal atom. Thus, within the context of metal CMP, if the dissolved metallic species is primarily an oxo anion, complexing agents can be expected to have little effect.

3.11 Surfactants and Inhibitors

In addition to oxidizers, buffers, and complexing agents, surfactants and inhibitors are also usually incorporated into polishing chemistry. By definition, a surfactant is a dissolved chemical agent that reduces the surface or interfacial tension of a liquid phase, typically water, in contact with some second phase. In general, the second phase can be a solid, liquid, or gas. In broad terms, this phenomenon is caused by accumulation of surfactant molecules at the liquid surface.

As a practical matter, surfactants are found to affect wetting of a liquid phase to a solid phase quite dramatically. Classically, wetting behavior is described by the geometry of a liquid drop resting on a solid surface. Furthermore, if one ignores any effect of gravity, i.e., the drop is relatively small, the drop shape can simply be taken to be that of a partial sphere. Phenomenologically, if wetting is poor, the drop "balls up" or "beads" and liquid is readily shed from the solid surface. In contrast, if the surface is well wetted, the drop "spreads" and liquid is not easily removed. Within this context, wetting behavior is conveniently described in terms of droplet contact angle, θ, which is illustrated by Fig. 3.26.
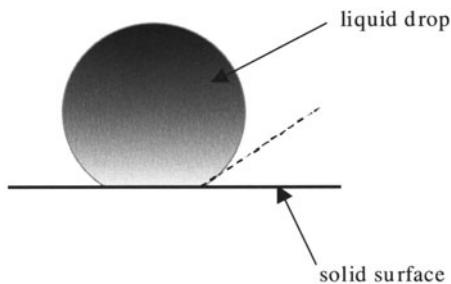

liquid drop

solid surface

Fig. 3.26. Definition of contact angle

Clearly, if θ equals π, i.e., $180°$, then the droplet touches the surface only at a single point, hence the liquid does not wet the solid at all. In contrast, if θ equals 0, then the liquid completely wets the solid surface. Naturally, partial wetting occurs for values of θ between 0 and π. Various methods are used for contact angle measurements, but all should give similar results.

If, as is usual, one assumes that when the liquid and solid materials are not in contact with each other, they are in contact with the atmosphere under standard conditions, then contact angle is determined by the values of air-solid, air-liquid, and liquid-solid interfacial tensions. This relationship is quantified by *Young's relation* [64]:

$$\cos\theta = \frac{\gamma_S - \gamma_{SL}}{\gamma_L}.$$

Here, γ_S and γ_L are air-solid and air-liquid surface tensions, respectively, and γ_{SL} is the liquid-solid interfacial tension. Generally, γ_S and γ_L are positively valued, but γ_{SL} is not necessarily positive and may become negative if the liquid strongly binds to the solid surface. Physically, interfacial tension is identified with the thermodynamic free energy of formation per unit area of interface. Therefore, an interface having a high interfacial tension is more difficult to form than one having a low interfacial tension. This is further made evident by explicit consideration of Young's relation. Clearly, if γ_{SL} is indefinitely large and positive, then the right hand side of Young's relation becomes less than -1, hence, θ is mathematically undefined. This means that a stable interface between the liquid and solid cannot be formed. In the case of water, such a solid surface is said to be *hydrophobic* and water de-wets from the surface under normal conditions. Conversely, if γ_{SL} is indefinitely large and negative, then the right hand side of Young's relation must be greater than 1. This means that not only is the liquid-solid interface stable, but that a liquid drop tends to spread indefinitely across the solid surface. Indeed, spreading will continue until the liquid forms an adsorbed molecular mono-layer. In this situation, bulk characteristics of the liquid phase are essentially lost. Of course, for water such a solid surface is said to be *hydrophilic*. Again, in a strict mathematical sense, θ is undefined, however, in practice contact angle is usually taken by definition to be 0 for a strongly hydrophilic surface. Similarly, θ is formally taken to be π for a strongly hydrophobic surface. Obviously, if γ_{SL} has intermediate values, then θ is between 0 and π and the solid surface is at least partially wetted.

As stated at the outset, the effect of a surfactant dissolved in a liquid is reduction of interfacial tension. Ideally, the surfactant concentration is so low that other physical properties of the liquid remain essentially unaffected. A large number of materials are now known that can act as surfactants when in dilute aqueous solution. However, the classical example remains that of ordinary soap, which is usually a sodium salt of a fatty acid, i.e., a long aliphatic chain carboxylic acid. Historically, fatty acids for soaps are derived from animal fats. More recently, corresponding alkali sulfates and sulfonates

or alkyl ammonium halides have been formulated as synthetic detergents. The main advantage of detergents over soaps is that unlike soaps, detergents do not form solid precipitates with alkaline earth and transition metal ions.

Irrespective of exact chemical formulations, compounds that act as aqueous surfactants have common structural characteristics. To be specific, they have a polar "head" group that has strong affinity to water due to dipole-dipole or ion-dipole interactions, but which is also attached to a relatively long (typically, eight or more carbon atoms), non-polar hydrocarbon chain or "tail". Thus, the polar head of a surfactant molecule is strongly solvated in aqueous solution, much like any other aqueous ionic species. However, the tail is not solvated readily and is similar to a non-polar hydrocarbon "oil" in its interaction with water. Clearly, if the head and tail groups could be separated as distinct chemical species, one would expect the head to remain dissolved, but the tail to separate into a separate phase. However, since head and tail groups are combined within a single chemical species, as illustrated in Fig. 3.27 surfactant molecules accumulate as a monolayer at solution interfaces with solvated head groups oriented into the bulk of the water and tails oriented outward.

Here, surfactant head groups are represented by filled circles, tail groups by irregular line segments, and the interface by the straight solid line. Now, it can happen that if the surfactant concentration is sufficiently large, a phase separation does occur. In this case, surfactant molecules form small globules called *micelles* that have head groups at the surface and tails in the interior. The chemical nature of the head group serves to classify surfactants as anionic, cationic, zwitterionic, or nonionic. These classifications are summarized in Table 3.2. By definition, anionic and cationic surfactants form negative and positive ions, respectively. Thus, the solution must also contain corresponding aqueous counterions such as alkali metal or ammonium cations (Na^+, K^+, NH_4^+, etc.) or halide anions (Cl^-, Br^-, etc.). In contrast, a zwitterionic surfactant combines anionic and a cationic groups within a single molecular species and, therefore, does not require separate counterions. Non-polar surfactants do not form aqueous ions even though they are solvated by water.
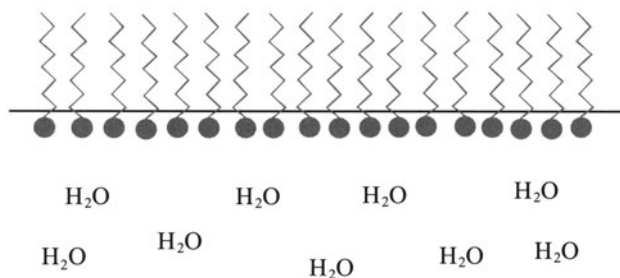
Fig. 3.27. Action of a surfactant

Table 3.2. Classification of Surfactants

Class	Example	Formula
Anionic	Alkali carboxylate	$R\text{-}COO^-\ M^+$ (M=Na, K, etc.)
	Sodium stearate	$H_3C(CH_2)_{16}\text{-}COO^-\ Na^+$
	Alkali alkyl sulfate	$R\text{-}OSO_2O^-\ M^+$ (M=Na, K, etc.)
	Sodium lauryl sulfate	$H_3C(CH_2)_{11}\text{-}OSO_2O^-\ Na^+$
	Alkali sulfonate	$R\text{-}SO_2O^-\ M^+$ (M=Na, K, etc.)
	Sodium lauryl benzyl sulfonate	$H_3C(CH_2)_{11}\text{-}C_6H_4\text{-}SO_2O^-\ Na^+$
Cationic	Alkylamine hydrohalide	$R\text{-}NH_3^+\ X^-$ (X=Cl, Br, etc.)
	Laurylamine hydrochloride	$H_3C(CH_2)_{11}\text{-}NH_3^+\ Cl^-$
	Tetraalkyl ammonium halide	$R\text{-}NR_3^+\ X^-$ (X=Cl, Br, etc.)
	Cetyl trimethylammonium bromide	$H_3C(CH_2)_{15}\text{-}N(CH_3)_3^+\ Br^-$
Zwitterionic	Alkyl betaine	$R\text{-}N^+(CH_3)_2CH_2COO^-$
	Lauryl betaine	$H_3C(CH_2)_{11}\text{-}N^+(CH_3)_2CH_2COO^-$
	Alkamidoalkyl betaine	$RCO\text{-}NH\text{-}R\text{-}N^+(CH_3)_2CH_2COO^-$
	Lauramidopropyl betaine	$H_3C(CH_2)_{10}CO\text{-}NH\text{-}(CH_2)_3\text{-}N^+(CH_3)_2CH_2COO^-$
	Alkamido 2-hydroxy propyl sulfobetaine	$RCO\text{-}NH\text{-}R\text{-}N^+(CH_3)_2CH_2CH(OH)CH_2SO_2O^-$
Nonionic	Polyoxyethylene alcohol	$R\text{-}(OCH_2CH_2)_n\text{-}OH$
	Alkylphenol ethoxylate	$R\text{-}C_6H_4\text{-}(OCH_2CH_2)_n\text{-}OH$

$R \equiv$ long chain hydrocarbon group (lauryl, cetyl, stearyl, etc.)
$R' \equiv$ short chain hydrocarbon group (methyl, ethyl, propyl, etc.)

Obviously, accumulation of surfactant molecules at solution interfaces can be expected to change interfacial free energy, that is to say surface tension. Furthermore, polar liquids, e.g., water, generally have significantly higher surface tensions than non-polar liquids, e.g., hydrocarbons, silicones, etc. Therefore, since non-polar tails of surfactant molecules are oriented outward, the corresponding interfaces are modified and take on some of the characteristics of those formed by pure non-polar liquids. Hence, surface tensions of aqueous media containing surfactant are significantly lowered in comparison to solutions containing no surfactant. This results in better wetting, formation of emulsions, as well as other associated phenomena. Of course, the degree of interface modification depends directly on surfactant concentration. Within the context of CMP, surfactants change the interaction between abrasive particles and the solution. This allows for better particle suspension and collateral reduction of particle agglomeration. This is particularly important for copper CMP since copper is soft and easily scratched.

As do surfactants, inhibitors also modify interfaces. However, the relevant modification is now not a reduction of surface tension, but reduction of some particular surface chemical reaction rate. Within the context of metal polishing, this typically means inhibition of oxidation of the metal surface, i.e., inhibition of corrosion. It can happen that the same chemical species acts both as a surfactant and as an inhibitor. However, more commonly these are different species. Generally, the common characteristic of corrosion inhibitors is that under favorable conditions they are adsorbed onto the metal surface to form a protective coating. Naturally, the structure of this coating is dependent on inhibitor concentration, electrochemical potential, pH, etc. One of the most widely known corrosion inhibitors is benzotriazole (BTA) and related derivatives, such as tolyltriazole (TTA, 5-methyl benzotriazole). These compounds are characterized by a five member triazole ring consisting of three nitrogen and two carbon atoms fused with a six member benzene ring. Molecular structures for BTA and TTA are given in Fig. 3.28. In addition to triazoles, other amines or amino acids can also serve as corrosion inhibitors.

In the case of copper and BTA, Vogt et al. [65, 66] have given a detailed description of copper corrosion and inhibition of corrosion by BTA in acidic media. In this work, rather direct observations of atomic scale surface struc-
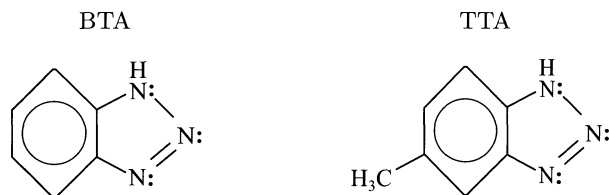

Fig. 3.28. Structure of triazoles

ture have been obtained by scanning tunneling microscopy (STM). Now, in the absence of BTA, for copper dissolution in sulfuric or hydrochloric acids, a pronounced difference in surface morphology was observed. This difference was attributed to complexation of oxidized copper by chloride ions, i.e., formation of the complex ion $CuCl_2{}^-$, in hydrochloric acid. According to Vogt et al., when BTA is directly adsorbed to copper the triazole group binds to the surface with the benzene ring oriented outward toward the solution and the ring plane nearly perpendicular. Of course, at low concentrations surface coverage must be incomplete. Clusters and chains of BTA molecules were observed by STM. As concentration is increased, a monolayer is formed and at even higher concentrations a relatively thick multilayer incorporating a small amount of oxidized copper is precipitated. Formation of the BTA layer is self-limiting and, hence, the copper surface is protected from further chemical attack. Clearly, the overall result is similar to natural passivation of a metal surface by the formation of an impermeable oxide layer. Indeed, the BTA layer can be considered as an artificial passivation. It was also found that the presence of chloride ion reduced the effectiveness of a complexing species such as chloride ion can reduce the effectiveness of a corrosion inhibitor.

3.12 The Future of Metal Polishing

No discussion of metal polishing could be complete without some prognostication as to future trends. It seems clear that tungsten and copper polishing will continue to be mainstream fabrication techniques for the foreseeable future. Within this context, equipment and consumables will be improved to support these processes. Although it has not been discussed previously, there continues to be a small amount of interest in damascene aluminum fabrication. However, aluminum is difficult to polish owing to the intrinsic softness of the metal and hardness of the oxide. In addition, dry etching processes for aluminum are well established and appear to be extendable to the required dimensional tolerances. Therefore, it is not likely that aluminum CMP will ever become a widely implemented process.

Of more importance is noble metal polishing for advanced nonvolatile memory device fabrication [67, 68]. Of particular interest are platinum and iridium. These metals are required as electrodes for perovskite materials such as PZT (lead zirconium titanate), etc. Currently, there is considerable interest in CMP of these metals because, as is the case with copper, they are nearly impossible to etch cleanly and anisotropically. Unfortunately, due to their general chemical inertness, they are also difficult to polish. However, reasonable results have been obtained with various proprietary slurries and removal rates as high as 200 nm/min have been observed. In addition, CMP of noble metal oxides is also of interest since these materials are also used in this application.

A closely related area is damascene gate fabrication. This technology is being driven by fundamental limitation of device performance due to carrier

depletion of doped polysilicon gate electrodes [69, 70]. At present there are many candidate metals for this application. However, this choice becomes complicated in a dual metal gate scheme, because for the p-channel devices a metal with a high work function is required. This severely limits possibilities and one is again forced to consider the noble metals such as platinum and iridium. In contrast, for n-channel devices a relatively low work function is required and many metals are applicable. Even so, optimization of CMP for such a process presents a serious challenge.

References

1. See also: S. Wolf and R.N. Tauber, "Silicon Processing for the VLSI Era", vols. 1–3, Lattice Press, Sunset Beach, CA, 1990–2000; S.M. Sze, "VLSI Processing", Wiley, New York, 1988.
2. N. Elbel, B. Neureither, B. Ebersberger and P. Lahnor, J. Electrochem. Soc., **145**, 1659, (1998).
3. T. Park, T. Tugbawa, J. Yoon, D. Boning, J. Chung, R. Muralidhar, S. Hymes, Y. Gotkis, S. Alamgir, R. Walesa, L. Shumway, G. Wu, F. Zhang, R. Kistler and J. Hawkins, Proceedings 1998 VMIC Conference, 437, IMIC, Tampa, 1998.
4. E.K. Broadbent and C.L. Ramiller, J. Electrochem. Soc., **131**, 1427, 1984.
5. See also: "Tungsten and Other Refractory Metals for VLSI Applications", parts I–IV, MRS, Pittsburg, PA, 1985–1988.
6. R.J. Saia, B. Gorowitz, D. Woodruff and D.M. Brown, J. Electrochem. Soc., **135**, 936, 1988.
7. J. Lee and D.C. Hartmann, Proc. IEEE VMIC IV, 193, 1987.
8. S. Mehta, S. Mittal, A. Haranahalli and D. Ranadive, *Proc. IEEE VMIC III*, 418, 1986.
9. R.C. Ellwanger, J.E.J. Schmitz, R.A.M. Wolters, and A.J.M. van Dijk, "Tungsten and Other Refractory Metals for VLSI Applications II", pg. 385, Ed.E.K. Broadbent, MRS, Pittsburg, PA, 1986.
10. R.H. Wilson, B. Gorowitz, A.G. Williams, R. Chow and S. Kang, J. Electrochem. Soc., **134**, 1876, 1987.
11. H.P. Hey, A.K. Sinha, S.D. Steenwyk, V.V.S. Rana and J.L. Yeh, IEDM Tech. Dig., 50, 1986.
12. S. Kang, R. Chow, R.H. Wilson, B. Gorowitz and A.G. Williams, J. Electron. Mater., **17**, 213, 1988.
13. R.V. Joshi, S.B. Brodsky, T. Bucelot, M.A. Jaso and R. Uttecht, *Proc. IEEE VMIC VI*, 113, 1989.
14. F. White, W. Hill, S. Eslinger, E. Payne, W. Cote, B. Chen, and K. Johnson, *IEDM Tech. Dig.*, 301, 1992.
15. C. Yu, S. Poon, Y. Limb, T.-K. Yu and J. Klein, *Proc. VMIC XI*, 144, 1994.
16. J. Givens, S. Geissler, O. Cain, W. Clark, C. Koburger and J. Lee, *Proc. VMIC XI*, 43, 1994.
17. C. Yi, W.C. Tu, K. Tsai, S. Hsieh and H.C. Chen, Proceedings 1997 CMP–MIC Conference, 107, IMIC, Tampa, 1997.
18. V. Blaschke, L. Witters, S.-W. Hsia, D. Dornisch and K. Rafftesaeth, Proceedings 1997 CMP–MIC Conference, 219, IMIC, Tampa, 1997.

19. M. Rutten, P. Feeney, R. Cheek and W. Landers, Proceedings 1995 VMIC Conference, 491, IMIC, Tampa, 1995.
20. K. Wijekoon, R. Lin, S. Yang, F. Redeker, S. Nanjangud, M. Bakshi and S. Ghaneyam, Proceedings 1998 VMIC Conference, 451, IMIC, Tampa, 1998.
21. G. Springer, Proceedings 1999 CMP–MIC Conference, 45, IMIC, Tampa, 1999.
22. L.-J. Chen, Y.-L. Huang, Z.-H. Lin and H.-W. Chiou, Proceedings 1998 CMP–MIC Conference,28, IMIC, Tampa, 1998.
23. H.W. Chiou, L.J. Chen and H.C. Chen, Proceedings 1997 CMP–MIC Conference, 131, IMIC, Tampa, 1997.
24. E. Sicurani, M. Fayolle, Y. Gobil, Y. Morand and F. Tardif, "Advanced Metallization and Interconnect Systems for ULSI Applications in 1996", 561, MRS, Pittsburgh, PA, 1997.
25. D.J. Stein, D.L. Hetherington and J.L. Cecchi, J. Electrochem. Soc., **146**, 376, 1999.
26. M.-C. Yang, F-Y. Shau, C-S. Huang, C. Yi and R. Tang, Proceedings 1998 CMP–MIC Conference, 216, IMIC, Tampa, 1998.
27. S.H. Li, H. Banvillet, C. Augagneur, B. Miller, M-P. Nabot–Henaff, and K. Wooldridge, Proceedings 1998 CMP–MIC Conference, 165, 1998.
28. D. Tamboli, V. Desai and S. Seal, Proc. Electrochem. Soc., **2000–26**, 213, 2001.
29. D.J. Stein, D.L. Hetherington, and J.L. Cecchi, J. Electrochem. Soc., **146**, 1934, 1999.
30. M Rutten, P. Feeney, R. Cheek and W. Landers, *Semicond. Int.*, **9**, 123, 1995.
31. H. van Kranenburg and P.H. Woerlee, Proceedings 1997 CMP–MIC Conference, 91, IMIC, Tampa, 1997.
32. S.R. Roy, "Advanced Metallization and Interconnect Systems for ULSI Applications in 1995", 733, MRS, Pittsburg, PA, 1996.
33. B. Luther, J.F. White, C. Uzoh, T. Cacouris, J. Hummel, W. Guthrie, N. Lustig, S. Greco, N. Greco, S. Zuhoski, P. Agnello, E. Colgan, S. Mathad, L. Saraf, E.J. Weitzman, C.K. Hu, F. Kaufman, M. Jaso, L.P. Buchwalter, S. Reynolds, C. Smart, D. Edelstein, E. Baran, S. Cohen, C.M. Knoedler, J. Malinowski, J. Horkans, H. Deligianni, J. Harper, P.C. Andricacos, J. Paraszczak, D.J. Pearson and M. Small, Proceedings 1993 VMIC Conference, 15, 1993.
34. W. Lee, H. Yang and J. Lee, Proc. Electrochem. Soc., **2000–27**, 63, 2001.
35. C. Verove, B. Descouts, P. Gayer, M. Guillermet, E. Sabouret, P. Spinelli and E. Van der Vegt, Proceedings 2000 International Interconnect Technology Conference, 267, 2000.
36. J.-F. Wang, A.R. Sethuraman, L.M. Cook, D.R. Evans and V.L. Shannon, Proceedings 1995 VMIC Conference, 505, 1995.
37. M. Hariharaputhiran, S. Ramarajan, Y. Li and S.V. Babu, "Chemical Mechanical Polishing-Fundamentals and Challenges", 129, MRS, Pittsburg, PA, 2000.
38. D.R. Evans, *Proc. of the Electrochem. Soc.*, **96–22**, 70, 1997.
39. J.A.T. Norman, B.A. Muratore, P.N. Dyer, D.A. Roberts and A.K. Hochberg, *Proc. IEEE VMIC VIII*, 123, 1991.
40. J.A.T. Norman, D.A. Roberts, A.K. Hochberg, P. Smith, G.A. Petersen, J.E. Parmeter, C.A. Apblett and T.R. Omstead, Thin Solid Films, **262**, 46, 1995.
41. N. Awaya, K. Ohno and Y. Arita, J. Electrochem. Soc., **142**, 3173, 1995.
42. K.H. Min, G.C. Jun and K.B. Kim, J. Vac. Sci. Tech., **B 14**, 3263, 1996.
43. J.O. Olowolafe, C.J. Mogab, R.B. Gregory and M. Kottke, J. Appl. Phys., **72**, 4099 (1992).

44. R. Contolini, S. Mayer and A. Bernhardt, Proceedings 1993 VMIC Conferernce, 470, 1993.
45. R.J. Contolini, L. Tarte, R.T. Graff and L.B. Evans, Proceedings 1995 VMIC Conference, 322, 1995.
46. T. Andryushchenko, W. Holtkamp, W.C. Ko, F. Lin, D. Papapanayiotou and C.H. Ting, Proceedings 1998 VMIC, 55, 1998.
47. J.M.E. Harper, C. Cabral, Jr., P.C. Andricacos, L. Gignac, I.C. Noyan, K.P. Rodbell and C.K. Hu, J. Appl. Phys., **86**, 2516, 1999.
48. H. Lee, S.D. Lopatin and S.S. Wong, Proceedings 2000 International Interconnect Technology Conference, 114, 2000.
49. S.H. Brongersma, E. Richard, I. Vervoort and K. Maex, Proceedings 2000 International Interconnect Technology Conference, 31, 2000.
50. U. Landau, *Proc. of the Electrochem. Soc.*, **2000–26**, 231, 2001.
51. S. Sankaran, W. Harris, G. Nuesca, E.O. Shaffer, S.J. Martin and R.E. Geer, Proceedings 2000 International Interconnect Technology Conference, 40, 2000.
52. D.R. Evans, M.R. Oliver and M. Kulus, *Proc. of the Electrochem. Soc.*, **2000–26**, 122, 2001.
53. S. Kondo, N. Sakuma, Y. Homma, Y. Goto, N. Ohashi, H. Yamaguchi and N. Owada, Proceedings 2000 International Interconnect Technology Conference, 253, 2000.
54. H. Yamaguchi, N. Ohashi, T. Imai, K. Torii, J. Noguchi, T. Fujiwara, T. Saito, N. Owada, Y. Homma, S. Kondo and K. Hinode, Proceedings 2000 International Interconnect Technology Conference, 264, 2000.
55. Y. Kamigata, Y. Kurata, K. Masuda, J. Amanokura, and M. Yoshida, *Chemical Mechanical Polishing 2001 - Fundamentals and Challenges*, PV-671, MRS, Warrendale, PA, 2001.
56. T. Laursen and M. Grief, *Chemical Mechanical Polishing 2001 - Fundamentals and Challenges*, PV-671, MRS, Warrendale, PA, 2001.
57. F. Preston, J. Soc. Glass Technol., **11**, 214, 1927.
58. F.B. Kaufman, D.B. Thompson, R.E. Broadie, M. A Jaso, W.L. Guthrie, D.J. Pearson, and M.B. Small, J. Electrochem. Soc., **138**, 3460, 1991.
59. A.R. Sethuraman and J-F. Wang, *Proc. of the Electrochem. Soc.*, **96–22**, 258 (1997).
60. D.J. Stein, "Mechanistic, Kinetic, and Processing Aspects of Tungsten Chemical Mechanical Polishing", Ph.D. dissertation, U. of N. Mex., Albuquerque, 1998.
61. E.A. Kneer, C. Raghunath, S. Raghavan and J.S. Jeon; J. Electrochem. Soc., **143**, 4095, 1996.
62. E.A. Kneer, C. Raghunath, S. Raghavan and J.S. Jeon; J. Electrochem. Soc., **144**, 3041, 1997.
63. L.M. Cook, J. of Non-Crystalline Solids, **120**, 152, 1990.
64. See also: J.A. Dean, "Lange's Handbook of Chemistry", 14th or 15th Ed., McGraw–Hill, NY, 1992–2000; R.C. Weast, Ed., "Handbook of Chemistry and Physics", CRC Press, Boca Raton, FL, *published annually.*
65. M.R. Vogt, F.A. Mller, C. M Schilz, O.M. Magnussen and R.J. Behm, Surf. Sci. **367**, L33, 1996.
66. M.R. Vogt, A. Lachenwitzer, O.M. Magnussen and R.J. Behm, Surf. Sci. **399**, 49, 1996.
67. D. Evans, U.S. Patent **6,290,736**, 2001.

68. T.K. Li, S.T. Hsu, B. Ulrich, H. Ying, L. Stecker, D. Evans, Y. Ono, J.-S. Maa and J.J. Lee, Appl. Phys. Lett., **79**, 1661, 2001.
69. D. Evans and S.T. Hsu, U.S. Patent **6,133,106**, 2000.
70. S.T. Hsu, D.R. Evans and T. Nguyen, U.S. Patent **6,274,421**, 2001.

4 Metal CMP Science

David Stein

4.1 Introduction

Tungsten CMP slurry chemistries are more reactive than oxide CMP slurries. Abrasive in a pH controlled fluid is not able to remove the tungsten metal, presumably due to the hardness and the relative inertness of tungsten. To overcome these difficulties, an oxidizer is incorporated in tungsten CMP slurries. Kaufman et al. [1] introduced the first widely accepted model of the tungsten CMP process; a part of that work included a model for the removal of tungsten during CMP. The model involves the formation, due to the oxidizing nature of the slurry, of a blanket tungsten oxide film (passivated layer) hypothesized to be softer than the un-passivated metal. Figure 4.1 depicts the model. According to this model, mechanical abrasion from the slurry and the polishing pad, where the pad contacts the tungsten film, removes the passivation layer. The bare metal, when re-exposed to the oxidizer, immediately oxidizes. The abrasion-passivation process is hypothesized to continue until a material that does not passivate and/or does not lend itself to mechanical abrasion (a stop layer) is reached. This mechanism requires that all tungsten removed be in an oxidized state.

The constituents of the aqueous polish slurry used by Kaufman were potassium ferricyanide (an oxidant), ethylene diamine (a possible complexant for tungsten), and potassium dihydrogen phosphate (a buffering agent, slurry pH ≈ 6). The proposed etching reaction was

$$W + 6Fe(CN)_6^{-3} + 4H_2O \rightarrow WO_4^{2-} + 6Fe(CN)_6^{-4} + 8H^+. \qquad (4.1)$$
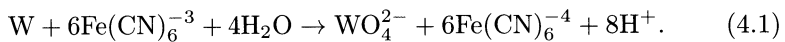
The competing passivation reaction was proposed as

$$W + 6Fe(CN)_6^{-3} + 3H_2O \rightarrow WO_3 + 6Fe(CN)_6^{-4} + 6H^+. \qquad (4.2)$$
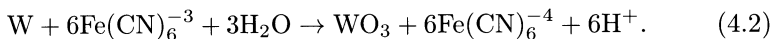
The authors suggested that the relative rate of each process was dependent on pH and table rotation rate due to the hydrogen ion formation and transport rates.
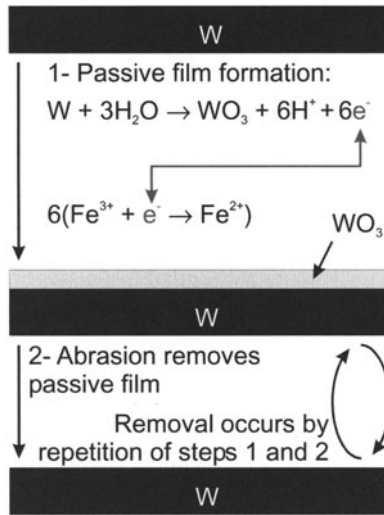
Fig. 4.1. The model proposed by Kaufman for tungsten removal during CMP. From [1]

4.2 Tungsten Experimental Data— Chemical and Electrochemical

The mechanism described above requires that all tungsten removed during the CMP process be in an oxidized state. The oxidation of tungsten in solution can be studied using electrochemical techniques. The electrochemical behavior of tungsten in hydrogen peroxide solutions had been investigated prior to the development of CMP applications [2]. Tungsten dissolves without inhibition in hydrogen peroxide and the rate of dissolution is proportional to the peroxide concentration. However, the dissolution rate does not increase with the addition of HNO_3, HF, H_2SO_4, H_3PO_4, acetic acid, or NaOH. The proposed dissolution mechanism is

$$W + 2H_2O_2 \rightarrow WO_2 + 2H_2O \tag{4.3}$$

$$2WO_2 + 6H_2O_2 \rightarrow H_2W_2O_{11} + 5H_2O \tag{4.4}$$

$$3H_2W_2O_{11} + 7H_2O \rightarrow 2H_2W_3O_{12} + 8H_2O_2 \ . \tag{4.5}$$
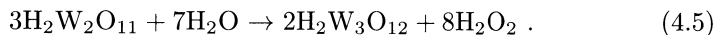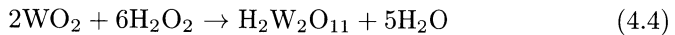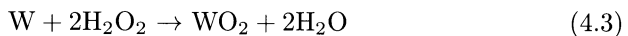
The long-term stable tungsten species in solution as a function of W potential and solution pH are shown in a Pourbaix diagram [m1](see Appendix). The diagram considers only water species and the concentration of tungsten species is constant at 10^{-4} M. These diagrams describe the system at equilibrium; they do not give any information about rates of change on species concentrations in the system. The lines labeled "a" and "b" represent the stability of water. Below a pH of 4 the non-soluble oxide WO_3 is stable. Above a pH of

4 the soluble oxide WO_4^{2-} becomes stable. Pourbaix mentions that tungsten forms numerous complexes that are not indicated on the diagram such as the hydrochloric complexes with trivalent tungsten and the cyanide complexes with pentavalent tungsten. The tungstates of the alkali metals are soluble while other tungstates are not. Also shown in the Appendix are the general regions of tungsten passivation (WO_3) and corrosion (WO_4^{2-}), as well as the solubility of WO_4^{2-} as a function of pH (the line labeled "4" is the solubility).

Kneer et al. [3] have investigated the electrochemistry of chemical-vapor deposited (CVD) W in aqueous solutions of interest in W CMP. They obtain polarization curves of CVD W in solutions of pH 2 and 4, with and without 5% hydrogen peroxide, which are shown in Fig. 4.2.

Their investigation was not performed under polishing conditions. The authors suggest that an existing surface oxide may affect the dissolution of the tungsten. They studied the electrochemical characteristics of an anodically grown W film. These results are shown in Fig. 4.3.

The values of the corrosion current, i_{corr}, measured without hydrogen peroxide are much lower than those measured with hydrogen peroxide. Figure 4.4 shows the values of i_{corr} obtained. Note that a current density of $10\,\mu A\,cm^{-2}$ is equivalent to a blanket removal rate of tungsten of $1\,\text{Å}\,min^{-1}$. They found that tungsten readily dissolves in hydrogen peroxide solutions at a pH of 2 or 4 and only a thin oxide film, less than 50 Å thick, was present on the surface after drying (measured *ex-situ* by X-ray photoelectron spectroscopy). They proposed that this oxide could easily have been formed during the cleaning and drying process used before the sample was placed into the high vacuum chamber of the XPS system. Figure 4.5 shows the XPS spectra obtained from W samples immersed for 48 hours. This data indicate that W, WO_2,
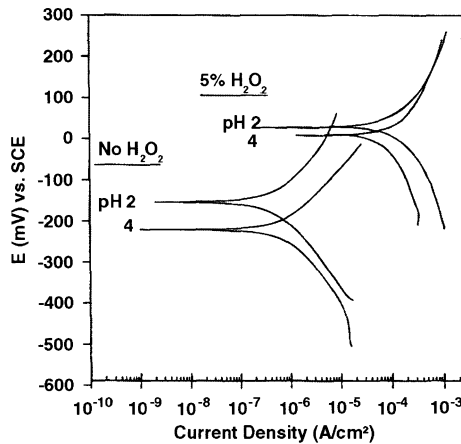
Fig. 4.2. Polarization curves of tungsten obtained without abrasion insolutions with and without H_2O_2 at pH 2 and pH 4. From [3]
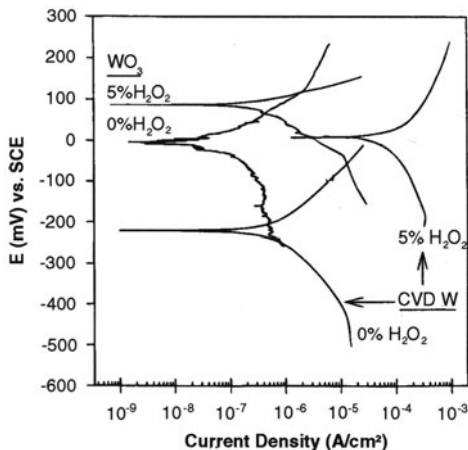
Fig. 4.3. Polarization curves of tungsten obtained without abrasion insolutions with and without H_2O_2 at pH 2 and pH 4. From [3]
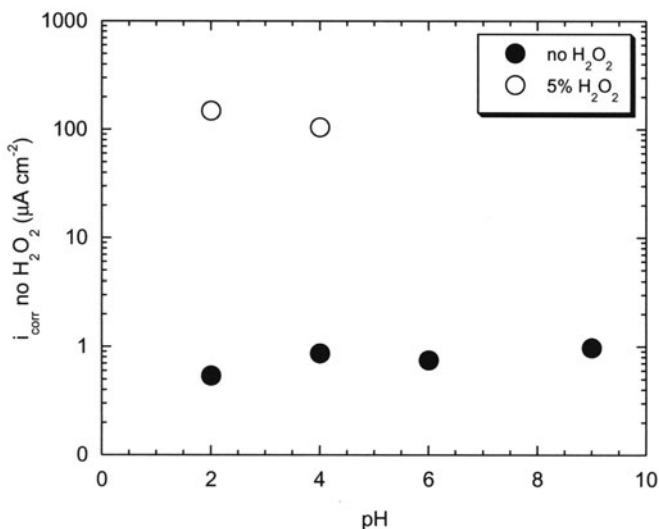
Fig. 4.4. Values of i_{corr} obtained from the polarization curves. Adapted from data presented in [3]

and WO_3 are present on the surface of each sample. They concluded that tungsten does not form a passive film in the solutions containing hydrogen peroxide. Thus the abrasion-passivation model proposed by Kaufman et al. was not applicable to CMP with hydrogen peroxide-based slurries.

In another study Tamboli et al. [4] obtained similar XPS data. Tungsten films were anodically polarized in solutions of KIO_3 and H_2O_2 at a pH
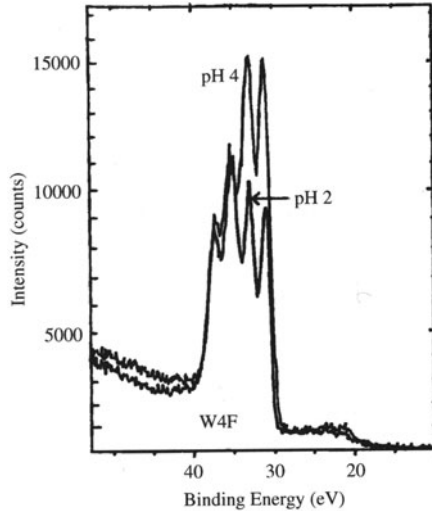
Fig. 4.5. The XPS spectra obtained from W samples immersed for 48 hours. From [3]
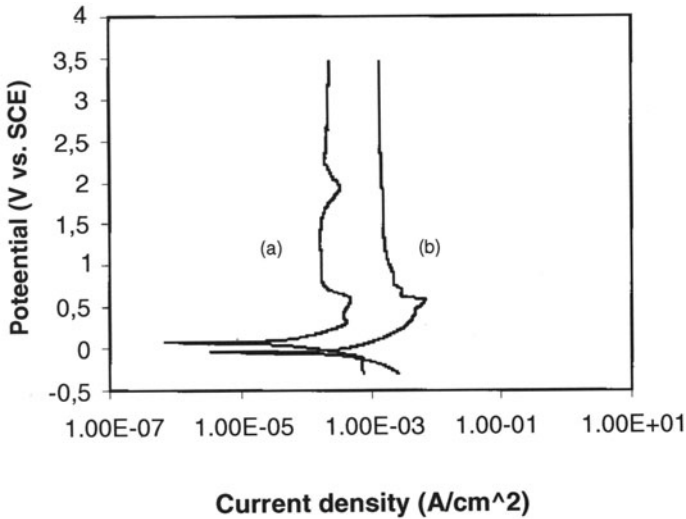
Fig. 4.6. The potentiodynamic polarization scans obtained in solutions of KIO_3 and H_2O_2 at a pH of 4. From [4]

of 4. Figure 4.6 shows the potentiodynamic polarization scans obtained. The authors note that W in both solutions shows a passivation regime (the indicated decrease in current density) but that the passivation regime for the W sample polarized in H_2O_2 is almost one order of magnitude greater than the

current density of the passivation regime in KIO_3. XPS spectra were taken on samples polarized at 0.5, 1.8, and 3.0 V with respect to a saturated calomel reference electrode (SCE). Figure 4.7 shows the variation in ratio of total oxygen content to tungsten content with polarization potential. The total oxygen content of the sample polarized in KIO_3 increases due to the growth of the oxide and additional chemisorbtion of OH groups. The total oxygen content of the sample polarized in H_2O_2 does not increase with applied potential, a trend that is contrary to the assumption that the change in current density shown in Fig. 4.7 is a passivation regime. Instead, the authors believe that what was initially assumed to be a passivation regime is actually a mass-transfer limited dissolution of tungsten oxide. At lower potentials the oxide growth is low enough that there is no interference of the direct dissolution of the W by the H_2O_2. This study does not investigate the nature or quantity of oxide that may grow on the surface of tungsten that is undergoing polishing.

Kneer et al. [5] also studied the re-passivation of W after polish. Figure 4.8 shows open circuit potential (OCP) lines as a function of polishing time for samples with an anodically grown tungsten oxide film. During polishing (starting at time 0) the OCP drops as the passivating film is removed. The OCP reaches a steady state when the passivating film has been removed and only bare metal is exposed to the slurry. They then stopped polishing and allowed a natural passivating film to reform on the tungsten. This resulted in an increase in the OCP. Note that the increase in OCP is large for the slurries without hydrogen peroxide and very small for the slurries containing hydrogen peroxide. This supports the conclusion stated above that
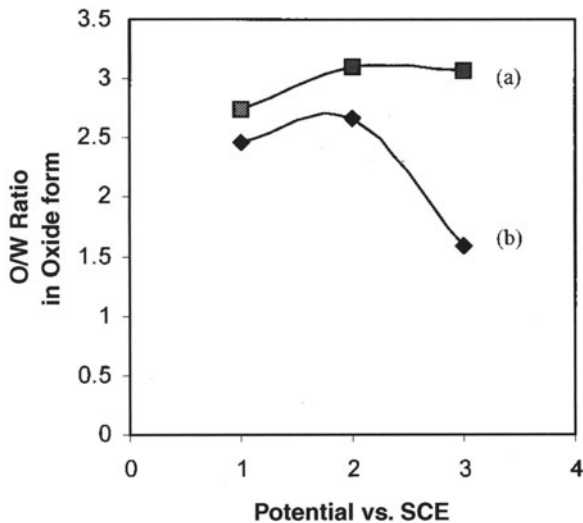
Fig. 4.7. The variation in ratio of total oxygen content to tungsten content with polarization potential. From [4]
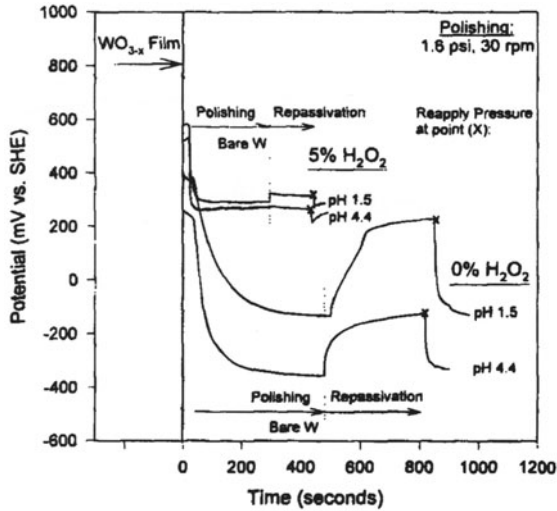
Fig. 4.8. Open circuit potential (OCP) lines as a function of exposure time for samples with an anodically grown tungsten oxide film. From [3]

the slurries containing hydrogen peroxide form little or no passivating film. Polishing began again after a stable OCP was obtained. The natural passive film that formed was removed much more quickly than the grown passive film as indicated by the quick return of the OCP to the steady state value during polish.

Kneer et al. also measured the corrosion current density of W during polish in various slurry chemistries. The highest corrosion rate measured is $30\,\text{\AA}\,\text{min}^{-1}$. The authors compared this number to typical values obtained during commercial utilization of CMP and concluded that, "[the corrosion rate] is miniscule when compared to the [removal] rates of 150 to 600 nm/min obtained under commercial CMP conditions. On this basis, it appears that the chemical oxidation and dissolution of CVD tungsten is not the primary removal mechanism in CMP."

Stein et al. [5, 6] measured the corrosion rate of tungsten with and without polishing and compared the electrochemically measured rates to the polish rate determined by weight loss of the same sample. Polish conditions in the experimental apparatus (pressure, speed, and slurry composition) were designed to be as close to commercial CMP as possible. Corrosion rates were calculated using a fit of the DC polarization data to the Butler–Volmer equation and relative corrosion rates were calculated from AC impedance spectroscopy data.

Figure 4.9a and 4.9b shows representative Tafel plots obtained by Stein et al. Figure 4.9a is data taken under static conditions (no polishing). Tungsten etch rates measured for each sample using weight loss are shown for the basic
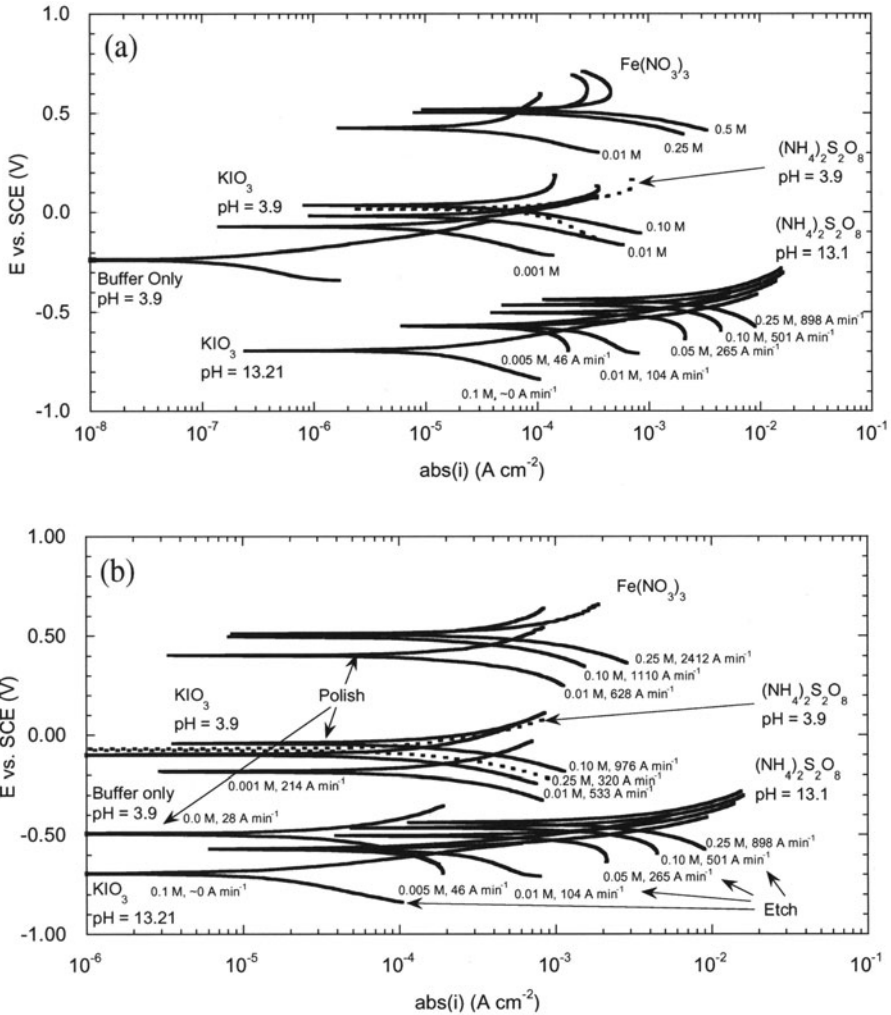
Fig. 4.9. (a) and (b). Representative Tafel plots. (a), top, is data taken under static conditions (no polishing). (b), bottom, shows representative Tafel plots obtained during polish. From [6]

persulfate-based etchant solutions. Etch rates for the four acidic slurries as well as the basic potassium iodate solution were below measurement limits indicating that a protective passive film had formed. The expected (since the corrosion rate increases) increase in the open circuit potential (OCP) and apparent corrosion current (i_{corr}) as a function of oxidizer concentration is visible for the persulfate-based solutions. The oxidizer-less (buffer only) slurry exhibits the lowest apparent i_{corr}, which is several orders of magnitude less than that of the other solutions or slurries. The OCP and apparent i_{corr}

of the pH 3.9 buffered persulfate- and iodate-based slurries are similar. The OCP of the ferric-based slurries is much higher than the OCP of any of the other solutions or slurries. These trends are expected since they reflect the relative strengths of these oxidizers.

Figure 4.9b shows representative Tafel plots obtained during polish. The tungsten removal rate due to polish is shown for all of the slurries. There are slight changes under polishing conditions but the changes are small.

Figure 4.10 is a plot of the i_{corr} vs. the tungsten polish rate. The polishing pressure was varied between 4.5 and 7.5 psi, and the speed between 0.35 and $1.18\,\mathrm{ms}^{-1}$. The values of i_{corr} measured during polish are 1 to 2 orders of magnitude lower than expected by complete oxidation. The calculated oxidation rate, based on the values of i_{corr}, does not account for the vast majority of the tungsten removed during polish, in direct contradiction with the passivation and abrasion mechanism of removal.

Figure 4.11 depicts the inverse of the calculated value of the charge transfer resistance (R_{ct}) determined from AC electrochemical impedance spectroscopy during polish vs. the measured removal rate. R_{ct} is inversely proportional to the corrosion current. The change in R_{ct} as a function of removal rate is strongest for the ferric slurry and weakest for the persulfate slurry, which is consistent with the trend in i_{corr} seen in Fig. 4.10. Figure 4.11 also shows calculated R_{ct} values obtained during etch in the high pH solutions. R_{ct} values during etch are much lower than those obtained during polish. The measured values of R_{ct} during polish are 1 to 2 orders of magnitude higher than expected based on the proposed mechanism. These data corroborate the data and conclusions drawn from the values of i_{corr} depicted in Fig. 4.10, i.e. the measured i_{corr} values are far too low to be consistent with a simple corrosion-abrasion model for tungsten CMP.
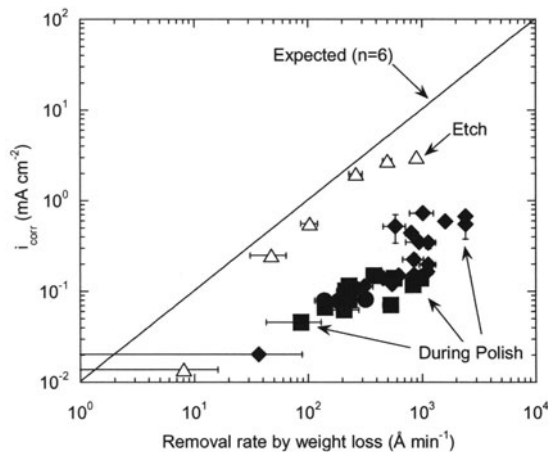
Fig. 4.10. Plot of the i_{corr} vs. the tungsten polish rate. From [6]
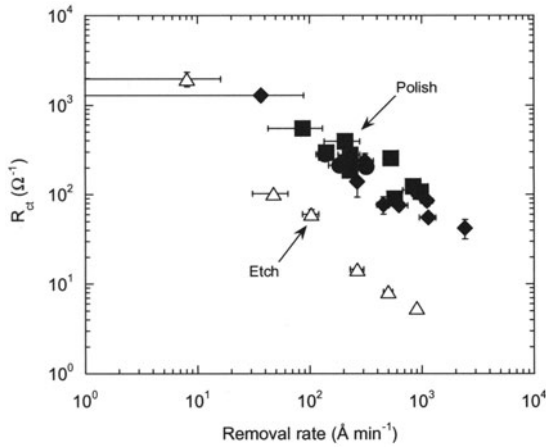
Fig. 4.11. Inverse of the calculated value of the charge transfer resistance (R_{ct}) determined from AC electrochemical impedance spectroscopy during polish vs. the measured removal rate. From [6]
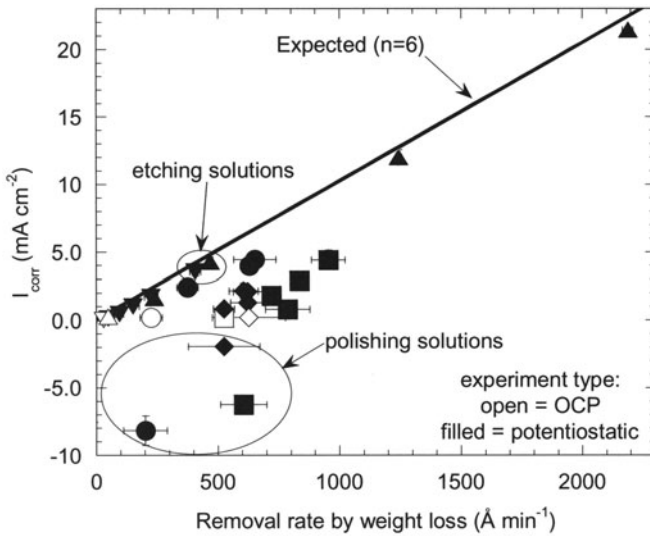
Fig. 4.12. The calculated values of i_{corr} (both averaged and calculated from Tafel plots) vs. the measured removal rate. From [6]

Stein et al. also measured the tungsten removal rate with and without polish at various constant overpotentials. This potentiostatic control can be used to force or suppress the oxidation of the metal. Figure 4.12 shows the calculated values of i_{corr} (both averaged and calculated from Tafel plots) vs. the measured removal rate. The values of i_{corr} for both the basic persul-

fate solution and buffer-only polish correlate well with the i_{corr} expected for a given removal rate. The values of i_{corr} measured during polish with the ferric-, iodate-, and persulfate-based slurries were much less than expected from the measured removal rate. The additional removal due to forced oxidation follows the expected trend of direct proportionality between i_{corr} and oxidation rate. Most important to note, however, is that the removal rate during polish did not significantly change when a cathodic potential (−0.5 V with respect to the OCP) was applied to the system. Cathodic potentials can suppress blanket oxidation of metal (similar to the sacrificial anode method of corrosion protection). These results further support the conclusion that blanket oxidation and passive layer formation do not play a significant role in the mechanism of tungsten removal during CMP.

Osseo-Asare et al. [7] investigated the role of tungstate ion on the electrochemical behavior of tungsten in potassium iodate solutions. Figures 4.13 and 4.14 show the combination of the cathodic polarization curve of IO_3^- on W and the anodic polarization curve of W at a pH of 2 and 4. The authors invoke mixed potential theory to suggest that the intersection of these 2 lines represent the rate of "W dissolution under CMP conditions." At pH = 4 the intercept current density for 10–3 M KIO_3 is lower than for 10^{-1} M KIO_3; hence the authors conclude that the concentration of KIO_3 will affect W removal rate at pH = 4. At pH = 2 the intercept current density for 10^{-3} M KIO_3 is the same as 10^{-1}M KIO_3 hence the authors conclude that the concentration of KIO_3 will not affect W removal rate at pH 2. This conclusion would only be valid if the passivation and abrasion mechanism was the dominant method of material removal during tungsten CMP. Note also that the maximum current density shown translates to a removal rate less
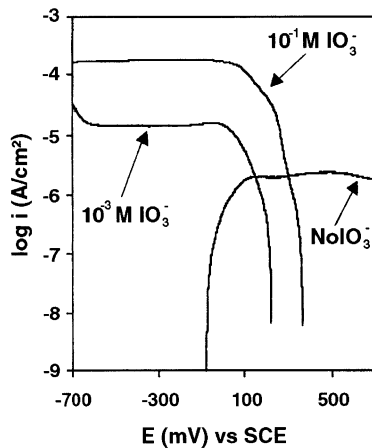
Fig. 4.13. The combination of the cathodic polarization curve of IO^{3-} on W and the anodic polarization curve of W at a pH of 2. From [7]
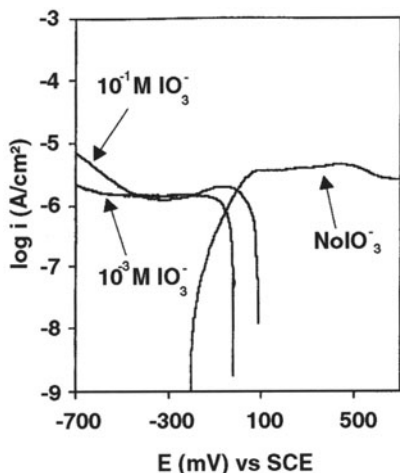
Fig. 4.14. The combination of the cathodic polarization curve of IO3- on W and the anodic polarization curve of W at a pH of 4. From [7]

than $1\,\text{Å}\,\text{min}^{-1}$. Gaffney et al. [8] measured the W CMP removal rates under similar conditions. Figure 4.15 shows the polish rates obtained. From these two sets of data it is apparent that the tungsten removal rate is dependent on KIO_3 concentration at both pH 2 and 4.

While it is clearly necessary to have an oxidant chemistry for tungsten CMP, the quantative matching of tungsten oxidation rates to tungsten CMP rates is very poor. Most electrochemical data implies a CMP rate that is over one magnitude lower than is observed. Thus, the actual removal mechanism must be more complex than simple oxidation and abrasion.
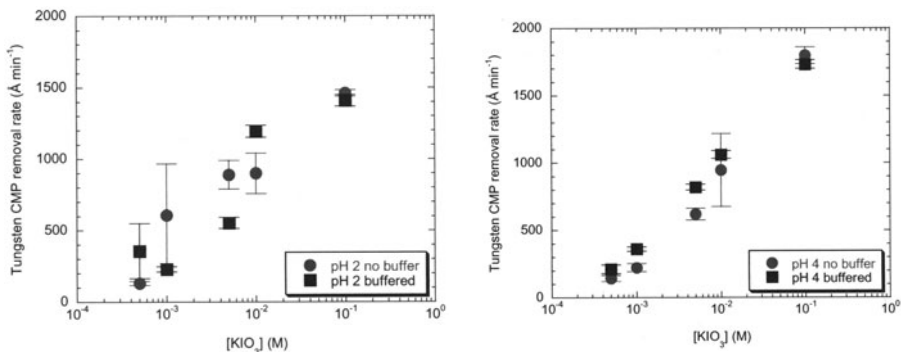
Fig. 4.15. The polish rates obtained from conditions similar to those in Figures 13 and 14. From [8]

4.3 Tungsten Experimental Data –
Role of Slurry Particle

Tungsten CMP slurries in practice have required abrasive particles. In commercial use, both silica and alumina have been both widely employed. Silica abrasives produce lower rates but also create lower surface damage. However, since high CMP removal rates are important in practical CMP, alumina-based tungsten slurries are also widely used and have been very extensively studied.

Bielmann et al. [9] investigated pH $4\,K_3Fe(CN)_6$ slurries containing 2 to 15 wt% alumina (Al_2O_3) particles of various sizes. Figure 4.16 shows the particle size distribution of the aluminas used. The smallest mean particle size was $0.25\,\mu m$ and the largest was $10.0\,\mu m$. Figure 4.17 shows the removal rate as a function of solids loading and particle size. Removal rate increases with solids loading but decreases with particle size. Figure 4.18 shows the roughness of the polished tungsten (measured using $5\,mm \times 5\,mm$ AFM images). The roughness is independent of the particle size used. No scratches were noticed on any sample except the one polished with the 10 mm alumina. The authors conclude that, "the removal rate mechanism is not a scratching type process, but is related to the contact surface area between particles and polished surface controlling the reaction rate."

Stein et al. [10] measured tungsten polish rates using various colloid particle species as abrasives. Table 4.1 shows the size (similar for all), phase, and species of each colloid studied. Figure 4.19 shows the tungsten polish rate using chemically identical slurries and process parameters. The polish rate is very dependent on the colloid phase and species. The alumina-based col-
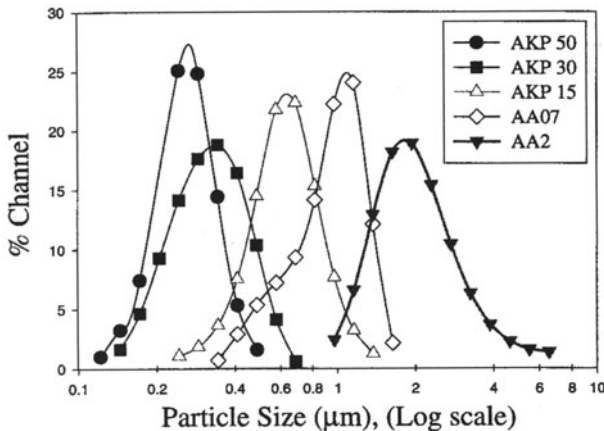
Fig. 4.16. The particle size distribution of various alumina particles used for tungsten CMP slurries. From [9]
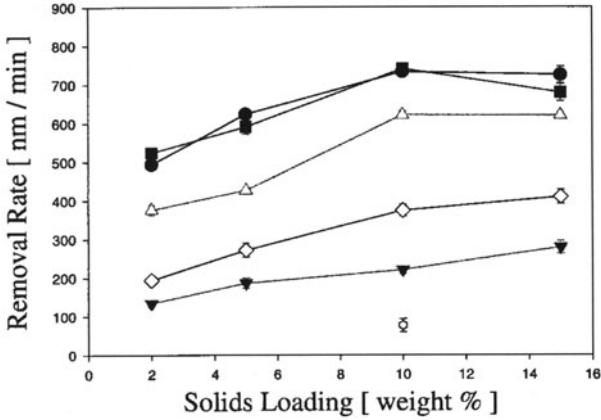
Fig. 4.17. The removal rate as a function of solids loading and particle size. From [9]
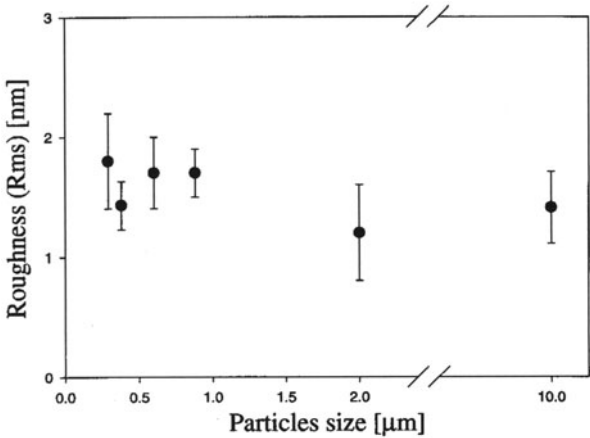
Fig. 4.18. The roughness of the polished tungsten (measured using $5\,\text{mm} \times 5\,\text{mm}$ AFM images). From [9]

loidal slurries, in general, showed the highest polish rates. The yttrium-based colloid showed the lowest polish rate.

Babu and coworkers have studied the role of the slurry particulate in both W and Cu CMP. The majority of their work has been on Cu CMP, which is discussed below. For comparison they also investigated W CMP. Ramarajan et al. [11] obtained alumina particles of varying bulk density. They made slurries containing the different aluminas and chemistries. Figure 4.20 shows the W polish rate in water and $0.1\,\text{M}\,\text{Fe(NO}_3)_3$ containing slurries made with these aluminas. The polish rate of tungsten is essentially zero without the added chemistry. Except for the lowest bulk density of $3.2\,\text{g/cc}$, the polish

Table 4.1. The size (similar for all), phase, and species of each colloid studied. From [10]

Colloid	Metal	Manufacturer	Brand name	Major phases	Size (Å)
1	yttrium	Nyacol	–	*	100
2	zirconium	Nyacol	–	Baddeleyite	500
3	cerium	Nyacol	–	Cerianite	200
4	aluminum	Nyacol	–	Bohmite (AlO(OH))	500
5	cerium	Nanophase	Nanotek Ceria	Cerianite	300
6	aluminum	Nanophase	Nanotek Alumina	$\gamma-Al_2O_3$	300
7	aluminum	Moyco	Planar W	Gibbsite $(Al(OH)_3)$, $\gamma-Al_2O_3$	–
8	aluminum	Solution Technology	MET202	$\delta-Al_2O_3$ Gibbsite $\theta-Al_2O_3$	500
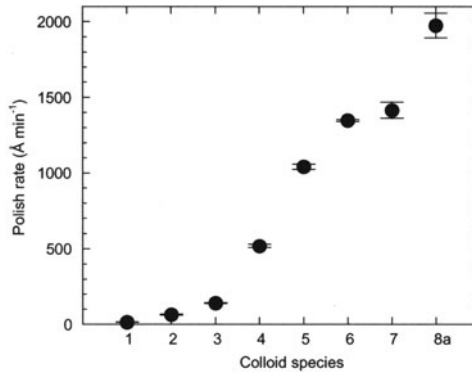
Fig. 4.19. The tungsten polish rate using chemically identical slurries and process parameters. From [10]

rate was independent of bulk density. They assumed that the particle bulk density was proportional to the "hardness/elastic modulus" of the particles. From the data shown and the hardness/elasticity assumption they conclude that a chemical aspect of the removal process is the rate-limiting step.

Stein et al. [12] measured the roughness of post-CMP tungsten as a function of CMP process parameters (pressure, velocity, and platen temperature). Figure 4.21 shows AFM scans of the post-CMP tungsten surfaces. The post-CMP roughness was found to be independent of process parameters. TEM
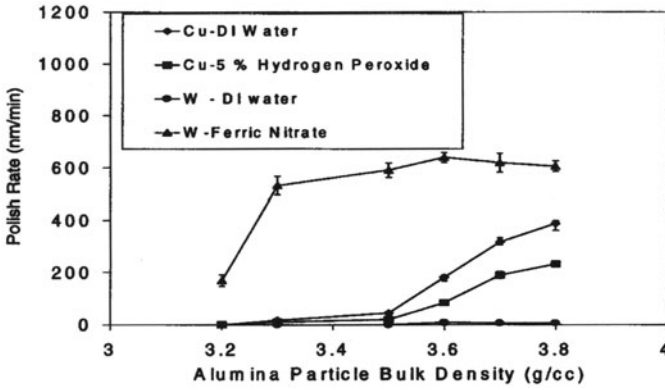
Fig. 4.20. The W polish rate in water and $0.1\,\mathrm{M}\,\mathrm{Fe(NO_3)_3}$ containing slurries made with aluminas of varying bulk density. From [11]
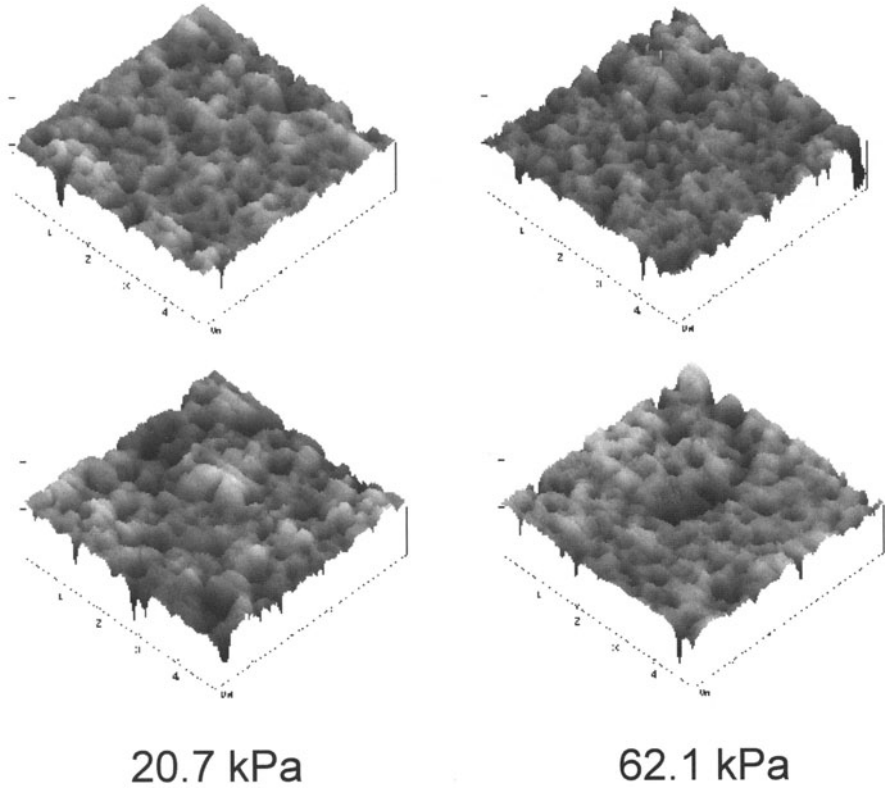
Fig. 4.21. AFM scans of the post-CMP tungsten surfaces. From [12]

investigations also indicated that the intragranular morphology of post-CMP tungsten was independent of process parameters and could not be distinguished from samples etched in H_2O_2.

Stein et al. measured the friction between an individual alumina particle and a CVD tungsten surface using lateral force microscopy (LFM) [12]. LFM is a type of atomic force microscopy (AFM) that measures the lateral deflection of a tip that is in contact with the sample as the tip is scanned across the sample. In this study, an alumina particle was glued to the tip and the measurements were taken in solution. Figure 4.22 shows the friction force as a function of KIO_3 concentration and applied pressure. The friction force is independent of KIO_3 concentration but is dependent on applied pressure and pH. Elsewhere, Stein et al. [13] also investigated the effect of alumina and KIO_3 concentrations on the tungsten removal rate. Figures 4.23 and 4.24 show the tungsten removal rate and recorded process temperature as a function of alumina loading in the slurry (at a constant KIO_3 concentration of 0.1 M). Both the polish rate and process temperature increase to a steady state value with increasing alumina loading (as well as polish pressure and velocity). Figures 4.25 and 4.26 show the tungsten removal rate and recorded process temperature as a function of KIO_3 concentration in the slurry (with constant alumina loading of 5 weight percent). The polish rate increases with KIO_3 concentration, pressure, and velocity. However, the process temperature remained constant regardless of KIO_3 concentration. Using a process energy balance the authors determined that the process temperature is controlled by friction and that any heat generated by chemical reaction is so small that it cannot be detected. This conclusion is in agreement with the conclusion drawn from the *in-situ* lateral force microscopy measurements.
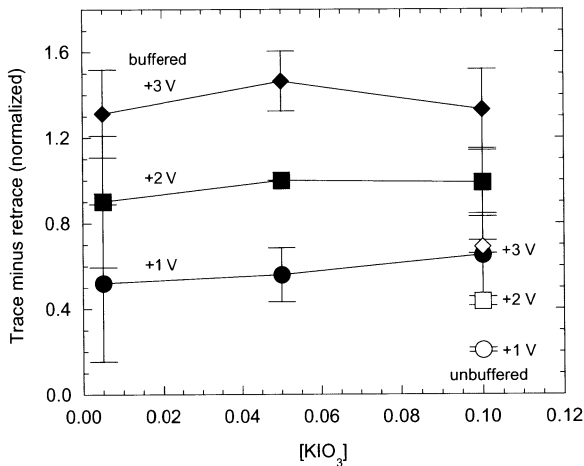
Fig. 4.22. The friction force (reported as AFM trace minus retrace deflection) as a function of KIO_3 concentration and applied pressure. From [12]
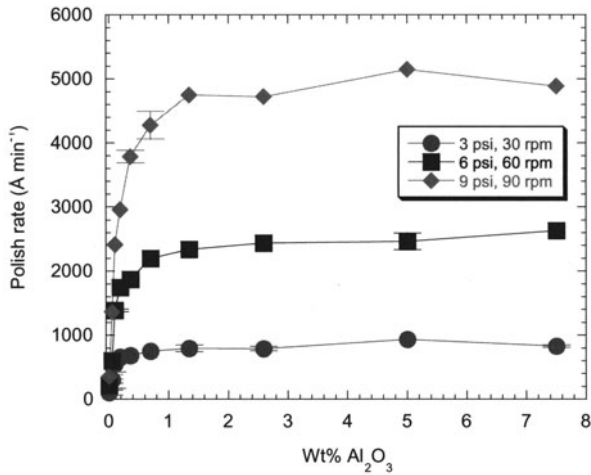
Fig. 4.23. The tungsten removal rate as a function of alumina loading in the slurry (at a constant KIO_3 concentration of 0.1 M). From [13]
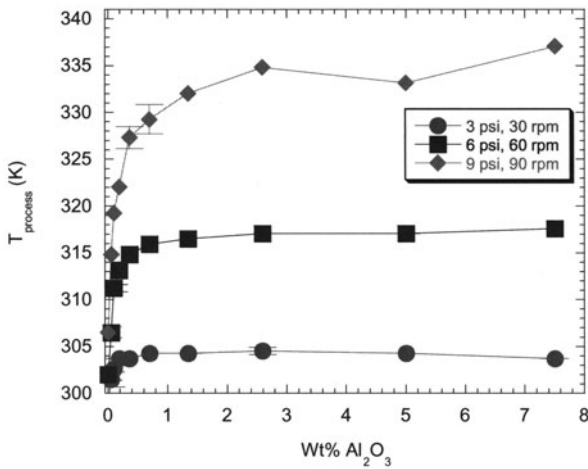
Fig. 4.24. The recorded process temperature as a function of alumina loading in the slurry (at a constant KIO_3 concentration of 0.1 M). From [13]

Another observation from Fig. 4.17 as well as Fig. 4.23 is that the polish rate saturates and depends only weakly upon solids content above a level of 1–2 wt% of alumina. This contrasts significantly with the case of silica concentration in glass polishing (see Chap. 3), where the polish rate is linear, or nearly so, with concentration to well over 10 wt% concentration.
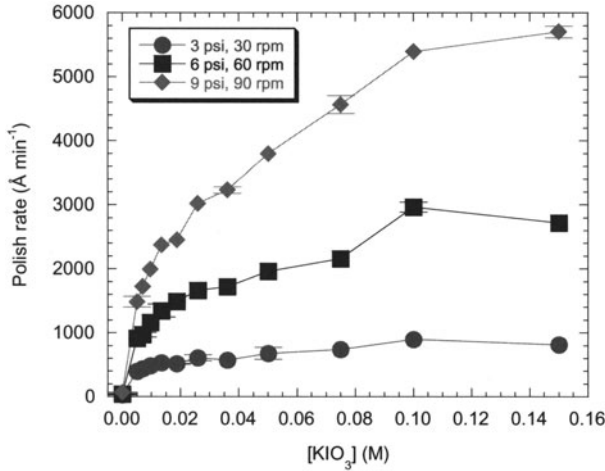
Fig. 4.25. The tungsten removal rate as a function of KIO_3 concentration in the slurry (with constant alumina loading of 5 weight percent). From [13]
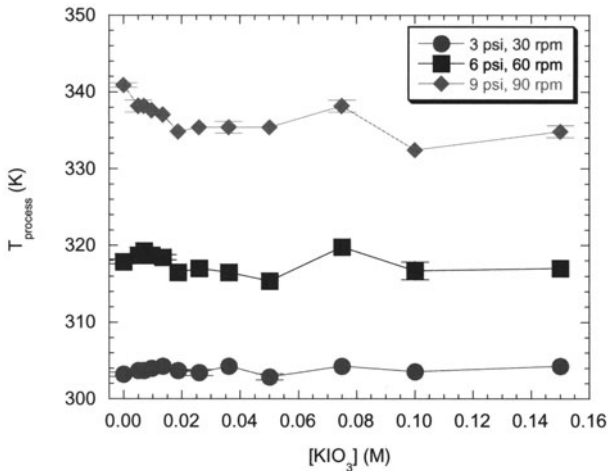
Fig. 4.26. The recorded process temperature as a function of KIO_3 concentration in the slurry (with constant alumina loading of 5 weight percent). From [13]

4.4 Conclusions on Mechanisms on W CMP

The passivation and abrasion model has been the usual starting point for research into the removal mechanisms at play during tungsten CMP. Electro-chemical experiments have been carried out to determine what role is played by metal oxidation during CMP. Several experiments have shown that either an oxide passivation layer does not form or that the rate of metal oxidation is at least an order of magnitude slower than the CMP removal rate. Many

of these experiments have been carried out under CMP conditions. Thus the passivation and abrasion mechanism is most likely not the removal mechanism for W during CMP. This statement naturally elicits the question "if the oxidation rate of the metal cannot account for the CMP removal rate, what is the purpose of the oxidizer?" Tungsten removal will not occur without it. Several models have been put forth but no consensus has been reached. These models are discussed at the end of the chapter.

The role of the particle in W removal has also been studied. The size of the particle, to an upper limit, only affects the removal rate. The upper limit may be the transition between polishing and grinding mechanisms. The particle size does not affect the surface finish (roughness) of the polished tungsten. Friction between the particle and the surface is dependent on the pH of the solution and the applied pressure but not on the concentration of oxidizer. Recent thinking suggests that the role of the particle is more than just mechanical abrasion. The surface chemistry of the particle may play a role in the removal mechanism. This most likely occurs in synergy with the chemistry in the slurry. Abrasive-free "reactive liquid" CMP systems exist for Cu CMP, as will be discussed below. W removal could possibly occur without an abrasive although there is, as yet, no data to support this. Hard pad asperities or surface chemistry may replace the silica or alumina abrasive in the reactive liquid system for Cu but for W a pad that fills the role of the abrasive has not yet been found.

4.5 Copper Experimental Data –
Chemical and Electrochemical

As with W CMP studies, the study of Cu CMP begins with a review of the appropriate Pourbaix diagram. A simplified Pourbaix diagram for Cu and Cu ions in aqueous solution is shown in the Appendix. As with W CMP investigations, the limitations of the Pourbaix diagram to Cu CMP studies becomes readily apparent (see Appendix).

In an early work, Steigerwald et al. [14] investigated Cu CMP using Al_2O_3-based slurries with various chemical reagents. Figure 4.27 shows the copper polish rate as a function of reagent. The samples were polished on a metallographer's wheel. Polish rates from 0.25 to greater than $1.5\,mm\,min^{-1}$ were obtained. These polish rates are exceptionally high compared to currently observed CMP removal rates on standard CMP tools. The difference is most likely due to the differences between the metallographer's wheel and a commercial CMP tool. Figure 4.28 shows the same data plotted as a function of slurry pH. Except for NH_4Cl the polish rate decreased as pH increased. The anomalous behavior of the NH_4Cl slurry was believed to be due to the electrochemical potential of the Cu in that solution. The authors measured the potential as $-1\,mV$ vs SHE. At a pH of 4.8 and this potential, the Pourbaix diagram indicates copper is immune to corrosion. The authors
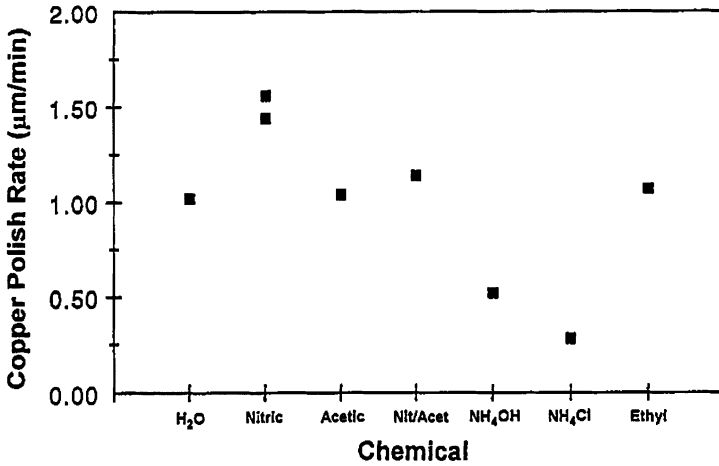
Fig. 4.27. The copper polish rate as a function of reagent. From [14]
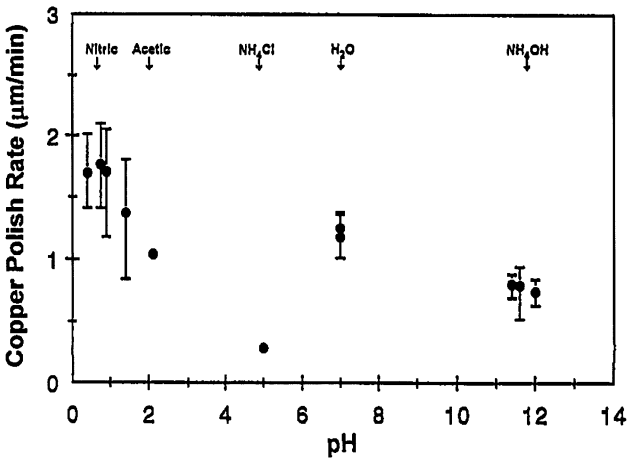
Fig. 4.28. This figure shows the same data as Fig. 4.27 but plotted as a function of slurry pH. From [14]

state, "Because of the immunity to corrosion, mechanically abraded copper does not dissolve in solution. Consequently, much of the abraded material redeposits onto the surface, and the efficiency of the mechanical abrasion decreases, yielding the reduced polish rate". They focused their studies on a HNO_3 based slurry. Figure 4.29 shows the polish rate for these solutions. From this data the authors conclude that "the dominant removal mechanism is mechanical abrasion of the surface followed by chemical dissolution of the
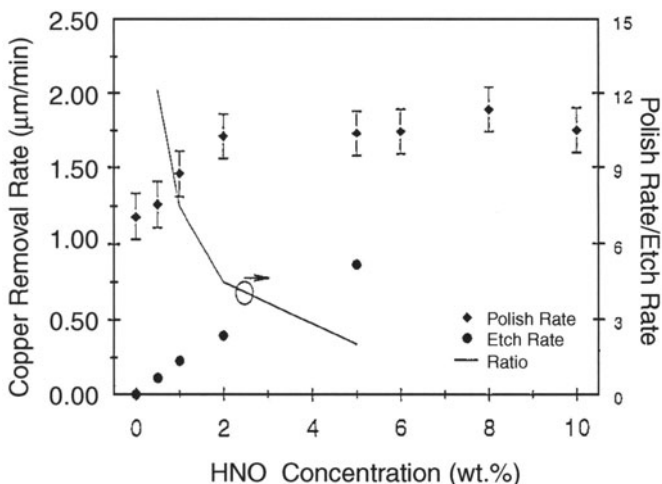
Fig. 4.29. Polish rate data for HNO_3-based slurries. From [14]
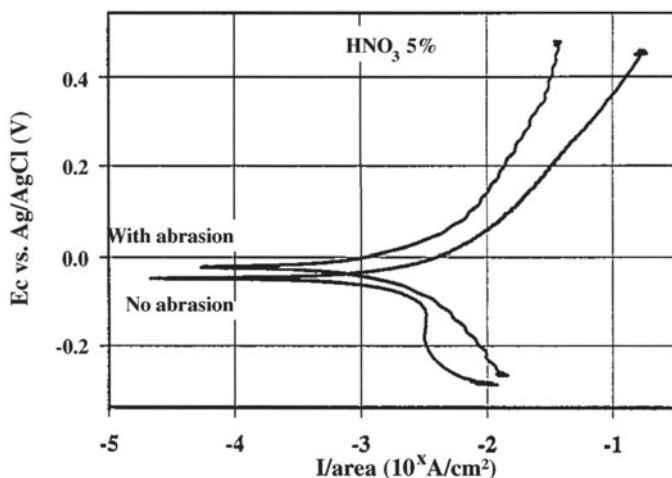
Fig. 4.30. The potentiodynamic curves obtained with and without abrasion in HNO_3 slurry. From [16]

abraded surface". This work was extended to include the effects on oxide and polymer inter-level dielectrics [15].

Carpio et al. [16] investigated variations of NH_4OH and HNO_3 slurry chemistries containing silica and alumina abrasives using D.C. polarization and A.C. impedance spectroscopy electrochemical techniques. Figures 4.30 and 4.31 show the potentiodynamic curves obtained with and without abrasion in HNO_3 and NH_4OH slurries and Table 4.2 shows the values of E_{corr} and i_{corr} calculated from the data shown in figures. HNO_3 is a strong copper
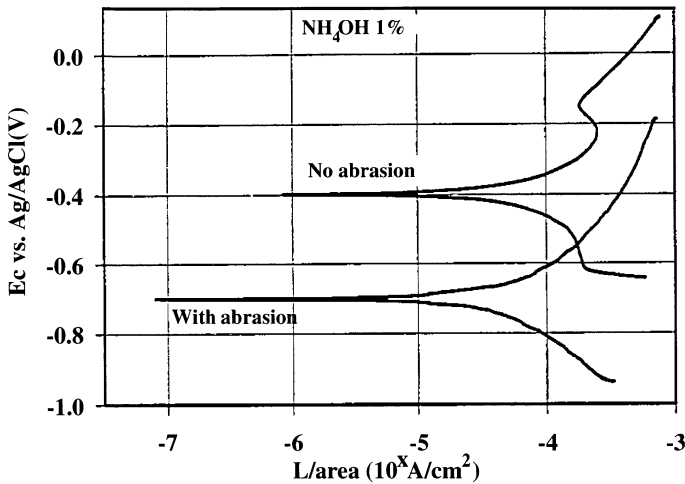
Fig. 4.31. The potentiodynamic curves obtained with and without abrasion in NH$_4$OH slurry. From [16]

Table 4.2. The values of E_{corr} and i_{corr} calculated from Figs. 4.30 and 4.31. From [16]

Solution	Abrasion E_{corr} (V)	Abrasion I_{corr} (mA cm^{-2})	No abrasion E_{corr} (V)	No abrasion I_{corr} (mA cm^{-2})
HNO$_3$ 1 wt.%	−0.0563	1.254	0.0556	1.408
5 wt.%	−0.022	11.15	−0.0467	6.101
NH$_4$OH, 1 wt.%	−0.6997	0.1026	−0.3985	*
5 wt.%	−0.668	0.3304	−0.4925	0.0864
KMnO$_4$, 3 wt.%	−0.485	1.238	−0.274	0.1182

etchant and this is evident in the data shown. The values of E_{corr} and i_{corr} do not significantly shift with abrasion and a corrosion current density of $1\,\mathrm{mA\,cm^{-2}}$ corresponds to Cu removal rate of $3.7\,\text{Å}\ \mathrm{s^{-1}}$. In contrast to the behavior of HNO$_3$, NH$_4$OH forms a passivating layer of the surface of the Cu without abrasion. With abrasion the value of E_{corr} drops approximately $300\,\mathrm{mV}$ and the value of i_{corr} increases from 0.0864 to $0.3304\,\mathrm{mA\,cm^{-2}}$. The authors note that the copper oxide passivation layer does not provide complete protection against corrosion and that the slurry pH needs to be as low as possible to enhance the selectivity between Cu and oxide polish rates. Their data also indicate that silica based slurries (stable at higher pH) do not polish Cu as well as alumina based slurries. Silica based slurries are typically stable at higher pH values (in general, this increases oxide removal rates and lowers selectivity). The authors state that a passivation type mechanism will
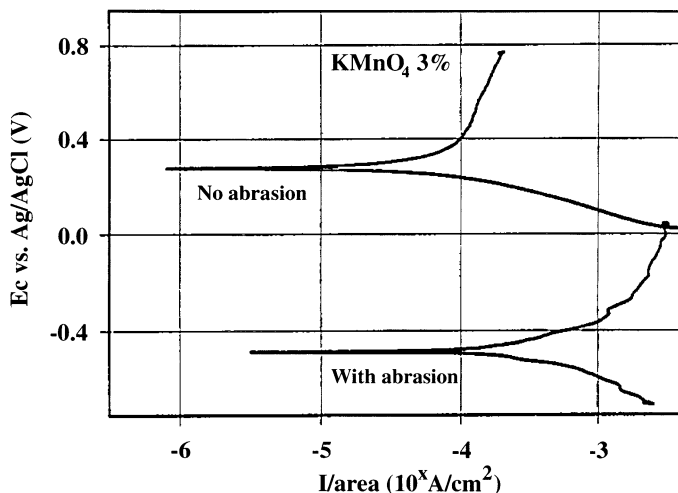
Fig. 4.32. The potentiodynamic curves obtained with and without abrasion in an unbuffered 3% solution of $KMnO_4$. From [16]

only work for CMP in slurries above pH = 6, which is too high for alumina suspensions to be stable. Thus alternative chemistries were investigated.

$KMnO_4$ was studied as a possible oxidizer for use in an alumina slurry. Figure 4.32 shows the potentiodynamic curves obtained with and without abrasion in an unbuffered 3% solution of $KMnO_4$. The removal of a passivation layer is clearly seen by the decrease in E_{corr} and increase in i_{corr} with abrasion. However, very low polish rates (less than $500\,\text{Å}\,\text{min}^{-1}$) were seen under both acidic and basic conditions hence this oxidizer was declared to have "limited application in copper CMP slurries". Figure 4.33 shows the data obtained from acidic and basic solutions of H_2O_2. The data is noisy perhaps because the authors used a platinum mesh counter-electrode (H_2O_2 spontaneously decomposes on platinum). H_2O_2 attacks copper without abrasion yet an unexplained drop in E_{corr} is seen with abrasion. Figure 4.34 shows potentiodynamic sweeps obtained with addition of BTA to a HNO_3 based slurry. The addition of benzotriazole (BTA, a corrosion inhibitor) was investigated because in other similar work it is an additive to slurries to prevent Cu corrosion in recessed areas. With addition of BTA, the value of i_{corr} without abrasion appears to be approximately one order of magnitude less than with abrasion while without the addition of BTA the value of i_{corr} with and without abrasion are similar.

Luo et al. investigated the CMP of Cu in both acidic and alkaline solutions [17, 18, 19]. The investigation in acidic media was carried out in solutions of $Fe(NO_3)_3$ that contained varying concentrations of BTA and ethylene glycol (used to stabilize the alumina abrasive). The addition of BTA reduced the zeta potential of the particles and therefore the settling rate of
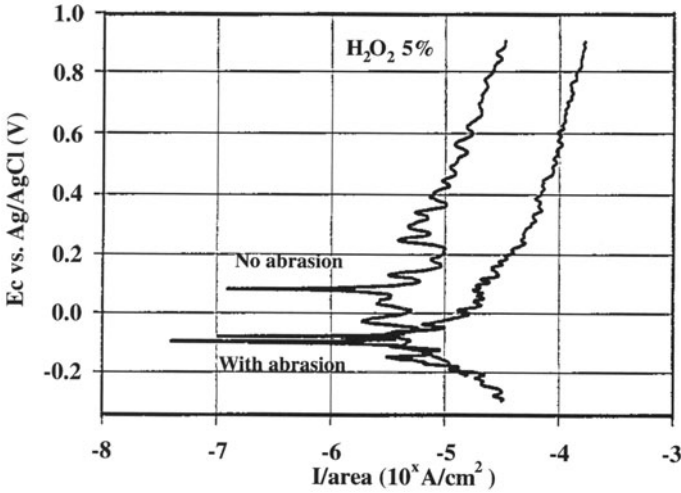
Fig. 4.33. Electrochemical data obtained from acidic and basic solutions of H_2O_2. From [16]
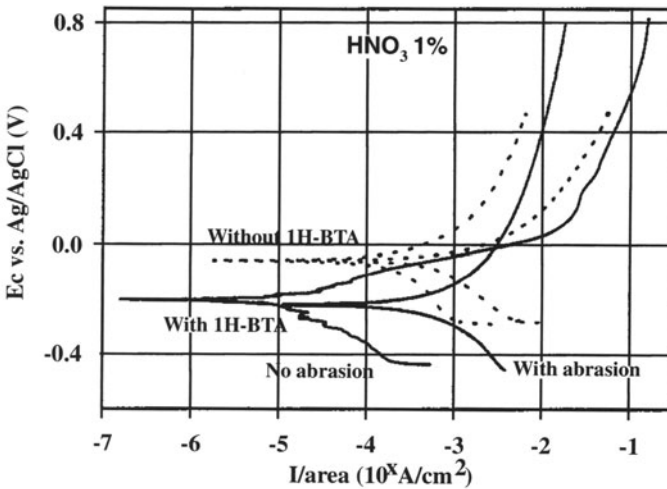
Fig. 4.34. Potentiodynamic sweeps obtained with addition of BTA to a HNO_3 based slurry. From [16]

the slurry particles in suspension increased. The zeta potential was believed to have been reduced by the adsorption of BTA onto the surface of the alumina, which resulted in some charge neutralization. The addition of very high molecular weight poly ethylene glycol (PEG) decreased the settling rate of the alumina suspension. The settling rate was hypothesized to have decreased

due to steric repulsion of the alumina particles by the adsorbed PEG groups on the particle surfaces.

The Cu removal rate and corrosion current increase with Fe^{3+} concentration and decrease with BTA concentration. Figure 4.35 shows the increase in both removal rate and i_{corr} (no abrasion). Figure 4.36 shows the decrease in removal rate and i_{corr} with BTA concentration. The authors note that the addition of PEG to the slurry was slightly detrimental to the corrosion resistance provided by BTA. They also obtained polish rates and i_{corr} values as a function of PEG concentration. The value of i_{corr} jumps with the initial addition of PEG but remains constant with increasing PEG addition. The Cu removal rate increases steadily with PEG addition. The authors conclude that Cu CMP can be performed in acidic media using a corrosion inhibitor but a balance between the BTA and PEG concentrations must be maintained.

CMP of copper in acidic media offers the benefit of high polish rate selectivity between copper and oxide films; however, the strongly acidic and oxidizing nature of the slurry can be detrimental to the polish tools and waste disposal systems. These issues were a motivation to study copper CMP in alkaline alumina abrasive ammonia containing slurries. Figure 4.37 shows the polish rate of the copper films as a function of ammonia concentration in the slurry. There is an initial increase from $130 \, \text{nm min}^{-1}$ at $0 \, \text{wt} \%$ ammonia to $210 \, \text{nm min}^{-1}$ at $0.3 \, \text{weight} \%$ ammonia. The polish rate remains roughly stable above ammonia concentrations of $0.3 \, \text{weight} \%$ though there is a slight increase from $3 \, \text{weight} \%$ to $6 \, \text{weight} \%$. Note that copper has a significant polish rate in a slurry containing only water and alumina (no oxidizers or
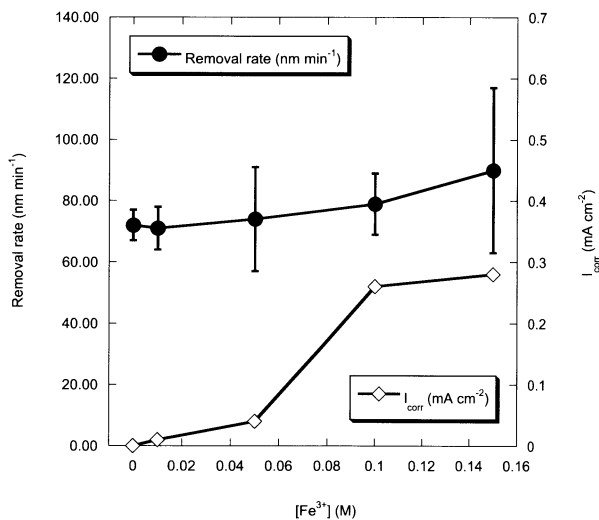
Fig. 4.35. The increase in both removal rate and i_{corr} (no abrasion) with Fe^{3+} concentration. From [18]
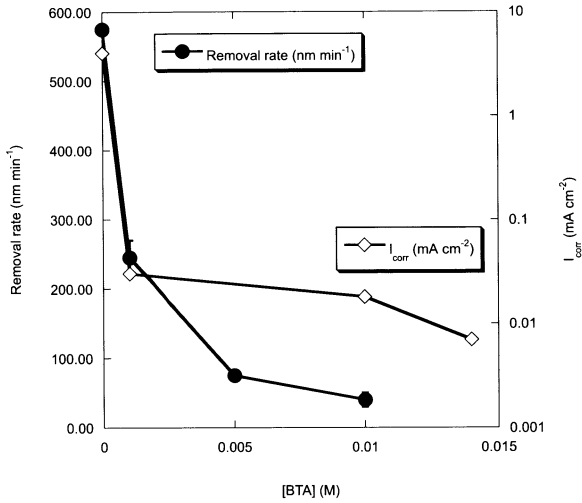
Fig. 4.36. The decrease in both removal rate and i_{corr} (no abrasion) with BTA concentration. From [18]
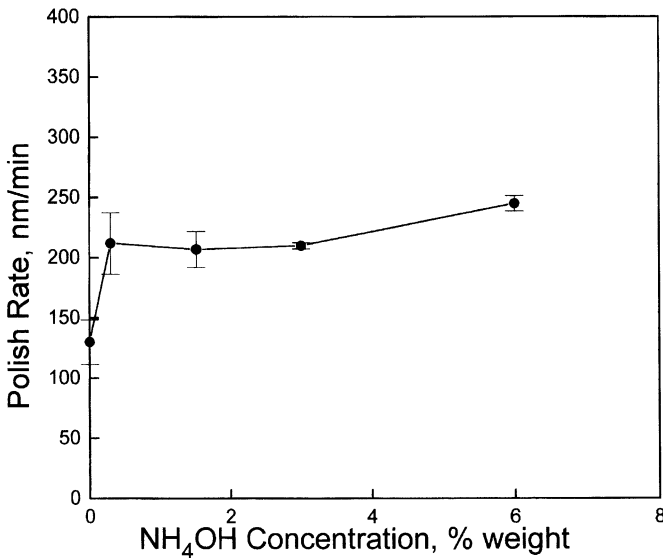
Fig. 4.37. The polish rate of the copper films as a function of ammonia concentration in the slurry. From [19]
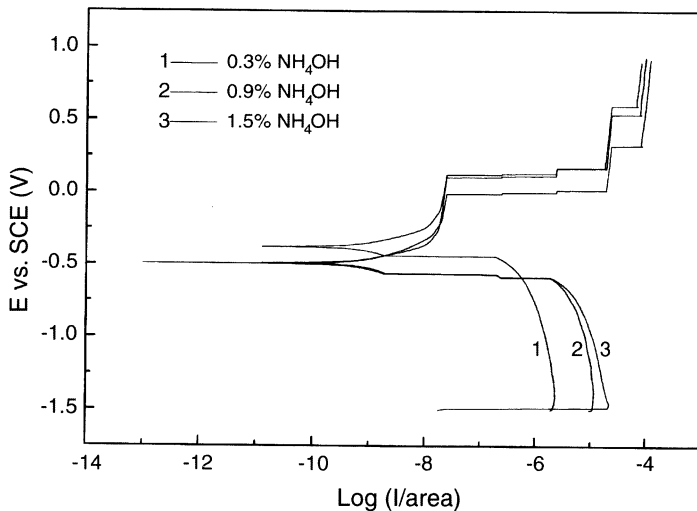
Fig. 4.38. The polarization curves obtained on the copper during polish. From [19]

pH adjusters). Figure 4.38 shows the polarization curves obtained on the copper during polish. The corrosion current densities are calculated to be around 1 nA cm^{-2} which corresponds to a corrosion rate at least an order of magnitude less than the CMP removal rate. The authors attribute most of
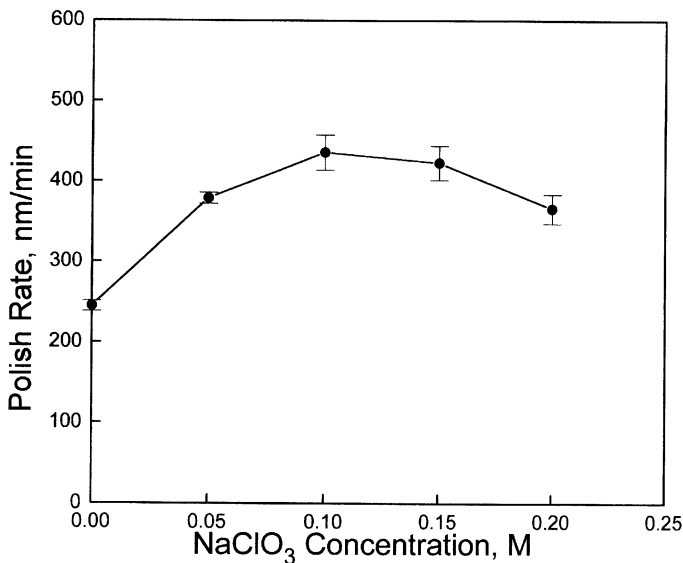
Fig. 4.39. The effect of added NaClO$_3$. From [19]

the CMP removal rate to a purely mechanical mechanism since the copper corrosion rate is insignificant compared to the CMP removal rate.

The effect of adding an oxidizer, $NaClO_3$, was also studied. Figure 4.39 shows the effect of added $NaClO_3$. The polish rate increases significantly (from 250 nm min^{-1} to 450 nm min^{-1}) until a concentration of 0.1 M $NaClO_3$ is reached at which point the polish rate decreases. Polarization data indicate that the corrosion current densities during polish with the $NaClO_3$-containing slurries do not go above 2 nA cm^{-2}. Based on these data the authors believe the removal mechanism is still predominantly mechanical in nature though they do not explain how the increase in $NaClO_3$ concentration has such a marked effect on the CMP removal rate. The copper CMP removal rate in other ammonia-containing salts NH_4NO_3 and $(NH_4)_2SO_4$ is also studied. Figure 4.40 shows the polish rates in slurries containing these salts. The polish rates at low pH levels and high pH levels are similar but around a pH of 10 the polish rates vary significantly. The authors believe that the oxidizing power of NO_3^- and dissolution inhibition from SO_4^{2-} may be responsible for the differences.

Table 4.3 demonstrates the effect of BTA on the static corrosion rate of the copper film and the CMP removal rate. The polish rate is only slightly negatively impacted by the addition of BTA while the corrosion rate drops by over a factor of 4. Table 4.4 shows the selectivity between the copper CMP removal rate and silicon dioxide CMP removal rate. The highest selectivity is 18:1, too low for commercial use. This low selectivity is generally observed
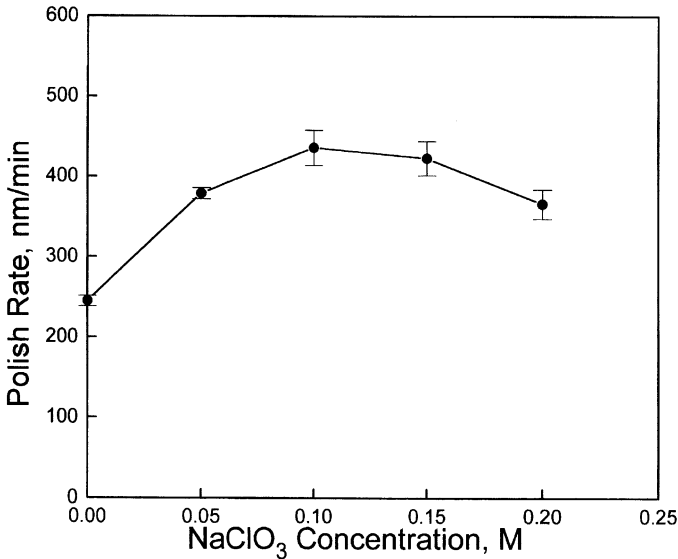
Fig. 4.40. The polish rates in slurries containing NH_4NO_3 and $(NH_4)_2SO_4$ salts. From [19]

Table 4.3. The effect of BTA on the static corrosion rate of the copper film and the CMP removal rate. From [19]

Copper polish rates and dissolution rates in various media						
NH$_4$OH concentration (%)	0.3	1.5	3	3	3	3
BTA concentration (M)	0	0	0	0.001	0.005	0.01
Polish rate (nm/min)	212	207	214	229	210	200
Maximum dissolution rate (nm/min)	13	30	29	8	3	3
Ration of copper polish rate to the maxiumum dissolution rate	16	7	7	29	70	67

Table 4.4. The selectivity between the copper CMP removal rate and silicon dioxide CMP removal rate. From [19]

Slurry formulation	Selectivity
0.3% NH$_4$OH, pH 10.5	9:1
1.5% NH$_4$OH, pH 11.1	11:1
3% NH$_4$OH, pH 11.3	11:1
0.5 M (NH$_4$)$_2$SO$_4$, pH 6.5	16:1
0.5 M (NH$_4$)$_2$SO$_4$, pH 9.2	6:1
1 M NH$_4$NO$_3$, pH 6.0	18:1
1 M NH$_4$NO$_3$, pH 8.6	8:1
3% NH$_4$OH, 0.001 M BTA	11:1
3% NH$_4$OH, 0.005 M BTA	9:1
3% NH$_4$OH, 0.01 M BTA	10:1

in the basic regime, and thus alkaline Cu CMP slurries are not considered commercially viable products.

Keleher et al. studied the corrosion of copper by hydroxyl radicals [20] and Hariharaputhiran et al. studied the formation and role of hydroxyl radical (*OH) formation in H$_2$O$_2^-$ and amino-acid based Cu slurries [21] (these references are related and will be discussed together). The authors noted that the *OH radical (uncharged species as opposed to OH$^-$ ion) is a much stronger oxidizer than H$_2$O$_2$ and may cause the majority of Cu corrosion that occurs during CMP. *OH radical concentration was measured using UV absorption at 440 nm in solutions containing p-Nitrosodimethylaniline (PNDA) which is an *OH trapping agent. Figure 4.41 shows the decrease in absorbance over time as the PNDA reacts with the *OH in solution. The authors measured the kinetics of the conversion of PNDA to PNDA-OH and were able to determine the initial *OH concentration in various slurries which are shown in
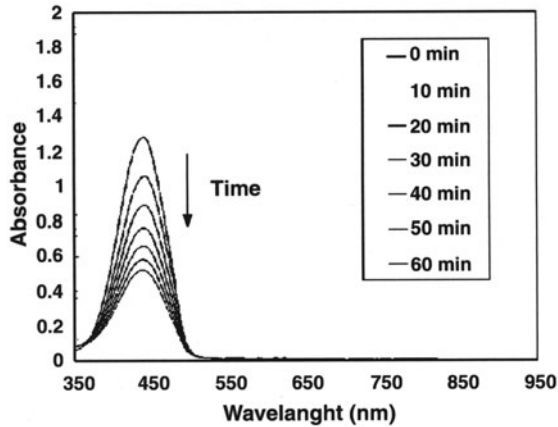
Fig. 4.41. The decrease in absorbance over time as the PNDA reacts with the *OH in solution. From [20]

Table 4.5. Although the addition of H_2O_2 and H_2O_2 with glycine increases the *OH concentration from $0.01 * 10^{14}$ M to $0.5 * 10^{14}$ M, the addition of 90 μM Cu^{2+} ions to a solution of H_2O_2 and glycine significantly increases the *OH concentration to $2.1 * 10^{14}$ M.

BTA is added to Cu CMP slurries to inhibit the corrosion of recessed regions while the protruding areas polish. Table 4.6 shows the Cu dissolution
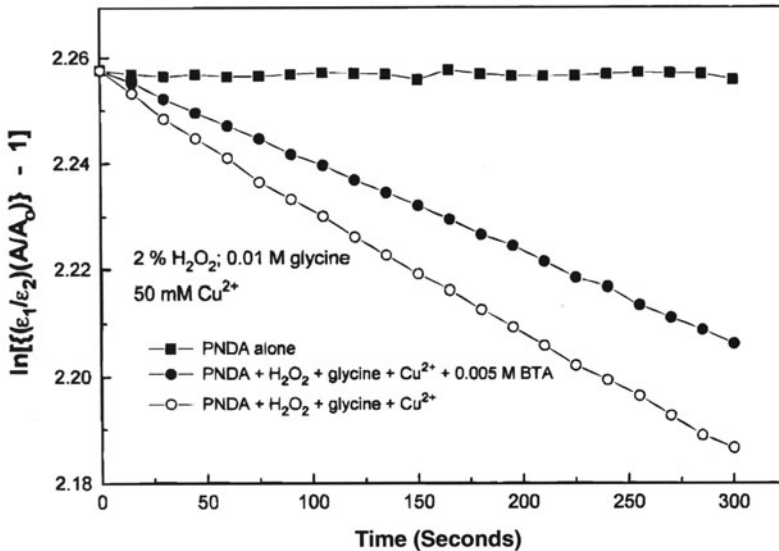
Fig. 4.42. The effect of BTA addition on *OH concentration. From [21]

Table 4.5. The initial *OH concentration in various slurries. From [20]

	Pseudo-first-order rate constant $(k \times 10^3)$ [1/min]		[*OH] $\times 10^{14}$ [M]
Solution composition	$T = 21°C$	$T = 50°C$	$T = 21°C$
1. PNDA only	0.1	0.1	0.01
2. PNDA + H_2O_2[a]	3.5	8.6	0.5
3. PNDA + H_2O_2[a] + glycine[b]	2.9	34.2	0.4
4. PNDA + H_2O_2[a] + Cu^{2+} (90 µM)	2.5	27.1	0.3
5. PNDA + H_2O_2[a] + glycine[b] + Cu^{2+} (90 µM)	15.8	68.8	2.1
6. PNDA + H_2O_2[a] + glycine[b] + Cu^{2+} (180 µM)	19.0	80.1	2.1
7. PNDA + H_2O_2[a] + glycine[b] + Cu^{2+} (270 µM)	28.0	88.9	3.7

[a] 2 wt % H_2O_2.   [b] 0.01 M glycine.

Amino acid present in solution[a]	Pseudo-first-order rate constant $(k \times 10^3)$ (1/min)	[*OH] $\times 10^{14}$ [M]
1. No amino acid	2.4	0.32
2. Arginine	2.5	0.33
3. Phenyl alanine	4.2	0.56
4. Glutamine	6.1	0.81
5. Glutamic acid	6.6	0.88
6. Glycine	13.8	1.84
7. Serine	15	2
8. Cysteine	24.7	3.4

[a] 2 wt % H_2O_2 + 50 µM copper acetate + PNDA + 0.013 M amino acid.

rate in a fixed concentration H_2O_2 and glycine solution containing varying concentrations of BTA. The Cu dissolution rate dramatically decreases with addition of BTA thus the authors conclude that the effect of BTA is not suppressed by the enhanced *OH concentration. The authors were concerned that BTA could quench the *OH concentration. Figure 4.42 shows the effect of BTA addition on *OH concentration. The addition of BTA at these concentration levels decreases the *OH concentration only 15% from the *OH concentration in a PNDA, H_2O_2, glycine, and Cu^{2+} based solution. The addition of BTA does decrease the polish rate by a factor of 2 and remains

Table 4.6. The Cu dissolution rate in a fixed concentration H_2O_2 and glycine solution containing varying concentrations of BTA. From [21]

Solution Composition	Dissolution rate at natural[a] pH (nm/min)	Dissolution rate at pH 8.4 (nm/min)
1. 5 wt % H_2O_2 in DI water	0.02	0.05
2. 1 wt % glycine in DI water	0.29	0.31
3. 5 wt % H_2O_2 + 1 wt % glycine (0.13 M) in Di water	162 ± 4	231 ± 7
4. 5 wt % H_2O_2 + 1 wt % glycine + 0.125 wt % $Cu(NO_3)_2$ in DI water	246 ± 8	290 ± 7
5. 5 wt % H_2O_2 + 1 wt % glycine + 1 wt % $Cu(NO_3)_2$ in DI water	394 ± 35	(no data)[b]
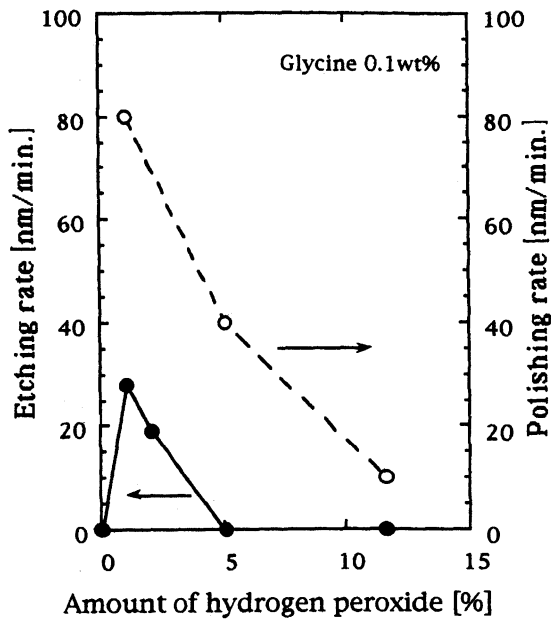
Fig. 4.43. The static Cu corrosion rate and Cu CMP removal rates in slurries of varying H_2O_2 concentration. From [22]

essentially the same with further addition of BTA. The authors do not speculate as to the role the *OH species could possibly play in the removal of Cu during CMP.
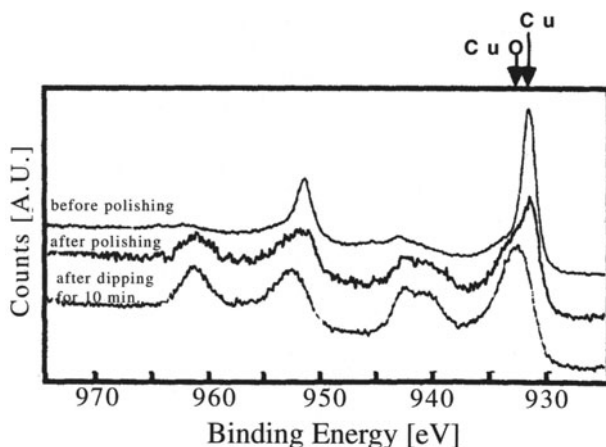
Fig. 4.44. The XPS spectra of samples before and after polish and also after static corrosion. From [22]

Hirabayashi et al. investigated Cu CMP slurries containing glycine and hydrogen peroxide [22]. Figure 4.43 shows the static Cu corrosion rate and Cu CMP removal rates in slurries of varying H_2O_2 concentration. Both corrosion rate and polish rate decreased with increasing H_2O_2 concentration at constant glycine concentration of $0.1\,wt\,\%$. With greater than $5\,wt\,\%$ H_2O_2 the corrosion rate is immeasurable but the polish rate still remains at 100 to $400\,\text{Å}\,min^{-1}$. Figure 4.44 shows the XPS spectra of samples before and after polish and also after static corrosion. The surface before CMP shows almost no copper oxide formation. The sample that underwent CMP shows a small amount of copper oxide but the authors believe this to be due to oxidation that occurred during cleaning, drying, and transfer to the XPS sample chamber. In a test to evaluate static corrosion, the surface grew a $300\,\text{Å}$ thick oxidized copper layer. The protective nature of the copper oxide was shown using dissolution experiments. The authors believe that their data is consistent with the abrasion and repassivation model of metal removal that was described at the beginning of this chapter.

4.6 Copper Summary

Copper CMP behavior contrasts in several ways with that of tungsten. Copper corrodes in almost all aqueous solutions and its oxides form slowly, and generally do not passivate the surface in the presence of strong oxidants needed for CMP. Thus other surface passivation approaches are necessary, such as the use of BTA or citric acid to suppress attack of the copper surface away from the plane of polishing.

In addition, copper CMP technology requirements are, in general, much more demanding that those of tungsten. In tungsten CMP, a large degree of post-CMP topography can be tolerated, as the subsequently deposited dielectric layer is planarized before tungsten is deposited. In copper dual-damascene processing, there is no replanarization before the copper is deposited at the next level (see Chap. 10). This leads to very stringent post-CMP topography requirements for copper CMP.

Typical commercial copper CMP is a two or more step process. The first step(s) is (are) designed to remove copper overburden and stop on the barrier material. The first step slurry is designed to be highly selective to the Cu barrier materials (TaN, TiN or WN). This slurry is generally acidic in nature to provide the high selectivity to the barrier and dielectric materials. There are several approaches for the second step slurry design. In one group, with selective second step slurries, the slurry removes the residual copper and barrier material, but stops on the dielectric. These slurries are usually acidic.

In the second approach, the residual Cu overburden, the barrier material, and small amount of dielectric are removed using a slurry with selectivities among the three materials (copper, barrier, dielectric) being within a factor of two or so of unity. This type of slurry is generally slightly basic. All copper slurries will contain an oxidizer such as hydrogen peroxide, a corrosion inhibitor such as BTA, and additional species such as glycine and poly ethylene glycol. Alumina or silica can be used as the solid constituent of the slurry. The decision of which material to use is made based upon polish rates and with an understanding of the colloidal stability of the particulate at the slurry pH.

4.7 CMP Removal Models

Preston [24] developed an empirical expression to predict material removal rate from polish pressure and relative velocity between the part surface and the surface of the polishing pad. The Preston equation is given as

$$RR = KPV, \tag{4.1}$$

where RR = removal rate, K is Preston's coefficient, a constant which is dependent upon the specific polishing system (see Chap. 2), P is the applied pressure between the part being polished and the polish pad, and V is the relative velocity between the surface being polished and a point on the polish pad. Note that, for most CMP systems, the motions are not linear so there is usually some variation in the local velocity as a function of time and position on the polish surface.

The Preston Equation (4.1) is discussed in Chap. 2 for the case of dielectric polishing. For metal CMP as practiced, there are often several differences in how well actual polishing behavior follows the Preston model.

As noted, tungsten was the first metal extensively polished using CMP. In addition, the first modeling efforts (Kaufman, et al. and others) were directed to tungsten CMP. There have been two types of models which have been developed which appear to fairly well describe the tungsten CMP process and these are covered here. Following the discussion of these two models, copper CMP will be considered.

4.8 Tungsten Model of Paul

Paul used standard reaction engineering techniques to model the passivation and abrasion mechanism presented at the beginning of the chapter [25]. He specifically modeled the removal mechanism of tungsten CMP but the author states that the model is applicable to any metal CMP process in which the passivation and abrasion mechanism holds. Tungsten forms an oxide slowly then forms a stoichiometrically incomplete oxide at a faster rate. The formation of the oxides can be described with the reactions:

$$\text{W} + \text{O}x \leftrightarrow \text{W}ox \rightarrow \text{W}oxn. \tag{4.2}$$
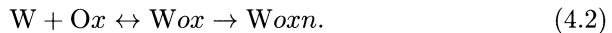
The general mechanism is given by three steps. First is the formation of the oxide (in the case of tungsten there is the formation, in series, of two oxides but for simplicity only one will be discussed). This is given as

$$\text{M} + \text{C} \leftrightarrow \text{MC}*, \tag{4.3}$$

where M is the material being removed (e.g. W or Cu) and MC* is the material-chemical complex. The second step is the removal of the oxide, which can occur by abrasion as well as dissolution (these were considered as independent phenomenon). The dissolution reaction is given by

$$\text{MC}* \rightarrow \text{MCaq} + \text{M}, \tag{4.4}$$

where MCaq is the aqueous dissolution product and M is a new, un-oxidized metal site. This reaction is assumed to be overall first order and the reaction rate constant (kD) includes the strength of the oxidizer in the slurry. The abrasion reaction is given by

$$\text{MC}*+\text{A} \rightarrow \text{MC–A} + \text{M}, \tag{4.5}$$
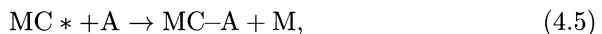
where A is the abrasive and MC–A is a material oxide-abrasive complex. This reaction is also assumed to be first order and the reaction rate constant (kM) includes terms for pressure, velocity, and abrasive diameter. The rate of formation of the MC* complex is assumed equal to zero at steady state.

The mechanical abrasion process requires an abrasive component in the slurry. The action of the abrasive is modeled by active sites on the pad that

the abrasive can occupy. The fraction of occupied sites is proportional to the number of available sites and the concentration of abrasive in the slurry. The rate at which abrasives leave is simply proportional to the number of abrasive sites on the pad occupied. Combining these gives the number of effective abrasives on pad area A as

$$Na = Ac_p[A]/([A] + KA), \qquad (4.6)$$

where c_p is the site density on the pad, [A] is the concentration of abrasives in the slurry, and K is the pad-abrasive equilibrium constant. Using the assumptions given and combining the above equations written in terms specifically for W CMP the rate of material removal per work piece area is given by

$$R = (R_d + R_m)/\theta A = (k_d + k_m c_p \theta A)\theta/d_m^2, \qquad (4.7)$$

where the R's are the individual reaction rates for the dissolution (D) and mechanical abrasion (M) reactions, θA is the fraction of pad sites occupied by abrasive particles, and theta is the fraction of surface sites covered by the MC* complex. The dependence on the concentration of chemistry in the slurry [C] is contained in theta while the dependence on particle concentration [A] is contained in θA. The rate given above can be transformed to show the specific dependence on [C] or [A] as shown in the following equations

$$R = \beta_1[C]/([C] + \beta_2); \qquad (4.8)$$

expressed as [C] dependence, and

$$R = \beta_3 + \beta_4[A]/([A] + \beta_5); \qquad (4.9)$$

expressed as [A] dependence.

The expressions for each beta are shown in Table 4.7 for the general case and the case where the rate of dissolution of MC* is insignificant ($k_D = 0$). Paul states, "It is helpful to note that all of the β_i have the form of a product of chemical and mechanical terms divided by their sum. This form forces a balance between the chemical and mechanical processes." Both rate expressions show an increase in removal rate as a function of concentration (chemistry or abrasive, respectively) and asymptotically reach a maximum polish rate limit as concentration increases. The removal rate approaches zero as [C] approaches zero and approaches β_3 (the static corrosion rate) as abrasive loading approaches zero.

Different sets of tungsten polish rate data obtained from the open literature were used to explore the developed model [26]. Figure 4.45 shows the polishing rate as function of chemical concentration for three different polish pressure and rotation rate settings. The rate initially increases rapidly with chemical concentration then begins to approach an asymptote. The fitting parameters β_1 and β_2 vary monotonically with the pressure and rotation rate. Figure 4.46 shows polish rate as a function of solids loading of the slurry.

Table 4.7. The expressions for each beta for the general case and the case where the rate of dissolution of MC* is insignificant ($k_D = 0$). From [25]

	General formula	$k_D = 0$
β_1	$\dfrac{k_2(k_D + k_{oM}c_P\theta_A Pv)}{d_W^2(k_2 + k_D + k_{oM}c_P\theta_A Pv)}$	$\dfrac{k_2(k_{oM}c_P\theta_A Pv)}{d_W^2(k_2 + k_{oM}c_P\theta_A Pv)}$
β_2	$\dfrac{k_D + k_{oM}c_P\theta_A Pv}{K_1(k_2 + k_D + k_{oM}c_P\theta_A Pv)}$	$\dfrac{k_{oM}c_P\theta_A Pv}{K_1(k_2 + k_{oM}c_P\theta_A Pv)}$
β_3	$\dfrac{k_D k_2 F_{Ox}}{d_W^2(k_D + k_2 F_{Ox})}$	0
β_4	$\dfrac{k_2 F_{Ox}}{(k_D + k_2 F_{Ox})}\dfrac{k_2 F_{Ox}(k_{oM}c_P)Pv}{d_W^2(k_D + k_2 F_{Ox} + k_{oM}c_P\theta_A Pv)}$	$\dfrac{k_2 F_{Ox}(k_{oM}c_P)Pv}{d_W^2(k_2 F_{Ox} + k_{oM}c_P\theta_A Pv)}$
β_5	$\dfrac{K_P(k_D + F_{Ox})}{(k_D + k_2 F_{Ox)}) + k_{oM}c_P Pv}$	$\dfrac{K_P k_2 F_{Ox}}{k_2 F_{Ox} + k_{oM}c_P Pv}$
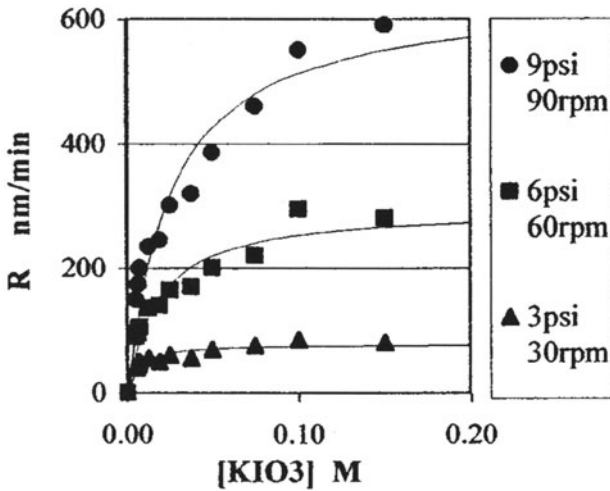
Fig. 4.45. The polishing rate as function of chemical concentration for three different polish pressure and rotation rate settings. From [26]

Again, the polish rate is low with low abrasive loading and increases to an asymptote as abrasive loading increases. The fit of the data show that β_3 equals zero implying that the static corrosion rate is negligible. The fit also shows that β_4 and β_5 increase monotonically with pressure and rotation rate. Figure 4.47 shows polish rate as a function of abrasive loading for five different sizes of alumina. The data fit again shows β_3 is zero. In this case β_4 and β_5 vary with particle size (pressure and rotation rate were not included
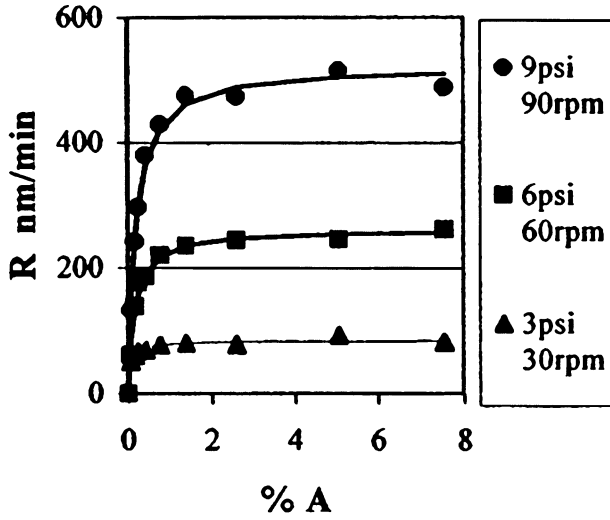
Fig. 4.46. Polish rate as a function of solids loading of the slurry. From [26]
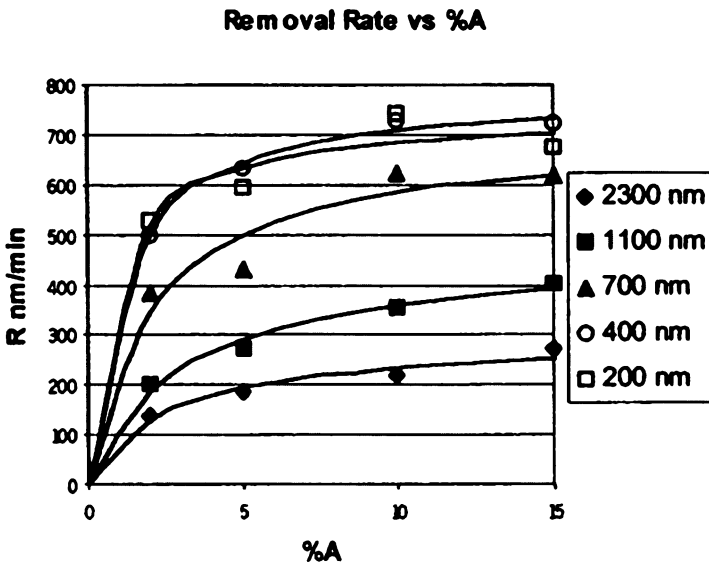
Fig. 4.47. Polish rate as a function of abrasive loading for five different sizes of alumina. From [26]

as factors in the original study). In all cases Paul's model visually appears to fit the data well.

4.9 Tungsten Model of Stein et al.

Taking another approach to metal polish modeling, Stein et al. [13] modeled the removal phenomena they reported using a heuristic surface reaction mechanism for illustrating possible fundamental processes that could occur during tungsten CMP. They assumed that the slurry colloid has sites (S) that will interact with the tungsten (W) to perform removal. The sites may be active (S_a) or inactive (S_i). Inactive sites are sites that participated in the removal of tungsten from the surface. Inactive sites are converted to active sites by the slurry chemistry (KIO_3, H^+). They assumed that all transport, adsorption, and desorption steps are rapid compared to the surface kinetics. The representative reaction on the colloid surface is

$$S_i + KIO_3 + H^+ \underset{k_{ar}}{\overset{k_{af}}{\leftrightarrow}} S_a, \tag{4.10}$$
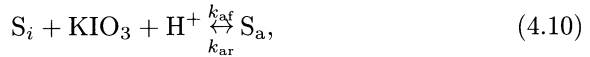
where k_{af} and k_{ar} are the forward and reverse reaction rate constants. The representative reaction on the tungsten surface is

$$S_a + W \underset{k_{wr}}{\overset{k_{wf}}{\leftrightarrow}} S_i, \tag{4.11}$$

where k_{wf} and k_{wr} are the forward and reverse reaction rate constants, and

$$k_{wf} = k'_{wf} Pv \frac{[colloid]_o}{[colloid]}. \tag{4.12}$$

Contact mechanics predicts that the area of contact between the colloid and the tungsten is proportional to the polish pressure (P). The numerical value of applied pressure assumes that the colloidal particle packing factor is unity, i.e. that the load per particle is independent of colloid concentration. The $[colloid]_o/[colloid]$ term is introduced to correct for packing factors below unity. $[colloid]_o$ is the colloid concentration at which further increases in colloid concentration do not increase the polish rate. The number of alumina sites that a particular tungsten atom could contact per unit time is directly proportional to the colloid velocity (v). The total number of surface sites is proportional to the colloid species concentration

$$k_c [colloid] = S_a + S_i, \tag{4.13}$$

where k_c is the constant of proportionality that includes the differences in site density between colloid species and phases. By taking reaction (4.11) to be at equilibrium, reaction (4.10) to be rate limiting, and assuming $k_{ar} \ll k_{af}$,

the best fit of the experimental data is obtained. With these assumptions, the rate of tungsten removal is given by

$$PR = \frac{k\text{Pv}}{1 + k\text{Pv}}, \tag{4.14}$$

where

$$k = \frac{k_{\text{af}} k_{\text{wf}} k_c^a \left[\text{colloid}\right]^{a-1} \left[\text{KIO}_3\right]^b \left[\text{H}^+\right]^c}{k_{wr}} \tag{4.15}$$

and

$$k'' = \frac{k'_{\text{wf}}}{k_{\text{wr}}} \frac{\left[\text{colloid}\right]_{\text{o}}}{\left[\text{colloid}\right]} \tag{4.16}$$

and a, b, and c are the apparent species reaction orders. Equation (4.14) predicts that the polish rate at low polish pressures and velocities is a function of the colloid species, colloid concentration, potassium iodate concentration, hydrogen ion concentration, polish pressure, and polish velocity. At high polish pressure and polish velocity, the polish rate should be dependent only on the colloid and slurry chemistry. Equation (4.14) also predicts that the polish rate in KIO_3 and colloid concentration limiting cases is a function of the concentrations of the limiting species.

Equation (4.14) also suggests that the reverse reaction of (4.11) proceeds at a significant, but slower rate than the forward reaction, since (4.14) was derived assuming the reaction of (4.11) was at equilibrium. This indicates that redeposition of tungsten from the colloid to the surface must be possible. Redeposition might be predicted from the mechanism for two reasons. First, redeposition might actually be occurring, as it does during glass polishing. Second, the strict limitations and simplifications placed on this heuristic model may prevent a more accurate description of the interaction between the colloid and the tungsten.

Figure 4.48 shows three polish rate data sets presented in Stein's work. These data show the polish rate in potassium iodate limited and alumina limited slurries as well as the polish rate in high concentration potassium iodate and alumina slurry. The data shown are fit to (4.14). The high R^2 values obtained indicate that (4.14), in a purely empirical form, describes the unlimited and iodate limited data with good accuracy.

As a heuristic model, (4.14) can be investigated using the values of k' and k''. Equation (4.16) predicts k' for the unlimited slurry should be larger than k' for the iodate limited slurry, regardless of the apparent reaction order. The unlimited data has a k' of 83.3, and k' for the iodate-limited slurry is 45.9. Thus, the fit data shows the expected trend in k'. k'' is independent of potassium iodate concentration, hence the unlimited and potassium iodate limited data sets shown should have similar values of k''. These data sets are well modeled using a k'' of 0.003. The alumina limited slurry should have a k'' significantly greater than 0.003 since, for this slurry, $[\text{colloid}]_{\text{o}}/[\text{colloid}]$
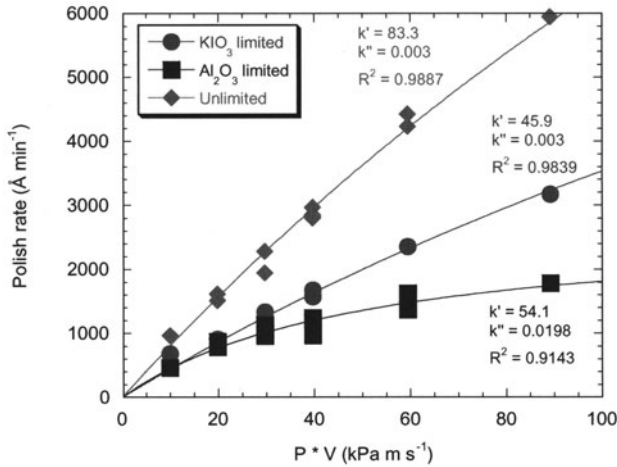
Fig. 4.48. Three polish rate data sets and the constants determined from them. From [13]

is approximately 10. The best fit k'' for the alumina limited case is 0.0198, hence the data shows the expected trend in k''.

Figure 4.49 shows polish rate versus potassium iodate concentration data taken from another work of Stein et al. [13]. An exponential fit to these data was performed to determine the apparent reaction order b. The best fit to all three curves occurs using the same value of b occurs when $b = 0.32$. Equation (4.16) predicts that the ratio of k' for the KIO$_3$ limited slurry
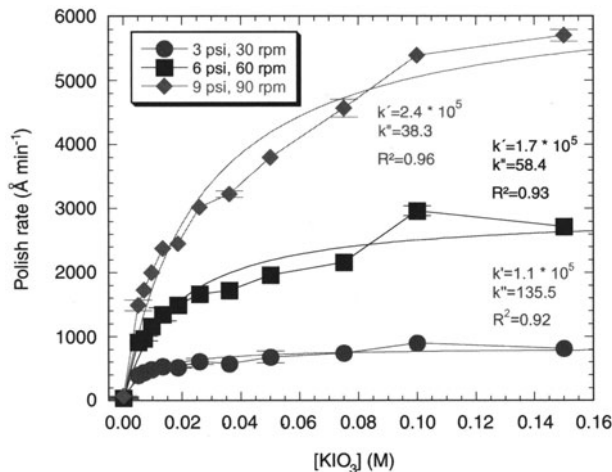
Fig. 4.49. Polish rate versus KIO$_3$ concentration data and the model constants determined from them. From [13]

(0.25 M) to the unlimited slurry (0.1 M) should be approximately 0.64. The fit value of k' is 45.9 for the KIO_3 limited slurry and 83.3 for the unlimited slurry, so the ratio is 0.56. Thus, the heuristic model predicts the polish rate response data well.

4.10 Copper Model of Babu et al.

Babu and coworkers have proposed a removal mechanism that is an extension of the Preston equation [27]. Figure 4.50 shows the Cu CMP removal rate as a function of table speed for 2 slurries. Figure 4.51 shows the Cu CMP removal rate as a function of downforce for the same two slurries. The first is a commercially available slurry that requires the addition of H_2O_2. The second slurry was made using 5 wt. % 100 nm α-alumina in a solution of 0.1 M $Fe(NO_3)_3$ and 0.005 M BTA. The scatter in the removal rate data was attributed to substrate wafer roughness.

The Preston equation is typically used as a first approximation to mathematically describe the removal rate of a CMP process. Note that the Preston equation does not include a term for an intercept with the y-axis (removal rate) for either or both P or v equal to zero. The best linear fit of the data shown in Figs. 4.50 and 4.51 indicate a non-zero intercept.

The authors explore two extensions of the Preston equation that might account for the non-zero intercept and then propose possible mechanisms for
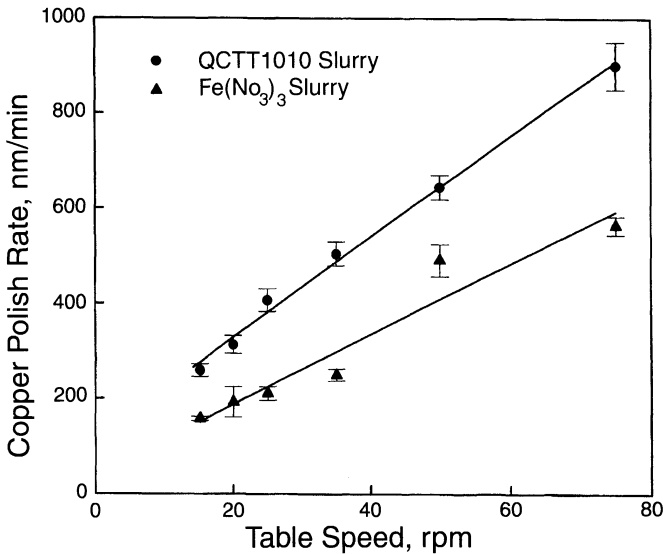
Fig. 4.50. The Cu CMP removal rate as a function of table speed for 2 slurries. From [27]
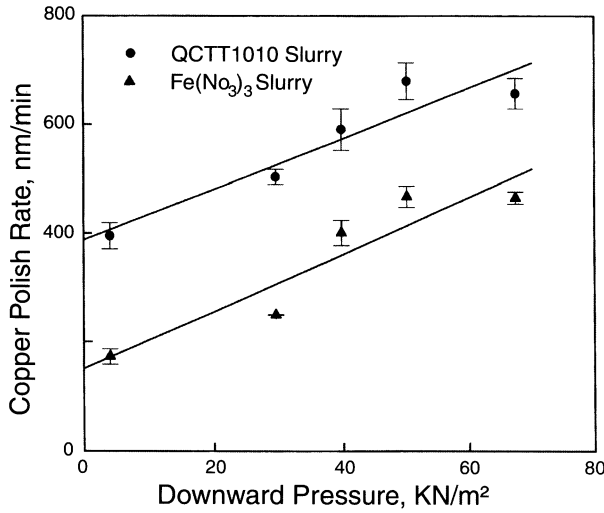
Fig. 4.51. The Cu CMP removal rate as a function of downforce for the same 2 slurries. From [27].
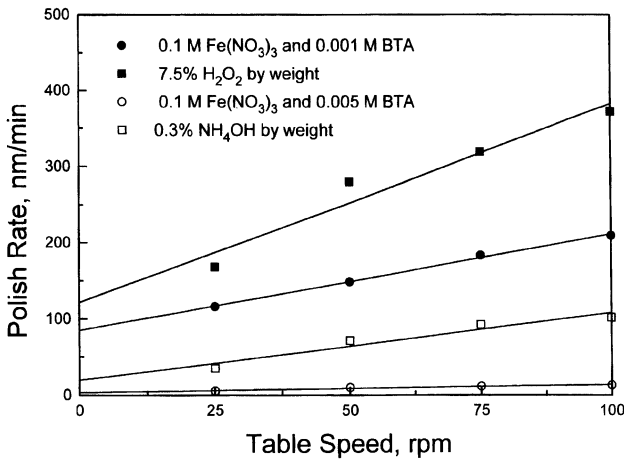
Fig. 4.52. The Cu CMP removal rate as a function of table speed at constant applied downforce (these experiments were performed using a small bench-top polisher to conserve materials). From [27]

the extra term(s). The first extension of the Preston equation is

$$RR = K(P + P_0)(v + v_0) = KPv + aP + bV + R_c. \qquad (4.17)$$

This equation predicts that the removal rate is proportional to applied pressure even at zero velocity. The authors cite unpublished data to indicate that this is not true. The second extension of the Preston equation is given by
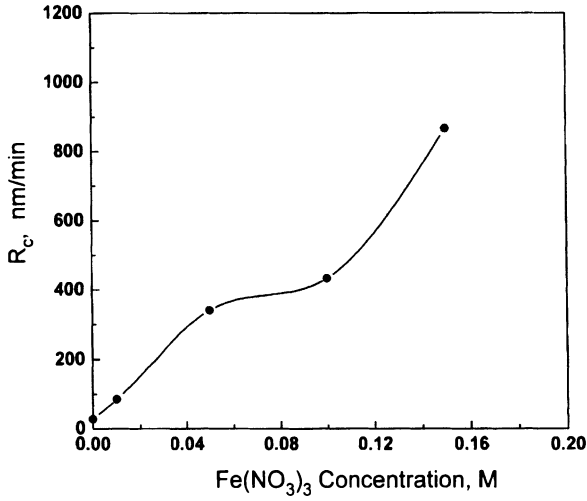
Fig. 4.53. The value of R_c as a function of $Fe(NO_3)_3$ concentration. From [27]

$$RR = (KP + B)v + R_c = KPv + Bv + R_c, \qquad (4.18)$$

where K, B, and R_c are constants. The authors interpret the constant Rc as representing the removal rate solely due to chemical effects. Figure 4.52 shows the Cu CMP removal rate as a function of table speed at constant applied downforce (these experiments were performed using a small bench-top polisher to conserve materials). Rc was determined from the intercept of each linear least squares fit line. It is evident that the value of Rc depends significantly on the slurry chemistry. Figure 4.53 shows the value of Rc as a function of $Fe(NO_3)_3$ concentration. Previous work has shown that the static corrosion rate of Cu in the presence of Fe^{3+} ions is between 60 and 100 nm min^{-1} for Fe^{3+} concentrations between 0 and 0.15 M Fe^{3+}. The value of Rc ranges between approximately 30 to 900 nm min^{-1} during CMP. The authors attribute the increase in the value of Rc to the "interaction of electric double layer and the reduction of copper surface hardness." The reduction of surface hardness was thought to occur through stress- induced corrosion.

4.11 Model Summary

The models presented above represent the second generation of theoretical investigation into removal mechanisms at play during metal CMP. First generation models, such as the Preston equation and its derivatives represent only mechanical removal. These models do not fit the experimental data well. The second-generation models, specifically those for tungsten CMP, take into account mechanical as well as chemical effects as well as the interactions between the two. The role of the particle as a pure abrasive is dropped and the

role of the particle as an abrasive and as a player in the chemical effect is taken up in these new models.

Though the tungsten CMP models require substantial further development in order to quantitatively understand the detailed chemical interactions at the tungsten surface, the copper CMP models are even less developed. The highly complex chemistry of copper and its oxides contributes to the challenge of developing a detailed model, since the reactions of interest must consider the suppressor and other chemistries introduced into the slurry.

The models needed for new generations of copper slurries will continue to be even more difficult to formulate since increasing complex slurry formulations are being developed for both first and second step copper slurries. Especially in copper CMP, the needed technology is moving well ahead of a detailed model of the actual physics and chemistry taking place in the process.

In summary, though the second generation models represent quite and advance from the first generation, an accurate molecular level description of the CMP removal processes(s) has still not been found.

4.12 Future Trends

Two new processes for metal CMP are on the horizon. The first is abrasive-free Cu CMP. This can be done either using a special pad that incorporates the abrasive into the polymer or by using specially designed reactive liquid systems and standard (non-abrasive, polyurethane only) pads. Most effects of process parameters when using abrasive-impregnated pads do not vary from traditional abrasive slurry-based CMP, though the consumable sets are distinctly different. For instance, when the polish pressure or rotation rates are increased, the polish rate increases. When the wafer carrier is set to the same or similar rotation rate as the platen, the within-wafer non-uniformity is lowest.

In contrast to abrasive-impregnated pad technology, reactive liquid Cu CMP does not use abrasive material at all [28], [29]. In this technology a specially formulated liquid and a regular polyurethane pad are used to remove the Cu overburden. The potential advantages are a reduced defect level and improved final topography control. Also, waste disposal can be substantially simplified. Some of the differences between standard, abrasive-based Cu CMP and reactive liquid CMP are that significant variations between carrier and platen rotation rate can decrease removal rate non-uniformity and that pad groove profiles become a very important factor in removal rate. This not yet well understood. However, the technology is very promising and should be the subject of strong development in the near future. The second challenge (or opportunity) on the horizon for metal CMP is for the removal and planarization of noble metals (see Chapter 10) such as gold, platinum, ruthenium, and iridium [30]. These materials will be used in non-logic IC technologies such

as DRAM and FeRAM. These materials pose unique challenges for CMP because of the difficulty of oxidizing or otherwise chemically reacting with these materials.

References

1. F.B. Kaufman, D.B. Thompson, R.E. Broadie, M. A Jaso, W.L. Guthrie, D.J. Pearson, and M.B. Small, J. Electrochem. Soc., **138**, 3460, 1991.
2. Gmelin Handbook of Inorganic Chemistry, 8th ed., Tungsten Supplement A7, Springer–Verlag, Berlin, 1987.
3. E.A. Kneer, C. Raghunath, S. Raghavan and J.S. Jeon, J. Electrochem. Soc., **143** (12), 4095, 1996.
4. D. Tamboli, S. Seal, V. Desai, and A. Maury, J. Vac. Sci. Technol. A **17** (4), 1168, 1999.
5. D.J. Stein, D.L. Hetherington, and J.L. Cecchi, Presentation at the 1st Annual Northern California AVS CMP User's Group Meeting, San Jose, November 1996.
6. D.J. Dtein, D. Hetherington, T. Gulinger, and J.L. Cecchi, J. Electrochem. Soc., **145** (9), 3190, 1998.
7. K. Osseo–Asare, M. Anik, and J. DeSimone, Electrochemical and Solid-State Letters **2** (3), 143, 1999 and M. Anik and K. Osseo–Asare, Electrochemical Society Proceedings **99–37**, 354, The Electrochemical Society, Inc., Pennington, 2000.
8. T. Gaffney, D.J. Stein, and D.L. Hetherington, Proceedings 2002 CMP-MIC Conference, 483, IMIC, Tampa, 2002.
9. M. Bielmann, U. Mahajan, and R.K. Singh, Electrochemical and Solid-State Letters **2** (8), 401, 1999.
10. D.J. Stein, D.L. Hetherington, and J.L. Cecchi, J. Electrochem. Soc., **146** (5), 1934, 1999.
11. S. Ramarajan, M. Haritharaputhiran, Y.S. Her, J.E. Pendergast, and S.V. Babu, Proceedings 1999 CMP–MIC Conference, 430, IMIC, Tampa, 1999.
12. D.J. Stein, J.L. Cecchi, and D.L. Hetherington, J. Mater. Res. **14** (9), 3695.
13. D.J. Stein, D.L. Hetherington, and J.L. Cecchi, J. Electrochem. Soc., **146** (1), 376, 1999.
14. J.M. Steigerwald, S.P. Murarka, R.J. Gutmann, and D.J. Duquette, Materials Chemistry and Physics **41**, 217, 1995.
15. R.J. Gutmann, J.M. Steigerwald, L. You, D.T. Price, J. Neirynck, D. Duquette, and S.P. Murarka, Thin Solid Films **270**, 596, 1995.
16. R. Carpio, J. Farkas, and R. Jairath, Thin Solid Films **266**, 238, 1995.
17. Q. Luo, D.R. Campbell, and S.V. Babu, Proceedings 1996 CMP–MIC Conference, 145, IMIC, Tampa, 1996.
18. Q. Luo, D.R. Campbell, and S.V. Babu, Thin Solid Films **311**, 177, 1997.
19. Q. Luo, M.A. Fury, and S.V. Babu, Proceedings 1997 CMP–MIC Conference, 83, IMIC, Tampa, 1997.
20. J. Keleher, E. Tyre, R. Her, S.V. Babu, and Y. Li, Proceedings 2000 CMP–MIC Conference, 66, IMIC, Tampa, 2000.
21. M. Hariharaputhiran, J. Zhang, S. Ramarajan, J.J. Keleher, Y. Li, and S.V. Babu, J. Electrochem. Soc., **147** (10), 3820, 2000.

22. H. Hirabayashi, M. Higuchi, M. Kinoshita, H. Kaneko, N. Hayasaka, K. Mase, and J. Oshima, Proceedings 1996 CMP–MIC Conference, 119, IMIC, Tampa, 1996.
23. T.C. Hu, S.Y. Chiu, B.T. Dai, M.S. Tsai, I.-C. Tung, and M.S. Feng, Materials Chemistry and Physics **61**, 169, 1999.
24. F. Preston, J. Soc. Glass Tech. **11**, 214, 1927.
25. E. Paul, Proceedings Spring 2000 MRS Symposium E, San Francisco, and E. Paul, J. Electrochem. Soc., **148** (6), G359, 2001.
26. The data from Figures 48 and 49 were obtained from D.J. Stein, D.L. Hetherington, and J.L. Cecchi, J. Electrochem. Soc., **146** (1), 376, 1999. The data from Figure 50 was obtained from M. Bielmann, U. Mahajan, and R.K. Singh, Electrochemical and Solid-State Letters **2** (8), 401, 1999.
27. Q. Luo, S. Ramarajan, and S.V. Babu, Thin Solid Films **335**, 160, 1998.
28. S. Kondo, N. Sakuma, Y. Homma, Y. Goto, N. Ohashi, H. Yamaguchi, and N. Owada, J. Electrochem. Soc., **147** (10), 3907, 2000.
29. R.E. Barker, G.C. Mandigo, and C.D. Lack, Proceedings 2001 CMP–MIC Conference, 144, IMIC,Tampa, 2001.
30. K. Moeggenborg, V. Brusic, I. Cherian, A.N. Powell, and W. Downing, Proceedings 2001 CMP–MIC Conference, 150, IMIC, Tampa, 2001.

5 Equipment Used in CMP Processes

Thomas Tucker

Chemical mechanical planarization is an enabling technology for fabrication of leading edge semiconductor devices. Originally considered to be "dirty" and incompatible with cleanroom processes, CMP has evolved into a critical process technology that includes not only the planarization step, but the post-CMP cleaning process as well. It is used not only in back end of the line interconnect processes, but is also used for critical process steps in the fabrication of transistors and other key device elements – where control of contamination is crucial if the device is to be functional. The evolution of dry-in dry-out CMP equipment platforms has been key to the introduction of CMP processes in mainstream semiconductor manufacturing beginning with the 0.35 µm technology node. If CMP is a key enabler in the manufacture of advanced semiconductor devices, then evolution of equipment platforms has enabled the use of CMP processes consistent with high volume manufacturing in a clean room environment. Going forward, automated process control for CMP will lead to further improvements in manufacturability, stability and predictive process results.

Equipment used in CMP processing evolved from tool sets used for production of polished semiconductor substrates. While silicon polishing requires a larger number of tools in order to meet the semiconductor industry's needs for starting wafers, the number of tools is dwarfed compared to the number used in advanced semiconductor processes that utilize CMP. As the requirements for CMP have evolved, CMP tools have become more sophisticated, with more features and processing capability incorporated into the toolset. At least four distinctive generations of tools can be identified, and a fifth is projected, as summarized in Table 5.1.

5.1 CMP Tool Requirements

The requirements for a robust CMP process are relatively easy to specify, but more difficult to implement in practice. The primary tool parameters that need to be controlled are those that determine the interaction between the wafer and the consumables set.

To a first approximation, removal rate is determined by mechanical abrasion (as first described in the Preston equation) and chemical activity. The

Table **5.1.** CMP equipment generations

Tool Generation	Time Period	Wafer Size, mm	Kinematics	Through-put, wph	Carrier Technology	Tool Process Capability	Metrology
1	1984–1994	100–200	Rotary	10	Fixed Plate	Dry-In Wet-Out 1 Step CMP + Buff Clean Single Wafer 10 mm Edge Exclusion	Motor Current EPD
2	1993–1999	125–200	Rotary Orbital Carousel	20	Fixed Plate Back Pressure Active Retaining Ring	Dry-In Wet-Out 1 Step CMP + Buff Clean Multiple Wafers 6 mm Edge Exclusion	Motor Current EPD In Line Optical Metrology
3	1998–2003	200–300	Rotary Orbital Linear	35	Active Retaining Ring Membrane Backing	Dry-In Dry-Out Multiple Step + Buff Clean Multiple Wafers 3 mm Edge Exclusion	In Line Optical Metrology Run-To-Run Control In-situ Optical EPD Acoustic Monitor EPD Chemiluminescence EPD
4	2001–	200–300	Rotary Orbital Linear Elliptical	50	Active Retaining Ring Membrane Backing Zone Pressure Control	Dry-In Dry-Out Multiple Step + Buff Clean Multiple Wafers ≤ 2 mm Edge Exclusion	In Line Optical Metrology Adaptive Process Control Coefficient Of Friction EPD In-situ Optical EPD Acoustic Emission EPD Chemiluminescence EPD Eddy Current EPD
5	2003–	300	Rotary Orbital Linear Elliptical	50	Active Retaining Ring Membrane Backing Zone Pressure Control	Dry-In Dry-Out Multiple Step + Buff Clean Multiple Wafers ≤ 2 mm Edge Exclusion Integration With Electro-plating Module And Rapid Thermal Anneal Electrically Enhanced CMP	In Line Optical Metrology Adaptive Process Control Coefficient Of Friction EPD In-situ Optical EPD Acoustic Emission EPD Chemiluminescence EPD Eddy Current EPD

chemical component results in conversion of surface layers into softer by-products that are more easily removed by mechanical abrasion, or results in etching of the film surface. In general, etching is undesirable, since it is more difficult to control and terminate. As a result, the CMP tool is actually a collection of subsystems, each with its specific purpose for conducting and controlling the process and the environment associated with the CMP process.

The major tool sub-systems include:

1. The **robotics system** provides wafer handling for routing wafers through the CMP tool to the different process stations in the required sequence. Pick and place robots, water tracks, and wafer carriers serve as components in the wafer transport system.

2. The **mechanical drive system** controls the surface velocity within one percent or better of the desired set point under full loads. Any variation in surface velocity, or deviation from the set point, will directly result in variations in the removal rate. Separate drive systems are used for the wafer carrier and for the platen, upon which the polishing pad is mounted.

3. The **down force system** is used to precisely control pressure within one percent or better during the CMP process. Any deviation from the required load will also result directly in variations in removal rates, and can potentially impact planarization length and efficiency if the deviations are sufficiently large. This system includes a mechanism for applying a down force on the workpiece, and sensors for measuring it.

4. The **thermal management system** is used to control the temperature of the CMP process. Because lateral frictional forces and mechanical abrasion result in material removal, heat is generated during the CMP process [1]. Oxide CMP processes have been reported to generate temperature rises of $5°C$ or more, while many metal CMP processes are exothermic and result in significantly higher process temperatures. In addition, frictional forces resulting from interactions between the film surface, pads and slurries used during metal CMP generally are higher than frictional forces in oxide processes, which results in greater heat generation [2]. Since chemical activity varies with temperature, it is beneficial to control the temperature to a desired set point. The temperature control can be achieved by using recirculating coolant through the polishing platen and a heat exchanger, or by recirculating coolant through the wafer carrier. Accurate control of the slurry temperature can also be used to assist in thermal management during CMP processing. It is desirable to control the CMP process temperature within $1°C$ of the setpoint.

5. The **pad conditioning system** is used to regenerate the polishing pad at frequent intervals. The pad conditioning restores asperities on the polishing pad surface, and eliminates glazing, or build-up of polishing by-products.

6. The **slurry distribution system** is used to deliver slurry, other chemicals and rinse water to the polishing platen at the proper locations, in the proper sequence, and at controlled flow rates. If the flow rate is too high, the process cost of ownership increases significantly because of wasted slurry. If the flow rate is too low, slurry starvation may result in deteriorating removal rate uniformity, and can increase defectivity. In some instances, it is desirable to utilize point of use mixing as the slurry components are dispensed on the polishing pad, since the slurry may be unstable due to settling of the abrasive, or may have a short shelf life because of loss of the oxidizer with time. This is particularly true with highly selective slurries used in some shallow trench isolation CMP processes.

7. The **wafer cleaning system** removes organic, metallic and particulate contamination from the wafer surface and prepares the wafer for the next step in the integrated process flow.

8. The **metrology system** provides either in-line or in-situ measurements that can be used for endpointing the CMP process, for providing data for use in run-to-run control, or for in-situ measurements used in closed loop, automated process control.

9. The **wafer carrier** is one of the most critical systems used on the CMP tool. The wafer carrier ensures that the wafer remains in place during the CMP process and intra-machine transfers, and determines the pressure distribution across the face of the wafer during the planarization step.

10. The **waste stream system** provides for management of the effluent from the CMP process.

11. The **air flow system** controls the pressure distribution within the CMP tool, and provides pressure balancing between the fab and the various CMP modules and subsystems. It is designed so that air flows from the cleanest area to the most contaminated region (the fab, then to the post-CMP cleaning module, and then to the CMP station) before it is exhausted from the tool. Depending upon the slurries used in the CMP process, the air may require treatment before it is discharged into the atmosphere.

12. The **control system** manages the operation of each of the other systems, stores and controls the process recipes and process sequencing, and communicates with the fab host computer. It also provides data logging functions, and can display the operating history of the tool. It also can be used for diagnostic purposes during maintenance procedures, and can be used to configure access levels for operators, process engineers and maintenance personnel.

Kinematics, carrier design, pad conditioning, cleaning and end pointing technologies are the primary differentiators for commercial CMP tools. Many different tool configurations have been commercialized. While tools using rotary kinematics are the most common, tools have also been developed using

orbital kinematics, linear kinematics, planetary kinematics, and variations of
modified back-grinding methods (sometimes referred to as "elliptical" since
the shape of the polishing pad does not need to be circular). The objective
is to try to obtain the same average velocity at every point on the wafer,
while limiting the velocity range across the wafer face. Each of the different
kinematic approaches can achieve the same average velocity profile across the
wafer face, and also can achieve nearly identical velocity ranges of approxi-
mately 15% at some point within the operating space of the tool. However,
the process latitude varies between the different commercial tools, and there
are other factors that must be considered such as throughput, the tool foot-
print, the platen size (which determines the pad size and therefore its cost),
and overall tool reliability. In addition, some platforms enable easier imple-
mentation of in-situ measurements that can be used for endpointing. The tool
architecture directly impacts the sampling rate for measurements because of
the tool design.

The following drawings illustrate elements in a typical rotary CMP plat-
form. Systems based on other kinematic approaches have similar characteris-
tics, although the method of implementation can differ because of constraints
imposed by the actual tool design.

As shown in Fig. 5.1, the wafer is located face down within a wafer carrier.
(In some implementations, the wafer surface is face-up, and the polishing pad
is located above the wafer). The wafer is mounted in the carrier through use
of a load cup, and application of a vacuum facilitates wafer loading and
seating within a retaining ring. The carrier is rotated by a spindle under
a controlled, vertical force that applies pressure across the wafer face against
a polishing pad located on the polishing platen. The carrier spindle provides
passages for pulling a vacuum and for delivery of fluids behind the wafer. In
some cases back pressure is applied behind the wafer in order to control the
pressure distribution across the wafer face. The polishing platen rotates using
a separate drive system. The platen spindle also has passages for providing
coolant flow within the platen in order to control the process temperature.
In addition, other components may use the channels in the platen spindle to
provide electrical connections to various sensors (not shown) which can be
used for in-situ monitoring of the CMP process; optical and electrical slip
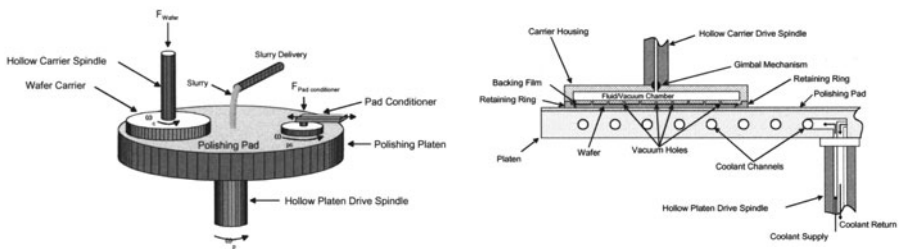
Fig. 5.1. Elements in a typical rotary CMP tool

rings may also be mounted on the spindle in order to transmit and receive electrical signals to or from a separate control system or to the tool's primary control system.

As the wafer is rotated at a rate ω_H under load F on a polishing pad rotating at a rate ω_P, polishing slurry is dispensed at the center of the pad at a controlled flow rate. Centrifugal forces distribute the slurry across the polishing pad, which carries it between the pad and the wafer surface. The combination of mechanical abrasion from the particles in the slurry and pad asperities provides a mechanical removal component, while active components in the slurry control the slurry pH, electrochemical properties, and chemical reactions occurring on the wafer surface.

As the polishing process proceeds, the asperities on the pad surface are degraded, leading to a progressive falloff in removal rates. In addition, polishing by-products and other debris can build up on the pad surface. The pad conditioning mechanism is used to restore the pad's surface texture and asperities, and to remove the accumulated debris. The most common method used is abrasive pad conditioning, using diamond-embedded end effectors on a conditioner arm that sweeps across the pad surface as it rotates under a controlled load. The pad conditioning can be performed in-situ as the wafer undergoes the CMP process, or it may be performed between wafers, while the wafer just polished is unloaded and a new wafer is loaded onto the carrier.

Following completion of the CMP process, the wafer is transported to a separate load/unload station. Pressure is applied behind the wafer, and the wafer is discharged from the carrier and transported into the post-CMP cleaning module. A new wafer is placed in the loading cup, the carrier is lowered over the wafer, and vacuum is applied to seat the wafer within a retaining ring on the carrier. The new wafer is transported within the carrier to the polishing platen, and the process sequence repeats.

Each of the tool designs based on the different kinematic approaches will be discussed in depth, along with the advantages and disadvantages of each approach.

5.2 Rotary CMP Tools

The earliest CMP processes were developed using single-wafer rotary polishing platforms originally used for silicon wafer polishing. However, significantly different requirements were identified early on with respect to machine and process operating parameters. These differences include:

 1. Removal rates for CMP processes are lower than those used in silicon wafer polishing, and the amount of material removed is typically on the order of 1 to 2 microns compared to material removal of 10 to 25 microns in substrate polishing.

2. Rotation rates during semiconductor CMP processes typically are in the range of 10–100 rpm, while silicon substrate polishing is typically performed with rotation rates up to 350 rpm.

3. In silicon wafer polishing, the system is designed to produce front and backside surfaces that are parallel within 0.5 to 2 μm. Some wedging, or taper, results from the polishing method used; generally, during polishing the backside is used as a reference for polishing the front side. During CMP, the front surface must be parallel to the polishing pad, which requires a gimballing or other self-leveling mechanism in the carrier, so that the CMP surface is front-surface referenced. For ILD processes, the remaining film thickness must be uniform across the wafer diameter, which is not achievable if back surface referencing is used. In metal damascene and STI processes, non-uniform clearing requires over-polishing, leading to unacceptably high dishing and erosion.

4. Most silicon wafer polishing processes use fixed mounting methods for attaching the wafer to a wafer carrier using a high quality wax. During CMP, waxless techniques are used almost exclusively.

5. Silicon wafer polishing is performed using batch processes with multiple wafers mounted on a metal wafer carrier. During CMP, each wafer is mounted on a separate carrier.

6. Slurries used in silicon wafer polishing are typically alkali based, with a pH between 10 to 11. Slurries used in CMP may have a pH as low as 2 for metal processes, and as high as 11 for dielectric processes. Many CMP slurries have a pH near 7. Materials of construction in a CMP tool must be compatible over the full pH range of 2 to 12.

5.3 Rotary Kinematics

The kinematics for a rotary CMP process have been described previously in the literature [3]. Using Fig. 5.2 as a reference, the velocity vector for the point Q on the wafer surface relative to the polishing pad can be readily described.

R_{CC} is the position vector from the center of the polishing platen to the center of the carrier, R_H is the position vector from the center of the polishing carrier to point Q on the wafer surface, and R_Q is the position vector from the center of the polishing platen to the point Q. The position vectors are related by the expression

$$R_Q = R_{CC} + R_H.$$

If the wafer carrier rotates about its axis with an angular velocity ω_P, the velocity of point Q with respect to the polishing pad can be described by the expression

$$V = -V_P + V_H = -(\omega_P \times R_Q) + (\omega_H \times R_H).$$
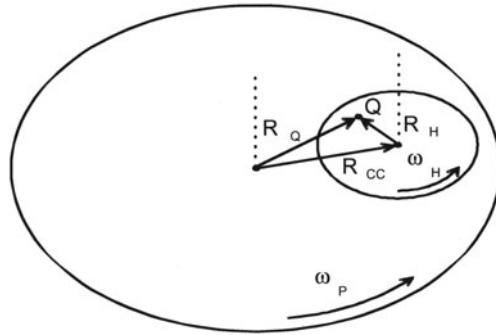
Fig. 5.2. Rotary kinematic diagram

Substituting for $\boldsymbol{R}_{\mathrm{Q}}$, the expression becomes

$$V = -(\boldsymbol{\omega}_{\mathrm{P}} \times (\boldsymbol{R}_{\mathrm{CC}} + \boldsymbol{R}_{\mathrm{H}})) + (\boldsymbol{\omega}_{\mathrm{H}} \times \boldsymbol{R}_{\mathrm{H}}).$$

Or, by rearranging,

$$V = -\boldsymbol{\omega}_{\mathrm{P}} \times \boldsymbol{R}_{\mathrm{CC}} - \boldsymbol{R}_{\mathrm{H}} \times (\boldsymbol{\omega}_{\mathrm{H}} - \boldsymbol{\omega}_{\mathrm{P}}).$$

For the case where the carrier and the platen have identical angular velocities, the second term is zero and the velocity is equal at all points on the wafer surface, leading to uniform polishing based on the kinematic analysis.

However, other factors confound removal rate uniformity. Because the inner part of the pad is utilized more than the outer part, frictional forces generate non-uniform temperatures across the pad surface [1]. In addition, the inner part of the pad is subjected to more wear, which results in non-uniform degradation of the asperities in the region of the pad used for polishing. Grooving and embossed patterns on the pad surface may further impact polishing behavior. As a result, achieving uniform average surface velocities is not sufficient to achieve uniform removal rates across the surface because of these effects.

In the simplest case, the vector $\boldsymbol{R}_{\mathrm{CC}}$ is fixed by the geometry of the polishing system. If the wafer rotates at the same rate as the polishing platen, or some integer multiple, points on the wafer will arrive at the same position on the polishing pad during each revolution (referred to as harmonic polishing). If any artifacts exist in the polishing pad, they can result in non-uniform polishing on the wafer. For this reason, the carrier rotation rate is generally varied slightly from the platen rotation rate in order to average out variations in the polishing pad-wafer interface. The wafer is also not rigidly mounted on the carrier, and can rotate and translate within the retaining ring during the process; the wafer may rotate one or two revolutions within the carrier during a process cycle. In addition, in order to utilize more of the pad, the wafer center is generally oscillated over the pad surface. While some systems

oscillate the wafer on the order of ± 1 cm or more from its average position along a platen radius or perpendicular to a platen radius, many different oscillation patterns have been employed. Although the velocity of the oscillating movement of the wafer center is small compared to the velocity of the platen and the carrier at the average wafer carrier center position, the net effect can be significant because of the change in direction and magnitude of the vector $\boldsymbol{R}_{CC}$. The velocity equation for point Q on the wafer surface is still valid, but the calculation of the velocity vectors over the wafer face must take into account the equation of motion of the wafer center. Figure 5.3 shows some of the many variations that have been implemented for changing the position of the wafer center during CMP processes.

Hocheng and co-workers [4] have discussed the simplest case where the center of the wafer is oscillated at a constant velocity U along a radius. In their work, they concluded that in order to maintain velocity non-uniformity of less than 3%, $3 < \omega_H/\omega_P < 1/3$.

While the discussion has been limited to a single wafer being polished on a polishing platen, other variations of rotary polishing have also been implemented. There is no fundamental limitation on the number of wafers that can be processed on a platen, other than geometrical constraints and the need to perform periodic maintenance functions such as pad conditioning. Systems have been commercialized with up to six wafers per platen. By increasing the number of wafers that are polished, the floor space utilization of the tool can be improved, and throughput also increases depending upon the efficiency of the wafer loading and unloading cycle time. Figure 5.4 shows some of the basic system layouts that have been commercialized.

CMP processes based on rotary kinematics are used on more than 80% of the installed base of spindles. In the early development of CMP, processes of record were established for rotary processes and extensive characterization
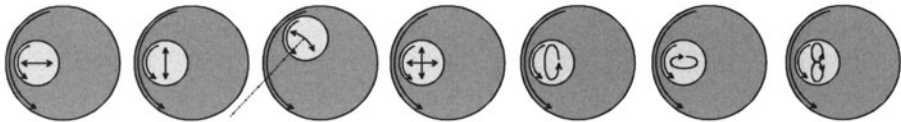
Fig. 5.3. Oscillation methods used during CMP processes
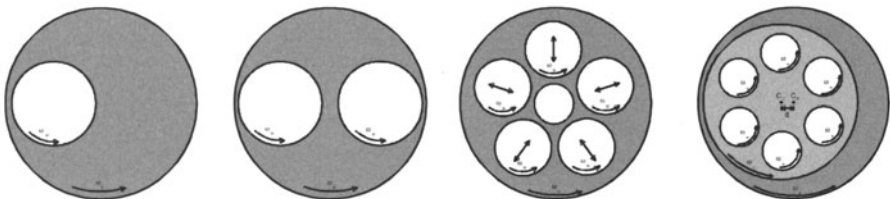
Fig. 5.4. Variations of the rotary process that have been commercialized

data was established. Consumables were readily available for rotary platforms as CMP processes were phased into production. Significant effort has been expended in optimizing consumables sets for rotary processes.

There are disadvantages with the rotary method, however. If a single wafer is polished on the platen, the space utilization is poor and the equipment does not scale well. In order to provide for the oscillation range of the wafer center and to increase the surface velocity for a given rotation rate, the platen diameter is 2.5 to 3.0 times larger than the wafer diameter. As a result, the diameter of the platen increases at least two times the increase in the wafer diameter. For single wafer/platen processes, the platen size for a 200 mm diameter wafer is typically 500 to 550 mm. For 300 mm wafers, the platen size increases to 750 mm to 830 mm. The average pad utilization rate is also relatively low, and pad wear is non-uniform. Space efficiency can be improved by polishing two or more wafers per platen, but at the risk of greater losses if a catastrophic event occurs, such as wafer breakage in one of the carriers. For multiple wafer processes, the wafers may not all reach endpoint at the same time, requiring independent control over the down force on each wafer carrier. Consumables costs are also relatively high, particularly because of poor utilization of slurry. Because of these constraints, other CMP platforms based on different kinematics have been developed and are slowly being introduced into volume production.

5.4 Carousel Systems

One of the alternative kinematic approaches is a carousel motion that is a derivative of the rotary process, as shown in Fig. 5.5. Wafer carriers are mounted on a carousel at equal distances from the carousel center, C_C.
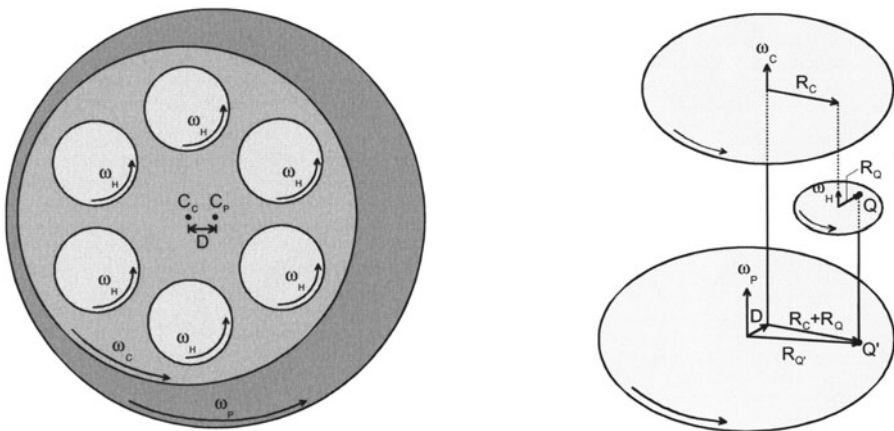
Fig. 5.5. Kinematic diagram for carousel polishing system

The center of the carousel is displaced a distance $\boldsymbol{D}$ from the center of the platen C_P, upon which the polishing pad is mounted. The polishing platen rotates with angular velocity $\boldsymbol{\omega}_P$, the carousel rotates about the displaced center C_C with an angular velocity $\boldsymbol{\omega}_C$, and the wafers rotate about their wafer centers with a velocity $\boldsymbol{\omega}_H$. The carousel motion effectively moves the wafers toward and away from the platen center. Since there are three angular velocities, the kinematic equation for the motion of a point Q on a wafer surface relative to the polishing pad differs from that of the pure rotary case. It has been shown [5] that the relative velocity of point Q is

$$\boldsymbol{V}_Q = \boldsymbol{\omega}_C \times \boldsymbol{R}_C - \boldsymbol{\omega}_P \times (\boldsymbol{R}_C + \boldsymbol{D}) + (\boldsymbol{\omega}_C + \boldsymbol{\omega}_H - \boldsymbol{\omega}_P) \times \boldsymbol{R}_Q,$$

where $\boldsymbol{R}_C$ is the position vector from the center of the carousel to the center of the wafer and $\boldsymbol{R}_Q$ is the position vector from the center of the wafer carrier to the point Q. For the condition

$$(\boldsymbol{\omega}_C + \boldsymbol{\omega}_H - \boldsymbol{\omega}_P) = 0$$

the instantaneous velocity is the same for all points on the wafer. The velocity will vary over a range during a complete cycle, however, because the position vectors $\boldsymbol{R}_C$ and $\boldsymbol{R}_Q$ are time dependent. By adjusting the process parameters, the velocity range can be reduced to a few percent.

5.5 Orbital Systems

CMP systems using orbital kinematics have also been commercialized. A system based on orbital kinematics has several advantages compared to rotary or carousel systems. If only an orbital motion is used, each point on the wafer surface has the same relative velocity on the pad surface. Because of the geometry of the system, the footprint of the tool can be minimized, resulting in improved floor space utilization. The system also enables simplified, through-the-pad slurry delivery since rotary couplings for fluid delivery are not required. As a result, the amount of slurry required for the process is minimized, and it is possible to rapidly change slurry formulations in a multi-step process. In-situ endpointing is simplified, since flexible cables can replace electrical and optical slip rings. Another advantage is that the size of the system scales directly with the increase in wafer diameter. Finally, the pad surface area is more fully and uniformly utilized.

The orbital kinematic system also has some disadvantages. In order to obtain surface velocities comparable to other kinematic systems, either the orbiting radius must be large, which increases the tool footprint, or a high orbiting frequency must be used, which may result in the creation of low frequency vibrations which can be transmitted to other process tools. Also, while the pad area is used more efficiently, pad changes may be required more frequently, resulting in lower tool utilization rates. If only an orbital motion is

used, the pad and wafer will always return to the same relative position during each orbit, resulting in possible artifacts or non-uniform removal rates on the wafer surface. For this reason, other motions are typically combined with the orbital motion, resulting in radial velocity variations across the wafer.

Figure 5.6 shows a typical system based on orbital kinematics. The orbiting axis is displaced a distance R_0 from the center of the pad, which travels with a orbital motion as shown, while the pad is constrained from rotating. One possible embodiment for creating an orbital motion is described by Breivogel, et al. [6]; however, orbital motions can also result from other mechanical designs such as those using gears.

In the event that the wafer does not rotate, the relative velocity of each point on the wafer surface is the same, and is expressed as

$$V = \omega_0 \times R_0.$$

In order to obtain a surface velocity of one meter per second (achieved in typical rotary systems), for a system with an orbital radius R_0 of 1.5 cm, the orbital frequency must be approximately 10.5 cycles per second, or 630 rpm – a factor of 10 higher than used in typical rotary systems.

In Fig. 5.7, the motion of the polishing platen with respect to the wafer is shown for orbit positions of 0, $\pi/2$, π, $3\pi/2$, and 2π. In the diagrams, the center of the wafer is represented as an open circle, the orbit center is represented by a filled circle, and the center of the polishing platen is
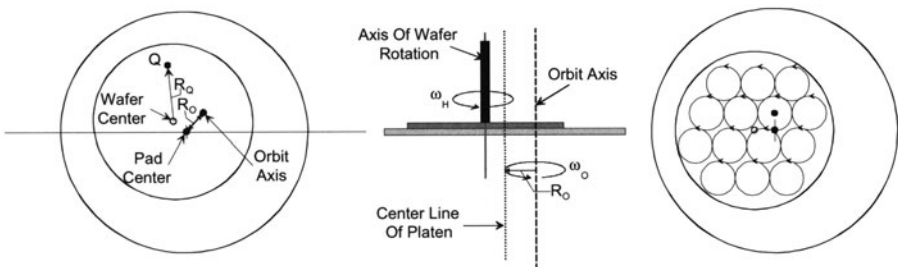
Fig. 5.6. Orbital kinematic system
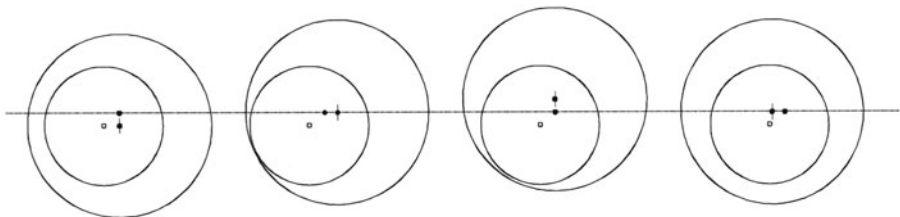
Fig. 5.7. Orbital motion when the wafer center, platen center, and orbit axis are not coincident

represented by a filled circle with a line through it. The center of the wafer, the center of the pad, and the orbit center do not coincide. If they are located as shown, the motion of the wafer does not fully utilize the pad surface and the area used is asymetrical with respect to the pad center. During the orbiting motion, the center of the polishing platen will be co-linear with the orbit center and the wafer center twice per revolution. If the platen center passes between the orbit center and the wafer center, the diameter of the polishing pad will be smaller than if the center of the wafer passes through the line connecting the orbit center and the platen center.

The most efficient orbital design occurs when the orbit axis passes through the wafer center. When this geometry is used, the diameter of the platen is minimized, and the wafer motion with respect to the pad center is symmetrical as shown in Fig. 5.8, which shows orbit positions of 0, $\pi/2$, π, $3\pi/2$, and 2π. As a result, smaller diameter polishing pads can be used. However, the wafer is susceptible to non-uniform polishing because the wafer and pad always return to the same relative position. As a result, the orbit axis is best displaced by a small distance from the wafer center.

In practice, to provide a more random orientation of any point Q with respect to the polishing pad, the wafer is rotated while the pad undergoes an orbiting motion. Referring again to Fig. 5.6, the kinematic equation for the relative motion of point Q with respect to the polishing pad becomes

$$\boldsymbol{V}_{\mathrm{Q}} = -\boldsymbol{\omega}_{\mathrm{H}} \times \boldsymbol{R}_{\mathrm{Q}} + \boldsymbol{\omega}_0 \times \boldsymbol{R}_0.$$

Now, however, all points on the wafer surface no longer travel at a constant velocity because of the increased velocity at the edge of the wafer compared to the wafer center as the wafer is rotated. For very low wafer rotation rates, the velocity variations are negligible, but for rates of 0.5 revolutions per second, the velocity range can be $\pm 30\%$ of $\boldsymbol{V}_{\mathrm{avg}}$ – significantly higher than in typical rotary systems.

Systems have been commercialized where the wafer is stationary while the polishing platen orbits [7], where the wafer rotates and the polishing platen orbits [6] where the wafer orbits and the polishing platen orbits [8], where the wafer orbits and the polishing platen rotates [9], and where the wafer rotates and orbits while the polishing platen orbits [10]. Some of these kinematic
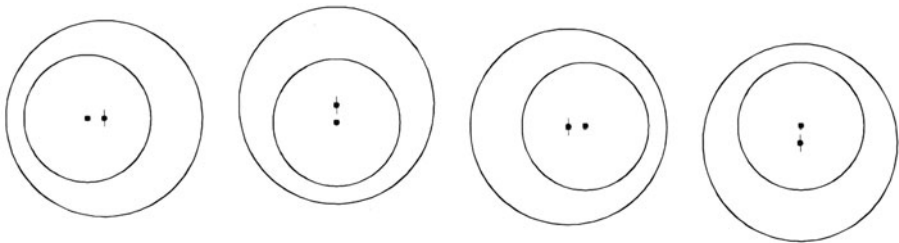
Fig. 5.8. Orbital motion when the wafer center and orbit axis are coincident

systems become very complex, and generally result in a much smaller process window compared to the simple case of a pure orbital motion by either the wafer or the polishing platen.

5.6 Linear Systems

Another system that has been commercialized uses linear kinematics for the CMP process [11]. One method of achieving linear planarization is to use a continuous belt as shown in Fig. 5.9. A motor drives the belt, and surface velocities greater than two meters per second can be achieved before the wafer starts to hydroplane. As a result of the higher surface velocities, equivalent removal rates can be achieved while using lower down forces, which results in improved within wafer removal rate uniformity and higher planarization rates [12].

The polishing belt is fabricated using conventional polishing pads mounted on a stainless steel belt. The belt drive is controlled so that the belt moves at a controlled, constant velocity. If the wafer does not rotate or translate, the velocity at all points Q on the wafer surfacc is the same and can be expressed as

$$\boldsymbol{V}_Q = \boldsymbol{V},$$

where $\boldsymbol{V}$ is the velocity of the belt. The wafer is located over a rigid plate that contains a fluid bearing which can be used to control the pressure distribution across the wafer face [13]. Since the fluid bearing controls the pressure distribution, the wafer can be mounted in a carrier with a rigid backing plate.
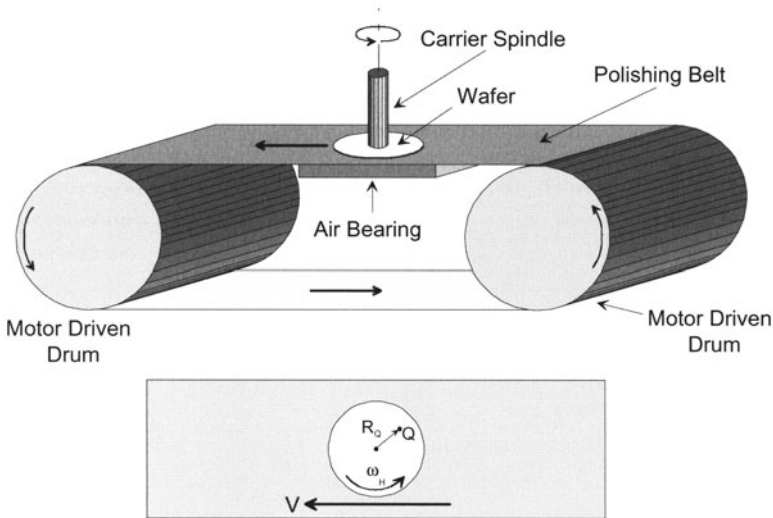
Fig. 5.9. Linear planarization system and kinematics

A gimballing mechanism is required to ensure that the wafer face remains parallel to the polishing belt. In practice, the wafer can be rotated to produce an averaging effect to compensate for any anomalies in the polishing pad, but it can not translate since it must remain in a fixed position over the fluid bearing in order to maintain the optimum pressure distribution for minimizing within wafer removal rate variations.

In the event that the wafer rotates, the kinematic equation for the velocity of any point Q on the wafer surface with respect to the polishing pad is expressed as

$$\boldsymbol{V}_Q = \boldsymbol{V} - \boldsymbol{\omega}_H \times \boldsymbol{R}_Q.$$

If the polishing belt is moving at 2 meters/second and the wafer is rotating at 10 rpm, the instantaneous velocity varies by $\pm 5\%$ from the average velocity at the wafer edge. The average velocity is the same at all points on the wafer surface, however.

Another method for achieving linear kinematics on a CMP polishing tool is shown in Fig. 5.10 [14]. This method uses a system of rollers and belts to generate a reciprocating motion on a polishing pad. As the fixed end of the pad passes over or under a roller the motion of the polishing pad is reversed, controlling the direction and range of motion. Separate drive systems can be used to rotate or translate the wafer. The kinematic equations describing the motion of any point Q on the wafer surface are the same as in the previous case. However, it is also possible to synchronize the wafer translation across the width of the polishing belt. By properly controlling the timing and direction of the translation of the wafer carrier across the
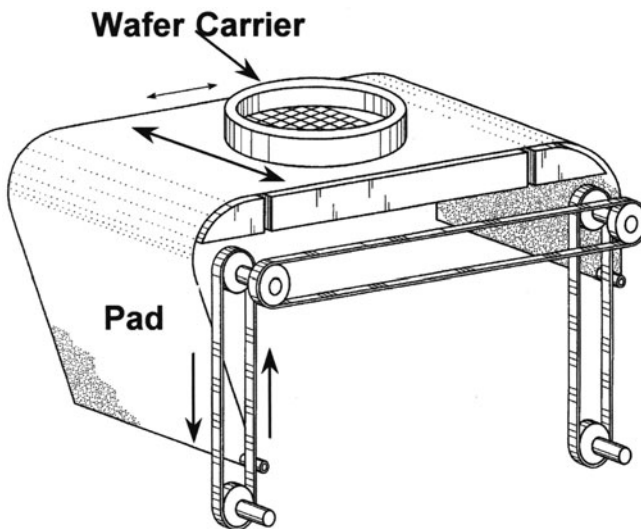
Fig. 5.10. Use of a reciprocating pad to create a linear kinematic system. From [14]

width of the belt, and by properly controlling the velocity and direction of the belt motion, the wafer can be made to create an orbital motion with respect to the pad surface.

5.7 Modified Grinding Systems

The majority of the CMP system architectures that have been described orient the wafer face down against the polishing pad. Alternative systems have been developed where the wafer is in a face up position, and a small wheel-shaped polishing pad is used to provide mechanical abrasion on the surface of the wafer. The polishing wheel rotates, and also scans across the wafer face. These systems are typically modifications of machine tools used for wafer back grinding and thinning prior to chip packaging. A system using this approach is shown in Fig. 5.11 [15]. A similar system which uses an index table for multi-station processing [16] has been developed and is now also being commercialized.

In these systems, the polishing wheel can be fabricated from a standard polishing pad, or can use a fixed abrasive embedded in a polymer matrix. The removal process can be facilitated by applying a chemical solution, or by using a standard free abrasive slurry. There are several advantages claimed for this approach which include: ease of access to the wafer face for in-situ measurements and endpointing; lower slurry consumption; use of smaller polishing pads which lowers consumables costs; and the ability to change process parameters in localized regions to compensate for variations in local removal rates or variations in the incoming material. These systems typically use high rotation rates on both the wafer and the polishing wheel. They have the potential to reduce the downforce on the wafer surface in order to reduce erosion, dishing and total metal loss. These systems are significantly more difficult to control, and in general have a narrower process window for controlling removal rates than offered by other systems. The system shown in

$$R = \int kP(\theta, x)\{V_{Wh}(x) + V_{Wa}(x)\}dt$$
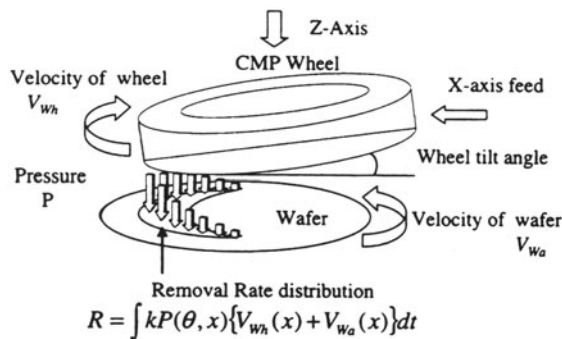
Fig. 5.11. Modified grinding wheel used for CMP processing. From [15]

Fig. 5.11 tilts the axis of the polishing wheel so that the pressure distribution varies from the leading edge of the wheel to the trailing edge of the arc of contact. The removal rate therefore varies as the pressure varies on the wafer face. The contact area using this approach has proved difficult to control because of variations in the polishing pad compressibility as the pad ages, or from pad to pad variations during extended runs.

5.8 Web Format Tools

CMP systems with a web format have also been developed. These systems can either use webs fabricated from typical polishing pad materials such as polyurethane in conjunction with conventional CMP slurries, or they can utilize fixed abrasive webs with chemical formulations which provide chemical activity to facilitate the removal process. These systems have the advantage of incrementally suppying new pad material with each wafer processed, which creates a steady state condition for multiple wafer processing. Pad changeouts and tool requalification time are minimized, since the web supply reel can provide sufficient material for a week or longer under continuous production conditions. Figure 5.12 shows the architecture of a typical web-based system [17, 18], while Fig. 5.13 shows the motions of the wafer during the CMP process.

As described in the references, the wafer orbits using x-y motion controllers to control the motion of the wafer while maintaining a constant velocity for all points on the wafer surface. The wafer does not rotate, which enables local pressure control on the wafer face through use of pneumatics. The web travels through a support with sloped sides which maintains alignment and also serves as a slurry containment device. Figure 5.14 shows the traces of successive wafers polished on the web as it moves through the CMP tool.
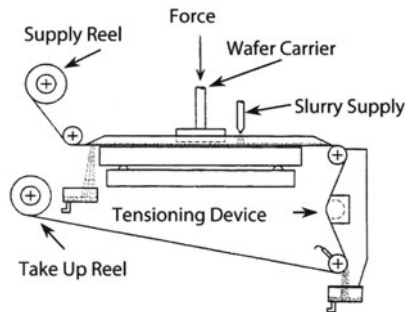
Fig. 5.12. Web supply and take up mechanism used on a CMP tool
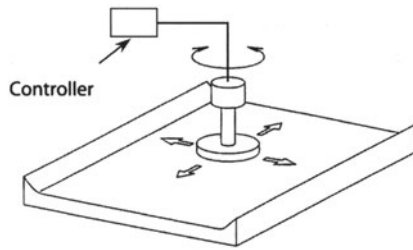
Fig. 5.13. Wafer motion on a web format CMP tool
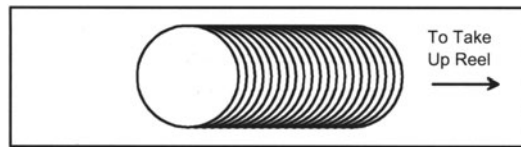
Fig. 5.14. Web area used by the motion of successive wafers

Many other architectures have been proposed for CMP tools, all which have the same objectives of providing a reliable design that is simple, which minimizes consumable use, which has a small footprint, which provides easy access to the wafer face for obtaining in situ measurements, and which enables accurate, local control of the pressure distribtution so that adaptive process control can be used. However, the systems that have been reduced to commercial practice are those that have been described in this chapter.

CMP processes have been used since the late 1980's and immense resources have been used to characterize processes using rotary systems. As a result, rotary systems still are the dominate architecture in use. Although orbital systems were introduced in the mid-1990's, they have primarily been used for tungsten CMP processes, and will most likely not be extended for use in copper CMP processes without significant modifications. Linear systems have represented the most radical departure from the processes of record used in the late 1990's time frame, but they offer potential compelling advantages for use in next generation processes; although the process data is not as voluminous for linear CMP systems, the potential footprint advantages and lower tool acquisition costs, coupled with improved planarization efficiency most likely will continue to attract significant interest in the industry.

The future of web format tools is still uncertain. While these systems offer some process advantages and enable direct polish STI processes, defectivity from use of fixed abrasives, along with the higher cost of consumables has detracted from wide spread use. If these issues are resolved, web format architectures may see increased use for CMP processing.

The future of the modified back grinding approach is also uncertain, In order to be successful, it must be clearly differentiated from existing systems

in terms of offering compelling, enabling process technology, or it must offer significant cost reductions compared to existing systems and processes of record. While the potential of significant improvements exist, convincing data has yet to be generated that verify that the anticipated process and cost advantages are real.

5.9 Electrochemical Mechanical Planarization

Copper is electrochemically more reactive than tungsten, aluminum, and insulating dielectric materials used in semiconductor manufacturing, and does not easily form passivating native oxides. During copper CMP and post-CMP cleaning, care must be taken to eliminate chemical, galvanic and photo-induced corrosion [19]. In order to minimize these effects, the wafer must be shielded from exposure to light during processing.

Because the thickness of electrochemically deposited copper is pattern sensitive and is influenced by additives in the electroplating electrolyte, over-polishing is required to ensure complete removal of copper and barrier films in the field areas. This results in excessive dishing for wide features, and erosion (loss of dielectric thickness) in dense areas with narrow lines and spaces. Isolated features also create challenges during copper polishing. There is also concern about shear forces used during CMP which can result in delamination, particulary when porous low-k dielectric materials are used.

In order to lower stresses during planarization and to reduce CMP process times, electropolishing or electrochemical mechanical polishing processes are being investigated. Methods are being developed to abrade the copper film during the electroplating process, which results in thinner copper films in the field areas, and significantly improved planarity across the wafer surface [20, 21]. The copper overburden can then be removed by spin etching [22], electropolishing [23, 24], or a short CMP process [25]. Other processes are being explored where the wafer is positively charged in an electric field in the presence of a conductive polishing liquid, which significantly increases the removal rate and permits significantly lower downforces to be used during the CMP process [26, 27]. Barrier films can be removed by reactive ion etching, eliminating the need for a separate CMP step. Using any of these processes, dishing and erosion are minimized and delamination of fragile low-k dielectric films can be avoided. These methods also have the potential to significantly reduce the CMP process time in back end of the line interconnect processes, resulting in the need for fewer process tools.

5.10 Carrier Technology

The wafer carrier is one of the most critical components on the CMP tool. It transmits the downforce to the wafer face, and determines the pressure

distribution over the area of the wafer. It also determines the interaction of the wafer with the polishing pad. If the carrier is not properly designed, it will result in non-uniformities in removal rate, particularly at the wafer edge.

The earliest CMP carriers relied on mechanical means to transmit force to the wafer and to control the pressure distribution. A rigid, fixed plate surrounded by a retaining ring and a compressible insert were used for holding the wafer during the CMP sequence. The insert is a soft buffed or unbuffed poromoric material with a pressure sensitive adhesive on the backside that can be used to mount the insert on the wafer carrier. The poromoric mounting material contacts the work piece within the retaining ring. Capillary effects from water in the elongated pores in the mounting material coupled with surface tension of the water film under the work piece act to hold it in place. The template or retaining ring provides a rigid structure around the work piece that eliminates lateral movement during the polishing operation.

One of the earliest carrier designs is shown in Fig. 5.15 [28]. The carrier design was effective in providing front surface referencing of the wafer to the polishing pad, and it was demonstrated that the carrier design did not introduce taper into a film or wafer during polishing. The design resulted in a gimbal point that was significantly above the plane of the wafer face undergoing planarization. As a result, the lateral forces on the carrier from the friction between the polishing pad and the retaining ring/wafer surface created a moment about the gimbal point which caused the leading edge of the wafer to "dive" into the pad, resulting in a non-uniform pressure distribution across the wafer face. Non-uniformities from this carrier design were on the order of 6% 1σ, using a 10 mm edge exclusion zone. In an attempt to
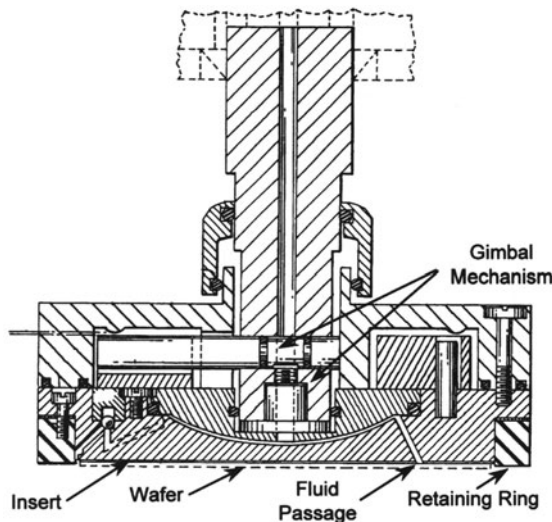
Fig. 5.15. Early mechanical design for a CMP carrier

compensate for the nonuniform removal rates, the carrier spindle was canted slightly toward the leading edge of the wafer, and a convex curvature of 1 to 5 µm was used on the rigid backing plate. The amount of curvature required was dependent upon the down force used during the CMP process, since mechanical deflections were present within the carrier, and the frictional forces generating the mechanical moment were dependent upon the relative velocity between the wafer and the polishing pad and the characteristics of the slurry used in the process.

The retaining ring was adjusted so that approximately one third of the wafer thickness protruded from the retaining ring. There was no capability for separately controlling the down force on the retaining ring. Because of the viscoleastic properties of the polishing pad and the tendency of the leading edge of the carrier to deflect downward into the pad (because of the mechanical moment about the gimbal point resulting from frictional forces), a significant non-uniformity resulted at the wafer edge that could not be eliminated.

The different generations of carrier design are shown in Fig. 5.16. One of the earliest improvements in carrier design was to use a rigid floating plate that permitted independent movement of the retaining ring. In this improved design, a rigid plate was coupled to the retaining ring using a rubber membrane so that the wafer face automatically adjusted to the same plane as the surface of the retaining ring that contacted the pad. As a result, the gimbal point for the carrier was located in the plane of the front surface of the wafer, eliminating the need for shaping an empirically derived curvature on the carrier face. Pneumatic pressure behind the plate resulted in improved pressure uniformity across the wafer. However, the pressure on the retaining ring was not independently controlled, and pad viscoelelastic properties still were a factor in determining within wafer removal rate uniformity. However, by increasing the width of the retaining ring, the viscoelastic effects could be mitigated [29].

The next improvement in carrier design permitted independent control of the force on the retaining ring and the force on the rigid plate used for mounting the wafer. This design improvement resulted in an additional degree of freedom in compensating for edge effects during the CMP process [30].

As carrier technology continued to evolve, a flexible membrane replaced the rigid plate behind the wafer. Pneumatic pressure behind the membrane then resulted in a uniform pressure distribution across the face of the wafer. These carrier designs continued to use independent control of the down force on the retaining ring [6, 31]. These changes resulted in significant reduction in edge effects; within wafer removal rate nonuniformity was reduced to less than 3% 1σ with a three mm edge exclusion zone.

The use of pneumatic pressure behind a flexible membrane and independent control of the down force on the retaining ring has been further extended. State of the art carriers now provide the capability to adjust the pressure in
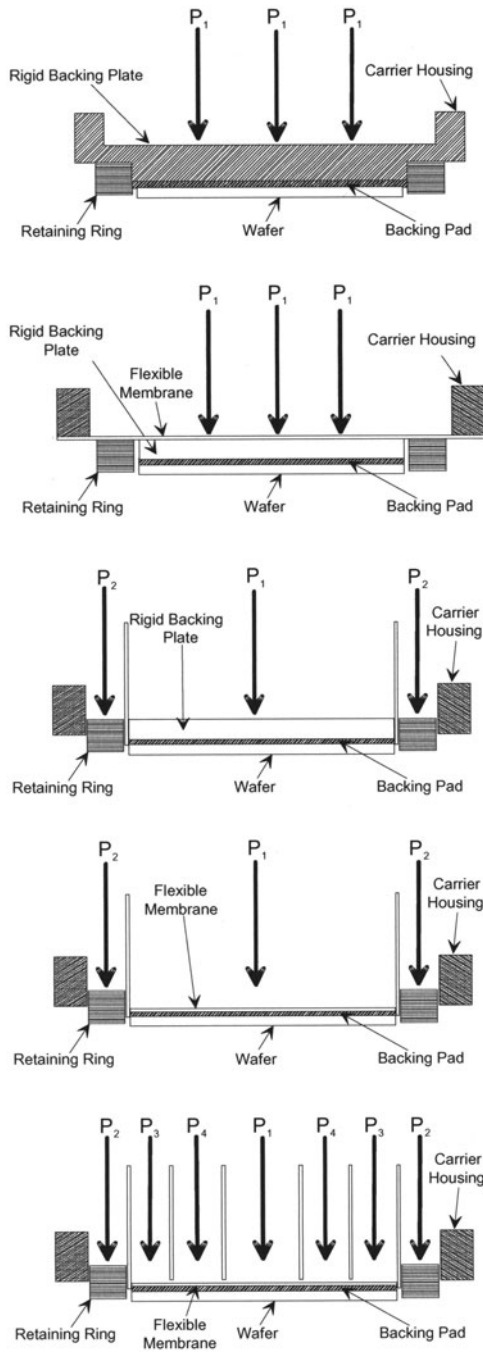
Fig. 5.16. Different CMP carrier design generations
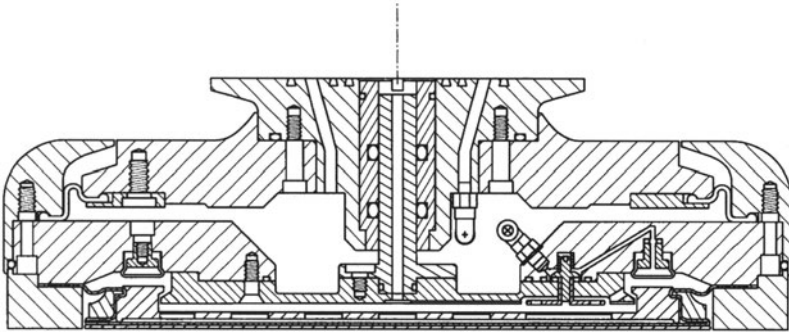
Fig. 5.17. CMP carrier design. From US patent 6,244,942; June 12, 2001

different zones across the wafer surface. With these features, it is now possible to decouple the pressure settings between zones, and to control the local removal rates. The carriers now can be used to compensate for incoming film thickness variations, resulting in a uniform film thickness across the wafer after the CMP process. When coupled with suitable in-situ metrology, automated, closed loop control of the CMP process is possible [32].

While Fig. 5.16 shows the basic design concepts used for CMP carriers as the technology has evolved, the actual carriers based on these designs are complex, and use a large number of components. Materials of construction are important, since fatigue can cause premature failure leading to a short carrier life before replacement or maintenance is required. The carrier assembly can become quite complex, increasing costs significantly. Figure 5.17 shows one of the least complex carriers used in a production CMP process [33].

5.11 Pad Conditioning

During the early development of CMP processes, stabilization of removal rates using cast polyurethane pads was a critical issue as shown in Fig. 5.18. The Rodel IC1000 series polishing pads were found to provide the most effective planarization characteristics compared to other commercially available pads. However, the removal rate for PECVD oxide films on the as-received pads was initially low, increasing to a maximum value within a few minutes of polishing, followed by an exponential decay in removal rate with increased polishing time. Various pad conditioning methods were used in an attempt to stabilize the removal rate. Abrasive pad conditioning using diamond-impregnated disks was found to be the most effective method. If the pad was conditioned under the proper conditions before the first wafer was polished, the initial removal rate doubled compared to removal rates on unconditioned pads. If pad conditioning was repeated after each wafer was polished, the removal rate remained stable for hundreds of wafers [34].
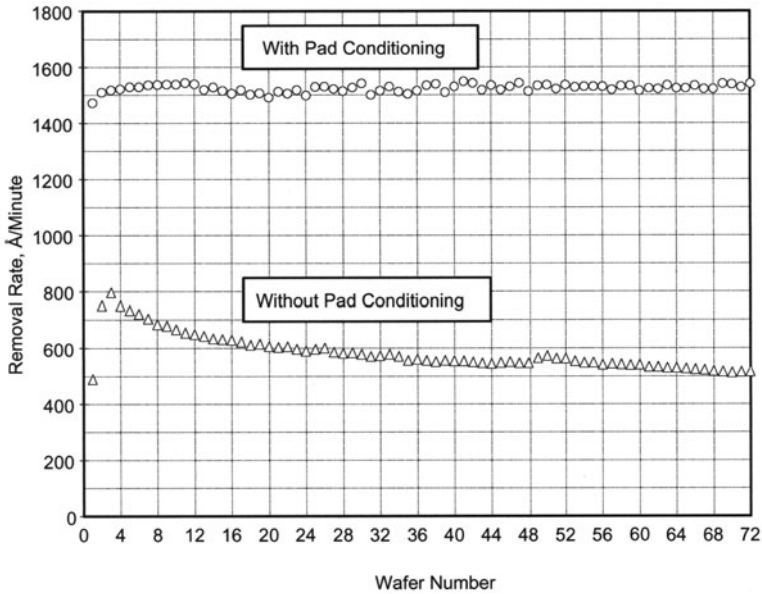
Fig. 5.18. Effect of pad conditioning on wafer to wafer removal rate

Subsequent development has shown that abrasive pad conditioning gener-
ates asperities on the pad surface, which significantly impact the removal
rate. As long as the asperities are controlled to a consistent height and den-
sity, the removal rate remains stable as additional wafers are processed [35].
Subsequently, the pad conditioning sub-system is one of the most important
modules on the CMP tool. Ideally, the pad conditioner and tool geometries
will permit in-situ pad conditioning while wafers are being processed.

The pad conditioning process can be modeled using rotary kinematics as
discussed earlier in this chapter. During pad conditioning, the pad becomes
the work piece, which is significantly larger in diameter than the end effec-
tor (see Fig. 5.1). Typical end effector diameters are on the order of 100 mm
compared to a polishing pad diameter greater than 500 mm. The end effector
rotates about its center during pad conditioning. It also is moved linearly
toward and away from the polishing pad center in order to condition the
working area of the pad. Because different regions of the polishing pad have
different utilization rates, and because the relative velocity between the end
effector and the polishing pad varies with radial position, it is common prac-
tice to segment the polishing pad into different radial zones. The dwell time
of the end effector is adjusted in each zone in order to control the amount of
pad conditioning and to maintain the thickness profile for the polishing pad.
The pad conditioning process has been modeled extensively using the rotary
kinematic equations [36].

Pad conditioning has a significant impact on the cost per wafer pass during CMP processing. An extensive supplier base for supplying diamond disk end effectors has evolved. Diamond conditioning disk costs are significant, and are only exceeded by slurry and pad costs. If the pad is under-conditioned, the process will not remain stable leading to extensive and expensive rework or scrapped wafers. If the pad conditioning is too aggressive, the life of the polishing pad is shortened, leading to higher costs and greater tool downtime as the pad is changed out and a new pad is qualified.

Initially, pad conditioning end effectors were manufactured by embedding diamond abrasive in an electroplated nickel matrix on a steel substrate. However, because of inconsistencies from disk to disk, it was found that the diamond shape and relative abrasiveness, as well as particle sizing were important for improving end effector performance. The earliest end effector manufacturing methods resulted in random orientation and placement of the diamond particles in the matrix. The nickel-plated method was satisfactory for manufacturing conditioning disks for oxide processes. For metal CMP processing with very low pH slurries, the nickel was chemically attacked, leading to short conditioner life and loss of diamond particles which resulted in high defectivity. Today, diamond end effectors are provided with protective coatings to increase their chemical resistance, and alternative methods such as
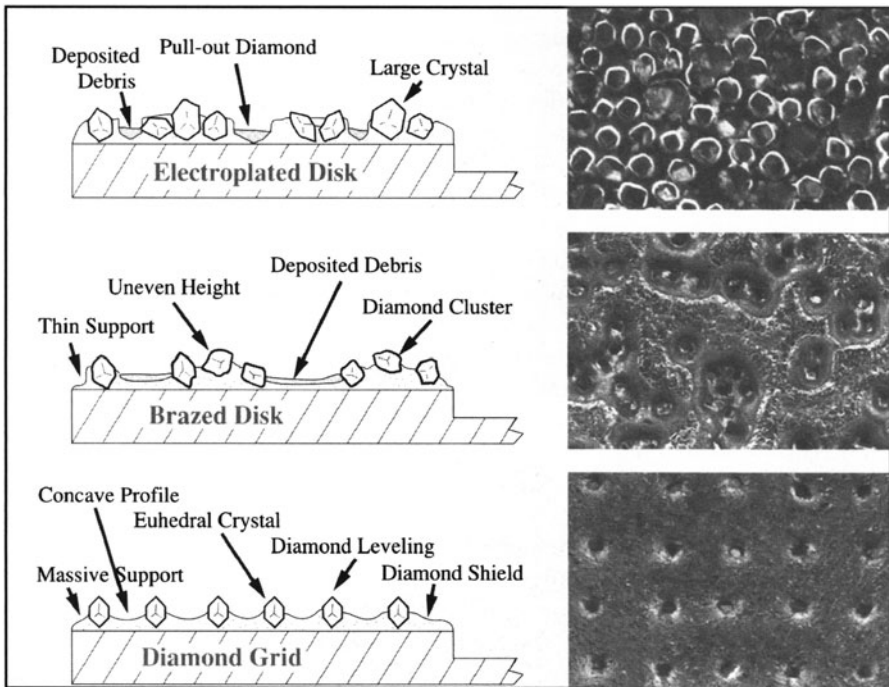
Fig. 5.19. Properties of diamond pad conditioning end effectors

brazing are used which result in a more chemically inert matrix. The diamond particles are carefully controlled with respect to orientation, sharpness of cutting edges, protrusion, placement in a controlled pattern, and adequate build-up of the bonding matrix on the side of the particle. Figure 5.19 shows a diagram of a pad conditioning disk for use in oxide processes provided by Kinik, a Taiwanese supplier of diamond pad conditioning end effectors.

CVD diamond coatings have been used to increase the conditioner disk lifetime, but the cost of these disks is significantly higher and the cutting characteristics are generally degraded. Diamond particles have also been embedded in a ceramic matrix for use as pad conditioning disks [37].

Today, a wide variety of diamond end effector disks are commercially available with a global supplier base. Disks in a wide range of diameters are available, and end effectors range from small pellets to disks to segmented wheels. Different matrix materials are used, and the pattern of the diamond particles can be controlled to individual specifications. As a result, the suppliers of CMP tools specify conditioning disks as part of their process of record; the tool suppliers also differentiate their tools based on pad conditioning capability.

5.12 Endpointing

Within the last few years, significant advances have been realized in endpointing technology. Initial oxide CMP processes relied on process time and control of removal rates in order to achieve target thicknesses. The CMP tool and process results in a hostile environment for in-situ metrology. It therefore is not surprising that the earliest endpointing techniques relied upon indirect methods for detecting changes in the CMP process.

The earliest technique used commercially for endpointing relied upon monitoring motor current changes [38] while driving the carrier or platen rotation. This technique is an indirect method of measuring changes in frictional forces between the pad and the wafer surface. Tungsten CMP processes on softer polishing pads generally have higher frictional forces than oxide films using the same consumables set. As a result, motor current monitoring has been effectively used for determining when tungsten overburden has been removed, exposing the oxide dielectric in a tungsten plug damascene process. Figure 5.20 shows a typical endpointing trace for a tungsten CMP process [39].

While attempts have been made to extend the motor current method for use in STI processes and copper dual damascene processes, the technique has been only marginally succesful because the frictional forces from the exposed films and the polishing pad do not differ significantly. As a result, other methods have been developed for use with these processes. More recently, direct monitoring of changes in the coefficient of friction has been proposed as an endpointing means [40].
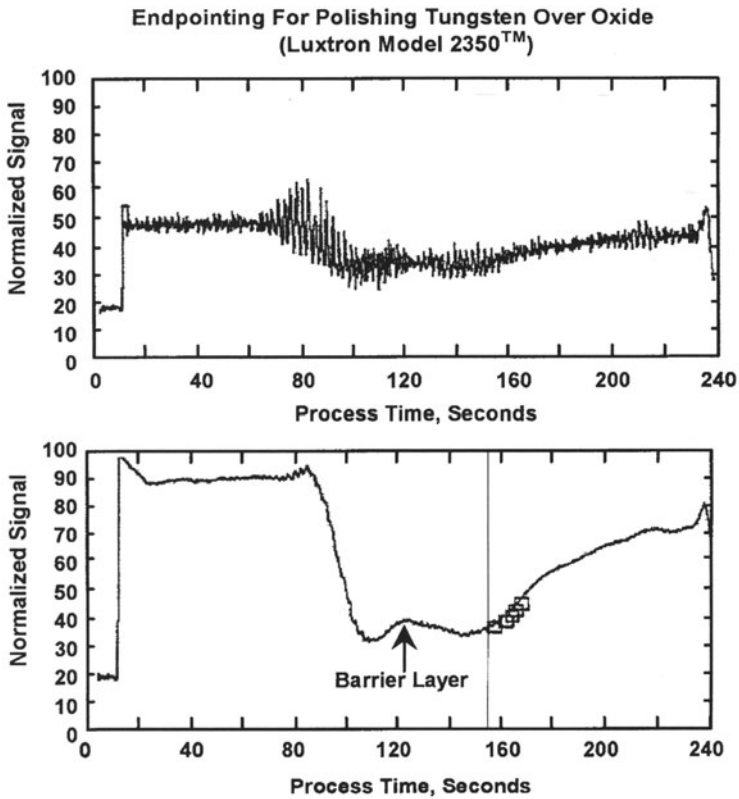
Fig. 5.20. Motor current endpointing traces. From [39]

A vibration monitoring method, which detects changes in frictional forces as film interfaces interact with the polishing pad, has been developed and is being used for CMP endpointing in high volume production [41]. Vibrations generated during the CMP process are detected using an accelerometer on the carrier behind the wafer. The method is reported to be sufficiently sensitive for production use in dielectric, STI and metal CMP processes. The method can detect delamination of low-k dielectric films in copper dual damascene processes. An endpointing method based on monitoring high frequency acoustic emissions generated during CMP processing has also been proposed [42].

A highly sensitive endpointing detection method is being used in production for shallow trench isolation processes. As the trench fill oxide film is removed from the field areas, trace amounts of ammonia are generated when the nitride film is exposed and polished. The ammonia is detected using chemiluminescence after it is catalytically converted to nitric oxide. The

technique can also be used for endpointing any other film stack containing a nitride layer [44].

Optical reflectrometry is the preferred method for endpointing CMP processes. However, the method is highly intrusive, requiring process compromises or extensive hardware modifications. While it is effective in detecting interface changes in thin film stacks, it has been less effective for in-situ monitoring of the thickness of dielectric layers in multilevel interconnect structures.

Because of the difficulty in using in-situ optical techniques for monitoring dielectric film thicknesses, in-line optical metrology has been used extensively. Wafers can be measured prior to processing, and again after processing is completed while the next wafer is undergoing CMP. The data can then be used for implementing changes in the process using run-to-run control methods for compensating for device pattern dependent and equipment induced disturbances. Nova Instruments was the first to develop an in-line thickness measuring system that was successfully integrated with CMP polishing equipment for oxide processes [45]. While it does not serve as an in-situ endpoint monitor, it can measure thin film thickness at multiple locations on the wafer surface after it is unloaded from the wafer carrier. The system has pattern recognition capability, and can complete thickness measurements at multiple pre-determined sites on a wafer while the next wafer is being polished. The method is particularly useful for monitoring removal rates and within wafer uniformity and is able to provide measurements on 100% of the wafers prior to and following CMP on even the highest throughput CMP tools. The wafer can be maintained in a wet environment during the measurement process. The spot size for the optical measurement is in the micron range. Data generated from the system can be integrated with the polisher control system to provide closed loop process control using run-to-run controller concepts [46]. Data is fed forward and backward from the CMP module in order to optimize the entire manufacturing module performance.

In-situ optical endpointing methods are being used on production CMP tools, but such systems are highly intrusive when used on rotary platforms and require placement of an illumination source and detector within the polishing platen as shown in Fig. 5.21 [47]. In order to transmit and receive data from the endpointing module, either electrical and optical slip rings must be used, or wireless data transmission methods must be employed. Optical techniques can be used to measure the thickness of both dielectric films, and very thin metal films ($< 1000\,\text{Å}$), or can be used to detect a change in the film interface properties when material is completely removed such as clearing oxide in a shallow trench isolation process, or in removing copper from the field areas in a metal damascene process. Figure 5.22 shows the Applied Materials Mirra$^{\text{TM}}$ CMP system with an transparent optical window in the polishing pad for providing access to the wafer surface.
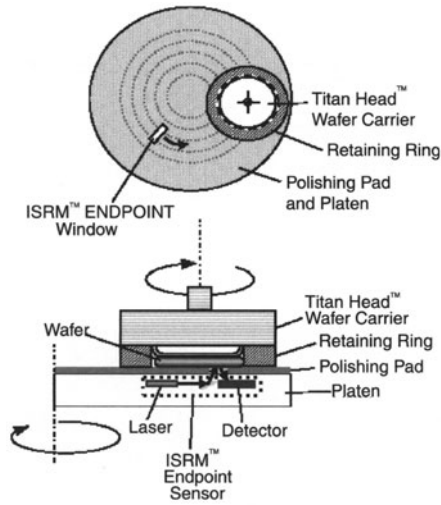
Fig. 5.21. Applied materials ISRM$^{\text{TM}}$ optical endpointing system. From [47]
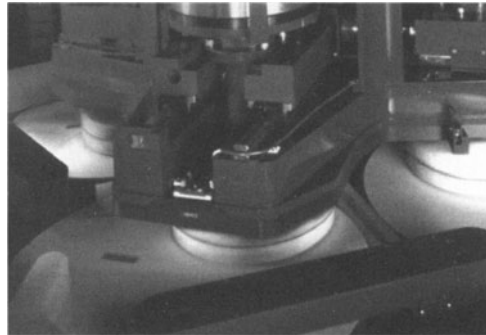
Fig. 5.22. Applied materials ISRM$^{\text{TM}}$ in operation. Note window in pad

Optical endpointing systems on orbital platforms are less intrusive, since flexible cables or optical fibers can be used to transmit and receive signals through the orbiting polishing pad. It is also possible to use multiple fibers so that the process can be monitored in different radial zones. When used with slow rotation of the wafer carrier, and with a sufficient number of probes, the entire surface of the wafer can be monitored [48].

Copper CMP processes present special challenges. The thickness of the incoming copper film is generally non-unifirm as a result of pattern-sensitive deposition rates during electroplating. Since dishing and erosion are exremely sensitive to the downforce and degree of over-polishing, it is desirable to use multiple process steps. The first CMP step rapidly reduces the thickness of

the copper film to a predetermined target value. When the thickness of the copper is reduced to about 1500 Å in the field areas, and prior to clearing of the film anywhere on the wafer, a low down-force, low removal rate process is initiated until clearing is complete and the barrier layer is exposed. A third process step is then used to remove the barrier layer in the field areas.

Endpointing technology for copper CMP processes generally employs two different methods. Eddy current measurements are used to monitor the copper thickness and thickness profile during the first CMP process step. The system must have extremely fast computational capability so it can measure
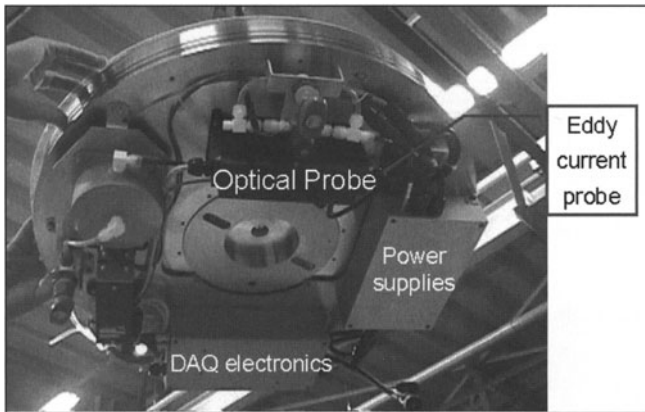
Fig. 5.23. KLA-Tencor Precice$^{\text{TM}}$ endpointing system
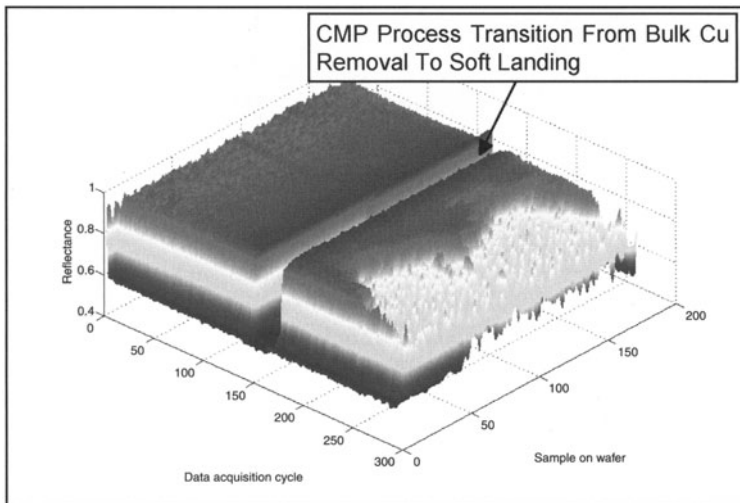
Fig. 5.24. Measurement results from KLA-Tencor Precice$^{\text{TM}}$ endpointing system

the film thickness uniformity as the sensor passes under the wafer, allowing the potential for adaptive process control. When the target copper thickness is obtained, the system switches to the optical monitoring mode where it can detect the start and completion of copper clearing, and the start and end of the barrier layer/adhesion layer clearing. Optical methods have the capability to detect copper and barrier layer residuals, permitting over-polish times to be minimized [49]. The optical system shown in Fig. 5.23 uses a single wavelength, multiple angle system, which overcomes problems with reflected signal quality resulting from scattering because of abrasive particles in the slurry. The system is highly intrusive, and requires extensive modifications of the CMP tool. Figure 5.23 shows components in the KLA-Tencor Precice system that are mounted on the underside of the polishing platen; probes are also embedded in the platen that can access the wafer surface. Measurement results are shown in Fig. 5.24.

5.13 Summary

Significant progress has been made in CMP tool technology since the first CMP processes were developed. CMP technology has moved from an empirically-based "black art" to stable, predictive processes that are used for both front end of the line and back end of the line processing; CMP is a key enabler for fabricating both leading edge transistors, and multilevel interconnect structures. Improvements in consumables technology, pad conditioning, post-CMP cleaning and endpointing technologies have resulted in the evolution of sophisticated CMP tools with the near-term potential to perform complex, fully automated processes. As the semiconductor industry transitions to new device structures, innovations in CMP tool design will continue to advance at a rapid pace.

References

1. H. Hocheng, Y.-L. Huang and L.-J. Chen, Journal Of The Electrochemical Society, **146** (11), 4236, November 1999.
2. D. White, D. Boning and A. Gower, Proceedings 2000 CMP–MIC Conference, 229, IMIC, Tampa, 2000.
3. W. Patrick, W. Guthrie, C. Standley and P. Schiable, Journal Of The Electrochemical Society, **138** (6), 1778, June 1991.
4. H. Hocheng and H.-Y. Tsai, Proceedings 1997 CMP–MIC Conference, 277, IMIC, Tampa, 1997.
5. R. Kolenkow and R. Nagahara, *Solid State Technology*, **35** (6), 112, June 1992.
6. J. Breivogel, S. Louke, M. Oliver, L. Yau and C. Barns, US Patent 5,554,064, September 10, 1996.
7. M. Tuttle, T. Doan, A. Fox, G. Sandhu and H. Stroupe, US Patent 5,232,875, August 3, 1993.

8. K.H. Lee, Y.B. Lee and S.W. Kang, US Patent 6,315,641, November 13, 2001.
9. N. Shendon, and D. Smith, US Patent 5,582,534, December 10, 1996.
10. K. Honda, US Patent 6,042,459, March 28, 2000.
11. H. Talieh, and D. Weldon, H., US Patent 5,692,947, December 2, 1997.
12. R. Jairath, S. Chadda, E. Engdahl, W. Krusell, T. Mallon, K. Mishra, A. Pant and B. Withers, Proceedings 1997 CMP–MIC Conference, 194, IMIC, Tampa, 1997.
13. A. Pant, D. Young, A. Meyer, K. Volodarsky and D. Weldon, US Patent 5,800,248, September 1, 1998.
14. H. Talieh, US Patent 6,207,572, March 27, 2001.
15. A. Yoshio, CMP Technology For ULSI Interconnection, Pp.K1 – K10, ISBN 1-892568-50-0, SEMICON West, San Francisco, July, 2000.
16. Y. Saimitsu, *Chemical Mechanical Planarization In IC Device Manufacturing III*, Electrochemical Society Proceedings, **99–37**, 546, Honolulu, Hawaii, October 1999.
17. T. Donohue, R. Williams, J. Barber, J. Hoshizaki, L. Lee, C.-L. Meng and P. Sommer, US Patent 6,312,319, November 6, 2001.
18. J. Hoshizaki, R. Williams, J. Buhler, C. Reichel, W. Hollywood, R. de Geus and L. Lee, US Patent 5,759,918, June 2, 1998.
19. Y. Homma, S. Kondo, N. Sakuma, K. Hinode, J. Noguchi, N. Ohashi, H. Yamaguchi and N. Owada, *Chemical Mechanical Planarization In IC Device Manufacturing III*, Electrochemical Society Proceedings, **99–37**, 83, Honolulu, Hawaii, October 1999.
20. H. Talieh, US Patent 6,176,922, January 23, 2001.
21. H. Talieh and C. Uzoh, US Patent 6,328,872, December 11, 2001.
22. R. Contolini, S. Mayer and L. Tarte, US Patent 5,486,234, January 23, 1996.
23. S. Mayer, R. Contolini and A. Bernhardt, US Patent 5,096,550, March 17, 1992.
24. H. Wang, US Patent 6,248,222, June 19, 2001.
25. M.H. Tsai, S.W. Chou, C.L. Chang, C.H. Hsieh, M.W. Lin, C.M. We, W.S. Shue, D.C. Yu and M.S. Liang, Tech. Dig. 2001 IEDM, 80, Washington, DC, December 2001.
26. C. Uzoh and J. Harper, US Patent 5,807,165, September 15, 1998.
27. S. Sato, Z. Yasuda, M. Ishihara, N. Komai, H. Ohtorii, A. Yoshio, Y. Segawa, H. Horikoshi, Y. Ohoka, K. Tai, S. Takahashi and T. Nogami, Tech. Dig. 2001 IEDM, 84, Washington, DC, December 2001.
28. G. Gill and T. Hyde, US Patent 4,944,119, July 31, 1990.
29. N. Shendon, K.C. Struven and R. Kolenkow, US Patent 5,205,082, April 27, 1993.
30. H. Kobayashi, H. Miyairi and O. Endo, US Patent 5,584,751, December 17, 1996.
31. N. Shendon, US Patent 5,624,299, April 29, 1997.
32. J. Schlueter, S. Schultz, N. Korovin and F. Elkhodr, CAMP 6th International CMP Symposium, Lake Placid, New York, August, 2001.
33. Steven Zuniga,US Patent 6,244,942; June 12, 2001.
34. Tom Hyde, Westech Systems, Private Communication.
35. M. Oliver, R. Schmidt and M. Robinson, *Chemical Mechanical Planarization In IC Device Manufacturing IV*, Electrochemical Society Proceedings, **2000–26**, 77, Fourth International Symposium On Chemical Mechanical Planarization In Integrated Circuit Manufacturing, Fall Meeting Of The Electrochemical Society, Phoenix, Arizona, October, 2000.

36. C.-Y. Chen, C.-C. Yu, S.-H. Shen and M. Ho, *Journal Of The Electrochemical Society*, **147** (10), 3922, October, 2000.
37. Noritake, Private Communication.
38. G. Sandhu, L. Schultz and T. Doan, US Patent 5,036,015, July 30, 1991.
39. Luxtron Corporation, Private Communication.
40. N. Gitis, M. Vinogradov and C. Gao, "Quantitative Evaluation Of CMP Processes And Materials Using A CMP Tester With Multiple Sensors", *Proceedings Of The Second International Conference On Microelectronics And Interfaces*, American Vacuum Society, Santa Clara, California, February, 2001.
41. A. Fukuroda, K. Nakamura and Y. Arimoto, "In Situ CMP Monitoring Technique For Multi-layer Interconnection", *Technical Digest 1995 IEDM*, Pp. 469–472, Washington, DC, December 1995.
42. T. Kojima, M. Miyajima, F. Akaboshi, T. Yogo, S. Ishimoto and A. Okuda, IEEE Transactions On Semiconductor Manufacturing, **13** (3), 291, August, 2000.
43. J. Tang, C. Unger, Y. Moon and D. Dornfeld, "Low-k Dielectric Material Chemical Mechanical Polishing (CMP) Process Monitoring Using Acoustic Emission", Abstract N3.7, Abstracts Of The Spring 1997 Meeting Of The Materials Research Society, San Francisco, California, March 31 – April 4, 1997.
44. L. Li, C. Wei, J. Gilhooly and C. Morgan, "End Point Detection In Metal And Nitride-Containing CMP Processes", *Proceedings Of The Second International Conference On Microelectronics And Interfaces*, American Vacuum Society, Santa Clara, California, February, 2001.
45. G. Dishon, D. Eylon, M. Finarov and A. Shulman, Proceedings 1998 CMP–MIC Conference, 267, IMIC, Tampa, 1998.
46. N. Patel, G. Miller, C. Guinn and S. Jenkins, "Device Dependent Control Of Chemical-Mechanical Polishing Of Dielectric Films", IEEE Transactions On Semiconductor Manufacturing, **13** (3), 33, August, 2000.
47. B. Adams, B. Swedek, R. Bajaj, K. Wijekoon, S. Nanjangud, A. Wiswesser, S. Tsai, D. Chan, F. Redeker and M. Birang, Proceedings Of 2000 CMP–MIC Conference, 267, IMIC, Tampa, 2000.
48. T. Laursen and M. Grief, *Chemical-Mechanical Polishing 2001 – Advances And Future Challenges*, Materials Research Society Symposium Proceedings, **671**, Pp. M7.6.1–6, San Francisco, California, April, 2001.
49. V. Bhaskaran, C. Chen, R. Allen, K. Lehman, H. Chen, D. Watts and B. Stephenson, "Advanced In-SituEnd-Point Control System For CMP Applications", CMP Users Group Meeting, American Vacuum Society, Santa Clara, California, October 2001.

6 CMP Polishing Pads

David B. James

6.1 Introduction

Although polishing is an old technology [1, 2, 3], it has become an enabling process for the manufacture of leading-edge semiconductor devices. In the manufacture of such devices, polishing is used to maintain planarity at each step in the process of depositing and photolithographically imaging sequential insulating dielectric and conductive metal layers. Also as noted in the first chapter, CMP is now also employed to remove a bulk film and then stop, as in damascene copper and tungsten polishing. As semiconductor devices become increasingly complex, requiring finer feature geometries and more metallization layers, greater demands are placed on the polishing consumables used in the CMP process to manufacture such devices [4, 5, 6].

Polishing consumables usually comprise a polymeric polishing pad used in conjunction with an aqueous based polishing slurry [7]. Conventionally, the slurry contains abrasive particles but, also recently, two variants on the standard consumables have been studied. In the first new approach, the abrasive particles have been incorporated into the pad and in the second, the pad is used with a particle-free reactive liquid. This chapter will focus on conventional polishing pads and, for completeness, will include only a brief section on slurryless pad technology. More specifically, this chapter will discuss polymer criteria for polishing pads, types of pads available and their manufacture, control of pad properties, and the relationship of those properties to polishing performance.

6.2 Polymer Requirements for Polishing Pads

Since polishing is both a mechanical and a chemical process, the polymeric polishing pad must have sufficient mechanical integrity and chemical resistance to survive the rigors of polishing. As discussed later in this chapter, important mechanical properties include high strength to resist tearing during polishing, acceptable levels of hardness and modulus selected based on the material being polished, and good abrasion resistance to prevent excessive pad wear during polishing.

Chemically, the pad must be able to survive the aggressive slurry chemistries used in CMP polishing without degrading, delaminating, blistering or warping. Slurry chemistries include either highly alkaline slurries near pH 11 used for polishing of inter-layer dielectric oxide layers or highly acidic, oxidizing slurries used for polishing metal films, such as tungsten and copper. Slurries for metal CMP can have pH values less than 2 and contain oxidizing agents such as hydrogen peroxide, ferric nitrate or potassium iodate. Typically, over the lifetime of a single pad, it is exposed to these chemistries for many hours and often several days, usually at temperatures above ambient.

A third criterion is that the pads must be sufficiently hydrophilic. The aqueous-based abrasive containing slurry or the particle-free reactive liquid must wet the surface of the pad and form a liquid film between the wafer and the pad. If the liquid does not wet but instead beads on the pad surface, it will be swept away by the wafer edge and the interior of the wafer will be starved of the necessary chemistry to enable effective polishing.

Pad hydrophilicity may be expressed in terms of the pad's Critical Surface Tension. This is defined as the wettability of a solid surface by noting the lowest surface tension a liquid can have and still exhibit a contact angle greater than zero degrees on that solid. In terms of polishing, this means that pads made from polymers with higher critical surface tension values are more hydrophilic and slurry will more readily wet these pads.

Table 6.1 shows critical surface tension values for several commercially available polymers [8].

From the Table 6.1, polymers with higher critical surface tension values correspond to those polymers which are useful as polishing pads. The minimum value is around 37 mN/m, and preferred values are in the mid 40's. This range includes polymers such as poly(methyl methacrylate), polycarbonates, nylons, polysulfones, and polyurethanes.

As the critical surface tension of the pad decreases, polishing performance will also decrease. Thus one would expect polyethylene and PTFE pads to perform poorly because slurry does not spread over the pad surface. However, a number of approaches have been used to enable less hydrophilic polymers to be used as polishing pads. These include adding high levels of wetting agents to the slurry, adding hydrophilic fillers such as silica to normally hydrophobic polymers, or by chemically modifying the hydrophobic polymer to make it more hydrophilic. Techniques used for the latter include plasma treatment, corona discharge or the chemical addition of polar groups to the pad surface. However, since pads are continually abraded during use, either by conditioning prior to polishing or by the polishing process itself, the treatment needs to be effective through the whole cross-section of the pad and not just the surface layer.

Since polishing is a wet process, the mechanical properties of the pad must be essentially retained, even when the pad is wet. However, the properties of hydrophilic polymers which are preferred for polishing pads will change

Table 6.1. Critical Surface Tension Values of Polymers

Polymer	Critical Surface Tension (mN/m)
Polytetrafluoroethylene	19
Polydimethylsiloxane	24
Silicone Rubber	24
Polybutadiene	31
Polyethylene	31
Polystyrene	33
Polypropylene	34
Polyacrylamide	35 – 40
Polyvinyl alcohol	37
Polymethyl methacrylate	39
Polyvinyl chloride	39
Polysulfone	41
Nylon 6	42
Polycarbonate	45
Polyurethane	45

during exposure to aqueous polishing slurries. Typically water acts as a plasticizer which decreases pad modulus and hardness, and increases ductility. Thus under polishing conditions, the pads will become more ductile and flexible. These changes are reversible and pad properties revert to the original values on drying, indicating that chemical attack of the pad is minimal.

The changes in pad properties during polishing are important for those researchers interested in modeling the mechanisms of CMP polishing. Often dry properties rather than the more appropriate wet properties are used in the models. Secondly, most properties measured are bulk properties. Since polishing is an interfacial process, in many cases the properties of the pad surface rather than the bulk should be used. Surface properties change rapidly during immersion in slurry and quickly reach equilibrium values [61].

The final criterion for the polymer is that the polymer formulation and morphology can be varied to give pads with specific, predictable properties for different polishing applications. Thus a family of pads is preferred, such that performance may be fine-tuned for the specific polishing application, polishing tool, wafers and slurry.

The types of polymers which best satisfy the criteria discussed above are polyurethanes. Polyurethanes combine good mechanical properties with excellent chemical stability and, as shown in the next section, their properties may be readily and precisely controlled. Furthermore, with polyurethane

technology it is possible to fabricate a wide range of pad microstructures including foams, impregnated felts and solid pads, and to use a variety of polyurethane manufacturing processes including casting, molding, extrusion, web-coating and sintering. These will be discussed in more detail later in the chapter.

6.3 Basics of Polyurethanes

Many books have been written covering the chemistry, morphology, properties and manufacture of polyurethanes, and the reader is referred to these [9, 10, 11] for in-depth discussions. This section will provide a brief overview of polyurethanes and cover those aspects most relevant to polishing pad technology.

6.3.1 Formulations

Polyurethanes are typically composed of at least three components:

1. Long chain polyol
2. Diiocyanate or isocyanate of higher functionality
3. Chain Extender.

Most polyols used in the manufacture of polyurethanes are polyethers with terminal hydroxyl groups. Hydroxyl-terminated polyesters are sometimes used to obtain polyurethanes with special properties. However, polyester diols tend to be more expensive and are less chemically stable than polyether diols, especially in basic solution. The choice of polyol, especially the size (molecular weight), flexibility of its molecular structure, and functionality, has a large effect on the properties of the resultant polyurethane.

A second method of varying the properties of a polyurethane is through the selection of the isocyanate. Several aromatic and aliphatic isocyanates are available, but 95% of all polyurethanes are based on either toluene diisocyanate (TDI) or diisocyanato-diphenylmethane (MDI) and its derivatives. MDI is often preferred over TDI because of its lower toxicity and greater chemical flexibility. Although pure MDI is a difunctional solid having two reactive isocyanate groups per molecule, producers of isocyanates have developed liquid MDI variants of higher functionality than two. This enables control of crosslinking reactions and hence of the resultant polyurethane properties.

The third important component in a polyurethane formulation are small molecules known as chain extenders. These are chemicals containing groups which can react with isocyanates and link such isocyanates together to introduce specialized polymer segments into the polyurethane backbone. Examples include molecules such as low molecular diols (e.g. ethylene glycol or butane diol), diamines, and water.

Fundamentally, there are three approaches to formulating polyurethanes, described by the urethane industry as "prepolymer", "quasi", and "one-shot". In the "one-shot" approach all the components are mixed together and reacted at one time. This can result in a highly exothermic reaction which is more difficult to control and can lead to reproducibility problems. In the "prepolymer" approach, the isocyanate is pre-reacted with the long chain diol to form a high molecular weight isocyanate terminated moiety. This can then be further reacted with diol or diamine curatives to complete the polyurethane formation. The advantages of this approach are greater control over the chemistry and a more consistent product. The "quasi" approach is intermediate between the two other approaches.

Additional ingredients commonly included in commercial polyurethane formulations for polishing pads are catalysts, fillers and blowing agents.

6.3.2 Chemistry

Polyurethanes are addition polymers [12] formed by reaction of di- or polyfunctional isocyanates with polyols:

$$Polyol + Isocyanate = Polyurethane.$$

However, urethane reactions are actually much more complex. This is a consequence of the high reactivity of isocyanates which can react with any molecules present containing active hydrogen groups, and by the fact that it is usual to use a stoichiometric excess of isocyanate with respect to diol. Thus, under suitable conditions, many secondary reactions are possible, such as:

$$Isocyanate + Urethane = Allophanate$$

$$2 \times Isocyanate = Uretidinedione\ (Dimerization)$$

$$3 \times Isocyanate = Isocyanurate\ (Trimerization)$$

$$Isocyanate + Water = Carbon\ Dioxide.$$
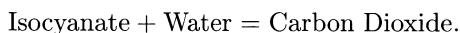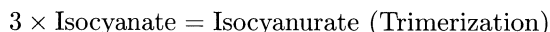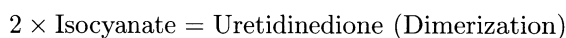
These reactions can be preferentially controlled through the reaction conditions. Thus one reaction over another can be favored through the use of catalysts and by controlling the reaction temperature. The first three of the above reactions create chemical cross-links between chains and a network structure. As will be illustrated later, as a consequence of the polyurethane cross-linking reactions, pad properties can be fine-tuned through control of the stoichiometric ratio of isocyanate to diol.

6.3.3 Morphology

Polyurethanes are multi-phase materials with complex morphologies [9, 10, 11]. Their molecular structures vary from rigid cross-linked polymers to lin-

ear, highly extensible elastomers. A common feature, however, of polyur-
ethanes is the presence of so-called "soft" and "hard" segments. The type
and relative amount of these in the polyurethane structure are major deter-
minants in controlling polyurethane properties.

Soft segments are quite mobile and are normally present in coiled forma-
tion. Chemically, they comprise the high molecular weight long chain diol
component of the formulation. The mobility of molecular chains in the soft
segment results in increased flexibility, toughness and impact resistance. Mo-
bility depends on the chemical nature and chain length of the soft segment.
Ideally, the soft segment should be amorphous and have a low glass tran-
sition temperature. Phase separation increases with increasing chain length
and decreasing polarity of the soft segment due to less hard segment/soft
segment interaction. Preferred molecular weights are in the 1000 to 4000
range. At higher molecular weights, especially at low hard segment amounts,
there is a tendency for the soft segments to crystallize which will reduce the
elastomeric benefits conferred by the soft segments.

Soft segments alternate with hard segments which are stiff oliourethane
units, principally composed of reacted isocyanate and chain extender moi-
eties. Hard segments act as pseudo cross-links and control the dimensional
thermal stability of polyurethanes. Thus properties such as strength and stiff-
ness at elevated temperatures are controlled by the hard segments. Above
a certain temperature, the hard segments "melt" and, in the absence of
chemical cross-links, the polyurethane becomes thermoplastic with greatly
reduced strength and stiffness.

6.4 Types of Commercially Available Polishing Pads and Their Manufacture

6.4.1 Types of Pads

This section will cover the types of pad that are commercially available and
currently used for CMP polishing. Following the classification of Cook [13],
the pads may be categorized into four types differentiated by their microstruc-
ture. The types are:

Type 1: Polymer Impregnated Felts
Type 2: Poromerics (synthetic leathers)
Type 3: Filled Polymer Sheets
Type 4: Unfilled Textured Polymer Sheets.

Table 6.2 summarizes the key features, properties, commercial trade
names, and typical applications for the different pad types.

Table 6.2. Key Features, Properties and Applications for Different Pad Types

	Type 1	Type 2	Type 3	Type 4
Structure	Felted fibers impregnated with polymeric binder	Porous film coated on a supporting substrate	Microporous polymer sheet	Non-porous polymer sheet with surface macrotexture
Microstructure	Continuous channels between fibers	Vertically oriented, open pores	Closed cell foam	None
Slurry loading capacity	Medium	High	Low	Minimal
Pad Examples	Pellon^{TM}, Suba^{TM}	Politex^{TM}, Surfin^{TM}, UR100^{TM}, WWP3000^{TM}	IC1000^{TM}, IC1010^{TM}, IC1400^{TM}, FX9^{TM}, MH^{TM}	OXP3000^{TM}, IC2000^{TM}
Compressibility	Medium	High	Low	Very Low
Stiffness	Medium	Low	High	Very High
Hardness	Medium	Low	High	Very High
Typical Applications	Si stock polish, Tungsten CMP	Si final polish, Tungsten CMP, post-CMP buff	Si stock, ILD CMP, STI, metal damascene CMP	ILD CMP, STI, metal dual damascene
Key US Patents	4,728,552 4,927,432	3,100,721 3,763,054 4,841,680 6,099,954	5,578,362 5,900,164	5,489,233 6,022,268

Note: Suba^{TM}, Politex^{TM}, UR100^{TM}, WWP3000^{TM}, IC1000^{TM}, IC1010^{TM}, IC1400^{TM}, MH^{TM}, OXP3000^{TM}, and IC2000^{TM} are trade-names of Rodel Inc., Pellon^{TM} and FX9^{TM} of Freudenburg, and Surfin^{TM} of Fujimi.

6.4.2 Methods of Manufacture

Although all the four pad types are polyurethane based, each is manufactured by a different process, illustrating the versatility of urethane chemistry. Table 6.3 provides a schematic outline of the manufacturing process for the different pad types. Types 1 and 2 pads are manufactured by a continuous roll or web process, Type 3 pads by a batch process, and Type 4 pads by either a batch process or by a unit operation, net shape process. Emerging, alternative manufacturing processes will be discussed later in this section.

Each manufacturing process has its own advantages and limitations, and the preferred process is largely dictated by cost and pad properties required for the specific polishing application. Each process also has its own set of

Table 6.3. Manufacturing Processes for Different Pad Types

Type 1	Type 2	Type 3	Type 4
Needle polyester fibers to form non-woven felt	Prepare supporting substrate using modified Type 1 process	Mix polyurethane precursors and pore forming agent	Form polymer sheet
1st Urethane Impregnation	Coat substrate with polyurethane solution	Cast into mold	Apply texture (e.g. grooves)
2nd Urethane Impregnation	Coagulate surface film	Cure at elevated temperature	Apply PSA
Split Impregnated Felt to required thickness	Buff to open pore structure and control thickness	Skive cake into individual pad slices	Laminate to Base Pad
Buff	Apply PSA	Perforate or groove	
Apply PSA		Apply PSA	
		Laminate to Base Pad	

process variables which can be used to control pad properties. The effects of manufacturing process variables on pad properties have been recently discussed by Cook [13] and the reader is referred to this reference for more information.

Currently, Type 3 pads, because of their higher stiffness and resultant ability to planarize versus Type 1 and 2 pads, are predominantly used for CMP polishing of ILD, W, Cu and STI. It is expected that Type 4 pads may emerge as the next generation pad of choice because of their potentially less complex manufacturing process, resulting in improved pad to pad consistency and more predictable polishing performance. Data illustrating the benefits of Type 4 pads will be discussed in more depth later.

The patent literature describes several novel approaches to manufacture pads for CMP polishing. The principal driving forces are to:

1. Simplify manufacturing processes to ensure increased pad to pad consistency
2. Eliminate skiving and machine grooving operation
3. Develop net shape manufacturing process capable of producing patterned polishing pads having a wide range of physical properties
4. Develop processes capable of linear pads for next generation polishers

Table 6.4 references some of the recently issued US patents.

Table 6.4. US Patents describing Alternative Manufacturing Processes

Manufacturing Approach	US Patents
Sintering of Polymeric Powders	6,017,265, 6,062,968, 6,106,754, 6,117,000, 6,126,532
Photopolymerization of Liquid Precursors	5,958,794, 6,036,579
Net-shape Molding	6,022,268
Extrusion of thermo-formable polymers	5,489,233, 6,022,268

Each of these approaches has strengths and weaknesses. The preferred process depends both on the type of material being polished and on the tool platform.

6.4.3 Pad Microstructures

Each type of pad has a unique microstructure, as shown in the following SEM photomicrographs. Commercially available polishing pads are composite materials with properties determined by both their microstructure and polyurethane formulation.

Figures 6.1 and 6.2 show the cross-section and surface microstructures respectively of a Type 1 pad, as exemplified by Suba 500^{TM}. The microstructure is characterized by non-woven polyester fibers, partially impregnated with polyurethane to leave open porosity throughout the pad. By controlling
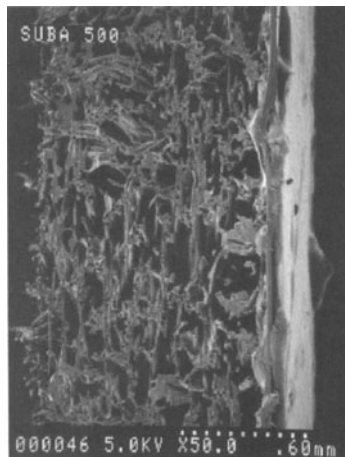
Fig. 6.1. Cross-section of a Type 1 Pad (Suba 500^{TM})

Fig. 6.2. Surface of a Type 1 Pad (Suba 500^{TM})

the type of polyurethane and the amount of residual porosity, a family of pads is possible differing in compressibility, hardness and stiffness.

Figures 6.3 and 6.4 show the cross-section and surface microstructures respectively of a Type 2 pad, as exemplified by UR100^{TM}. This type of pad has the most complex microstructure consisting of a porous layer on a supporting substrate similar in structure to a Type 1 pad. Type 2 pads are among the earliest pads used for polishing and their origin dates back to the Corfam^{TM} process developed by Dupont in the 1950's to make microporous, permeable artificial leather [14]. It is interesting to note that real leather was used during the early days of semiconductor technology to polish silicon wafers. Type 2
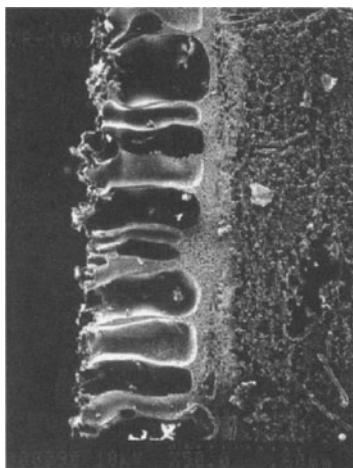
Fig. 6.3. Cross-section of a Type 2 Pad (UR100^{TM})
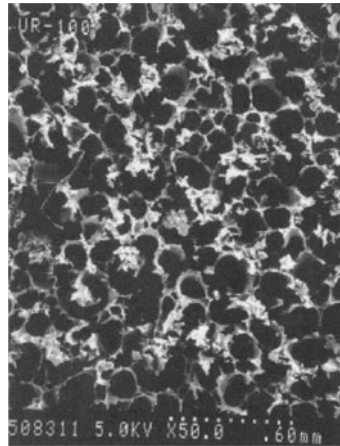
Fig. 6.4. Surface of a Type 2 Pad (UR100$^{\text{TM}}$)

pads are generically known as "poromerics" arising from the fact that these materials are *por*ous and poly*meric*. The surface of the pad consists of open pores which can hold and transport slurry across the pad surface.

Figures 6.5 and 6.6 show the cross-section and surface microstructures respectively of a Type 3 pad, as exemplified by IC1000$^{\text{TM}}$. This type of pad is essentially a closed cell foam, where the pores are created either by blowing agents or by the addition of micro-balloons. Since the pads are porous and made by mechanically machining individual pads, the pad surface has significant texture even prior to conditioning.
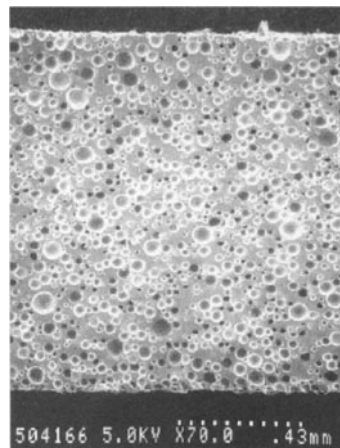
Fig. 6.5. Cross-section of a Type 3 Pad (IC1000$^{\text{TM}}$)
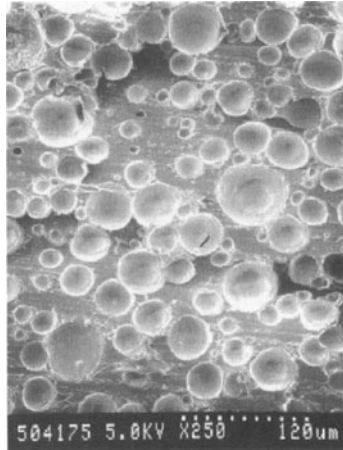
Fig. 6.6. Surface of a Type 3 Pad (IC1000$^{\text{TM}}$)
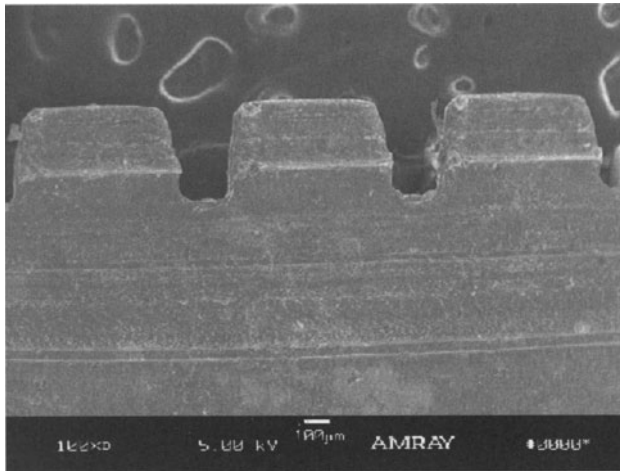
Fig. 6.7. Cross-section of a Type 4 Pad (OXP3000$^{\text{TM}}$)

Figures 6.7 and 6.8 show the cross-section and surface microstructures respectively of a Type 4 pad, as exemplified by OXP3000$^{\text{TM}}$. This type of pad has the simplest microstructure, being non-porous and unfilled. With such pads, it is known [15] that it is essential to have both macro- and microtexture to achieve acceptable polishing. As shown in Fig. 6.7, macrotexture consists of grooves formed in the pad surface either by mechanical machining as a post-processing step or by net-shape processing.

Type 3 and Type 4 pads are usually used in combination with an underlying base-pad. The function of the base-pad is to reduce polishing non-
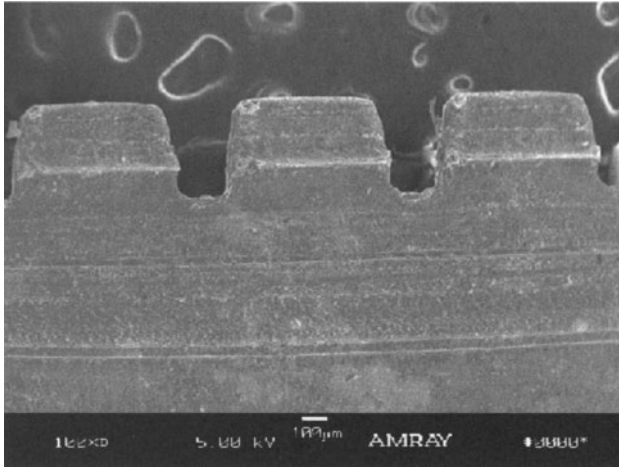
Fig. 6.8. Surface of a Type 4 Pad (OXP3000$^{\text{TM}}$)

uniformity across the wafer caused by non-planarity of the top-pad and pol-
ishing tool deficiencies. Base-pads are typically of higher compressibility and
lower stiffness than the top-pads and thus act essentially as supporting "cush-
ions" for the top-pad. Figures 6.9 and 6.10 show cross-sections of two types
of base-pads commonly used for CMP polishing. Figure 9 is an example of
an impregnated polyester felt (SubaIV$^{\text{TM}}$), similar to Type 1 pads discussed
above but with higher compressibility and porosity. Figure 10 is an example
of a polyurethane closed cell elastomeric foam.
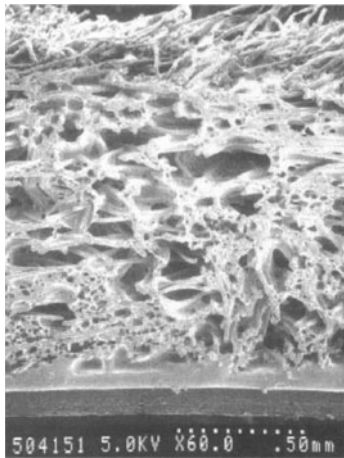
Fig. 6.9. Cross-section of a SUBA IV$^{\text{TM}}$ Base Pad
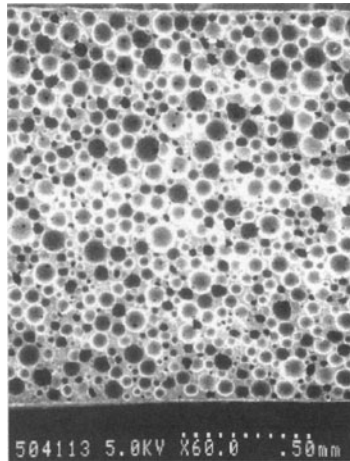
Fig. 6.10. Cross-section of a Closed Cell Foam Base Pad

6.5 Control of Polyurethane Pad Properties

This section will explore approaches available to control the properties of polishing pads. The focus will be on Type 3 and Type 4 pads, since they are commercially of most significance, but the approaches are also applicable to the other pad types. Approaches are:

  i. Control of hard and soft segments
 ii. Urethane stoichiometry
iii. Pad thermal history
 iv. Amount of porosity.

Each of these will be discussed in detail and illustrated with examples.

6.5.1 Hard and Soft Segments

Previously in this chapter, the concept of hard and soft segments in polyure-thanes was discussed. The type and concentration of these segments are major factors in controlling pad properties. Typically, increasing the soft segment concentration increases toughness and flexibility but reduces modulus and hardness. The hard segments, which usually soften at temperatures above ambient, improve high temperature properties and increase properties such as stiffness and strength.

The ability to manipulate pad properties through control of hard and soft segments is illustrated in Figs. 6.11 and 6.12. These are ternary phase diagrams for a three component polyurethane formulation. Each point within the triangle represents a specific composition. Figures 6.11 and 6.12 show hardness and elastic modulus respectively.
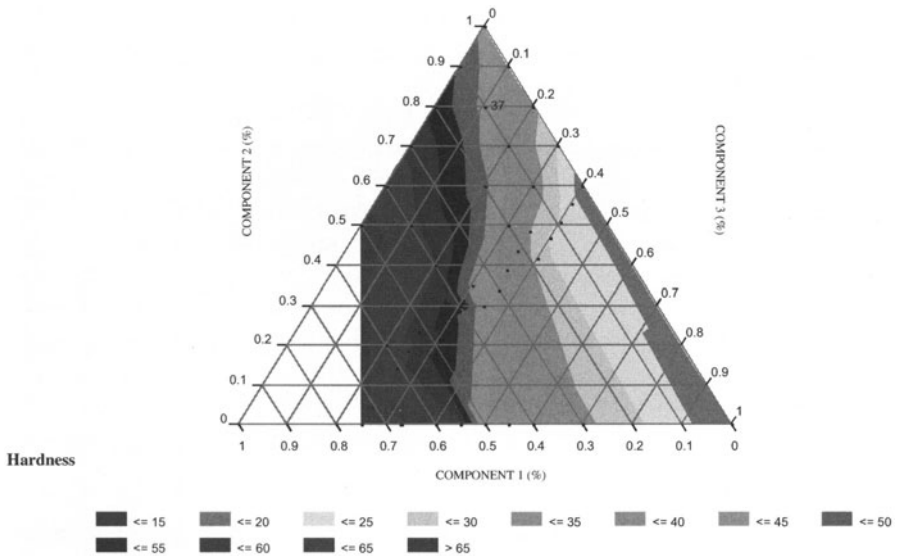
Fig. 6.11. Hardness as a Function of Composition
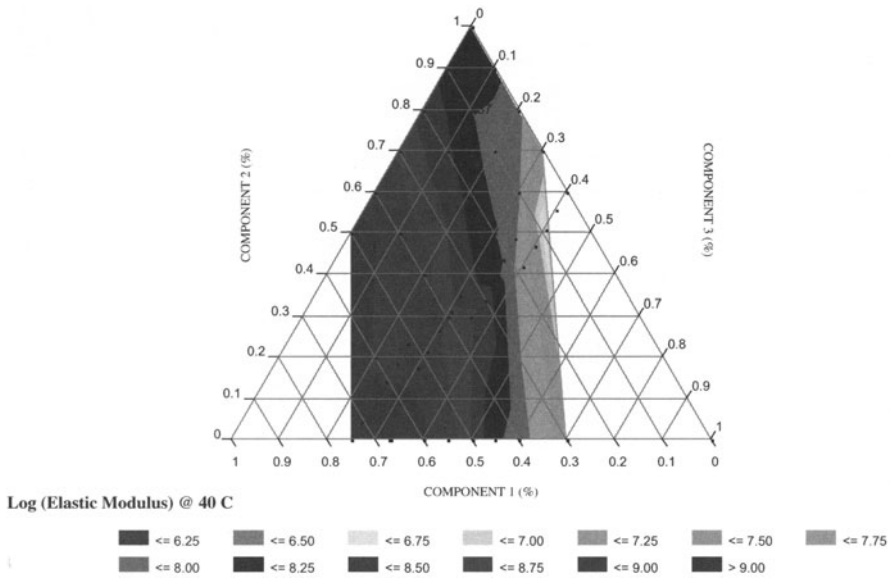
Fig. 6.12. Modulus as a Function of Composition

Figure 6.11 shows that it is possible to vary hardness from Shore D values of less than 15 to greater than 65. This covers the practical range of interest for all types of polishing pads. In general, harder pads are used for planarization of oxide dielectric layers, shallow trench isolation, and tungsten plugs

and conductors. Slightly softer pads are used for polishing copper damascene features, and still softer pads are used in final buff polishing to remove defects from the earlier steps.

Figure 6.12 shows a similar diagram for modulus. In fact, the diagram shows the logarithm of modulus, again illustrating the broad range of properties achievable through hard and soft segment control. In Fig. 6.12, modulus varies by three orders of magnitude from rigid, stiff pads to pads that are very flexible and elastomeric. As will discussed later, stiffer pads are preferred for those polishing applications where planarization is important. A comparison of Figs. 6.11 and 6.12 shows that hardness and modulus follow similar behavior with composition. Although it is often the case that hardness increases with increasing modulus, as will be seen in the next section, it is also possible to control them somewhat independently.

6.5.2 Urethane Stoichiometry

Stoichiometry refers to the ratio of reactive groups, usually diol or diamine moieties, to isocyanate groups. A stoichiometry of 100% indicates a perfect balance between isocyanate and diol groups. However, as mentioned earlier, it is customary with polyurethanes to use an excess of isocyanate and values of 85 to 95% are more typical. The reason for this is to take advantage of the side reactions that can take place when excess isocyanate is present. These are used to further control pad properties. A ratio of exactly 100% would result in a linear high molecular polymer [16]. Since isocyanate groups also react with moisture, exact stoichiometric balance is difficult to achieve and consequently this makes molecular weight control difficult, especially in a production environment.

Table 6.5 illustrates the effect of stoichiometry for an experimental, nonporous Type 4 pad. As the stoichiometric ratio increases, the amount of excess isocyanate and the number of side reactions decrease. Since many of these side reactions result in chemical bonds (known as cross-links) between polymer chains, the polyurethane transitions from a predominantly thermosetting network polymer to a more linear polymer, having more thermoplastic behavior, as stoichiometry increases.

Table 6.5. Effect of Stoichiometry on Pad properties

	Stoichiometry (%)		
Property	75	85	95
Hardness (Shore D)	60.5	61.2	63.4
Tensile Strength (MPa)	65.0	67.1	80.5
Elongation to Break (%)	281	298	459
Modulus (MPa)	554	518	514

As stoichiometry increases, hardness, tensile strength and elongation increase, consistent with a decrease in cross-links and an increase in ductility. Modulus decreases since cross-links increase rigidity. A comparison of the trend of hardness and modulus with stoichiometry shows that this is an example where it is possible to control hardness and modulus independently.

6.5.3 Pad Thermal History

The properties of polyurethanes are strongly influenced by their thermal history. In this context, thermal history includes the temperatures and times at which the pads are initially cured and, where appropriate, subsequently post-cured. Optimum times and temperatures will depend on the specific urethane system. However, for thermosetting polymers, which includes polyurethanes, it is generally accepted [16] that increasing baking temperatures and time at temperature will increase the degree of cure and change physical properties.

Table 6.6 illustrates the effect of thermal history on polyurethane properties for an unfilled Type 4 pad. Baking conditions have been designated as low, medium and high. High signifies a longer bake at higher temperature than medium, and medium versus low have the same relationship to one another. As baking conditions increase, pad properties such as hardness and modulus increase and elongation decreases. These trends are consistent with the formation of more cross-links between polyurethane chains producing a network structure and, relatedly, to an increase in the glass transition temperature (T_g) of the polyurethane.

T_g is the temperature at which the polyurethane softens appreciably. One method to measure T_g from the peak value of the tan delta damping curve as measured by dynamic mechanical analysis [17]. Tan delta is a measure of the damping ability of a material and is discussed in more detail later in this section. Figure 6.13 shows the effect of baking conditions on the tan delta curves for this urethane system. As baking conditions increase, the peak in tan delta

Table 6.6. Effect of Thermal History on Polyurethane Properties

Property	Baking Conditions		
	Low	Medium	High
Density (g/cm^3)	1.183	1.184	1.184
Hardness (Shore D)	70.5	71.2	73.9
Yield Strength (MPa)	29.4	31.6	36.1
Tensile Strength (MPa)	63.7	62.1	68.0
Elongation to Break (%)	340	320	250
Glass Transition Temperature (°C)	65	68	77
Modulus (MPa)	757	806	1050
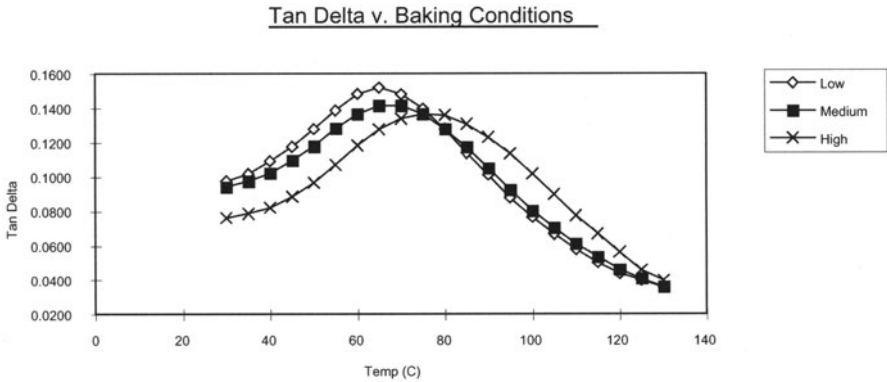
Fig. 6.13. Effect of Baking Conditions on Tan Delta

shifts to higher temperatures and slightly broadens. At more aggressive baking conditions, the tan delta curves will broaden further, signifying the onset of thermal degradation and a deterioration in pad properties. When selecting baking conditions, it is important to select times and temperatures which optimize properties for the specific application without causing degradation.

It is appropriate at this point to make a brief mention of the technique of dynamic mechanical analysis to study polishing pads. Since polishing is a dynamic process involving cyclic motion of both the polishing pad and the wafer and since polymeric polishing pads are viscoelastic materials, a valuable method of studying pad properties is by dynamic mechanical analysis [17]. In this technique, a cyclic deformation is applied to the sample at temperatures and frequencies that correspond to typical polishing conditions.

Viscoelastic materials exhibit both viscous and elastic behavior in response to an applied deformation. The resulting stress signal can be separated into two components: an elastic stress which is in phase with the strain, and a viscous stress which is in phase with the strain rate but 90 degrees out of phase with the strain. The elastic stress is a measure of the degree to which a material behaves as an elastic solid; the viscous stress, the degree to which the material behaves as an ideal fluid. The elastic and viscous stresses are related to material properties through the ratio of stress to strain, the modulus. Thus, the ratio of elastic stress to strain is the storage (or elastic) modulus and the ratio of the viscous stress to strain is the loss (or viscous) modulus. When testing is done in tension or compression, E' and E'' designate the storage and loss modulus, respectively.

The ratio of the loss modulus to the storage modulus is the tangent of the phase angle shift (δ) between the stress and the strain. Thus,

$$E''/E' = \tan \delta$$

and is a measure of the damping ability of the material.

  With specific reference to polishing, energy is transmitted to the pad during the polishing cycle. A portion of this energy is dissipated inside the pad as heat and the remaining portion of this energy is stored in the pad and is subsequently released as elastic energy during the polishing cycle.

6.5.4 Pad Porosity

Although the morphologies of Type 1, 2 and 3 pads differ, they all contain porosity which plays a key role in determining pad physical properties and polishing functionality. For these pads, porosity is needed to retain slurry on the pad surface and to distribute slurry uniformly across that surface. In this section, the importance of porosity will be illustrated using data for IC1000^{TM} and IC2000^{TM}.

  Table 6.7 shows typical physical properties of IC1000^{TM}, Style 5.

  Many of these properties are strongly dependent on the porosity present in the pad. For these pads, porosity is difficult to measure directly with a high degree of accuracy. Instead, porosity is most conveniently determined quantitatively by measuring pad density, since the two are related by the equation:

$$\text{Volume Fraction Porosity} = \frac{\text{Density of Bulk Polymer} - \text{Density of Pad}}{\text{Density of Bulk Polymer}}.$$

  Table 6.8 shows the statistical correlations between pad density and other physical properties. It is apparent that the correlations are very high (± 1.00 being a perfect correlation). Since many properties depend on pad porosity and are not independent of one another, as will be seen in a later section,

Table 6.7. Typical Physical Properties of IC1000^{TM}

Property	Value
Porosity (# of cells)	880 ± 120
Density (g/cm^3)	0.748 ± 0.051
Hardness (Shore D)	52.2 ± 2.5
Shear Strength (MPa)	51.2 ± 4.1
Proportional Limit (MPa)	9.1 ± 1.3
Tensile Strength (MPa)	21.6 ± 2.8
Elongation to Break (%)	175 ± 20
Storage Modulus (MPa)	310 ± 40
Loss Modulus (MPa)	28.0 ± 4.5
Tan Delta	0.090 ± 0.005

Table 6.8. Relationship between Pad Density and other Physical Properties

Pad Physical Property	Correlation Coefficient
Porosity	−0.77
Hardness	0.94
Shear Strength	0.84
Proportional Limit	0.88
Tensile Strength	0.96
Elongation to Break	0.80
Storage Modulus	0.89
Pad Stiffness	0.88

relating specific pad properties to polishing performance is made even more complex.

Figures 6.14 and 6.15 graphically show the linear relationships of density with hardness and storage modulus respectively.

It has been shown above that porosity has a strong influence on other pad properties. A consequence of this relationship is that porous pads have more variability in physical properties than equivalent non-porous pads. This may be illustrated by comparing the physical property distributions of IC1000^{TM} and IC2000^{TM}. IC2000^{TM} is very similar to IC1000^{TM} and made by the same manufacturing process but is non-porous. Typical physical properties of IC2000^{TM} are shown in Table 6.9.

Figure 6.16 compares the key physical properties for the two pad types in terms of the ratio of standard deviation to property average [18, 19]. It is
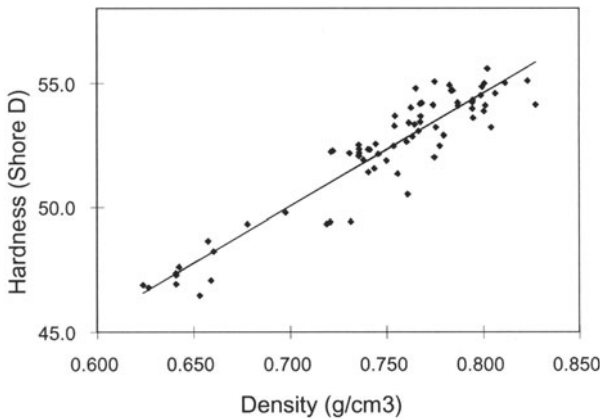
Fig. 6.14. Hardness versus Density for IC1000^{TM}
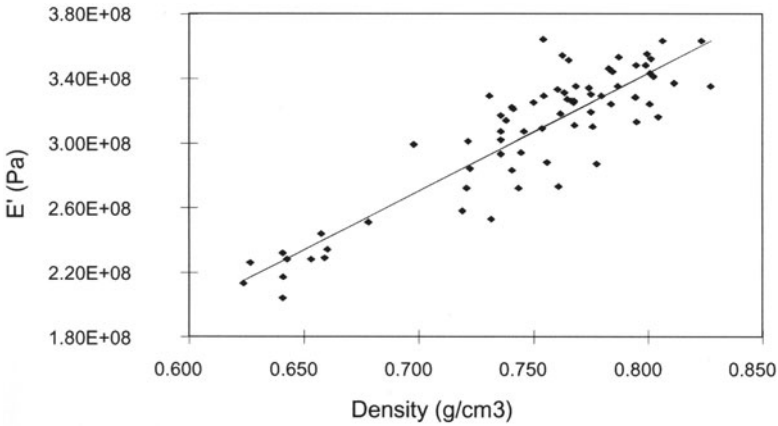
Fig. 6.15. Storage Modulus versus Density for IC1000^{TM}

Table 6.9. Typical Physical Properties of IC2000$^{\text{TM}}$

Property	Value
Porosity (# of cells)	0
Density (g/cm^3)	1.180 ± 0.002
Hardness (Shore D)	73.0 ± 1.0
Yield Strength (MPa)	33.4 ± 1.4
Tensile Strength (MPa)	75.0 ± 2.5
Elongation to Break (%)	335 ± 20
Storage Modulus (MPa)	850 ± 61
Loss Modulus (MPa)	87.0 ± 4.0
Tan Delta	0.103 ± 0.005

clearly evident that for non-porous pads the variability in density is significantly less and that a similar trend is present in the other properties shown.

An advantage of non-porous pads is thus a much tighter distribution of physical properties. Although the relationships between polishing performance and pad physical properties are not well understood, a tighter property distribution should translate into more consistent polishing performance. This may be illustrated for oxide polishing using the two pad types [18]. Figure 6.17 shows the distribution of removal rate for a large number of wafers polished using either IC1000$^{\text{TM}}$ or IC2000$^{\text{TM}}$. Conditioning was kept constant for the two pad types. The distribution of removal rate is much tighter for the non-porous pad, consistent with the above discussion.

In addition to the benefit of more consistent pad physical properties and less variability in polishing removal rate, there are other benefits for non-

Physical Property Variability of IC1000 and IC2000
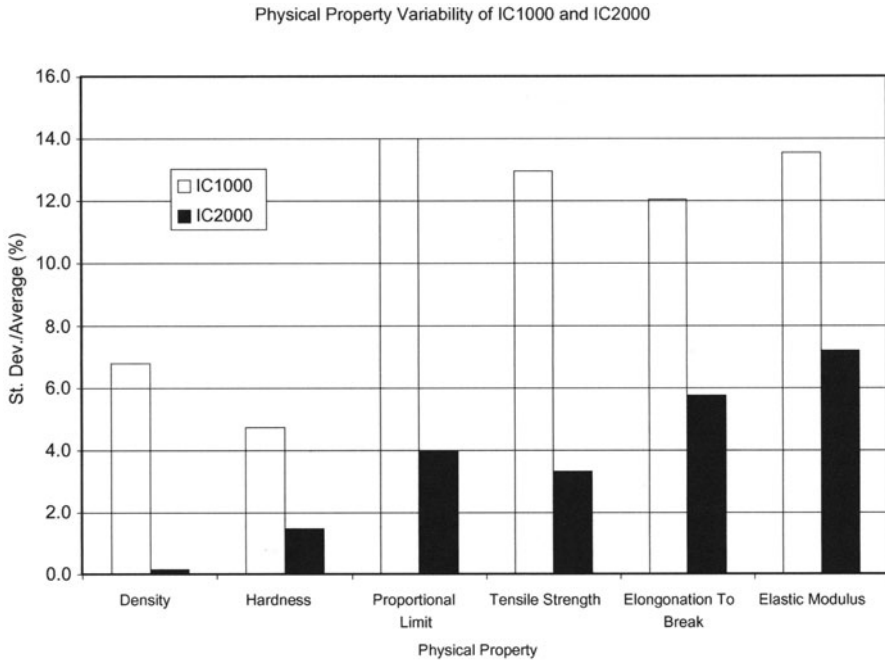
Fig. 6.16. Effect of Porosity on Variability of Pad Properties
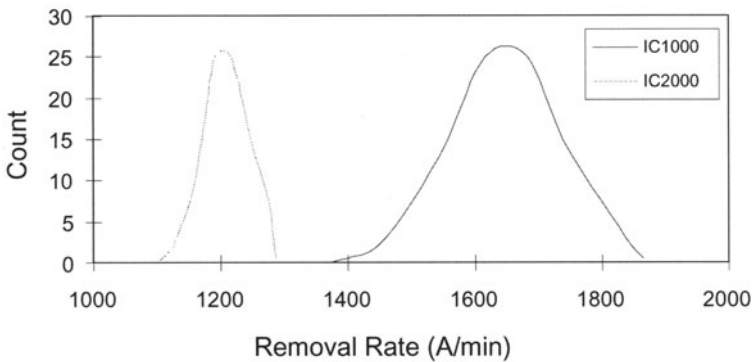
Fig. 6.17. Effect of Pad Porosity on Polishing Removal Rate Distribution

porous pads. Typically, as may be seen by comparing the property data for IC1000^{TM} and IC2000^{TM} shown in Tables 6.7 and 6.9 respectively, non-porous pads have a higher modulus. This translates to improved planarization of die-scale features during polishing. Abrasion resistance of non-porous pads also tends to be higher which results in less pad wear during polishing and hence longer pad-life. Another benefit of non-porous pads, which has

been demonstrated for copper CMP polishing, is reduced defectivity [20, 21]. Pores on the pad surface can trap polishing debris which can either scratch the wafer features, especially if the features are soft materials like copper, or leave residual particles on the wafer surface.

There are also challenges with non-porous pads. The primary challenge is that non-porous pads require more rigorous pad break-in to achieve optimum polishing performance. It is well-known [15] that for effective polishing, the pad surface must have both micro- and macro-texture. The latter is discussed in detail in the next section. Micro-texture refers to the roughness of the pad surface and the presence of asperities which contact the wafer during polishing. During pad break-in, the surface of the pad is roughened using typically a diamond impregnated conditioning tool. With porous pads, the pores in the pad surface inherently create a surface which is initially rougher than the surface of a non-porous pad, so conditioning is effectively given a head-start. In contrast, with non-porous pads, the initial surface is much smoother and all the micro-texture must be created by the conditioning process. Thus, non-porous pads require a longer conditioning break-in cycle and the removal rate decay observed when abrasive conditioning is stopped during polishing is higher [18].

The second challenge is "edge effects" during polishing which increase removal rate non-uniformity across the wafer surface by leaving a thicker ring of polished material around the wafer edge. This reduces the usable area of the wafer surface. Edge effects have been discussed by Baker [22], and related to the physical properties of the top and base layers of a polishing pad. Baker's modeling work, and subsequent experimental validation, shows that as the stiffness of the top pad increases, edge effects also increase. Since non-porous pads are stiffer than corresponding porous pads, edge effects would be expected to be more of a problem with the non-porous pads and this has been verified experimentally using IC2000^{TM}. However, edge effects can be largely eliminated by the use of retaining rings on the wafer carrier that are essentially coplanar with the wafer surface. This approach has been implemented with beneficial results by several polishing tool manufacturers. A second approach to minimize edge effects is through the judicious use of macro-grooves in the top pad which reduce pad stiffness in a controlled way without compromising die-scale feature planarity [62].

6.6 Control of Pad Properties Through Pad Geometry

The previous section discussed control of pad properties through the pad urethane chemistry and formulation. This section discusses the manipulation of pad properties by physical control of the pad geometry. This includes consideration of pad thickness, groove designs, pad shapes, and base pad.

6.6.1 Pad Thickness

Top pad thickness is important, since it determines the stiffness of the pad, given that pad stiffness is proportional to the product of pad modulus and cube of the thickness [23]. Thus doubling the pad thickness increases stiffness eight-fold. Polishing pads used for CMP are typically about 1.3 mm (50 mil) thick. Pad stiffness controls several important polishing parameters, including uniformity of removal rate across the wafer, die level planarity, and to a lesser extent dishing and erosion of features within a die. In order to planarize next generation devices, pad thicknesses greater than 1.3 mm (50 mil) are preferred, and 2.0 mm (80 mil) thick pads are now being used routinely by several major semiconductor manufacturers. Above about 5 mm (200 mil), polishing uniformity may suffer because of the inability of the pad to conform to variations in global wafer flatness.

As a polishing pad wears, the overall pad thickness and corresponding stiffness decrease. This again argues for a high initial pad thickness, as the change in stiffness with polishing time will be relatively less for a thicker pad. Additionally, since stiffness is less dependent on the grooved thickness, which is removed during pad use, rather than thc underlying ungrooved region, high thickness for the underlying ungrooved layer and for the overall pad are preferred.

For a given pad thickness, increasing pad modulus will increase pad stiffness and the ability of the pad to planarize. Thus, as mentioned previously, unfilled pads will planarize more effectively than filled pads. However, it is important to recognize that stiffness is proportional to the cube of thickness compared to only the single power of modulus, so that changing pad thickness can have a more significant impact than changing pad modulus.

Figure 6.18 [64] shows the effect of pad thickness on oxide planarity for a Type 4 experimental pad (OXP3000$^{\text{TM}}$ from Rodel). The polisher was a Strasbaugh 6DS-SP and the slurry was ILD1300$^{\text{TM}}$ also from Rodel. The figure shows planarity expressed as "planarity quotient" for different feature sizes. Planarity quotient is defined as the oxide removal rate at the bottom of a feature trench divided by the oxide removal at the top of a feature. Thus, the smaller the planarity quotient, the more efficient the planarization process. Since smaller features are more readily planarized, planarity quotient increases with increasing feature size. For a given value of planarization quotient, the feature size corresponding to that value may be defined as the "planarization distance". Longer planarization distances indicate enhanced ability of a pad to planarize.

Figure 6.18 shows that for a given value of planarization quotient, polishing with a 2.0 mm (80 mil) thick pad results in a higher planarization distance versus the 1.3 mm (50 mil) thick pad. For example, at an arbitrary planarization quotient of 0.3, the 2.0 mm (80 mil) thick pad has a planarization distance of about 40% longer than the distance for the 1.3 mm (50 mil) thick pad (3.9 versus 2.8 mm).
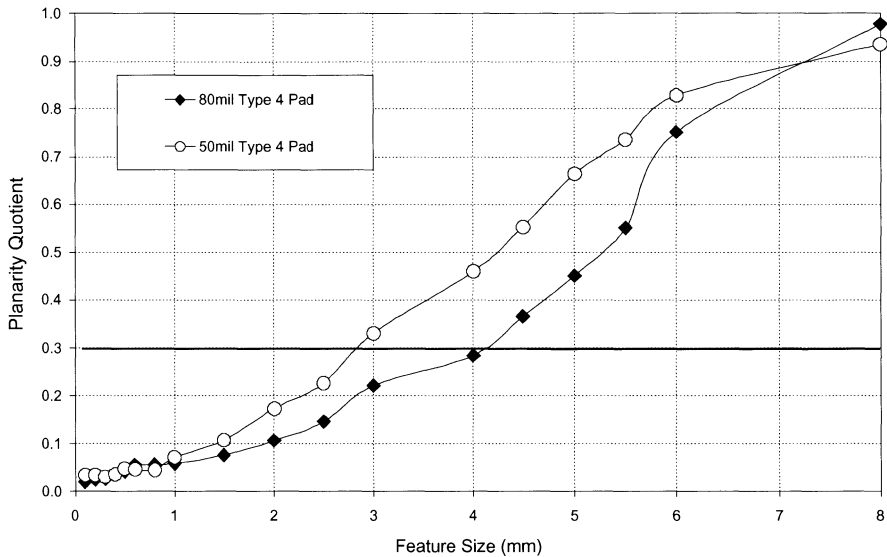
Fig. 6.18. Effect of Pad Thickness on Planarity

As discussed in the next section, grooving the surface of polishing pads also affects pad stiffness. In general, the deeper the grooves with respect to pad thickness and the closer the grooves are to one another, the more flexible the pad.

6.6.2 Groove Designs

Polishing pads used for chemical mechanical polishing typically have macro-texture. This can be either perforations through the pad thickness or surface groove designs. Such surface designs include, but are not limited to, circular grooves which may be concentric or spiral grooves, cross-hatched patterns arranged as an X-Y grid across the pad surface, other regular designs such as hexagons, triangles and tire-tread type patterns, or irregular designs such as fractal patterns, or combinations thereof. The groove profile may be rectangular with straight side-walls or the groove cross-section may be "V"-shaped, "U"-shaped, triangular, saw-tooth, etc. Further, the geometric center of circular designs may coincide with the geometric center of the pad or may be offset. Also the groove design may change across the pad surface. The choice of design depends on the material being polished and the type of polisher, since different polishers use different size and shape pads (i.e. circular versus linear).

Table 6.10 provides an overview of the multitude of groove designs that have appeared in the US patent literature over the last ten years.

Table 6.10. US Patents covering Groove Designs

US Patent Number	Description of Groove Design
5,020,283	Achieve constant surface area across pad using circular voids, holes, squares and ray designs
5,177,908	Sun-burst pattern of non-tapered rays
5,216,843	Circumferential triangular macro-grooves and conditioning microgrooves
5,297,364	Circular voids to control non-uniformity
5,329,734	Regions of different circular pore density
5,394,655	Annular rings with tapered cross-section
5,489,233	Non-porous pads having macro and micro-channels
5,578,362	Type 3 pads with circular and X-Y groove designs
5,609,719	X-Y channels with slurry recesses at intersections to minimize edge exclusion
5,628,862	Molded pad surface comprising channels between hemispherical features
5,645,469	Radially extending tapered channels and circumferential grooves
5,650,039	Off-center circular spiral groove
5,690,540	Inward spiral from pad periphery to center
5,725,420	Combination of holes and X-Y grooves where groove pitch > hole pitch
5,778,481	Spiral, swirl or concentric raised areas direct fluids from pad center to edge
5,842,910	Circumferential grooves offset from pad geometric center
5,888,121	Regions with different circular and X-Y groove dimensions
5,900,164	Type 3 pads with circular and X-Y groove designs
5,921,855	Circular, spiral or concentric grooves
5,984,769	Regions with different circular groove dimensions

Grooves are added to polishing pads used for CMP for several reasons:

1. To prevent hydroplaning of the wafer being polished across the surface of the polishing pad. If the pad is either ungrooved or unperforated, a continuous layer of polishing fluid can exist between the wafer and pad, preventing uniform intimate contact and significantly reducing removal rate.
2. To ensure that slurry is uniformly distributed across the pad surface [24] and that sufficient slurry reaches the interior of the wafer. This is espe-

cially important when polishing reactive metals such as copper, in which the chemical component of polishing is as critical as the mechanical. Uniform slurry distribution across the wafer is required to achieve the same polishing rate at the center and edge of the wafer. However, the thickness of the slurry layer should not be so great as to prevent direct pad-wafer contact.

3. To control both the overall and localized stiffness of the polishing pad. This controls polishing uniformity across the wafer surface and also the ability of the pad to level features of different heights to give a highly planar surface. Different regions of the pads may have different groove designs.

4. As a subset of 3, to reduce edge effects.

5. To act as channels for the removal of polishing debris from the pad surface. A build-up of debris increases the likelihood of scratches and other defects. This is related to 2, since as new slurry replaces the old, the old removes the entrained debris.

It is known [15, 25] that one factor determining pad-life of grooved pads is the depth of the grooves, since acceptable polishing performance is possible only until the pad has worn to the point where grooves have insufficient depth to distribute slurry, remove waste, and prevent hydroplaning. For Type 3 pads containing porosity, the minimum groove depth to prevent hydroplaning is about 3 mil and for Type 4 non-porous unfilled pads about 5 mil. In order to achieve the combination of acceptable pad stiffness and long pad-life, it is necessary to have deep grooves but also sufficient remaining pad to provide stiffness. As groove density and groove width increase, pad stiffness becomes more dependent on the thickness of the remaining ungrooved layer of the pad, rather than on groove depth alone.

6.6.3 Pad Shapes

The patent literature contains many examples of novel pad shapes which have been proposed to improve polishing performance. Although many of these concepts are untested, they illustrate some of the options that may be used to optimize polishing performance through control of pad geometry. Table 6.11 summarizes some of the key patents covering pad shape. This list is by no means complete but illustrates some of the approaches that are being considered.

6.6.4 Base Pads

Ideally, for uniform polishing, removal rate should be the same at all points on the wafer surface. This would suggest that the pad needs to be in contact with the whole wafer surface with the same contact pressure and relative velocity between pad and wafer at all points. Unfortunately, wafers are not

Table 6.11. US Patents related to Pad Shape

US Patent #	Assignee	General Area	Brief Description
5,310,455 5,516,400 5,624,304	LSI	Extended Pad Edge	A lower pad is mounted to the platen and is trimmed to the size of the platen. An upper pad is mounted to the lower pad, and is sized so that an extreme outer edge portion of the upper pad extends beyond the trimmed outer edge of the lower pad. The outer edge portion of the upper pad is deformed downwardly, towards the lower pad. In this manner, polishing slurry is diverted from the pad-to-pad interface. Additionally, an integral annular lip can be formed on the front face of the upper pad, creating a reservoir for slurry to be retained on the face of the upper pad for enhancing residence time of the polishing slurry prior to the slurry washing over the face of the upper pad.
5,234,867 5,421,769	Micron	Non-circular Pads	A polishing head mechanism moves the polishing head and semiconductor wafer across and past a peripheral edge of a non-circular pad to effectuate a uniform polish of the semiconductor wafer surface.
5,558,563	IBM	Raised Areas	A polishing pad which includes raised portions is used to apply varying amounts of pressure. In addition, the position, size and height of the raised portions are used to affect the amount of pressure applied. There are several possible methods to produce raised areas in the polishing pad. In particular, shims can be added to the polishing table or the polishing table can be machined so that the polishing table includes raised portions.
5,785,584 5,934,977	IBM	Raised Areas	The rotating polishing pad is caused to flex upward as it passes over a discrete, non-continuous "bump formed in the surface of an underlying stationary platen. This deflection in any given portion of the pad is a transient condition as the raised portion of the pad will fall back from to its original configuration once it clears the underlying bump. Thus, portions of the pad will be continuously progressed up, over and back down to create a discrete raised portion in the polishing pad surface at a fixed location.
5,888,126	Ebara	Raised Areas	The abrasive cloth has a projecting region on a surface thereof for more intensive contact with the workpiece than other surface regions of the abrasive cloth. Projecting regions on the pad surface are effected by creating projecting regions in the underlying upper surface of the turntable using actuators which may be electromagnetic, piezoelectric or compressed air.
5,769,699	Motorola	Multiple Regions	The polishing pad has a first region that is closer to the edge of the polishing pad and a second region adjacent to the first region and further from the edge of the polishing pad. The polishing pad is configured, so that the second region is thicker or less compressible compared to the first region.
5,738,567 5,910,043	Micron	Multiple Regions	The polishing pad has a polishing body and a cleaning element positioned in the polishing body. In operation, the cleaning surface periodically engages the wafer when the wafer is engaged with the pad to remove residual materials from the surface of the wafer.

Table 6.12. Patents describing Multiple Layer Pads

Patent #	Assignee	Brief Description
US5,212,910	Intel	Three layer composite pad, comprising a top polishing layer, an intermediate stiff layer which may be segmented to create a "bedspring" effect, and a bottom elastic cushioning layer.
US,5257,478	Rodel/ Westech	Pad has at least two layers, wherein the base layer is more compressible than the top planarizing layer.
US5,287,663	National Semiconductor	Three layer composite pad, comprising a top polishing layer, an intermediate rigid layer and a bottom resilient layer.
US5,564,965	SEH	Three layer composite pad, comprising a top polishing layer, an intermediate rigid layer and a closed-cell foam of soft rubber as the bottom layer.
US5,664,989	Toshiba	Two layer polishing pad, wherein the base layer consists of fine bags hermetically sealed with fluid to provide uniform pressure distribution across the wafer.
US5,871,392	Micron	Base pad contains a plurality of thermal conductors to conduct heat from the polishing layer to the platen.
US5,876,269	NEC	Two layer polishing pad comprising an upper polishing layer harder than the lower layer
US5,893,755	Komatsu	Discloses silicon rubber base pad to reduce waviness during polishing.
US5,899,745	Motorola	Base pad has an edge and central portion, such that the compressibility of the center portion is higher than that of the edge.
US5,899,799	Micron	Grooved base pad which creates depressions in the upper pad for control of slurry flow across the polishing pad.
EPA845,328	Sumitomo	Three layer composite pad, comprising a top porous polishing layer, an intermediate support layer and a bottom resilient layer.
EPA919,336	Speedfam	Two layer pad comprising a top pad over a more compressible base pad, such that both are uniformly permeable to polishing fluid.

perfectly flat and typically have some degree of curvature resulting from the stresses of manufacture and differing coefficients of thermal expansion of the various deposited oxide and metal layers. This requires the polishing pad to have sufficient flexibility to conform to wafer-scale flatness variability. One solution to this problem is to laminate a stiff polishing pad to a flexible underlying base pad, which is typically a more compressive, foam-type polymeric material. This improves polishing uniformity across the wafer without unduly compromising the stiffness of the polishing top pad.

Several patents have recently been issued disclosing the use of polishing pads, wherein the top pad, which contacts the wafer, is laminated to at least one underlying pad. This base pad is typically more compressible than the top pad. In some cases, a thin but stiffer intermediate layer is sandwiched between top and base pads. Table 6.12 summarizes some of the main patents covering multiple layer pads.

Table 6.13 shows physical properties of two commonly used base pads. The microstructures of these base pads have been shown previously in Figs. 6.9 and 6.10. Suba IV^{TM} is a polyurethane impregnated felt and the other base pad is a closed cell polyurethane foam.

As discussed previously, Suba IV^{TM} and foam base pads have very different microstructures. The effect of microstructure on physical properties is that the properties of Suba IV^{TM} are anisotropic, whereas those of the foam are isotropic. The properties shown above are measured through the thickness of the pad rather than in the plane of the pad. The porosity in Suba IV^{TM} is essentially open which can lead to wicking of slurry into the base pad.

Figure 6.19 [64] shows the effect of base-pad on the planarity of oxide features, polished with a Type 4 experimental pad (OXP3000^{TM} from Rodel). The polisher was a Strasbaugh 6DS-SP and the slurry was ILD1300^{TM} also from Rodel. Three sets of data are shown, corresponding to no base pad, Suba IV^{TM}, and an experimental foam base pad similar in microstructure and properties to the foam base pad discussed above.

It is clearly evident that the best planarity is achieved with no base pad. However, polishing uniformity across the wafer is poor caused by imperfect

Table 6.13. Typical Physical Properties of Base Pads

Physical Property	Suba IV^{TM}	Foam Base Pad
Thickness (mm)	1.3−1.4	1.3−1.5
Density (g/cm^3)	0.27−0.32	0.43−0.53
Hardness (Shore O)	65−75	45−55
Compressibility (%)	6−10	2−6
Rebound (%)	90−95	>90
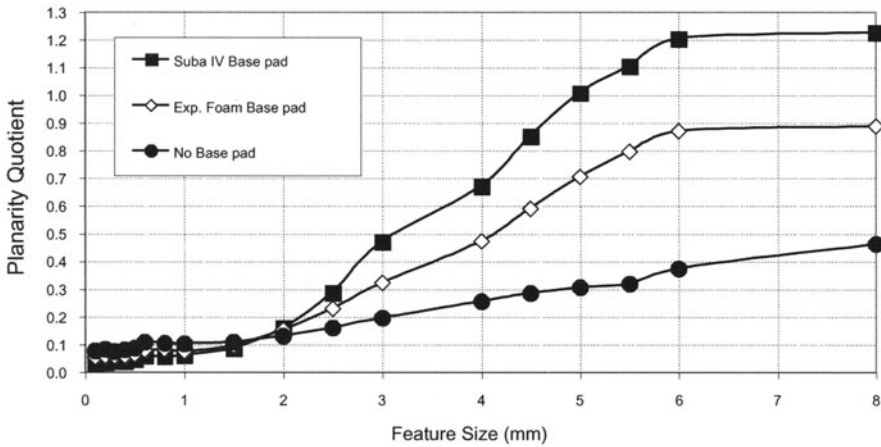
Fig. 6.19. Effect of Base Pad on Planarity

wafer flatness and thickness, as well as physical limitations of the top pad and polishing tool. Using a base pad compromises planarity but improves uniformity across the wafer by reducing the impact of the other problems. The experimental foam base pad has lower compressibility than Suba IV^{TM} which increases the ability to planarize. Thus at a given planarization quotient, the planarization distance of the foam base pad is longer than that of Suba IV^{TM}.

The base pad also has an impact on polishing non-uniformity at the edge of the wafer. As mentioned previously, polishing non-uniformity at the wafer edge depends on both the stiffness of the top pad and also on the compressibility of the base pad. The so-called "edge effects" result in less material being removed near the wafer edge. The problem becomes worse as the stiffness of the top pad increases and the compressibility of the base pad increases [22]. In order to achieve good planarization behavior with minimal edge effects, the preferred pad structure will comprise a stiff top pad over a base pad with just sufficient compressibility to eliminate polishing tool imperfections.

6.7 Relationships Between Pad Properties and Polishing Performance

The relationships between pad properties and polishing performance are complex and not fully understood. One reason for the complexity is that polishing performance is not categorized by a single parameter but by several parameters which are dependent on the scale of the features being polished. Table 6.14 categorizes polishing into three levels – wafer-, die- and feature-scale.

Wafer-scale refers to polishing across the whole wafer surface. Currently wafers are typically 200 mm in diameter but the industry is beginning to

Table 6.14. Characterization of Polishing Performance Parameters

Polishing Performance Scale		
Wafer-Scale	Die-Scale	Feature-Scale
Removal Rate (RR)	Planarization (P)	Conductor Dishing (CD)
Non-Uniformity (NU)	Defectivity (D)	Oxide Loss (O)
Edge Effects (EE)		Selectivity
Macro-scratches (MS)		Defectivity (D)
Pad Life (L)		Roughness (Rg)

gradually transition to 300 mm. Die-scale polishing applies to polishing across the area of a die, many of which populate the surface of a wafer. The third level, designated feature-scale, refers to the polishing of conductor lines, bond pads, posts and other features within the die. Such features have dimensions measured in sub-microns and microns. For each scale level, Table 6.14 shows the polishing parameters which are most important for that level. In most cases, the meaning of each of these parameters is self-evident and further definition will not be provided here. The reader is referred to other sections of this book for more details.

A second factor complicating an understanding of the relationships between pad properties and polishing performance is the strong dependence of one pad physical property on another. For example, it has been discussed previously that pad density strongly correlates with other properties such as hardness and modulus. Thus, it is very difficult to isolate one property and look at its effect on polishing performance without simultaneously changing other pad properties.

Thirdly, polishing performance depends on many factors other than just the pad. These include polishing tool set-up (e.g. platen and carrier speeds, down-force, back-pressure, carrier head design and carrier film, etc.), slurry-related variables (e.g. abrasive type and loading, pH, flow-rate, etc.), and pad conditioning (e.g. diamond density, exposure and placement, sweep profile, and down-force, etc.). How the pad conditions and the resultant topography of the pad conditioned surface also depends on the properties of the pad itself, as well as the design of the conditioner. In determining relationships between pad properties and performance, the other polishing variables must be kept constant, since many of these have a major impact on polishing results, which in some cases may be greater than the contribution from the pad itself.

Given the above caveats, this section attempts to relate the polishing performance characteristics described in Table 6.14 with specific pad properties. Only the major relationships will be discussed, as summarized in Table 6.15.

Table 6.15. Relationship between Pad Properties and Polishing Performance

| Pad Property | Polishing Scale | | | Conditionability |
	Wafer	Die	Feature	
Density (Porosity)	RR, NU	D	CD, O	Yes
Hardness	MS	D	D, Rg, CD, O	Yes
Tensile Properties	L			Yes
Abrasion Resistance	L			Yes
Modulus (Stiffness)	EE, NU	P		Yes
Thickness	L			
Top Pad Compressibility		P	CD	
Base Pad Compressibility	EE, NU	P		
Pad Texture (Grooves)	L, RR, NU, EE			
Pad Roughness	RR, NU	P	CD, O	Yes
Hydrophilicity	RR			Yes

6.7.1 Removal Rate

The rate of removal of the material being polished depends on many fac-
tors including both the macro and microtexture of the pad, and also the pad
physical properties. Macrotexture is achieved by either punching perforations
through the pad or by forming grooves into the pad surface. The latter has
been extensively discussed in an earlier section of this chapter. A key role of
macrotexture is to prevent hydroplaning of the wafer across the pad surface,
arising from an excessively thick layer of polishing fluid under the leading
edge of the wafer which limits intimate contact between pad and wafer and
significantly reduces removal rate. Levert et al. [26] have studied the thickness
of the slurry film for different pad types and shown that hydroplaning is po-
tentially more of a problem with impermeable (Type 4) and semi-permeable
(Type 3) pads, rather than with permeable pads (Types 1 and 2). Thus
macrotexture becomes more critical for Type 3 and is an absolute necessity
for Type 4 pads. Indeed, it has been shown that lack of macrotexture with
Type 4 pads results in no removal [15]. It has also been reported [27] for
Type 3 pads that grooves give higher removal rates and more stable removal

than perforations. The reason was attributed to grooves in the polishing pad surface facilitating slurry flow to the wafer surface during polishing. Grooves also enable slurry to be squeezed back out from the leading edge of the wafer, thus eliminating the potential for hydroplaning.

Removal rate also depends on pad microtexture which may be loosely defined as the localized roughness of the pad surface. Microtexture comes from both pad conditioning and from porosity within the pad. Thus other factors being held constant, Type 4 pads which have no inherent porosity give lower removal rates than Type 3 pads [18, 28].

Pad conditioning has a strong effect on removal rate and removal rate stability and has been studied extensively [13, 28, 29, 30, 31, 32, 33]. As discussed in the first chapter, it is known that either *in-situ* or *ex-situ* conditioning of the pad surface is required to achieve and maintain stable removal rates. The role of conditioning is to create asperities on the pad surface which contact the wafer. Without conditioning, the asperity population decays during polishing and removal rates rapidly decline. The rate of decline is usually higher for Type 4 than Type 3 pads, since in Type 4 pads there are no inherent asperities from a pore structure, such as that found in Type 3 pads.

Cook [13] and Hetherington [32] have independently used scanning electron microscopy to study changes in the surface of IC2000^{TM} (Type 4) and IC1400^{TM} (Type 3) pads respectively during polishing in the absence of conditioning. For both pad types, the photomicrographs clearly showed a progressive decrease in pad asperities and the increasing formation of a smooth mesa-type structure. Hetherington postulated that the asperities were deforming under the shearing conditions associated with polishing, whereas Cook favored a progressive abrasive smoothing of the asperities. It is probable that given the conditions of polishing both shear deformation and abrasive loss are occurring.

Bajaj et al. [28] and Oliver et al. [33] have also studied the relationship between pad properties and removal rate decline in the absence of conditioning. During polishing, pad surface roughness decreases because the shearing forces of polishing reduce the height and number of pad asperities. The rate at which the asperity population declines was shown to be inversely proportional to the shear modulus of the bulk polymer.

6.7.2 Non-Uniformity

Removal rate uniformity across the wafer surface (within wafer non-uniformity) also depends on the macro and microtexture of the surface of the polishing pad in contact with the wafer. However, as previously discussed, because of edge effects non-uniformity also depends on the base pad and the balance of properties between top pad stiffness and base pad compressibility [22].

Goetz [34] has studied the effect of the base pad construction on planarization at intra-die, die, and wafer length scales using contact and plate bending mechanics to determine the pressure variations due to loading. Both

single-layer and two-layer sub-pad constructions were analyzed. As expected, the effect of sub-pad construction was greatest in the multi-millimeter length scales (die to wafer scale) and attenuated by the stiffness of the upper polishing layer. Single and two-layer sub-pads responded in different ways to loading, such that the response of the single-layer sub-pad was dominated by contact mechanics, while the two-layer sub-pad response was dominated by plate bending.

Polishing non-uniformity across the wafer is also strongly dependent on the polisher design, especially the carrier head. Base pads were originally used to overcome limitations with the polishing tool. As polisher designs become more sophisticated, the need for base pads will decrease and polishing uniformity across the wafer surface, including the edge exclusion region, will improve. This trend will be especially important as the industry transitions to larger diameter wafers.

6.7.3 Pad Life

Pad life is a complex issue since several factors can determine the useful life of a pad. Also, different manufacturers have different criteria for determining pad life. However, it is usually determined by the point at which some specific polishing parameter exceeds the control limit for that parameter. Such parameters can be removal rate, non-uniformity or, for shallow trench polishing, selectivity of oxide to nitride. Two events which commonly limit pad life are glazing of the pad surface and loss of macrotexture. Glazing is the build-up of polishing debris on the pad surface which clogs the pores [28, 33, 35, 36] and becomes increasingly difficult to remove by further conditioning. This problem is often seen, especially with Type 1 and 3 pads. The second event involving loss of macrotexture results from the abrasive wear of the pad during polishing, especially from the pad conditioning process, which reduces the overall pad thickness and the depth of the grooves in the pad surface. Once the grooves have been abraded, hydroplaning will occur for the reasons discussed above. Secondly, as the thickness of the top pad decreases during polishing, the pad will become less stiff and polishing planarity will decrease. Pad life may be increased by cutting deeper grooves [15, 25] and by increasing pad thickness so that pad stiffness is not compromised.

6.7.4 Planarity

At the die-level, achieving planarity of the features is an important concern and a main driving force for CMP polishing as a device fabrication step. It has been recognized for sometime [37, 38, 39, 40, 41] that stiffer and harder pads give improved planarity. As has been discussed previously, for Type 3 pads stiffness and hardness are often strongly correlated to one another and it is not easy to determine which has the greater impact on planarity.

However, it is becoming generally accepted that stiffness of the polishing pad is the more important property determining planarity. As discussed earlier, pad stiffness depends on both the pad modulus and thickness.

Experimentally, the relationship between planarity and pad stiffness has been independently verified by several researchers [42, 43, 44]. For example in an oxide CMP process, the planarization performance of a Type 3 pad (IC1000^{TM}) was compared to that of a Type 4 pad (IC2000^{TM}). With other polishing parameters kept constant, the polishing data clearly demonstrated that the stiffer IC2000^{TM} pad produced improved planarity. Other work [41, 44], and data presented earlier in this chapter, have shown that increasing the pad thickness produces a similar response and that less compressible or thinner base pads also improve planarity.

Simplistically, the improved planarity achieved by using stiffer pads comes from the reduced bowing of the pad as it planarizes. This phenomenon has been modeled by several groups [41, 44, 45], basically by treating the polishing pad as a beam in deflection. Nanz and Camilletti [46] have summarized the various modeling approaches and critiqued the strengths and weaknesses of each model. Steigerwald et al. [7] have also reviewed models describing planarity.

Grillaert et al. [41] have analyzed oxide thickness variation after CMP within a die (WIDNU) for two extreme cases: perfect pad bending and no pad bending. The perfect pad bending model consisted of a thin top pad on a soft bottom pad. In this case, the top pad bends easily and large oxide variation was found. The no pad bending case, consisting of a thick inflexible top pad, significantly reduced oxide thickness variation. Experimentation confirmed that WIDNU for a real stack pad always falls between these two extreme cases.

There is also growing evidence that pad roughness has an effect on planarity. Cook et al. [47], in attempting to analyze the physics, mechanics and chemistry of the CMP process, have concluded that the best descriptor of that process is an asperity contact model. The authors found that pad roughness influences the limits to achievable planarity, such that planarity was degraded as pad roughness increased.

Renteln and Coniff [48] have related pad roughness back to pad modulus by hypothesizing that roughness created on the surface of the pad can be treated like an additional elastic layer. This can be modeled as a two-layer composite structure where the overall modulus depends on the volume fractions (thicknesses) and moduli of the individual layers. Since the surface layer will have a more open pore structure than the bulk, it is reasonable to assume that its modulus will be lower than that of the underlying layer. Using this model, Renteln and Coniff were able to explain the experimental data. In the same paper, they also postulated that planarization rate depends solely on the elastic modulus characteristics of the pad and that anelastic effects of the pad do not play a direct role.

Using a different approach to explain the dependence of planarity on pad roughness, Yu et al. [49] have developed a statistical asperity model which relates the time-dependent deformation of asperities to the planarization of device features. The model assumes that the applied load is distributed between the asperities and the fluid film which fills the gap between the polishing pad and the wafer. At low platen speeds the hydrodynamic pressure of the fluid film is small and the load is principally carried by the asperities. The model also relates asperity size to trench widths. For wide trenches where the trench is larger than the asperity, the asperity can contact and remove material from the trench bottom. However, for narrower trenches, device geometry restricts asperity contact and polishing is selective. For wide trenches, geometry effects become insignificant and polish selectivity disappears. In order to match better the model predictions with experimental data, the authors proposed that, since the pad material is viscoelastic, an asperity does not deform instantaneously as it enters or leaves a trench. In consequence, contact to the trench bottom is reduced and polish selectivity is increased.

Grillaert et al. [50] have developed a model which attempts to explain why step height as a function of polishing time initially decreases linearly followed by an exponentially decreasing reduction as planarity is achieved. Also, their model predicts the influence of pattern dependence on step height reduction. The model assumes that the applied load is only supported by the features in contact with the pad, the pad has limited compressibility, the pad bends perfectly, Preston's equation [51] can be extended to local removal rates and pressures, and that the pad is perfectly smooth. Since the pad has limited compressibility, during initial polishing the entire load is supported by the up features only and removal rate of these features is high. As long as the step height is high enough the pad does not touch the down features and no material is removed there. As polishing proceeds step height reduces. At some point, the step height is small enough so that the pad comes into contact with the down features. When this occurs the pressure will be distributed over both up and down features and the rate of further step height reduction will decrease. Since the pad surface is not perfectly smooth but is comprised of asperities, the protruding asperities can remove material from the down features before the average pad surface is in contact with the down features.

From the above discussion, the three pad properties determining planarity are pad stiffness, pad compressibility and surface roughness from pad asperities. Also the time dependence of these properties is important and must be considered. It is also probable that the relative importance of each depends on the material being polished. To date, no published model satisfactorily combines these into a single, working, predictive model.

6.7.5 Macro-scratches and Defectivity

Although pad hardness is believed to have only an indirect impact on planarity, hardness remains an important physical property in terms of macro-

and micro-scratches, and other defects involving physical damage of the device being polished. In general, to avoid scratches, softer materials such as aluminum [52] or copper are polished with softer pads, and harder materials like oxide and tungsten use harder pads. However, there are many exceptions to this rule. For example, very soft Politex^{TM} pads are used to polish tungsten plugs. One reason to use harder pads to polish softer materials is the need for good planarity and, as previously discussed, higher hardness and stiffness are often linked.

Defectivity also appears to be related to pad porosity. One explanation is that pores in the pad surface become clogged with polishing debris which can cause scratching and leave residual particles on the wafer surface. For copper CMP, it has been shown that an experimental unfilled Type 4 pad gave a significantly lower level of defects after polishing than a standard Type 3 IC1010^{TM} pad [21].

6.7.6 Conductor Dishing and Oxide Erosion

Conductor dishing and oxide erosion are feature-level polishing parameters applicable to metal CMP and are especially important for next generation devices using copper metallization. As feature sizes decrease and conductors become smaller, copper is emerging as the preferred metallization because of its superior electrical conductivity over aluminum and tungsten. Copper damascene processes are being developed as the preferred method of manufacturing [53]. In these processes, trenches are created in the oxide dielectric layer. Copper is then deposited by electroplating, filling these trenches and covering the surface of the rest of the chip. CMP polishing is used to polish away the excess copper, leaving only the inlaid interconnect lines.

The objective of the copper polishing step is to achieve planarity of features with only minimal loss in the cross-sectional area of the conductor lines, so that high conductance is maintained. However during polishing, conductor dishing and oxide erosion occur which produce device feature deviations from the idealized damascene structures. Conductor dishing is defined as the difference in elevation of the insulator region to the metal line and oxide erosion is the loss of dielectric within the conductor feature array. Additional oxide loss, known as field oxide loss, occurs globally because the oxide polish rate is non-zero during over-polishing. Total conductor thickness loss which must be minimized is the sum of field oxide loss, local oxide erosion and conductor dishing. Figure 6.20 further illustrates the terms dishing (D), oxide erosion (E) and field oxide loss (F).

Several authors [53, 54, 55, 56, 57, 58] have independently developed models to explain the phenomenon of dishing. In general, their models are extensions of those used for within die planarity as discussed earlier, and take into account pad properties such as pad stiffness, pad compressibility, hardness and surface roughness.
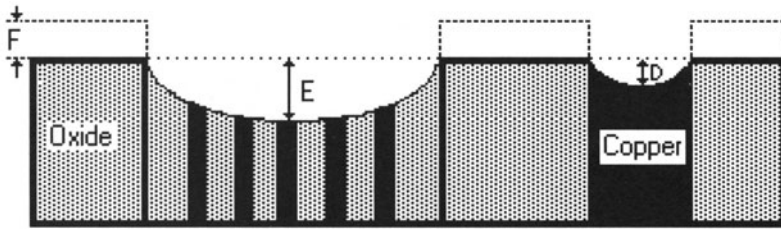
Fig. 6.20. Schematic Illustration of Dishing, Oxide Erosion and Field Loss

In an early paper on dishing of copper damascene structures, Steigerwald et al. [53] extended the pad deflection beam model of Sivaram et al. [59] to include the effects of both pad compressibility and roughness. Pad compression was taken into account by assuming that only the surface layer of the pad compresses as it pushes against the wafer and, as the pad rides over a recess, that some of the compression is relieved. The assumption was made that if pad deflection is due to pad compression, then only the near-surface layer is assumed to bend. Pad roughness was incorporated into the model using the concept of Renteln and Coniff [60], in which the rough surface layer is assumed to have a lower elastic modulus than the bulk material. In agreement with experimental observations, the model predicts that stiffer, smoother pads of low compressibility will give lower dishing.

In a more recent paper, Yang [57] has attempted to develop a quantitative model of dishing based on a Prestonian analysis of different removal rates for high and low copper features. The model incorporates terms for pad bending and pad compressibility. Calculated results of dishing are in good agreement with experimental data, such that the model accurately predicts copper dishing as a function of line-width and over-polish. The model further predicts an increase in copper dishing as feature size increases due to pad bowing. The wider the metal feature or the more compliant the pad, the more the pad can deform to remove metal within the dish. Thus the model predicts that both low pad compressibility and low conformity are desirable for reducing copper dishing.

Nguyen et al. [58] have developed an alternative model to explain dishing based on the assumption that material removal occurs predominantly at the pad/wafer contacts. Their statistical contact mechanics model assumes: a) material removal occurs at the mechanical contact between pad asperities and wafer, b) distribution of pad asperity contact size is Gaussian, c) different removal rates occur for asperities with contact size smaller and larger than the line width, d) Preston's law is valid, e) dishing occurs during over-polishing, and f) oxide erosion is neglected. Excellent agreement is obtained between experimental data and model predictions for both time dependence and feature size dependence of dishing. Their model predicts that pad surface morphology has the greatest impact on asperity contact size distribution and

hence on dishing. This intuitively makes sense given that the typical rough-
ness of a pad surface is comparable to the size of the conductor features being
polished. Also it predicts the importance of pad conditioning in producing
an optimized surface topography for polishing.

Closely related to dishing is the phenomenon of oxide erosion. This is
a thinning of the oxide layer resulting from a non-zero oxide polish rate
during the over-polish step. Like dishing, it causes a reduction in conductor
cross-section and is equally undesirable. As discussed above, less compressible
and stiffer pads are preferred to minimize dishing. However, such pads also
tend to be harder which can lead to increased oxide erosion, since harder pads
tend to have higher oxide polish rates [53]. Thus from a pad perspective, there
are opposing needs in terms of pad hardness. One solution to the problem
is to use a fairly hard pad in combination with a slurry which has a much
higher removal rate for copper than oxide. Such slurries are discussed in detail
elsewhere in this text.

Yang [57] has extended his quantitative model of dishing to describe oxide
erosion. The model accurately predicts oxide erosion as a function of pattern
density and over-polish time.

6.7.7 Pad Conditioning

As mentioned already several times in this chapter, pad conditioning has
a major impact on all aspects of polishing performance. Many pad properties
affect and are influenced by pad conditioning. As an example of the former, it
is found that pads containing porosity (Type 3) are more readily conditioned
than Type 4 pads. This is because porosity reduces the abrasion resistance
of the pad which facilitates the creation of microtexture by the diamond
conditioner. The mechanism of microtexture creation is also affected by the
tensile, hardness and modulus properties of the pad. For pads which have high
modulus and hardness with low ductility, the conditioning disk diamonds cut
channels by preferentially fracturing and removing pad material. In contrast,
for softer, lower modulus and more ductile pads, the diamonds cause plastic
flow and plough the material aside rather than by physically abrading it to
create channels. So in the former cause, pad conditioning results in higher pad
wear, whereas in the latter case pad loss during polishing is significantly less,
since the surface is essentially being rearranged. This is especially true when
polishing softer materials such as copper which do not themselves significantly
wear the pad surface.

Examples of pad properties which are affected by conditioning are, of
course, pad roughness but also pad hydrophilicity. The process of roughening
the pad surface by conditioning in the presence of either water or polishing
fluid increases the critical surface tension of the pad and makes it more hy-
drophilic. This aids the wetting of the pad surface by the polishing fluid and
allows the fluid to spread uniformly across the pad surface and under the
wafer.

6.8 Slurryless Pad Technology

Although "Slurryless Technology" is covered in a separate chapter in this book, for completion given its emerging importance, it is appropriate to include a brief discussion of Slurryless Pad Technology in this chapter.

Conventionally, CMP polishing is accomplished by using an abrasive particle-containing slurry in combination with a polishing pad. While such slurries are universally employed, it is recognized that their use gives rise to significant problems:

1. Although polishing is practiced in a clean room environment, the particles themselves are a serious source of contamination when polishing semiconductor devices and can leave residual particles on the polished wafer surface.

2. The quality of the surface produced during polishing is highly dependent upon the particle size distribution and composition of the particles in the slurry. Anomalously large particles, even in extremely small concentrations, are commonly responsible for scratches and other post-polish mechanical defects. These are highly deleterious to the yield of semiconductor devices processed by polishing. At the solids content of these polishing slurries (typically >12%) it is practically impossible to use filtration to remove the oversize particles due to clogging effects on the filter medium. Thus expensive and time consuming efforts have been made to control and reduce oversize particles in the slurries employed. However, there are few practical safeguards against their accidental introduction.

3. A desirable practice in CMP polishing is to reuse or recirculate the polishing slurry to reduce manufacturing cost and the quantity of waste products from the operation. However, the activity of polishing slurries is commonly observed to vary with time when recirculated. This may be due to the addition of dross, or polishing byproducts from the substrate into the slurry solution, attrition or breakdown of the polishing particles themselves during use, or chemical changes in the particles which reduce activity. This level of variation in recirculated slurries is unacceptably high for processing semiconductor devices. Recirculation of this system is exceedingly difficult because the byproducts of the polishing process are often practically indistinguishable from the original slurry particles, and it is equally impossible to control their size or remove them from the solution. In consequence, the solid particle content of the recirculated slurry continuously increases with time. As the polishing rate is directly proportional to the solids content of the slurry, practical control of the polishing rate is difficult. A serious additional problem is the accidental incorporation of oversize contaminant particles into the recirculating slurry, often due to substrate breakage. The aforementioned difficulties in filtering slurries make it virtually impossible to remove these contaminants.

Because of the above concerns, recirculation of slurry is not widely practiced in the polishing of most semiconductor devices because of the need to control activity precisely and the avoidance of damage by contaminants. Slurry is usually simply used once and disposed of as waste. As a result, the cost of slurry and slurry waste disposal is the single largest contributor to the cost of polishing semiconductor devices.

From the above, it is clear that a polishing process which uses a particle-free liquid which can be readily recirculated and kept in a constant particulate-free state, would be preferred in the processing of semiconductor wafers over conventional particle-containing slurries. Such particle-free liquids, commonly known as "Reactive Liquids" are beginning to find acceptance for CMP polishing [65, 66, 67]. Typically, reactive liquids are based on similar chemistries to those of conventional particle-containing slurries and are used in combination with polishing pads which themselves contain abrasive particles incorporated into the pad matrix.

Table 6.16 shows recent patents describing reactive liquid chemistries and types of slurryless pads. The reader is referred to these patents and the "Slurryless Technology" chapter of this book for a more detailed discussion of polishing using slurryless technology.

6.9 Future Trends in Polishing Pads

For the foreseeable future, CMP polishing will remain as an integral manufacturing operation in the production of semiconductor devices. As CMP polishing evolves and semiconductor devices become more complex with finer feature geometries and more metallization layers, the need for more reproducible, precision polishing performance will increase. This will necessitate improved polishing pads with less "pad to pad" variability which provide more predictable and consistent polishing performance. CMP will become more of a standard process and semiconductor manufacturers will continue to expect reduced cost of ownership for polishing consumables. This will initially impact the more established CMP polishing operations, such as oxide and tungsten polishing, but will eventually also impact still emerging processes, such as copper CMP.

With respect to polishing pads, it is expected that polyurethanes will remain as the preferred polymer type because of their attractive combination of properties, their versatile chemistry, the ability to control pad properties over wide ranges, and the ability to form pads by a wide range of manufacturing processes. Pad suppliers will continue to develop alternative, simpler pad manufacturing processes with the goals of more consistent pad properties at reduced cost.

Work will continue by pad suppliers, semiconductor manufacturers, and academics to develop a more comprehensive understanding of the relationships between pad properties and polishing performance. This work will by

Table 6.16. Patents describing Reactive Liquids and Slurryless Pads

Patent #	Assignee	Brief Description of Invention
US4,393,628, US4,466,218, US4,613,345	IBM	Slurryless pads for disk polishing, comprising a polyurethane foam containing fixed alumina abrasive
US5,152,917, US5,304,223, US5,435,816, US5,454,844, US5,549,961, US5,692,950, US5,958,794, US6,007,407	Minnesota Mining and Manufacturing Company	Extension of 3M's Microreplication technology. Photopolymerizable slurryless pads comprising a three-dimensional fixed-abrasive element
US5,725,417, US5,782,675	Micron	Conditioning and refurbishment of fixed-abrasive pads
US5,759,918	Obsidian	Linear polishing tool for fixed-abrasive pads
US5,932,486, US6,030,899	Rodel	Slurryless polishing using reactive liquids free from particulate matter in combination with polishing pads having a multiplicity of surface nanoasperities.
US6,022,264, US6,099,394, US6,069,080	Rodel	Slurryless polishing pads comprising a high modulus phase and a low modulus phase
WO99/55493	Ebara	Grinding wheel for slurryless polishing comprising abrasive grains and pores in a polymeric binder
EP 0 874 390 A1	Hitachi	Grindstone for slurryless polishing comprising abrasive grains and a bonding resin
US6,117,775	Hitachi	An abrasive-free reactive liquid for polishing metal films comprising an oxidizer and a substance which renders oxides water-soluble

necessity encompass all aspects of the polishing process, including interactions of the pad with the polishing fluid, the role of conditioning and polishing parameters, and of polishing tool design. From such work, it is expected that polishing consumable suppliers will be able to offer manufacturers a total polishing solution consisting of, for example, an optimized pad, fluid, condi-

tioner combination. Such combinations will enable synergistic performance advantages for the customer.

Industry trends that are currently in their infancy which will impact pad suppliers are the shift to 300 mm. wafers, polishing with abrasive-free reactive liquids, the increasing use of copper metallization in combination with low K dielectrics, and the growing importance of linear polishing tools. All four are driven by the need to reduce semiconductor device manufacturing costs, improve yields and performance, and to increase throughput. These issues are separately discussed in detail elsewhere in this book. The important point to note is that the pad can no longer be developed in isolation from the rest of the polishing system but must be developed and optimized with respect to the polishing tool, the type of wafers being polished, and the polishing fluid.

6.10 Acknowledgements

References

1. T. Izumitani, "Polishing, Lapping and Diamond Grinding of Optical Glasses", in "Treatise on Material Science and Technology", **17**, eds. M. Tomozawa and R. Doremus, Academic Press, New York, 1979.
2. L. Cook, J. Non-Crystalline Solids, **120**, 152, 1990.
3. A. Kaller, Glastech. Ber., **64** (9), 241, 1991.
4. S. Murarka, J. Steigerwald and R. Gutmann, "Inlaid Copper Multilevel Interconnections using Planarization by CMP", MRS Bulletin, 46, June 1993.
5. L. Shon-Roy, "CMP: Market Trends and Technology", *Solid State Technology*, 67, June 2000.
6. A. Braun, "CMP Battles Low-K, Fundamental Barriers", *Semiconductor International*, 66, October 2000.
7. J.M. Steigerwald, S.P. Murarka and R.J. Gutmann, "Chemical Mechanical Planarization of Microelectronic Materials", Wiley, 1997.
8. "Polymer Handbook", Second Edition, J. Brandrup and E.H. Immergut, Editors, Wiley and Sons, 1975.
9. G. Woods, "The ICI Polyurethanes Book", J. Wiley & Sons, 1990.
10. G. Oertel, "Polyurethane Handbook", Second Edition, Hanser, 1994.
11. M. Szycher, "Handbook of Polyurethanes", CRC Press, 1999.
12. R.W. Lenz, "Organic Chemistry of Synthetic High Polymers", Interscience, 1967.
13. L.M. Cook, "CMP Consumables II: Pad", Semiconductors and Semimetals, **63**, Chapter 6, 155-181, Academic Press.

14. H.F. Mark and N.G. Gaylord (editors), "Encyclopedia of Polymer Science and Technology", **8**, "Leather-Like Materials", 210, 1968.
15. L. Cook, J. Roberts, C. Jenkins and R. Pillai, US Patent 5,489,233, "Polishing Pads and Methods for Their Use", 1996.
16. F. Billmeyer, "Textbook of Polymer Science", J. Wiley, 1972.
17. J. Aklonis, W. MacKnight and M. Shen, "Introduction to Polymer Viscoelasticity", Wiley-Interscience, 1972
18. M. Fury and D. James, "Relationships between Physical Properties and Polishing Performance of Planarization Pads", SPIE Microelectronic Manufacturing Symposium, Austin, October 16, 1996.
19. P. Freeman, D. James and L. Markert, "A Study of the Variation of Physical Properties in Random Lots of Urethane Polishing Pads for CMP", Surface Technology, Vol. 2, Issue 6, June 1996, Rodel Publication.
20. A. Vishwanathan and D. James, "Next Generation Polishing Pad for Copper Damascene CMP", Internal Rodel Communication.
21. D. James, "Control of Polishing Pad Physical Properties and Their Relationship to Polishing Parameters", CAMP Fifth International Symposium on CMP", Lake Placid, August 13, 2000.
22. R. Baker, Electrochemical Soc. Proc., **96-22**, 228, 1996.
23. Machinery's Handbook, 23rd Edition, 297.
24. I. Sohn, B. Moudgil, R. Singh and C. Park, "Hydrodynamics of a CMP Process", Mat. Res. Soc. Symp. Proc., **566**, 181, 2000.
25. S. Huey, S. Mear, Y. Wang, R. Jin, J. Ceresi, P. Freeman, D. Johnson, T. Vo and S. Eppert, "Technological Breakthrough in Pad Life Improvement and its Impact on CMP CoC", Tenth Annual IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop, 54, 1999.
26. J. Levert, F. Mess, L. Grote, M. Dmytrychenko, L. Cook and S. Danyluk, "Slurry Film Thickness Measurements in Float and Semi-Permeable Polishing Pad Geometries", Proceedings of the International Tribology Conference, Yokohama, 1995.
27. M. Weling, C. Drill, W. Parmantie and G. Fawley, Proceedings 1996 CMP-MIC Conference, 40, IMIC, Tampa, 1996.
28. R. Bajaj, M. Desai, R. Jairath, M. Stell and R. Tolles, "Effect of Polishing Pad Material Properties on CMP Processes", Mat. Res. Soc. Symp. Proc., **337**, 637, 1994.
29. R. Baker, Proceedings 1997 CMP-MIC Conference, 339, IMIC, Tampa,1997.
30. D. Stein, D. Hetherington, M. Dugger and T. Stout, J. Elect. Mat., **25**, 1623, 1996.
31. R. Baker and S. Lane, "Pad Conditioning for Next Generation CMP Consumables", CMP Technology for ULSI Interconnection, SEMI 1998, M1-10.
32. D. Hetherington, "Tribological Evaluation of Polyurethane CMP Pads", CMP Technology for ULSI Interconnection, SEMI 1998, L1-19.
33. M.R. Oliver, R.E. Schmidt and M. Robinson, "CMP Pad Surface Roughness and CMP Removal Rate", ECS Meeting, Phoenix, October 2000.
34. D. Goetz, Mat. Res. Soc. Symp. Proc. **566**, 51, (2000)
35. S. Sivaram, H. Bath, R. Leggett, A. Maury, K. Monnig and R. Tolles, *Solid State Technology*, 87, May 1992.
36. I. Ali and S. Roy, "Pad Conditioning in Interlayer Dielectric CMP," *Solid State Technology*, 185, June 1997.

37. R. Kolenkow and R. Nagahara, "Chemical Mechanical Wafer Polishing and Planarization in Batch Systems", So*lid State Technology*, 112, June 1992.
38. I. Ali, S. Roy and G. Shinn, "CMP of Interlayer Dielectric: A Review", *Solid State Technology*, 63, October 1994.
39. R. Jairath, M. Desai, M. Stell, R. Tolles and D. Scherber-Brewer, "Consumables for the CMP of Dielectrics and Conductors", Mat. Res. Soc. Symp. Proc., **337**, 121, 1994.
40. R. DeJule, "CMP Challenges below a Quarter Micron", *Semiconductor International*, 54, November 1997.
41. J. Grillaert, M. Meurix, E. Vrancken, N. Heylen, K. Devriendt, W. Fyen and M. Heyns, "Modeling the Influence of Pad Bending on the Planarization Performance during CMP", Mat. Res. Soc. Symp. Proc., **566**, 45, 2000.
42. S. Pangrle, I. Salugsugan, A. Dangca, M. Segovia, D.M. Schonauer and D. Erb, "Polishing Performance of the Rodel EX2000 Pad", Proceedings 1996 CMP-MIC Conference, 47, IMIC, Tampa, 1996.
43. J. Grillaert, H. Meynen, J. Waeterloos, B. Coenegrachts and L. Vandenhove, "Minimizing Within Die Non-Uniformity in CMP by Optimizing Polishing Parameters and Consumables", Advanced Metallization and Interconnect Systems for ULSI Applications in 1996, MRS, 525, 1996.
44. J. Grillaert, M. Meurix, E. Vrancken, N. Heylen, K. Devriendt, W. Fyen and M. Heyns, Mat. Res. Soc. Symp. Proc., **566**, 45, 2000.
45. J. Warnock, J. Elecrochem. Soc., **138** (8), 2398, 1991.
46. G. Nanz and L. Camilletti, IEEE Trans. on Semi. Manuf., **8** (4), November 1995.
47. L. Cook, J. Wang, D. James and A. Sethuraman, *Semiconductor International*, 141, November 1995.
48. P. Renteln and J. Coniff, Mat. Res. Soc. Symp. Proc. **337**, 105, 1994.
49. T. Yu, C. Yu and M. Orlowski, "A Statistical Polishing Pad Model for CMP", Tech. Dig. 1993 IEDM, 865, 1993.
50. J. Grillaert, M. Meuris, N. Heylen, K. Devriendt, E. Vrancken and M. Heyns, Proceedings 1998 CMP-MIC Conference, 79, IMIC, Tampa, 1998.
51. F. Preston, J. Soc. Glass Tech., **11**, 214, 1927.
52. V. Sachan, N. Chechik, P. Lao, D. James and L. Cook, Proceedings 1998 CMP-MIC Conference, 401, IMIC, Tampa, 1998.
53. J. Steigerwald, R. Zirpoli, S. Murarka, D. Price and R. Gutmann, J. Electrochem. Soc., **141** (10), 2842, 1994.
54. G. Wu and L. Cook, "Mechanism of Copper Damascene CMP", Proceedings 1998 CMP-MIC Conference, 150, IMIC, Tampa, 1998.
55. J. Grillaert, E. Vrancken, W. Fyen, M. Meuris and M. Heyns, "The Basic Mechanism of CMP of Cu and TaN", IMEC 1999.
56. J. Pan, P. Li, K. Wijekoon, S. Tsai and F. Redeker, "Copper CMP Integration and Time Dependent Pattern Effect", Proceedings 1999 International Interconnect Technology Conference, 164, 1999.
57. L. Yang, "Modeling CMP for Copper Dual Damascene Interconnects", Solid State Technology, 111, June 2000.
58. V. Nguyen, P. Van der Velden, R. Daamen, H. Van Kranenburg and P. Woerlee, "Modeling of Dishing for Metal CMP", 46th Annual IEEE International Electron Devices Meeting (IEDM), San Francisco, December 11, 2000.
59. S. Sivaram, H. Bath, E. Lee, R. Legett and R. Tolles, Advanced Metallization for ULSI Applications in 1991, MRS, Pittsburgh, 511, 1992.

60. P. Renteln and J. Coniff, Proceedings of the MRS 1994 Spring Meeting, **337**, 105, 1994.
61. A. Krishnashree, "Evaluation and Characterization of Polyurethane CMP Polishing Pads", Ph.D. Thesis, Clarkson University, Potsdam, NY, 1998.
62. K. Torii, "Polishing Device having a Pad which has Grooves and Holes", U.S. Patent 5,725,420, March 1998.
63. T. Zhang, "Report on Thicker OXP3000^{TM} Pad Evaluation", Rodel Internal Communication, 2000.
64. T. Zhang, "Effects of Polishing Down-force and Sub-Pad on Planarization Performance of OXP3000^{TM} Pad", Rodel Internal Communication, 1998.
65. M. Oliver, "Copper CMP with Reactive Liquid Technology", CAMP Fifth International Symposium on CMP", Lake Placid, August 13, 2000.
66. Y. Shimamura and Y. Kamigata, "Newly Developed Abrasive Free Slurry for Copper Interconnections", CAMP Fifth International Symposium on CMP", Lake Placid, August 13, 2000.
67. J. Amanokura, Y. Shimamura, H. Habiro, Y. Kamigata, H. Suzuki, and M. Hanazono, "Advanced Cu CMP Slurry for Sub-micron Generation", CAMP Sixth International Symposium on CMP", Lake Placid, August 12, 2001.

7 Fundamentals of CMP Slurry

Karl Robinson

7.1 Introduction: Basic Components of CMP Slurries

Slurries are not new. In some form or another they have been around since individuals began polishing surfaces. The reasons vary depending upon the times, such as inlaid metal on Roman shields or Galileo's first lenses. However they all consisted of the same components, a solution and an abrasive. It was a natural technology transfer of the lens polishing slurry to semiconductor wafer polishing slurries in the early 1980's. The reasons were the need to planarize what was then advanced next-generation IC devices in a manner more reproducible and reliable than existing technology allowed. Although the application of the slurry was an advance in the semiconductor technology, the slurry composition remained unchanged from the composition used in lens polishing. However the devices soon became more complex and CMP related defects became more apparent as yield limits. This resulted in more attention to the finer details of CMP, in particular the slurry.

Realistically there is a large amount of interaction between the components in CMP slurries. However, taken individually the basis of each component can be broken down as such:

a. Role of Abrasive

The abrasive component, whether delivered in a solution phase or abraded from a solid phase, is generally thought to provide the mechanical part of CMP [1]. The abrasive particle impacts the surface and abrades the chemically treated surface exposing new material for chemical attack. At this point the first distinction between metals CMP and dielectric CMP needs to be made. Metals CMP does not require, to date, chemical activity from the particle. Dielectric CMP does require chemical activity along with the mechanical abrasion. Regardless of the CMP process, there is a considerable amount of chemistry involved in the particle technology. To begin with, the particles must have the correct surface charge to stay suspended, the correct hardness to impact the wafer surface, the correct chemical properties to not dissolve in the solution and, particularly for dielectric CMP, the correct chemical bonds to adhere to the wafer

surface. The result of these needs has been the introduction of various particles and methods of making particles.

b.   Role of Solution

The solution component plays several roles in CMP. Foremost is the role of providing chemical agents that attack the surface to be polished. These agents can be in the form of pH adjusters, oxidizers, catalysts and inhibitors. Additionally, the solution must provide an electrostatic or steric balance that stabilizes the abrasive suspension. Electrostatic stabilization comes from the repulsive forces resulting from like charges in a solution. Steric stabilization comes from the introduction of secondary components such as particles or high-molecular weight organic compound that physically intervene between two particles. The solution also plays a significant mechanical role in CMP being a lubricant between wafer and pad, transporting waste material and controlling temperature. The interaction between particle and solution can greatly change the overall CMP process. To address many of these demands, the concept of in-line mixing of components has become accepted. This approach is believed to minimize interaction times between various components and reduce the negative side effects.

c.   Pad Interaction. Although the pad is discussed in detail elsewhere, it is worth noting that the pad is the primary means of transporting slurry to the wafer surface. Although there have been several novel equipment modifications to minimize the impact of pad variations in this role, the pad still provides the "reaction chambers" within which CMP occurs. In this sense, the pad must be inert to the abrasive and solution chemistries and their by-products. Also, the pad must provide a steady-state environment for the CMP reactions, not just during an individual CMP wafer step but also wafer-to-wafer, lot-to-lot, tool-to-tool, fab-to-fab and device-to-device.

CMP is now used in several process modules. Each of the currently important modules, ILD CMP, STI CMP, tungsten CMP and copper CMP, requires very different particle, chemical and process properties to be effective. Correspondingly, there is also an increasing degree of complexity in the slurry as these processes are integrated into device manufacturing. ILD CMP requires the ability to polish oxide films (PTEOS, BPSG, HDPTEOS) to planarity but need not provide any selectivity to other materials as this is a stop-in-film polish. STI represents a new ILD process that still planarizes oxide films, but stops at a silicon nitride interface, thus requiring selectivity between two dielectric materials. Tungsten CMP is used to isolate conductive plugs and vias through a damascene process. Here the films are very different and consist of the W metal and a liner, often Ti/TiN. The CMP step must

uniformly polish through the metals and stop on the supporting dielectric film. This requires a high selectivity, as measured by removal rate, between metal and dielectric. Copper CMP is similar except that the copper does not self-passivate and is chemically very active, thus limiting the available oxidizers in the solution. Unfortunately for the CMP process this chemically active copper film is usually deposited on an electrochemically inactive tantalum or tantalum nitride film. This combination has proven very difficult to polish and meet the integration needs of a copper device.

This increasing complexity of the integration needs has provided a huge challenge to slurry suppliers. The demand for improved uniformity, planarity and defectivity along with new processes, has pushed slurry development from an empirical study into the fundamentals of surface science, colloidal science, electrochemistry and rheology. In this chapter, the fundamental role of the various components in slurry technology is described in relation to the fundamental sciences. Although the components will be covered separately, it will become obvious that they are very inter-related.

The perspective of this chapter is that of a CMP process engineer with a fundamental understanding of the interactions within CMP slurries. The principles discussed here form a basis for a dialogue between the process engineer who must work with the slurries and the specialty chemist who creates them. There is a large amount of specialty proprietary chemistry that goes into the choices of chemical compounds in making slurry, as demonstrated by the large number of referenced patents as opposed to refereed publications, that is not covered in this section.

7.2 Surface Science and Electrochemistry in CMP Slurry

7.2.1 Dielectrics

The dielectric surface of most current interest in semiconductor is that of silicon dioxide (oxide). Low-k dielectrics will become employed more as technology performance continues to improve and will eventually need to be processed by CMP. Currently there is still a lot of learning to be gained on oxide CMP that can be later applied to low-k materials. There are several types of oxide that are subject to CMP, such as P-TEOS, HDP-TEOS, and BPSG. The chemical nature of CMP is similar for each oxide although resultant performance can be very different. Oxide surfaces typically exhibit a termination structure as shown in Fig. 7.1. This surface structure changes with pH of the solution. At a pH = 2.2 the point of zero charge, pzc, is reached in which the oxide surface switches from a positively charged surface to a negatively charged surface [1, 2]. This negative charge is due to the accumulation of OH-groups.
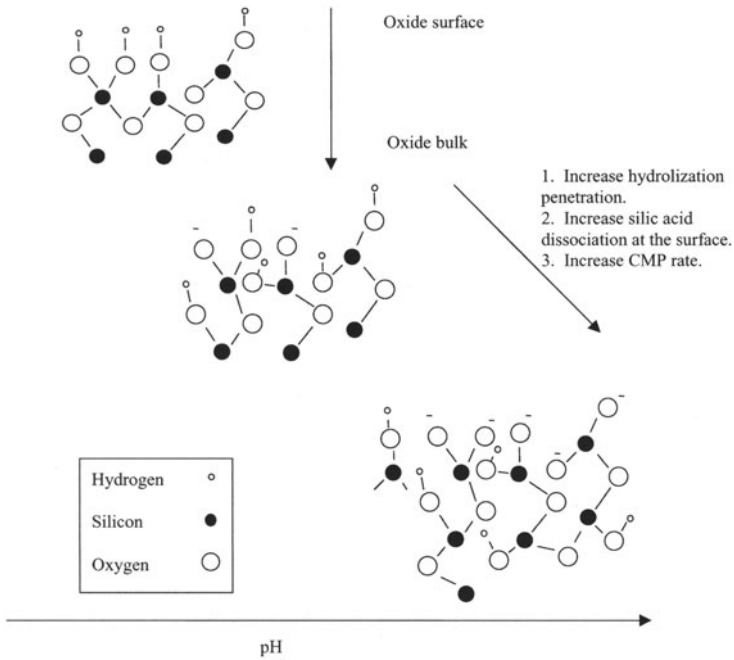
Fig. 7.1. Oxide surface changes with pH. Increasing pH causes increase in hydrolization penetration and dissociation of the silic acid and charging of the surface

As the pH increases the increasing concentration of the OH- groups, via attachment with water molecules diffusing into the oxide, weakens the oxide surface in a hydrolyzing process [1]. Eventually at a pH $\sim$ 12, the silicon oxide begins to dissolve in the solution. Since dissolution of the surface would result in an isotropic etch of the surface and not provide any planarization CMP chemistries are kept below a pH of 12. Particles are used to impact the surface and remove hydrolyzed oxide such that topography is removed in an anisotropic method [2]. A balance of hydroxylation rate and removal rate by CMP provides a pseudo-steady state relation in which the surface is weakening at a rate equivalent to the effective number of particle impacts. The term effective alludes to the indication that not every particle impact results in a contribution to the overall removal rate [1, 3].

In the case of thermal silicon nitride, Si_3N_4, the surface reaction is two-step reaction. First the nitride reacts to form a SiO_2 interface followed by a subsequent oxide removal via hydrolization and particle impact.

$$Si_3N_4 + 10H_2O \rightarrow 3SiO_2 + 4NH_4OH$$

$$SiO_2 + OH^- \rightarrow SiO_3H^-.$$
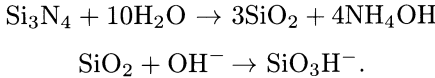
In typical oxide slurries the nitride reaction is the limiting factor giving a selectivity of oxide: nitride $\sim$ 4. As the pH is reduced, the reaction rate
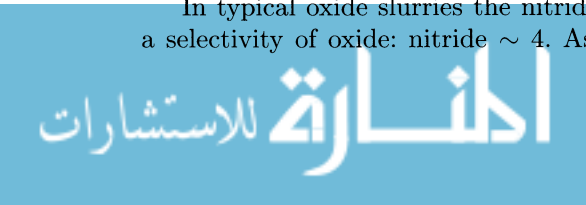
of the conversion of silicon nitride to silicon oxide is reduced more rapidly than the CMP removal rate of the silicon oxide which results in a slight increase in the selectivity. Improvements in selectivity between silicon oxide and silicon nitride can also be achieved by inhibiting the initial conversion of silicon nitride to silicon oxide. Although there are several possible approaches to inhibit this conversion the most common methods are to add blocking agents or inhibitors to the slurry. Inhibitors, such as fluorinated organics [4, 5, 6], reduce the oxide formation reaction by binding to the silicon nitride surface and chemically inhibit the reaction further increasing the selectivity. Binding agents, such as low-molecular weight polymers, bind to the surface and physically (sterically) inhibit the formation of the oxide. The results of these additives are high selective slurries that enable stop on nitride STI CMP in a one-step process as opposed to the more expensive reverse mask process as described in the integration chapter.

7.2.2 Metals

CMP of metals in semiconductors was first introduced as a method of isolating interconnects, or vias, between metal layers. The process replaced metal etchback that left the dielectric surface rough and the metal recessed. The approach is simple; etch a contact, overfill the metal and polish back to stop on the oxide and isolate the contact. The details of tungsten CMP are discussed elsewhere in this book, however it is a good example of the use of electrochemistry in CMP [7, 8, 9]. In addition, several common IC device films and their half-cell reaction and resulting oxidation potential as well as several oxidizers and their half-cell reaction and potentials are summarized in the Appendix.

Since it is not feasible, to date, to directly charge a wafer surface on a production CMP tool, the slurry must contain the oxidizing agents, which control the local electrochemical potential. A key element in the choice of an oxidizer is the operational pH of the oxidizer and the resultant metal oxide film. For example, a weak oxidizer with a potential in the 0.2–0.4 V range can still etch Cu at low pH, however as the pH increases the Cu first forms a passivating film of Cu_2O which is stable to very high pH. A strong oxidizer, like hydrogen peroxide with a potential of 1.7 V, again etches Cu at low pH. However the passivation region is smaller at a higher pH. This may lead to an incomplete passivation layer leaving the surface vulnerable to further attack. This additional attack under non-passivating conditions is termed corrosion and results in a non-controlled oxidation and dissolution of the copper layer. This phenomena is fundamentally related to the thermodynamic stability of the $Cu-H_2O$ system which is conveniently summarized as a Pourbaix diagram [10] in the Appendix. However, it cannot be overemphasized that the kinetics of the oxidation or passivation reactions is not related to the Pourbaix diagram, which is only a description at equilibrium. Of course the addition

of other slurry components will affect the Pourbaix diagram according to the Cu-component interaction.

In general, most metals readily oxidize in the presence of hydrogen peroxide, potassium iodate or ferric nitrate. These are the most widely used oxidants for metal CMP. The slurry must compensate for competing oxidation reactions of etching and passivation. Etching results in the direct dissolution of the metal into the solution as an ionic species. Passivation results in an oxidation film that coats the metal surface and protects it from corrosion, a form of rapid oxidation and dissolution. A third but less common reaction is the direct oxidation into a volatile oxide, such as Ru into RuO_4. This should be avoided as some metal oxides and in particular RuO_4, have toxic properties. Care must be taken to not use an oxidizer that produces these volatile metal oxides directly.
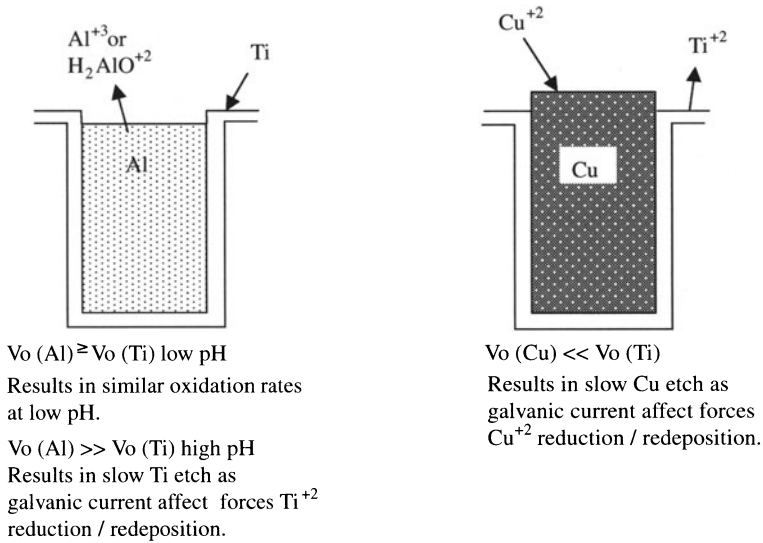
In the simplest version of tungsten CMP the resultant metal oxide passivates the metal protecting the W film from corrosion. Initial work suggested that the passivation layer is then removed by means of particle impact similar to the hydrolyzed oxide layers [11]. Subsequent work with more accurate electrochemical cells has shown the initial work accounts for only a percentage of the removal rate [9]. In fact the electrochemical data suggests that the process does not require a blanket passivation film for high CMP rates. Corrosion appears to play a role via grain boundary corrosion that results in unoxidized W crystallite removal during CMP. The oxidation reactions proceed readily at low pH which not only facilitates the W oxidation reactions but also provides selectivity to the supporting dielectric which has a minimum of hydrolization at low pH. The typical W liner films of Ti and TiN behave in a similar to W with even higher oxidation rates.

Metal CMP slurries also must prevent the formation of undesired particles formed during metal oxidation. Unlike copper and aluminum CMP, which produce soluble ions or compounds upon oxidation, tungsten CMP produces low solubility oxides in solution. Due to this difference in polishing by-product, tungsten slurries must also prevent the agglomeration of tungsten oxide particles in the solution. Such large agglomerates may result in increased defects on the post CMP wafer. One mechanism is to include organic additives in the slurry that adsorb on the metal particle surface and prevent further reaction with the wafer interface inhibiting further growth. An alternative approach to reducing agglomeration of insoluble metal oxides is to use additives that inhibit formation of the oxide. In the case of tungsten, chelating agents can be added that prevent formation of a stable metal oxide compound [9, 11] which can precipitate out of solution to form particles. The by-product in this case is a soluble form of tungsten oxide that does not precipitate out of solution to form agglomerations.

The electrochemical learning associated with tungsten CMP has been applied to copper CMP slurry development. However substantial differences exist in their electrochemical behaviors. Unlike tungsten, copper films do not

form self-passivating oxide films. The copper oxide films are porous in nature and are soluble in aqueous solutions exposing areas of bare Cu film for further oxidation. This non-grain boundary oxidation leads to rapid film corrosion in solution requiring the presence of oxidation inhibitors in the slurry to prevent Cu corrosion post CMP. The integration requirements of copper and barrier CMP are so demanding that the first generations of copper slurries generally have been two-step slurries [12]–[16]. Typical first step slurries rely on the ability to distinguish between the oxidation of Cu and Ta or TaN to improve selectivity [17]. The "stop-on-barrier" scheme is followed by a second polish that use slurries that match the barrier polish rate to either that of the Cu film (selective approach) or that of the dielectric (non-selective approach). A more advanced interconnect replaces the barrier metal with a material that oxidizes and polishes at similar rates as copper providing a one-slurry approach. More details of advanced copper integration are discussed elsewhere and note that no one chemical and abrasive solution will meet all the integration methods.

It is important to consider that the exposure of two metals in contact with an electrolyte sets up a galvanic cell. With the different half-cell reactions, a potential is setup between the two metal films forcing one film to be the cathode and the other the anode. In an electrolyte that would normally corrode Cu, the exposure of the Ta barrier metal and the resultant oxidation forces the Cu to become electron rich. To complete the circuit, Cu ions in the electrolyte are reduced to Cu metal on the wafer. Addition of complexing and inhibiting agents reduces the galvanic reactions by either inhibiting the

V_o (Al) $\geq V_o$ (Ti) low pH
Results in similar oxidation rates at low pH.

V_o (Al) $>> V_o$ (Ti) high pH
Results in slow Ti etch as galvanic current affect forces Ti^{+2} reduction / redeposition.

V_o (Cu) $<< V_o$ (Ti)
Results in slow Cu etch as galvanic current affect forces Cu^{+2} reduction / redeposition.

Fig. 7.2. Galvanic affect of barrier and line metals during CMP due to the electrochemical half-cell reaction potential (V_o)

Ta oxidation or shielding the Cu ion from the reduction reaction. In either case, the galvanic cell provides a method for producing very high selectivity between the Cu and the Ta barrier film.

Ti barrier metals produce similar properties, as pictured in Fig. 7.2, at a much wider pH condition than do Ta barrier metals.

7.3 Slurry as a Suspension

As discussed in the previous section, ionic species are present in both oxide and metal CMP slurries. In general, the multiple chemical components can create a wide range of slurry properties that control the CMP process. The suspension plays multiple roles. First, the slurry provides the chemical and mechanical action of planarizing the wafer. Second, the slurry acts as a lubricant to reduce the friction between the pad and the wafer. Third, the slurry acts to dissipate the heat generated by the friction of the CMP process. Lastly, the slurry acts as a transport medium for moving reactants and particles to the wafer interface and then the byproducts of the CMP process away from the wafer interface. The pad plays the role of ensuring the uniformity of each of these roles across the wafer surface. It is also very important to consider the by-products of the surface reactions. Their presence changes the balanced electrostatics of any colloidal suspension and at the most critical point, the wafer surface during CMP.

7.3.1 Colloid Science and Rheology in Slurry

One visually striking slurry property is its settling rate. In a typical settling rate experiment a column of slurry is allowed to sit while the rate of particle settling is measured. The slower the settling rate, the longer the slurry stays in suspension and the longer the properties remain uniform. A higher settling rate is indicative of an unstable suspension. Without going into the full derivation, the mass and density of the particles are proportional to the settling rate, v, by [18]

$$m(1 - \rho_1/\rho_2)g = fv, \tag{7.1}$$

where m is the mass, ρ_1 is the density of the aqueous phase, ρ_2 is the particle density, g is the gravity constant and f is the friction factor. This can be related to the diffusion coefficient by

$$f = kT/D, \tag{7.2}$$

where k is the Boltzmann constant, T is temperature and D is the diffusion coefficient. The settling rate can also be related to the particle size, assuming a spherical particle and little or no dissolution in the aqueous phase, by

$$R = (9\eta v/2(\rho_2 - \rho_1)g)^{1/2}, \tag{7.3}$$

where R is the radius, and η is the viscosity. Although simple and not completely applicable to complex slurry systems, these expressions can be used to provide some measure of slurry stability and quality. A standard use of the settling phenomena is in slurry decanting. In particle manufacturing it is possible sometimes to form large hard agglomerates which prior to polishing must be removed from the smaller desirable particles. Decanting allows the larger particles time to separate from the slurry through the faster settling rate.

Equation (7.3) contains another important parameter, η, the viscosity of the solution. Viscosity is actually a measurable coefficient describing the response of a fluid to an applied force per unit area. Figure 7.3 shows several relationships between the applied force and the rate of shear for several materials [18]. Most solutions are Newtonian in that viscosity, measured as the slope of the force-shear relation, is constant. In slurry behavior, viscosity is important to understand when one considers that a CMP tool can be considered a giant plate viscometer. However most viscometers work in the range of a 100 s^{-1} to 1000 s^{-1}, while the CMP tool is estimated to produce shear rates greater than 10^6 s^{-1} [19]. At these shear rates, the applied forces to the CMP process are large and any fluctuations in this force, due to viscosity changes, can influence the overall stability of the CMP removal rates. Thus the effect of slurry particles on viscosity is important to understand.

Viscosity also plays an important role in determining the flow of slurry under the wafer. The Reynolds number is a widely used dimensionless term re-
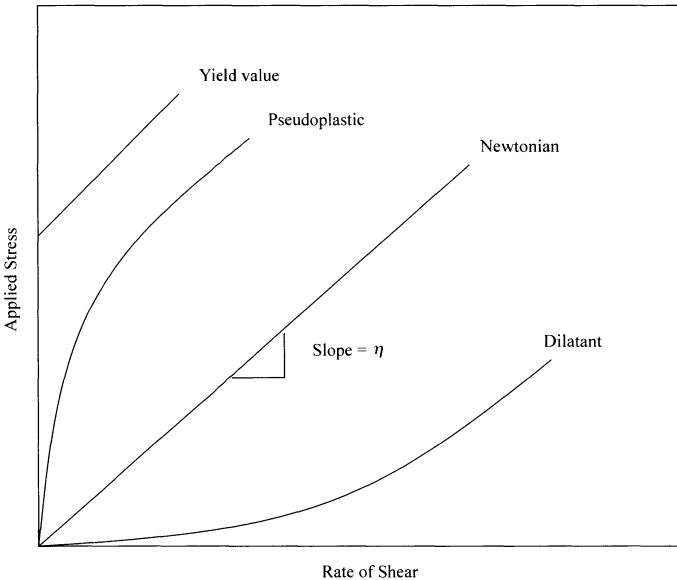

Fig. 7.3. Response of applied force to shear rate for various fluids. From [18]

lated to the flow characteristics. In a parallel plate relationship, the Reynolds number is defined as $LV\rho/\eta$, with L being the gap between the plates, V being the velocity of the medium, ρ, the fluid density and η, the viscosity. As the Reynolds number increases, the fluid flow changes from laminar to turbulent. The two parameters most difficult to calculate are L and η. Laminar flow is described as a smooth transition with little or no mixing of fluid layers. In laminar flow particles in the bulk are not transported to the interface of the wafer and that the film layer is relatively thick compared to the particle size. Turbulent flow is described as a chaotic mixing at the interface with a thin boundary layer. Which flow pattern is occurring during CMP is currently under investigation by several groups and is not resolved [19, 23]. The pad complicates the Reynolds number calculation as the contours of the pad, asperity, groove and soak-length, are large enough to move the Reynolds number from laminar to turbulent flow. The gap also affects the viscosity of non-Newtonian fluids as the shear rate increases. To make some sense from this an understanding of how slurry viscosity is affected is in order.

The simplest relationship is Einstein's law of viscosity [18],

$$\eta/\eta_o = 1 + 2.5\omega, \tag{7.4}$$

where η_o is the viscosity of the aqueous phase, η is the viscosity of the slurry and ω is the volume fraction, which is related to % solids, φ, as $\omega = \varphi\rho_1/\rho_2$. It is important to point out several assumptions in (7.4), mainly that the particles are monodisperse spheres, the slurry is stable (particles are not flocculating or dissolving), the flow rates (or shear rates) are low and the slurry is Newtonian. Equation (7.4) is valid to $\varphi \sim .1$, above which the dispersion viscosity takes on more complexity. For a typical silica particle, $\rho \sim 2.2$, this translates into a wt.% solids number of ~ 20 wt.%. Most fumed silica slurries report wt.% solids below 15 wt.%, except when delivered as a concentrate, within the limits of (4). Precipitated silica reports values of ~ 30 wt.% solids, making the viscosity of this slurry very susceptible to minute changes in the particle concentration. By contrast, ceria, $\rho \sim 7.1$, may contain up to 70 wt.% solids and still have a predictable viscosity, provided it will stay in suspension.

Of particular concern in the use of the Einstein relationship are the very high shear rates in CMP. Although most slurries are Newtonian in behavior, flocculated colloidal particles are a class of slurries that can exhibit non-Newtonian behavior [18] at various pH levels [2]. In these slurries, the apparent viscosity changes with the rate of shear. In particular, certain precipitated slurries, which have symmetrical particles, exhibit increased viscosity under high shear, a behavior termed dilatancy [18]. At very high shear, these particles show a tendency to gel producing a network of loosely bound particles, which leaves a residue on the wafer surface. Fortunately this gel is readily cleaned with typical post CMP HF based cleans. How this gelation affects planarization and uniformity is dependent upon the wafer topography, film

hardness, polish pressure, particle surface charge and particle size distribution. Slurry suppliers can do little about the first conditions, however surface charge and particle size distribution are results of the solution pH, particle manufacturing process and the mixing mechanism of the slurry. For example, a bimodal mixture of symmetrical particles can be used to target a certain concentration of solids. Dependent upon the size difference between the two particle groups, the surface area will be dominated by the smaller particles. The surface area is critical in gelation as the smaller particles can fill the void between large particles increasing the strength of the gel. This leads to the possibility that minimal variations in solids concentrations can result in very large variations in gelation.

As much as the solution modifies the surface to be polished, the solution also greatly affects the particles suspended in the solution. Slurries are composed of both ionic species and organic species, all of which must interact in a manner to stabilize the suspension, promote the CMP process and be reproducible. The next two sections cover these two classes independently.

7.3.2 Ions in Slurry

To fully comprehend the effects that ionic species have on slurries, it is important to understand how particles stay in suspension. The settling rate expressions relate particle size to settling rate. Below some particle size, the gravitational settling force is less than the diffusion force and the particles stay suspended. In a simple model, a suspension becomes unstable when particles begin to flocculate and reach a critical accumulated size at which settling commences. There are two primary methods for preventing flocculation, electrostatic and steric stabilization. Electrostatic stabilization is the use of repulsive electrical fields between like charged particles to keep particles from agglomerating. Steric stabilization is the use of organic networks, or polymers, adhered on particles to physically keep particles from agglomerating. Although there is some indication that surfactants and other large molecular weight molecules contribute to steric stabilization, in conventional slurries the dominant method is electrostatic.

A particle in contact with an ionic solution exhibits several "layers" depicted in Fig. 7.4 [18]. The Stern layer accounts for molecular adsorption of ionic species at the particle interface. The shear layer accounts for the beginning of fluid flow in the boundary layer around the particle. The electric double layer is the projected electric field from the particle. A frequently quoted particle property is the zeta-potential and is often misrepresented as the charge at the particle surface. In reality, the zeta-potential is the charge at the shear surface. The magnitude of the zeta-potential is dependent upon the particle type and the solution pH. It can vary from 0 mV at the pzc up to true particle surface charge in a vacuum. There are expressions that relate the true surface charge to the zeta-potential however they contain several parameters that are not experimentally measurable.
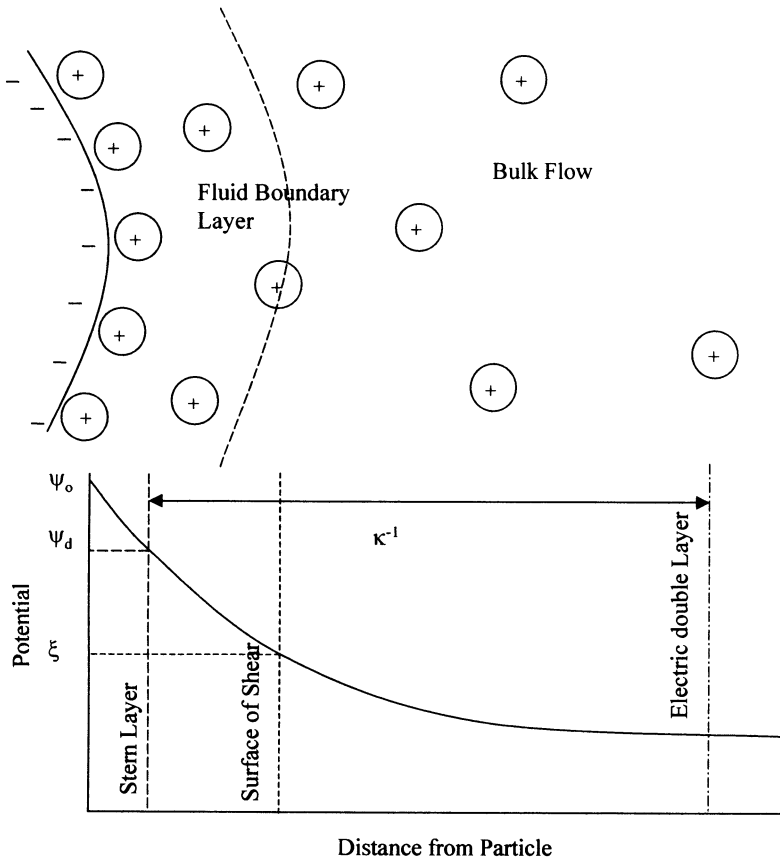
Fig. 7.4. Description of the electrostatic layers and their respective potentials of a particle in an electrolyte. The double layer extends beyond the hydrodynamic boundary layer at this ionic strength. From [18]

The electric double layer dimension, κ^{-1}, is directly calculated from the ionic strength of the solution by [18]

$$\kappa = \left(\frac{1000 e^2 N_A}{\varepsilon k T} \sum_i z_i^2 M_i \right)^{1/2}, \tag{7.5}$$

where e is the electron charge, N_A, is Avogadro's number, ε, is the dielectric constant of the medium, z_i is the ionic valence and M_i the molar concentration of species i. A high ionic strength produces a small double layer, typically on the order of the particle radius also known as the Helmholtz–Smoluchowski limit. Conversely, a weak ionic strength produces a large double layer, typically 100 times the particle radius, also known as the Huckel limit. When two particles approach each other, the double layers begin to overlap. At this point competing forces of attraction, Van der Waals forces, and repulsion,

ionic forces, determine if the particles repel and remain in suspension or begin to flocculate. The combination of these forces is known as the DLVO (short for Derjaguin, Landau, Verwey and Overbeek) theory [18],

$$\Phi = \frac{64 n_o k T \Upsilon_0^2}{\kappa} \exp(-\kappa d) - \frac{A}{12\pi} d^{-2}. \tag{7.6}$$

The first term in (7.6) is the repulsive force due to overlapping double layers and contains, n_o, the ionic content per cubic meter, Υ_0 is a function of the surface potential and d is the separation distance. The second term in (7.6) is the attractive force where A is the Hamaker constant of the system. The Hamaker constant is a complex coefficient describing the molecular level attractive forces in solution. More details are provided in the next section.

The slurry stability can be easily seen in light of (7.6). Looking first at the particle surface charge one can deduce that adsorption of specific ions on the particle surface changes the value of Υ_0. For example, as pH is reduced the adsorption of OH^- on the surface of silica particles is reduced, decreasing the surface charge, as measured by the zeta-potential. Recall that the zeta-potential is not the true surface charge but the charge at the surface of shear and is less than the true charge. By reducing the surface charge, the repulsive forces are reduced allowing particles to come closer together and eventually flocculate. Similarly if the ionic strength is increased by addition of acids, bases, salts, ionic-surfactants or buffers, the electric double layer is reduced and κ is increased. Again this results in a reduction of the repulsive forces. It is worth noting that the valence of the ionic species is very important to this destabilization process. As flocculation begins when $\Phi \leq 0$, (7.6) can be rearranged to determine the ionic strength at this point. The result is the concept of a critical flocculation concentration, CFC, at which enough ions have been added to cause the particles to begin to flocculate. The result, called the Schultz–Hardy rule, for the ionic concentration at which the onset of flocculation occurs is proportional to the ion valence by [18]

$$M_{\mathrm{CFC}} \propto z^{-6}. \tag{7.7}$$

Thus very small concentrations of high valence ions, such as Fe^{+3}, Zr^{+4} and Al^{+3} are capable of causing flocculation. This indicates that the ionic purity, particularly of high valence ions, of the slurry, from distributor through to the dispenser, is critical for stability and uniformity of the slurry during CMP. There are many sources of such high valence contaminants and will be discussed presently.

7.3.3 Organics in Slurry

The other method of affecting the stability of the slurry according to (7.6) is to effectively modify the Hamaker constant of the system. Recall that

Table 7.1. Organic compounds present in various advanced CMP slurries

Role	Additives	Processes
Surfactants	Fluorinated carboxylic acids [4], Polyacrylic acids [20], carboxylic acids [44], quartenary ammonium salts (CTAB) [20], Polyvinyl alcohols (Triton)	STI, CuCMP
Inhibitors	Benzotriazole [14], hydrogen phthalate salts [4, 44]	CuCMP, WCMP, STI
Complexing Agents	hydroxylamines [13], ammonium hydroxlate [14], citric acid [14], lactic acid [14], tartaric acid [14], succinic acid [14], amino acids [14], acetic acid [15], EDTA [20]	CuCMP
Oxidizers	Oxalic acid [12], peroxides [14]	CuCMP
Microemulsions	Iso-propyl alcohol [18], glycerol [44]	STI, WCMP
Catalysts	Organic iron compounds [45]	WCMP

the Hamaker constant is a measure of the attractive forces between particles. It is composed of the integrated molecule-molecule Van der Waals attraction forces in particle-particle geometry in solution. Van der Waals forces between molecules are composed to varying degrees of permanent-permanent, permanent-induced and induced-induced dipole interactions. In most solvents, except water, the induced-induced or dispersion component, is the strongest contributor. Water molecules have a permanent dipole resulting in the permanent-permanent dipole as the larger contributor to Van der Waals forces.

The most common method to change the Hamaker constant is to add a soluble organic, such as isopropyl alcohol (IPA), to the system. In the simplest case, if IPA were to completely disperse in the aqueous phase, then A would be less than that of water alone. The reduction would be dependent upon the concentration. The result would be an overall decrease in the attractive forces resulting in improved stabilization.

Organic species however can be used for several purposes beyond changing the attractive forces between particles. Most organic additives serve other purposes but their effect on the electrostatic stability of the slurry should always be accounted for. Referring to Table 7.1, typical organic additives can be classified into six categories, surfactants, inhibitors, complexing agents, oxidizers, microemulsions and catalysts with the first four being the most significant for CMP slurries.

Surfactants/Dispersants

Surfactants are a class of chemicals that contains both hydrophobic and hydrophilic groups. The hydrophobic section typically consists of long chain saturated organic "tails". The hydrophilic section typically consists of either an anionic, cationic or non-ionic "head", as pictured in Fig. 7.5. Some typical slurry surfactants are listed in Table 7.1. By the nature of their structures,
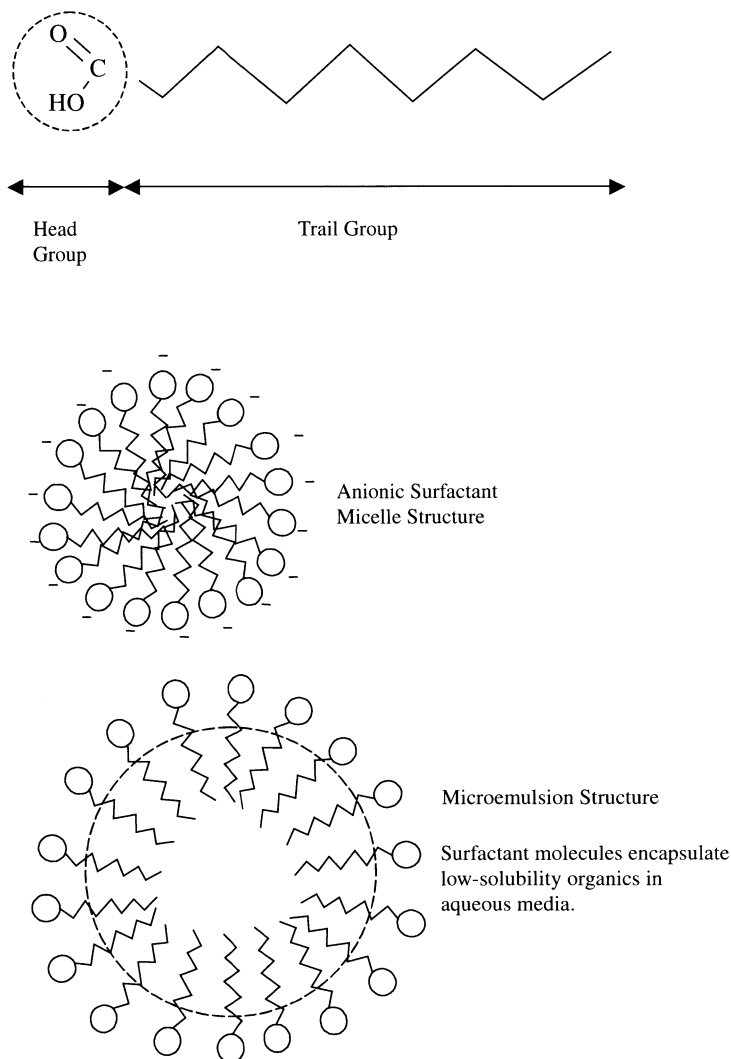

Fig. 7.5. Surfactant molecule formations from typical molecular structure to micellar structure in solution to a microemulsion containing low-solubility organic compounds

surfactants do not readily disperse in aqueous media but go through a process of surface accumulation before dispersion. The surface accumulation has the effect of reducing surface tension. When surface saturation is reached, known as the CMC, the surfactants create colloidal sized particles in solutions called micelles, also shown in Fig. 7.5. These micelles interact with the surrounding medium very differently than do solid particles. However, like solid particles, they have a surface charge and can interact to stabilize or destabilize a suspension. Most relevant to slurry chemistry is the increase in solubility of non-soluble compounds to form emulsions and particle/surface adsorption to change the CMP interaction.

An emulsion is the mixing of two normally immiscible liquids under agitation. When surfactant is absent, the liquids quickly separate. However, with a surfactant present, one phase can disperse, via the micelle structures, into the continuous phase. When a fourth material, such as an alcohol is added, the system forms a microemulsion with the surfactants forming colloidal size micelles that are swollen with the immiscible fluid, as pictured in Fig. 7.5. This increases the transport of this immiscible fluid throughout the slurry and to the wafer interface. A good example of this is the dispersion of a low solubility complexing agent, KHP [4], in slurry using surfactants and an alcohol. This mixture increases the probability and uniformity of the complexing agents reacting with silicon nitride to protect the layer from forming silicon dioxide and being polished during STI. Since it is demonstrated that fluorides also improve the selectivity to nitrides, another, more direct method of increasing selectivity is to use a surfactant based on fluorinated organic compounds and eliminates the need for the alcohol to carry the complexing agents [4, 5, 6]. In either case, the micelle structure acts as a continuous supply of chemicals for enhancing nitride selectivity. Any destabilization of this micelle structure changes the ability to deliver the needed chemistry to the nitride surface in sufficient quantity or uniformity to maintain the high selectivity.

Inhibitors

Inhibitors are organic species that disperse throughout the slurry medium and are employed to interact directly with the wafer surface. The purpose is to reduce or minimize a negative side reaction. In the case of STI CMP, this could be the formation of silicon oxide from silicon nitride. In the case of copper CMP, it is necessary to control corrosion, or passive etching, in the presence of the film oxidizing chemicals. This is particularly important when the wafer is moved from the polishing pad and not immediately cleaned of oxidizer chemistry. Such a situation occurs frequently in multiple wafer polishing when the wafers are placed in a wet hold tank prior to being removed from the tool for post CMP cleans. The inhibitor must be able to protect the film from further oxidation and corrosion before wafer cleaning. A good example is the use of benzotriazole, (BTA) in copper CMP[16]. BTA is often used to
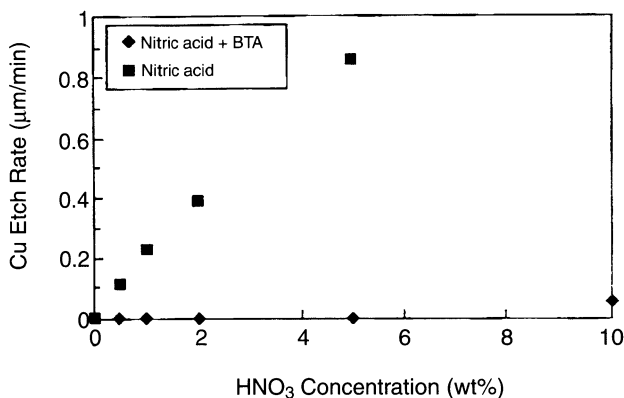
Fig. 7.6. Effect of benzotriazole on Cu film etch rates in HNO_3. From [21]

bind with the surface and inhibit rapid oxidation of the Cu film preventing corrosion. The inhibitor forms a copper complex film that acts as a barrier for oxidating compounds such as urea hydrogen peroxide [14, 15]. BTA is restricted to low pH slurries as it becomes insoluble for pH $\geq$ 5. However, this insolubility can be overcome by use of surfactants and microemulsions as previously discussed. The copper-BTA film can also provide topographical selectivity by being able to withstand the stresses of CMP in the Cu troughs. Figure 7.6 shows the etch rate of Cu with increasing oxidizer concentration and how the presence of BTA inhibits this etch. The CMP action effectively removes this film on the top surface leaving the copper surface free to oxidize. There must be an appropriate balance between oxidizer and inhibitor such that there is enough inhibitor to provide sufficient selectivity without reducing the overall CMP removal rate. One major affect not accounted for here is the nature of the topography and how this ratio of oxidizer to inhibitor should cover the wide range of topography encountered with different mask designs.

Oxidizers

Oxidizing agents were discussed previously in terms of their effects on metal films. Inorganic oxidizers predominate in tungsten metal CMP, however organic oxidizers are gaining use in Cu CMP. The reasons for the investigation of other oxidizers are varied but relate primarily to controlling the oxidation rate and reducing the damage resulting from competitive side reactions. One such oxidizer, urea hydrogen peroxide, [14, 15] has already been mentioned while others come from the family of hydroxylamines [13]. To be effective, similar to inorganic oxidizers, the compound must be soluble over a wide range of pH and result in the passivation or oxidative film, according to the Pourbaix diagram, needed for that process.

Complexing Agents

Complexing agents go by several names, such as chelating agents or bind-ing agents. As the name implies, these compounds bind with partial or fully charged species in the solution or at the wafer interface. This has a two fold effect one being the reduction of the charged species' contribution to the dou-ble layer thickness on the particles and, in some metals, preventing further oxidation to more stable metal oxide compounds. The second effect, in CMP slurries, is to bind with the wafer film inhibiting the reverse reduction reaction at the wafer surface resulting in an increased removal rate. The "binding" process can be accomplished by direct bonding of the species (ligand for-mation), ion pairing of the species (organic salt precipitation) or steric ionic shielding of the ion (high molecular weight chelation). Some agents complex one-to-one with ionic species, while others can complex with multiple ionic species. EDTA is a common complexing agent with three acid groups [20]. The acid groups enfold the positively charged metal ions in solution effectively shielding them from further reaction.

7.4 Solids Content

To this point only the effects of solution chemistry have been addressed. Although complex, the chemistry makes up only one of the two components in slurry. The particles that abrade the surface are clearly necessary to perform conventional CMP [1]. Table 7.2 contains some properties of the particles most often included in CMP slurries. The choice of particle properties has been fairly empirical. There has been some explanation as to the performance of ceria and silica in oxide polishing based on optical lens data [1]. Particle size selection is usually an empirical compromise between polish rate and defects. Over some range, large particles give higher polish rate, but create more defects too. Conversely, smaller particles result in fewer defects but do not polish at acceptable rates. There are exceptions to this, particularly alumina particles in oxide polishing [21]. In reality the only four particle

Table 7.2. Particle properties for common abrasives in CMP slurries

Property	Silica	Alumina	Ceria
Particle Structure	amorphous	Poly-crystalline	Poly-crystalline
Crystal Structure		Orthorhombic (corundum)	Cubic
Density (g/cc)	2.2–2.6	3.97 (alpha)	7.13
Hardness (Mohs)	6–7	9	
PZC (pH)	2.2	9	7

parameters readily changeable in the CMP process are flow rate, particle size, concentration of particles and particle type.

Flow rate is not in reality a particle property but it does affect the overall performance of the particles by maintaining a continuous media in which to polish as well as sufficient quantity of particles for polishing. The limitations on flow rate are dependent mainly upon the platen speed. High platen speeds require more slurry as the centrifugal action spins most of the slurry off the pad. The presence of a bow wave during CMP is an indication of sufficient slurry to provide a continuous media for polishing. Additionally, similar to a centrifuge, the distribution of particles and particle size may differ from center to edge of the platens with increasing rotational speed. Recall that particle settling rates contain a gravitational coefficient. As the platen spins, the g-force created along the radial axis of the pad will increase settling in the lateral direction. Pad conditioning and grooving plays a role in minimizing the "settling" of slurry particles across the pad, however within a groove the outer edge will experience a build up of particles that may lead to defects if not properly cleaned between wafers. Low speeds and large platens do offer the benefit of low rotational speeds while providing high linear speeds resulting in more uniform particle distribution based on settling rates. Figure 7.7 shows the effective g-force on a particle at variable speeds and radii. The flow rate must compensate for the particle loss or distribution by flooding the pad.

Adjustment of the particle size in CMP is not as simple as adjusting the flow rate. In general for end CMP users modifying particle size can only be achieved by working closely with a particle supplier. In particular precipitated silicas suppliers offer a wide range of available particle sizes for testing with near identical solution chemistry. Figure 7.8 depicts the removal rate variation with solids concentration and particle size [22]. Both increased particle size and concentration increase the bulk removal rate. The mechanism of increase removal rate with particle size, r, is less clear from this data as more particle
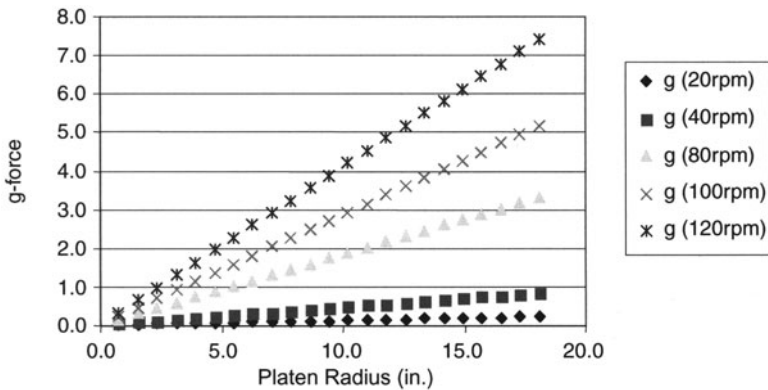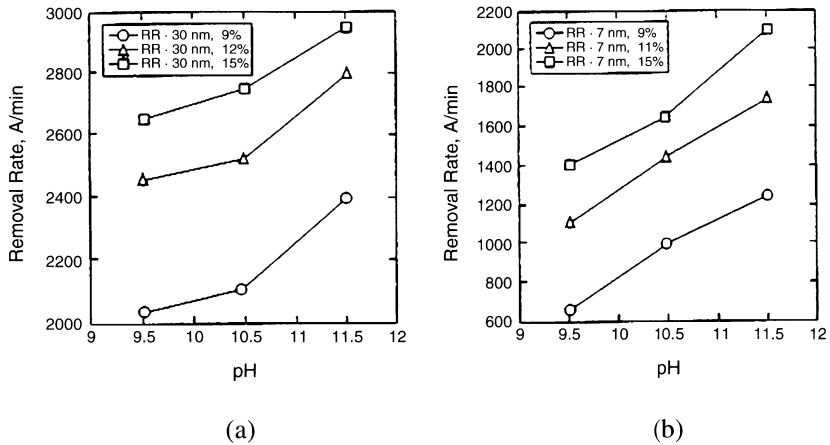

Fig. 7.7. G-Forces experienced on a CMP platen

Fig. 7.8. Removal rate dependence on pH and % solids for (**a**) 30 nm particles and (**b**) 7 nm particles. From [22]

sizes are needed to separate particle mass contributions, varying with r^3, from surface area contributions, which vary with r^2. Intuitively there is also a planarization optimization experiment that can be performed by varying concentration and particle size but the results will be topography and film type dependent. It is more often that control of the polydispersity, or size range, of the solution is more critical than the nominal, or mean particle size.

Dilution is the simplest method of changing particle concentration in the solution. In mixed component slurries, this is a straightforward procedure. The particle containing component can be mixed at varying ratios to the chemistry component. This offers the benefit of maintaining the chemical concentrations while understanding the affects of particle concentration on rate, planarization and defectivity. In single component slurries, dilution of the % solids also results in the dilution of the solution chemistry, which can effect both CMP rate and the suspension stability as the ionic strength changes. The combination of flow rate, particle size and concentration can be tailored to individual process step needs. For example the needs of ILD CMP are different from those of STI CMP. Dilution and flow rate changes may also offer the benefit of reduced cost of ownership.

A preferred method for changing a CMP property with the solids is to change the chemical makeup of the particle, either by surface adsorption to change the zeta potential or to different particle material. Although there is much industry interest in novel particle types, this section will discuss briefly the properties of the four most commonly used CMP slurry particles, fumed silica, precipitated silica, alumina and ceria.

7.4.1 Silica

Silica comes in multiple forms, however for CMP there are two of primary interest, fumed and precipitated (often referred in the industry as colloidal). Both of these types are amorphous in nature, i.e. they are not crystalline. The fuming process involves the gas phase formation of small particulates that are then fumed into larger chain particles. The small particulates are formed by the following reaction:

$$SiCl_4 + 2H_2O \rightarrow SiO_2 + 4HCl.$$
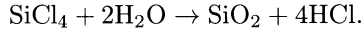
The resultant particles go through several procedures that modify the particle size, shape and surface charge before being shipped to the slurry supplier. Subsequent processing at the slurry production site includes milling and mixing. Figure 7.9 is a SEM of typical fumed silica particles.

Precipitated silica is formed in solution. As one example of precipitated particle formation, silicates are reacted in an acid according to:

$$K_2SiO_3 + H_2SO_4 \rightarrow SiO_2 \downarrow + 2K^+ + H_2O + SO_4^{-2}.$$
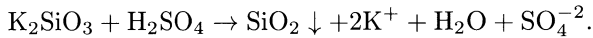
There are several other precipitated reactions often involving oxy-silane intermediary products, but the resultant particles share several common properties. The particles are spherical and virtually monodisperse.

There is an ongoing discussion in the CMP industry of the relative merits of fumed versus precipitated silica. Contamination is one area of difference between the two that involves trade-offs between particle purity and solution purity. The solution phase generation of precipitated particles reduces problems associated with the mixing and grinding processes of working with dry
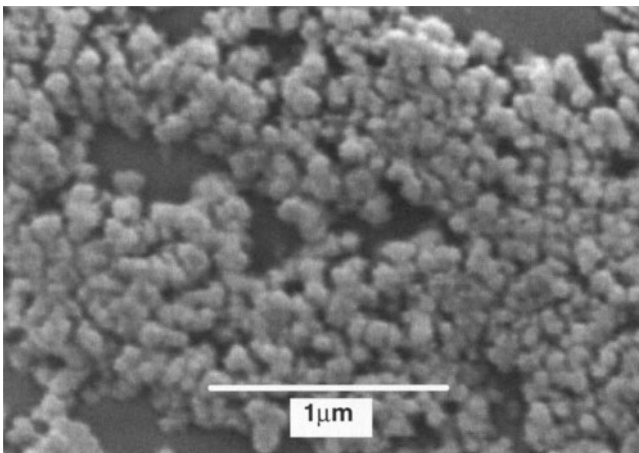
Fig. 7.9. Fumed silica particles pulled from a high pH slurry. [Courtesy Robert Schmidt of Rodel]

fumed silicas. Fumed silicas are often ball-milled which may lead to metallic debris from the balls being incorporated into the slurry. However, the precipitated silica may entrap ionic species, such as Na^+ or K^+, in the particle. There is no direct evidence that incorporated contaminants have caused any problems with mobile ions finding their way into the gate oxides of a device. Claims of reduced defectivity with the monodisperse precipitated particles are true when compared to simple fumed slurries. However filtration and decanting of fumed products have demonstrated improved defectivity comparable to that of precipitated slurries. Improved planarity using fumed slurries has been demonstrated however pad conditions, due to the higher solids content and viscoelastic response, at the slurry wafer interface can not be verified as similar enough to isolate the particle contribution as the source of improved planarity. The selection of slurry particulate, including silica type, depends upon the specific process module needs and the integration requirements.

7.4.2 Alumina

Currently, alumina is the most widely used particle type for metal CMP. Alumina is one of the hardest materials used for polishing, as measured by the Mohs index, see Table 7.2, and is second only to diamond in hardness. Mined alumina is processed to high purity through several processes. In this process, raw alumina is dissolved by a strong base and separated from impurities via decanting.

$$Al_2O_3 + 6OH^- + 3H_2O \rightarrow 2Al(OH)_6^{-3}.$$
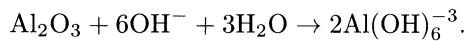
The liquid solution is then adjusted to neutral pH to form a hydrated alumina ion precipitate. The precipitate is crystallized and calcined to form pure alumina powder, Fig. 7.10 [24, 25, 26, 27]. The reaction sequence is:

$$2Al(OH)_6^{-3} + 6H^+ \rightarrow 2Al(OH)_3(H_2O)_3 \downarrow,$$

$$2Al(OH)_3(H_2O)_3 \rightarrow Al_2O_3 + 9H_2O .$$
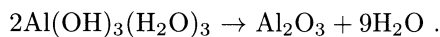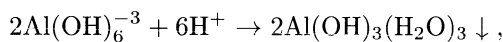
In contrast to silica, the alumina particles formed are polycrystalline in nature, not amorphous. The alumina crystallite size, phase, shape and surface adhesion are characteristics of the crystallization and calcination processes. Subsequent milling will also influence the alumina properties. Due to the inherent hardness of alumina, softer pads or lower down force processes are frequently used to reduce scratching of the ILD exposed during polishing. Although these processes are more conformal to wafer contour they provide more flexibility to adsorb the impact of a large particle trapped between the pad and the ILD surface. In a hard pad process, many defects are created by the particle being forced to grind or bounce along the ILD surface creating Hertzian type scratches. The ILD surface is not as hard as the alumina particle and is predicted to break before the alumina particle is ground into smaller crystallites. The ability to self-destruct into smaller particles in high
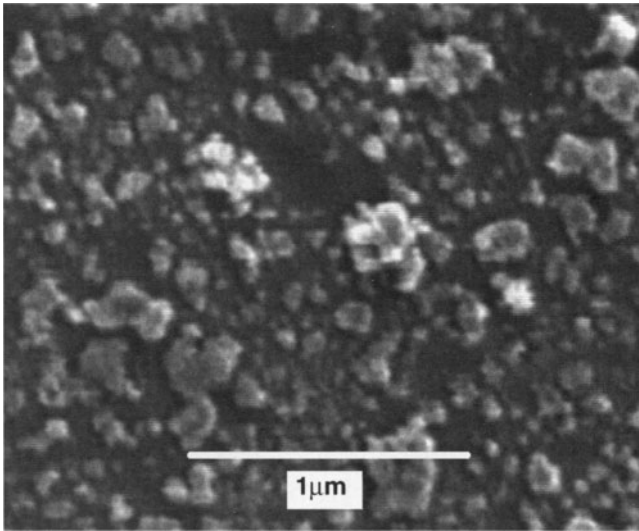
Fig. 7.10. Alumina particles pulled from a low pH tungsten slurry. [Courtesy Robert Schmidt of Rodel]

stress situations also plays a role in reducing wafer defects. The strength of the large polycrystalline particle is determined by its surface adhesion post manufacture and may vary as the manufacturing processes vary with differing raw materials. How this could be measured prior to polishing is not currently understood.
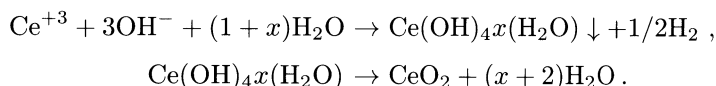
Unlike silica, alumina used in CMP has a high pzc, pH ~ 9. This means alumina has a positive surface charge in most CMP slurries. Most of the metals in these systems have a negative surface charge or at best a weak surface charge. This greatly affects the cleaning process post metal CMP with alumina [28, 29]. The typical DI water or ammonia scrub that is used for ILD cleans would enhance the electrostatic attraction between the metal and the particle reducing the effectiveness of the cleans.

7.4.3 Ceria

Ceria has been used in glass polishing and exhibits much higher chemical interaction with glass than silica particles [1]. It is used mainly in STI high selective slurries as it provides reasonable removal rates at the moderate pH of these slurries. The higher chemical activity may be related to the calcination procedure during manufacturing of the particle [1, 3]. The details of the improved polishing rate are discussed elsewhere [1]. Ceria is produced similar to alumina. In one method [30, 31], cerium nitrate salt is oxidized in a strong base to produce hydrous cerium oxide. The hydrous oxide precipitates out of solution, the the following sequence:
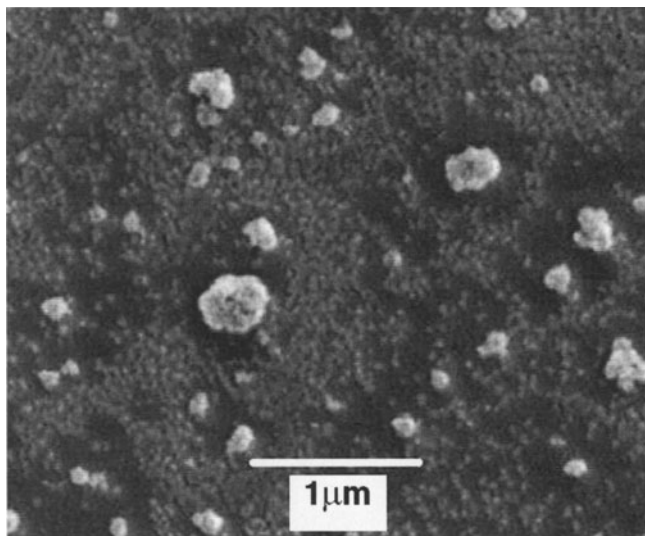
Fig. 7.11. Ceria particles pulled from a neutral pH STI slurry. [Courtesy of Robert Schmidt of Rodel]

$$Ce^{+3} + 3OH^- + (1+x)H_2O \rightarrow Ce(OH)_4 x(H_2O) \downarrow +1/2H_2 \ ,$$

$$Ce(OH)_4 x(H_2O) \rightarrow CeO_2 + (x+2)H_2O \ .$$

The hydrous oxide is crystallized and calcined to form ceria particles, as shown in Fig. 7.11. The degree of calcination is one of the factors defining the chemical activity and crystallinity of the ceria. Other methods involve either direct oxidation in acid media or other cerium salts such as carboxylates or sulfates.

Ceria based slurries suffer from colloidal stability for two reasons, density and surface charge. Ceria has a density of 7.1 g/cc compared to silica of 2.2 g/cc which according to equation 1, means ceria settles ~ 4 times more quickly. Ceria also has a pzc at a pH ~ 7 and a weak zeta potential at higher pH.

Again referring to the stability equations, ceria is difficult to stabilize in suspension via electrostatics. As the surface chemistry of ceria provides the high CMP removal rates for polishing ILD oxides, modification by ionic adsorption may be detrimental to CMP performance. Another method would be the adsorption of organic compounds to sterically stabilize the suspension. In either case, proper mixing of the ceria and solution is critical to maintaining a consistent suspension in the distribution systems and at the wafer interface.

7.4.4 Particle Sizes and Detection

Apart from the chemical nature of the particles in CMP, the most often referenced parameters of a slurry are the mean particle size and the particle size distribution. Figure 7.12 shows how large particle count is reduced as particle size distribution is improved as the process matures.

As discussed earlier, the blanket wafer removal rate is dependent upon the particle size. Therefore the mean particle size is set as a specification out of particle manufacturing. Most often, particle size distribution specifications are met by subsequent milling to the desired particle size distribution at the slurry supplier site and not the particle manufacturer's site. Wide variations in the particle manufacturing process require different milling times to reach the specified final mean particle size. As noted already, milling can introduce contaminants in the slurry and over time, the concentration of these contaminants will vary. As most of the contaminants from grinding are multi-valence ions, such as Zr^{+4} from the milling balls, the degree of destabilization of the slurry will vary according to the Schultz–Hardy rule.

The particle size distribution is very dependent upon the manufacturing method of each type of particle. In general, the ultimate preference for CMP particle size is a uniform mono-disperse particle distribution. In a particle size measurement the output from several commonly used tools is a Gaussian distribution about a mean. The measured size distribution is also dependent

SC-1 Manufacturing, LPC > 0.56
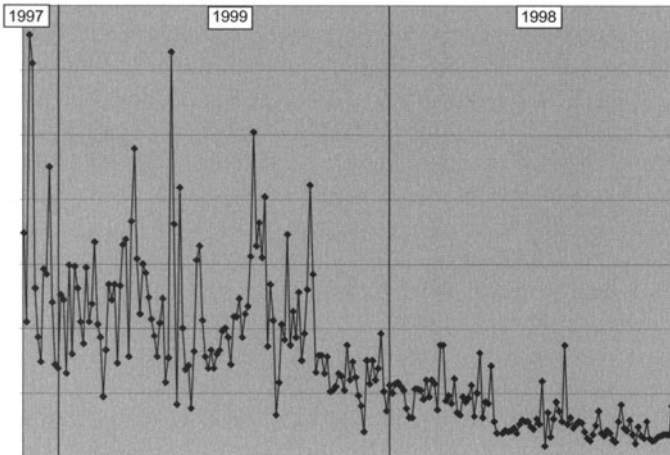
Fig. 7.12. Large particle count, LPC, reduction with slurry maturity. The decrease in LPC's leads to direct reduction in post CMP scratch defectivity. [Courtesy of Cabot Microelectronics]

Table 7.3. Particle size detection methodologies and particle size range. From [46, 47]

Methodology	Range (mm)
Capillary Hydrodynamic Fractionation	0.015 – 1.1
Classical Light Scattering	0.02 – 1000
Dynamic Light Scattering	0.003 – 5
Single Particle Light Extinction	0.5 – 400
Single Particle Light Scattering	0.5 – 20
Acoustic Spectroscopy	0.010 – 10

upon the method of particle detection. Table 7.3 shows some typical particle sizing methods and the particle size ranges associated with those methods. Increased concentrations of large particles are associated with increased defects [32] so a choice of a tool that is accurate in the measurement of large particle concentration, also known as the "tail of the distribution," is more appropriate for defect control. CMP removal rate is related to mean particle size [22] indicating that understanding removal rate variations as a function of particle size would require a tool that is more sensitive to the mean of the particle size distribution. As Table 7.3 indicates, this implies different measurement equipment for different process control needs.

Large particles associated with CMP defects can be filtered at the CMP production site. In this way, particles from the original manufactured slurry as well as particles produced by shipping and handling, or in the slurry tank and distribution system, can be removed before the slurry is used in CMP [34]. Filtration can occur at many different locations in the slurry distribution loop. Bulk filters at the mix stations or main slurry lines can filter out large agglomerates caused by particle aggregation due to line contaminants or pulling from the bottom of a decanted drum. These particles are not inherent in the normal particle distribution but are a result of shipping, contaminants or debris. Tool filters, installed at either the line head or the dispense tubes, remove the large particles inherent in a particle distribution, independent of their source. Practical limitations of filter clogging and line pressure drop dictate a lower limit on filtration size. As an example, spherical silica particles in neutral pH solution with insufficient electrostatic stabilization will gel with increased shear [2]. Pressure drop in the line, as a result of small filters, can induce sufficient shear in the liquid to cause gelation. Slurry gelation in the line can result in a wide range of CMP related concerns from variable loss of particle concentration and resultant CMP removal rate variation, to short filter lifetimes and clogged dispense lines with increased cost of consumables and maintenance.

7.5 Slurry Handling

Despite all the complications of manufacturing a slurry, balancing the chemistry, forming a stable suspension and controlling particle properties, the slurry must eventually be dispensed on the CMP platen and polish wafers acceptably. The general method of shipment of slurries involves large barrels or totes of slurry components that are mixed at the user site, either in a large distribution system or at the individual tool. The process of on-the-pad slurry injection has also been proposed [34]. The relative merits of specific mixing procedures is dependent upon the needs of individual fabs as no one system is universally optimal given the differences between fab infrastructures, local climates, slurry volumes and process advances. However there are some basic concerns relevant to all CMP users.

7.5.1 Shelf Life

Descriptions of shelf life are varied in the industry. In this text shelf life is described as the time over which the slurry properties are sufficiently stable to not contribute to observed CMP process variations, such as increased defects, removal rate variations or uniformity variations. It is also assumed for this discussion, that the manufacturing induced distribution about any specific particle property, such as surface oxidation state, are accounted for in the reported slurry shelf life. In mixed component slurries, the shelf life is considered two-fold. The first is the shelf life of the individual components, and the second is the shelf life of the mixed components.

In general, no solution chemistry and particle solutions are in equilibrium or steady state. Degradation reactions proceed continually, and many reactions are accelerated with increased temperature. Some effects, such as the physical effect of settling, proceed with time, but more slowly with increased temperature. Most chemical effects, governed by activation energies, proceed with time but much more quickly as the temperature is increased.

Due to the common use of hydrogen peroxide in tungsten and copper CMP slurries, it provides an excellent study of the time dependence of slurry. Other oxidizers and active ingredients will have different competing side reactions and respective shelf lives. The oxidation process for the passivation of a W film, whether it be WO_3 formation or WO_4^{-2} ion formation involves the following half-cell reaction:

$$H_2O_2 + 2H^+ + 2e^- \rightarrow 2H_2O.$$
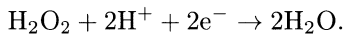
This reaction occurs at the film-solution interface and is highly oxidizing. However there is a competing reaction for peroxide in solution:

$$H_2O_2 \rightarrow H_2O + 1/2O_2(g).$$
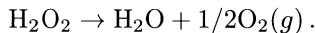
This reaction occurs naturally and is catalyzed by the presence of certain ionic species particularly halides. In a closed cell, this reaction would reach
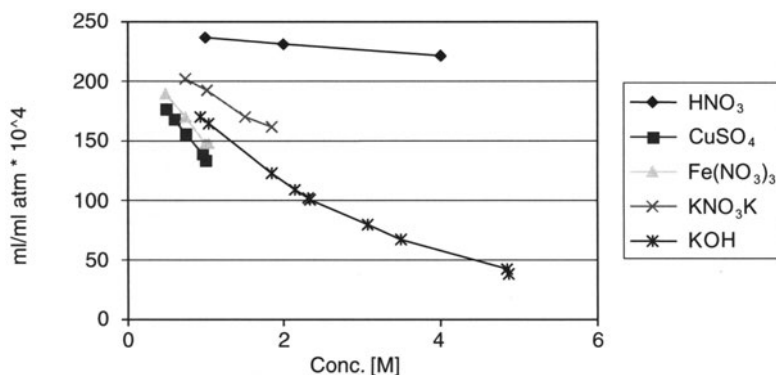
Fig. 7.13. Solubility of O_2 with concentration of various electrolytes in slurry. From [48]

equilibrium with the O_2 in the vapor phase. The equilibrium concentrations will depend upon multiple variables such as ion concentration. Figure 7.13 shows the O_2 solubility in aqueous electrolytes with ion concentration.

More important to the CMP user is the non-equilibrium reaction rate when the system is not in a closed cell, such as a vented mixing tank. In this situation, the O_2 will be continuously displaced and the peroxide concentration will drop accordingly with time and reduces the oxidative strength of the slurry. This is of little consequence to the CMP user with sufficient slurry usage that the peroxide is rapidly replaced. However systems that use small volumetric flow rates in large mix tanks, resulting in large gas overheads, can rapidly deplete the peroxide concentration. The above reaction is also increased by UV exposure. Most in house slurry systems adequately shield slurry from UV and shipment is in dark or UV protected drums to reduce this affect.

Time also plays an important role in the solids content of the slurry. It has already been discussed how some slurries may require decanting. This is nothing more than providing enough time for an acceptable amount of large particles to settle allowing easy removal before use. In terms of shelf life, the difficulty in settling rate comes in the form the particles take during settling. The DLVO theory discussed above leads to two possible forms of particle growth and settling, flocculation and aggregation. Flocculation is the formation of a particle network, or floc, in which the double layers overlap but the particles do not actually come in contact. Not all particles in solution form these flocs as this process is dependent upon forming a minima in the balance between attractive and repulsive forces at a distance from the particle surface. The "soft-flocs" can be broken back down into the primary particles and redistributed with agitation. Agglomeration, or "hard-flocs", is similar except the particles come into actual physical contact and are bound together. Agglomeration is an indication of insufficient electrostatic repulsive forces in

the solution allowing the attractive forces to dominate the particle-particle interactions. These "hard-flocs" can be very difficult to disperse and often requiring a highly energetic agitation such as that provided by high frequency sonication to break up the particles to their original manufactured size.

The other important shelf-life concern is temperature. Again hydrogen peroxide is an excellent case study of the effect of temperature. There are two types of temperature effects in hydrogen peroxide decay. The first is the Arrenhius type increase in the rate constant for a first order reaction, the second is the increase in the vapor pressure of the hydrogen peroxide. Again, in a closed cell, these affects reach equilibrium however in a vented system, the increased reaction rate produces more O_2 which is removed from the system while the increase vapor pressure pulls peroxide directly from the solution. Figure 7.14 shows the temperature dependence of the vapor pressure for various oxidizers.

Temperature also affects the particle stabilization. Increasing the temperature decreases the settling rate by means of increased Brownian motion. In essence the system becomes more agitated. Increased temperature also increases the repulsive strength term in the DLVO theory by a direct linear temperature dependence and by increasing the double layer thickness of the particle. Conversely a drop in temperature can result in increased agglomeration and particle settling. Dependent upon the nature of the particle stabilization, aggregation is accelerated at low temperatures; in practice slurry is exposed to low temperatures during shipment or outside storage. Without the use of on barrel temperature recorders, a low-temperature problem is very difficult to diagnose. One problem is that partial redistribution of the aggregates occurs as the slurry is warmed and shaken during staging such that there is no apparent increase in the agglomeration to an observer. Unless in-line particle size monitoring is employed, the first indication of an issue with the slurry is during the CMP process when an increase in defects, a shorter large particle filter life or variation in the removal rate that the increased number of aggregates is observed.
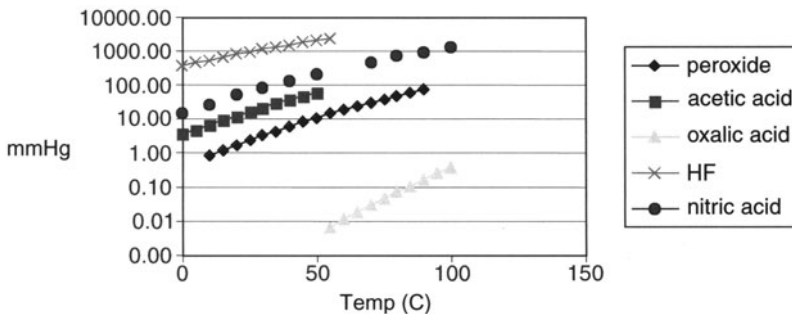

Fig. 7.14. Vapor pressure for various oxidizers in CMP. From [48]

Other slurry oxidizers have different shelf life behaviors. Each chemical has its own temperature limitations and chemical reactions that need to be evaluated. However shelf life is defined for a particular slurry, careful monitoring of the solids concentration of the supernatant (the clear solution left on top in a slurry that has settled), particle size distribution shifts and titration of the oxidizer can be used to verify the slurry properties. The problem of variable slurry properties is often addressed by the implementation of multiple in-line monitors, such as in-line titrators, particle sizers and pH monitors in the slurry distribution system.

7.5.2 Mixing of Slurry Components

The most common approach for increasing slurry shelf life has been to ship components separately. The components are then mixed on site prior to use at the tool. This has led to the report of the slurry "mix-life", similar to the shelf life but often measured in hours or days not months. The mix-life is defined to be the time from mixing to some percent reduction in CMP performance, such as rate, defectivity or selectivity. All of the parameters that may affect shelf life also affect mix life. In particular, oxidizers can react readily with the complexing agents and inhibitors used in more advanced slurries. One additional parameter that affects mix life is the method of mixing.

An example of the difficulty of mixing components is with precipitated silica slurries. Precipitated slurries are usually shipped as a single component system. However, the slurry can be modified in-line or at a mix station by the addition of carboxylic acid type surfactants with the intent of increasing selectivity between silicon oxide and silicon nitride [4, 35]. Precipitated silica goes through a several phases of stability as the pH decreases, full suspension, rapid aggregation and particle [2]. In this example, the problem arises as the pH drops during mixing of the surfactant into the slurry. Excessive agitation in the mix station, with pH change, induces a viscosity increase that can potentially lead to a gelation of the slurry in-line. Gelation of the slurry would result in a lower than expected flow rate, loss of particle concentration and variable chemical concentration at the wafer surface. Any of these would affect the outcome of the selectivity experiment described. Of course, such non-Newtonian dilatancy, as described previously, is dependent upon the ionic concentrations, pH and shear rate.

There are three basic methods of mixing multi-component slurries, bulk mixing, point-of-use mixing and on the platen mixing. From a semiconductor manufacturing point of view, bulk mixing is the most desirable. Bulk mixing provides fewer additional parameters to be factored into any CMP process and removes tool-to-tool variation by delivering the same slurry at the same concentrations to every tool. From a slurry developers point of view, on the platen mixing is the most desirable. The activity of each component can be individually tailored to meet the need of the particular process without worrying about side reactions with other components or particles. Point of

use mixing, or injection [34] is a compromise between these two extremes allowing the components to be kept separated right up to the dispense tube to minimize side reactions, but with carefully calibrated pumps and mix stations, the same concentrations can be delivered to various tools. The dwell time during mixing can be varied to balance undesirable side reactions with necessary adsorption and stabilization reactions.

7.6 Future Trends in Slurry

As with all semiconductor processes, new methods and ideas are continuously being introduced. CMP has progressed from a planarization process to enable lithography and dry etch improvements to a metal isolation technique to enable Cu metallization. It should be expected that eventually CMP will be improved to a form not requiring slurry. Several novel patents involving super-critical fluids, spin-on polyimide and float glass techniques have shown some new ideas in the area of dielectric deposition. New particles have also been tested. One in particular, manganese oxide, has shown some very interesting results for CMP [36, 37]. Closer to the core of CMP, three new technologies have been introduced that show some advancement from conventional CMP and overcome the limitations of particle size, mixing and competing side reactions discussed.

7.6.1 Fixed Abrasive

Fixed abrasive polishing appeared as a potential alternative to conventional slurry based CMP [49]. The general premise is that the entire section concerning particle suspension discussed above can be bypassed by incorporating the particles directly into the pad or some other polymeric matrix. There have been two approaches in the literature one incorporates the particles into a thin matrix best suited for a web format while the second incorporates particles into a more conventional rotary pad format. Both processes would open up the possibility of varying the chemistry available for preparing the surface to be polished. It should be noted that if the method requires the formation of localized slurry by pad erosion or residue formation, then these CMP processes are subject to the same colloidal science limitations as conventional slurry. The long term viability of fixed abrasives is still uncertain.

7.6.2 Abrasive-Free Slurries

Abrasive-free slurry processes differ from fixed abrasives in that the chemistry/pad interaction alone is sufficient for CMP. Most noteworthy has been in developing a process for copper CMP on a conventional pad and platform [38, 39]. In this pseudo-conventional CMP process, aggressive chemistry is used to complex the copper surface followed by abrasion with the

soft polymeric pad. The process stops readily on the barrier material with minimal dishing of the large copper pads. Subsequent barrier polishing with conventional slurry leaves a very planar surface. The biggest concern in such a process is the use of the aggressive oxidizing chemistry and its affect on tools and pads. By eliminating the particle from the CMP process, the entire approach for development of new oxidizing agents, passivating agents and binding agents is free from the confines of destabilizing the suspension.

7.6.3 Complex Particles: Hydrothermal Particles

Although there has been continuous improvements in particle manufacturing, there has been little shift from the core methods. Hydrothermal processing represents such a shift. This method is not new to the particle industry but may represent a huge opportunity in CMP for fine tuning CMP needs. Hydrothermal processing is described in several patents and publications [40, 41]. Briefly, the process involves a continuous flow of aqueous based reactants through high temperature and pressure reactor. The resultant particles are spherical in shape and suspended in solution. The unique capability of hydrothermal processing is not just in particle shape and crystallinity but also in surface coatings [42]. By changing reactants, different coatings and partial coating can be added to the particle surface. An example is in the formation of TiO_2 coatings on SiO_2 particles [42]. In this way, the chemical qualities of a metal oxide that makes a poor abrasive can be combined with the bulk properties of a suitable abrasive. One need not stop at metal oxide coatings but extend the concept into organic coatings with reactive end groups however this would not by hydrothermal processing.

7.7 Summary

CMP has rapidly gained acceptance in IC manufacturing despite identifiable problems directly associated with slurry [43]. Continuous improvements in the control and manufacturing of slurry have aided in the acceptance of CMP in new applications, such as polysilicon CMP, platinum CMP, low-k dielectric CMP and other processes in the IC roadmap. Each of these will bring more challenges as the materials in the IC world change from traditional dielectrics and metals. It is already apparent that what works for tungsten does not work for copper and, in the case of ruthenium oxide polishing, may create hazardous environments if the chemistry is not better understood. Will the same be true of copper versus platinum or iridium?

Future improvements in CMP slurries will be a result of more research into the mechanism of CMP and implementation of fundamental surface chemistry, electrochemistry, fluid dynamics and particle suspension science. Although each scientific domain is rich in opportunities for improvements in

slurry performance, the need to balance these opportunities with the possible adverse effects on the other incorporated domains still exists. Regardless of the needs of the CMP process there are technological limitations, as covered in this chapter, to producing, transporting and controlling chemically active particle suspensions. For example, the best additive or particle for preventing copper dishing is useless if it can not be delivered to the wafer surface without impeding the oxidative chemistry or destabilizing the particle suspension. In parallel to the slurry development, advances in new semiconductor materials, new integration schemes and new CMP tools will continue to change the process specifications needed in CMP slurries.

References

1. L.M. Cook, J. Non-cryst. Solids, **120**, 152, 1990.
2. R.K. Iler, *The Chemistry of Silica*, John Wiley & Sons Inc., NY, 1979.
3. T. Izumitani, Paper TuB-A1, Tech. Digest, Topical Meeting on the Science of Polishing, Optical Society of America, 17 April 1984.
4. S.D. Hosali, A.R. Sethuraman, J. Wang and L.M. Cook, "Composition and Method for Polishing a Composite of Silica and Silicon Nitride", US Patent # 5,738,800, April 1998.
5. J. Prasad, A. Misra, J. Sees, B. Morrison and L. Hall, "Mechanism of Chemical Mechanical Polishing Process for Oxide-Filled Shallow Trench Isolation Applications" in *Chemical Mechanical Planarization I*, I. Ali, S. Raghavan eds., **96-22**, 36, The Electrochemical Society, 1997.
6. D. Cossaboon, J. Wang and L.M. Cook, "Compositions and Methods for Polishing Silica, Silicates and Silicon Nitride", US Patent # 5,769,689, June 1998.
7. C. Rahunath, K.T. Lee, E.A. Kneer, V. Mathew and S. Raghavan, "Mechanistic Aspects of Chemical Mechanical Polishing of Tungsten Using Ferric Ion Based Alumina Slurries", in *Chemical Mechanical Planarization I*, I. Ali, S. Raghavan eds., **96-22**, 1, The Electrochemical Society, 1997.
8. E.A. Kneer, C. Raghunath, V. Mathew, S. Raghavan and J.S. Jeon, J. Electrochem. Soc., **144**, 3041, 1997.
9. D.J. Stein, D. Hetherington, R. Guilinger and J.L. Cecchi, J. Electrochem. Soc., **145**, 3190, 1998.
10. M. Pourbaix, Atlas of Electrochemical Equilibria in Aqueous Solutions, NACE, Houston, TX, 1975.
11. F. Kaufman, D. Thompson, R. Broadie, M. Jaso, W. Guthrie, D. Pearson and M. Small, J. Electrochem. Soc., **138**, 3460, 1991.
12. K. Yang, S. Avanzino and C. Woo, "Slurry for Chemical Mechanical Polishing of Copper", US Patent # 6,143,656, Nov. 2000.
13. R. Small, L. McGhee, D. Maloney and M. Peterson, "Chemical Mechanical Polishing Composition and Process", US Patent # 6,117,783, Sep. 2000.
14. V. Brusic and R.C. Kistler, "Chemical Mechanical Polishing Slurry useful for Copper Substrates", US Patent # 5,954,997, Sep. 1999.
15. V. Brusic, R.C. Kistler and S. Wang, "Chemical Mechanical Polishing Slurry useful for Copper Substrates", US Patent # 6,126,853, Oct. 2000.

16. V. Brusic, R.C. Kistler and S. Wang, "Chemical Mechanical Polishing Slurry useful for Copper/Tantalum Substrate", US Patent # 6,063,306, May 2000.

17. A.E. Braun, *Semiconductor International*, **22** (14), 54, 1999.

18. P.C. Heimenz, Principles of Colloid and Surface Chemistry 2nd ed., Marcel Dekker, Inc., NY, 1986.

19. C.H. Yao, D.L. Feke, K.M. Robinson and S. Meikle, J. Electrochem. Soc., **147**, 1502, 2000.

20. J.H. Golden, R. Small, L. Pagan, C. Shang, S. Raghavan, *Semiconductor International*, **23** (12), 85, 2000.

21. J.M. Steigerwald, S.P. Muraka and R.J. Gutmann, *Chemical Mechanical Planarization of Microelectronic Materials*, John Wiley & Sons, Inc., NY, 1997.

22. R. Jairath, M. Desai, M. Stell, R. Tolles and D. Scherber-Brewer, Mat. Res. Soc. Symp. Proc., **337**, 121, 1994.

23. C.H. Yao, D.L. Feke, K.M. Robinson and S. Meikle, J. Electrochem. Soc., **147**, 3094, 2000.

24. H. Erickson, "Method for Producing Alumina", US Patent # 4,066,740, Jan. 1978.

25. M. Mohri, Y. Uchida, Y. Sawabe, "Process for Producing Alpha-Alumina Powder", US Patent # 5,538,709, July 1996.

26. M. Mohri, Y. Uchida, Y. Sawabe and H. Watanabe, "Alpha-Alumina Powder and Process for Producing the Same", US Patent # 6,159,441, Dec 2000.

27. T. Harato, T. Furubayashi, T. Ashitani, T. Ogawa, "Process for Preparation of Alumina", US Patent # 5,302,368, April 1994.

28. L. Zhang, S. Raghavan, S.G. Meikle and G. Hudson, J. Electrochem. Soc., **146**, 1442, 1999.

29. F. Zhang, A.A. Busnaina and G. Ahmadi, J. Electrochem. Soc., **146**, 2665, 1999.

30. J. Wang, "Oxide Particles and Method for Producing Them", US Patent # 5,389,352, Feb. 1995.

31. C. David, C. Magnier and B. Latourrette, "Novel Ceric Oxide Particulates and Process of Making", US Patent # 4,859,432, Aug 1989.

32. G.B. Basim, J.J. Adler, U. Mahajan, R.K. Singh and B.M. Moudgil, J. Electrochem Soc., **147**, 3523, 2000.

33. "Post-dilution Filtration", www.millipore.com/micro/cmp.nsf.

34. F.C. Chou, M.N. Fu and M.W. Wang, J. Electrochem. Soc., **147**, 3873, 2000.

35. Y.Z. Hu, R.J. Gutmann and T.P. Chow, J. Electrochem. Soc., **145**, 3919, 1998.

36. S. Kishii, K. Nakamura and Y. Arimoto, Symposium on VLSI Technology Digest of Technical Papers, 3B-2, 27, 1997.

37. S. Kishii, K. Nakamura, Y. Arimoto, A. Hatada, R. Suzuki, N. Ueda and K. Hanawa, "Slurry Containing Manganese Oxide and Fabrication Process of a Semiconductor Device Using such a Slurry", US Patent # 6,159,858, Dec. 2000.

38. S. Kondo, N. Sakuma, Y. Homma, Y. Goto, N. Ohashi, H. Yamaguchi and N. Owada, J. Electrochem. Soc., **147**, 3907, 2000.

39. S. Kondo, N. Sakuma, Y. Homma, Y. Goto, N. Ohashi, H. Yamaguchi and N. Owada, IEEE, 253, 2000.

40. J.G. Darab and D.W. Matson, J. Elec. Mats., **27**, 1068, 1998.

41. S. Bruno, "Hydrothermal Process for Making Ultrafine Metal Oxide Powders", US Patent # 5,776,239, July 1998.

42. T. Noguchi, K. Iwasa, R. Anselmann, M. Knapp and M. Loch, "Coated Spherical SiO2 Particles", US Patent # 5,846,310, Dec. 1998.
43. A.E. Braun, *Semiconductor International*, **23** (12), 66, 2000.
44. S. Avanzino, C. Woo, D.M. Schonauer and P.A. Burke, "Chemical-mechanical polishing slurry formulation and method for tungsten and titanium thin films", US Patent # 5,916,855, June 1999.
45. S. Grumbine, C. Streinz and B. Mueller, "Composition and Slurry Useful for Metal CMP", US Patent # 5,980,775, Nov. 1999.
46. J.P. Bare, MICRO, 53, Sept. 1997.
47. A.S. Dukhin and P.J. Goetz, "Acoustic and Electroacoustic Spectroscopy" in *Ultrasonic and Dielectric Characterization Techniques* Ed.V. Hackley and J. Texter, American Ceramic Society, 77, (1998).
48. *Handbook of Chemistry and Physics 70*th ed, R. Weast and D. Lide, Eds., CRC Press, Boca Raton, Fl, D274-275, 1989.
49. T. Vo, T. Buley and J.J. Gagliardi, Solid State Technology, **43** (6), 123, 2000.

8 CMP Cleaning

John de Larios

8.1 Introduction

Chemical Mechanical Planarization is an enabling technology that has rapidly spread throughput the semiconductor manufacturing process. CMP is now used from FEOL applications such as shallow trench isolation and polysilicon contacts to BEOL including planarization of dielectrics and conductors. From a CMP cleaning perspective, the ubiquitousness of CMP requires that cleaning technology is capable of cleaning a wide range of contaminants and materials. CMP cleaning is called on to remove slurry abrasive materials, such as silica and alumina, from Si, silicon oxide, low-k dielectrics, nitride, tungsten, and copper surfaces, not to mention the barrier and blocking layers required for CMP integration. In addition to meeting a vast array of technology hurdles, the CMP clean must also target manufacturing requirements driven by cost of ownership considerations. It is no small wonder then that many different cleaning technologies, both complimentary and competitive, are found in semiconductor manufacturing facilities.

It is important to understand that wafer cleaning traditionally plays a secondary role to "value added" processes such as etch and CVD, as well as CMP. It is mainly in the area of pre-gate or diffusion cleans that cleaning is considered a critical step. The CMP development engineer is driven to create a process that meets planarization requirements such as removal rate, uniformity, dishing, and erosion targets. The optimization of the process for low defects is typically relegated to a secondary status. Despite the fact that defects are polish dependent, high defect levels are often considered a cleaning issue rather than a polishing issue. The successful cleaning process engineer, therefore, often becomes an expert in optimizing the CMP process for defect reduction. This often takes the form of suggesting pad conditioning changes or modifying the slurry and DIW rinse flow to the primary or secondary pad.

CMP cleaning is differentiated from other types of wafer cleaning in several aspects. Wafer cleaning is normally a multi-step process with each step targeting one or more types of contamination. Most variations of the classic RCA clean [1], whether performed on a wet bench or spray processor, include steps to remove organics, particles, and metals. CMP cleaning is primarily focused on particle removal with some emphasis on metals removal. Organic contamination, other than pad material debris, is seldom a concern with oxide

and tungsten CMP. However, the organic corrosion inhibitors used during Cu CMP, such at benzotriazole (BTA), can cause the formation of organic defects on the surface.

8.1.1 Background on CMP Defects and Cleaning Issues

Wafer surfaces exposed to the aggressive chemistries and pressures present during CMP processing are contaminated with slurry residues, trace metals, and mobile ions. The resulting modification and damage of the surface and near-surface region of the dielectric layer are inherent to CMP [2]. While the subsurface damage can adversely affect device properties, a greater concern is the substantial and unavoidable levels of surface contamination [3]. The main goal of CMP cleaning [4, 5] is the reduction of slurry residues that can potentially reduce device yields. Slurry remaining on the surface (front, back, or bevel edge) of a wafer can cause patterning and/or deposition errors in subsequent process. These errors can translate directly into shorts and opens in the interconnect conductors. Proper polishing techniques, however, and an appropriate clean, can get slurry contamination levels well within the requirements for integration into a manufacturable process. Fundamental studies on the interactions between particles and surfaces during cleaning have been carried out [6] and analysis specific to CMP cleaning exist in the literature [7, 8]. For example, predicted adhesion forces for alumina-based slurry particles is calculated to be 16 to 20 times that of silica-based slurries [9]. These studies and similar work are important in understanding the various forces between the particle and the wafer surface and the forces on the particle arising from the cleaning step and should provide a guide for the process engineer in optimizing the parameters of the clean. Despite the existence of theoretical studies, most work done on the cleaning of planarized substrates is empirical in nature. Since semiconductor device manufacturers do not generally release yield data to the public, there is little information in the literature that correlates defect studies directly with wafer yield, with few exceptions [10]. However, it is clear that interactions between polishing and cleaning [11] exist and the quality of the polish and clean will impact yield. The amount and distribution of slurry residue is influenced by details of the polishing process such as primary platen pressure, pad conditioning, and the buff process.

   In addition to the slurry residue contamination, but of secondary importance, is contamination due to trace metals and mobile ions. These contaminants have the potential for moving through the dielectric and causing significant changes in the electrical properties of the devices. If the contamination lies below the surface of the wafer, it is often necessary to remove the topmost layer of the dielectric during the cleaning step. Many device fabrication requirements stipulate that trace metals and ionic contamination must be below TXRF detectability limits. Cleaning with HF removes the damaged near-surface region [2] and much of the trace metallic contamination [12].

While HF is beneficial for removing many types of contamination, there is always a risk of damage to the dielectric or exposed metals. These issues will be discussed in the section on oxide CMP.

Finally, the CMP clean must not degrade any material properties as it removes surface contamination. In the case of the polishing of polycrystalline silicon, it is often important that the surface roughness is not adversely impacted by the clean. For Cu CMP cleaning, the clean must prevent corrosion of the metal lines.

8.1.2 Overview of Cleaning Process Used for CMP

Contact cleaning [13, 14] is the primary means of CMP cleaning in the manufacturing of semiconductor devices. Double sided scrubbing in particular has historically commanded a large percent of the standalone CMP cleaning equipment market. As polisher companies have developed their own integrated cleaning systems, they have continued to rely on brush scrubbing to perform the bulk of the cleaning.

Non-contact cleaning [15, 16] utilizing megasonics or spray processing [17] is a viable means of cleaning following CMP. This is particularly true for dielectric CMP, since silica particles are considerably easier to remove relative to the alumina slurry particles used for tungsten and Cu CMP. Many suppliers of cleaning equipment offer systems that utilize combinations of contact and non-contact cleaning. Most mechanical brush scrubbers can be configured with a megasonic option, and several non-contact cleaners have found it advantageous to include a contact cleaning module for particularly tenacious contamination. A study of combinations of contact and non-contact cleaning has shown that essentially all surface contamination can be removed [18]. Other cleaning technologies have been proposed for CMP cleaning such as CO_2 snow cleaning and laser ablation.

Remarkably little attention has been given to wafer drying in the context of CMP cleaning with a few notable exceptions [16]. Nearly all CMP cleaners rely on high speed spinning to dry the wafer surface for both batch and single wafer processing. The dry process following CMP cleaning can be considered a non-critical process relative to a pre-gate process that may benefit from a Marangoni or IPA dry. The process following CMP cleaning, typically a CVD step, is not as sensitive to low levels of contamination as the active region of a device.

8.1.3 Dry-in/Dry-out Processing vs. Stand-Alone Polishing and Cleaning

As CMP technology developed, it became apparent that there are significant advantages to integrating the cleaning system to the polisher, providing so-called Dry-in/Dry-out processing. In contrast, Dry-in/Wet-out processing

presents certain challenges for device manufacturers. The polished wet wafers are a nuisance for the operators to move through the cleanroom. The wafers at this stage are contaminated with chemicals from the slurry that can drip on the cleanroom floor. It is accepted practice for the wafers to be kept wet after polishing and before cleaning since dried on slurry is more difficult to remove. Therefore, it is often necessary to add holding tanks for storage if the cleaner is not immediately available. However, these tanks add to the total equipment cost and require additional expensive cleanroom floor space. In general, it is best to minimize the time between polishing and scrubbing. This lessens the risk of slurry drying on the wafer and, in the case of Cu CMP, reduces corrosion of the Cu lines.

8.1.4 Metrology of CMP Contamination and Defect Identification

The proper use of metrology techniques is a key to successful CMP cleaning [19]. Standard bright-field and dark-field semiconductor defect metrology equipment for patterned and blanket wafers all have their place in the measurement of CMP defects. Detecting and accurately quantifying different types of defects following CMP is complicated by the fact that there are significant film thickness differences across a wafer and between wafers [20]. Bright-field imaging tools are hampered by non-uniform thickness because the amount of reflected light and the color of the light are thickness sensitive. This variation in reflected light can result in a background noise that reduces the sensitivity of the defect counter. For laser particle counters, the thickness variation causes a problem since the scattering of light from a particle on the surface is partially determined by the surface's local reflectance, which is film thickness dependent. Laser scattering particle counters provide a good means of locating defects, however, additional review using techniques such as SEM or AFM are required to properly classify CMP Defects. For mass production, several pattern wafer inspection system are available for locating and classifying defects.

TXRF and SIMS measurement techniques [21] are used to monitor trace metal contamination following CMP cleaning. While it is well accepted that slurry abrasive removal is critical, there is not complete agreement on the importance of trace metal contamination for BEOL CMP applications. Many fabs remove trace metals as part of the BEOL CMP slurry removal clean. Others do not use a specific post-CMP process to remove trace metals but may depend on a subsequent cleaning step to obtain the desired level of cleanliness. In certain fabs, trace metals removal is driven by the fear that contamination from CMP can spread beyond the CMP bay to other manufacturing process with possible disastrous results. This is particularly true for Cu CMP. Other fabs are driven by metrology considerations where the philosophy is that any measurable contaminant should be removed. This level of cleanliness is the most difficult to achieve because the target metals level is the detectability limit of the metrology technique employed. This approach

Table 8.1. Classification of post CMP defects

Class	Type	Before	NH4OH Scrub	HF Scrub
A	Scratch	< 5	< 5	< 5
B	Area Defect	20–500	10–50	0–10
C	Large Particle	$> 10^5$	< 150	< 50
D	Small Particle	10^3–10^9	10^4–10^8	$< 10^3$
E	Bevel Edge	10^5–10^9	10^2–10^4	10^2–10^3

also has the drawback that the cleanliness requirements are tied to metrology capabilities rather than yield.

CMP related defects are typically measured on the front side of a wafer using laser scattering instruments. While this well established technology offers reproducible and meaningful particle information, it has significant limitations. The main limitation of laser scattering tools is that they cannot detect all particles depending on their size, morphology, or location. For example, particles located in the edge exclusion area or on the bevel edge of the wafer cannot be identified. This type of contamination can have a deleterious impact on die yield since the contamination can transfer to the front of the wafer during downstream processing [10]. Other types of defects cannot be detected using particle counters because of their size or morphology considerations. These types of contamination are often easily visible with dark-field microscopy, scanning electron microscopy (SEM), or atomic force microscopy (AMF). It is, however, difficult to quantify defects with these methods since they do not scan the entire wafer surface.

Since CMP defects cover a broad range of sizes and morphologies, it is convenient to divide them into several classifications. For convenience, one method of classification is suggested in Table 8.1 [4]. The nature of these defects ranges from simple scratches to less understood defects caused by small clusters of slurry particles. In Table 8.1, defects are classified based on several distinguishing characteristics: 1) Are they damage related or true particles, 2) their absolute size, 3) their thickness relative to their lateral dimension, and 3) the defect density. Oxide CMP defects such as scratches are classified as Class A defects. These defects are a direct result of a failure of the polish and are caused by one of several problems [22]. Any large and hard substance, such as a chunk of dried slurry, falling on the polishing pad can lead to the semicircular scratch pattern shown in Fig. 8.1. These defects can be several microns wide and many millimeters long. Since Class A defects cannot be removed by any cleaning technique, they must be minimized or eliminated by proper control of polishing, the environment, and pad-conditioning techniques. This class of defects can cause shorts between metal lines during subsequent processing resulting in severe yield loss.
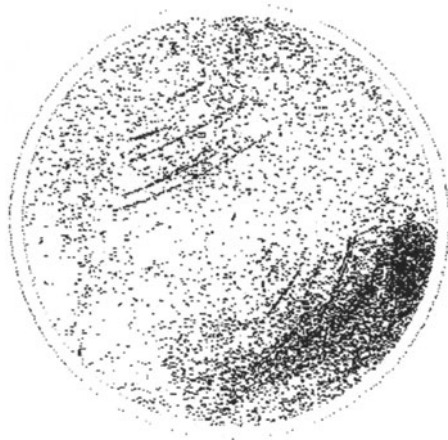
Fig. 8.1. True scratches on polished wafers have characteristic shapes. Shown is a typical semicircular-shaped scratch defect classified as a Class A defect.

Class B defects are seldom found in great numbers. On a laser scattering particle counter, they can appear as short area defects and may be misinterpreted as small scratches. However, as depicted by AFM in Fig. 8.2, many of these defects are clearly identified as slurry that appears to be smeared across the wafer surface. This type of defect can be several microns wide and tens of microns long. Class B defects are difficult to remove during the cleaning step unless HF is used to undercut the slurry. These defects may be caused by a poor control of the rinsing step on the polisher.
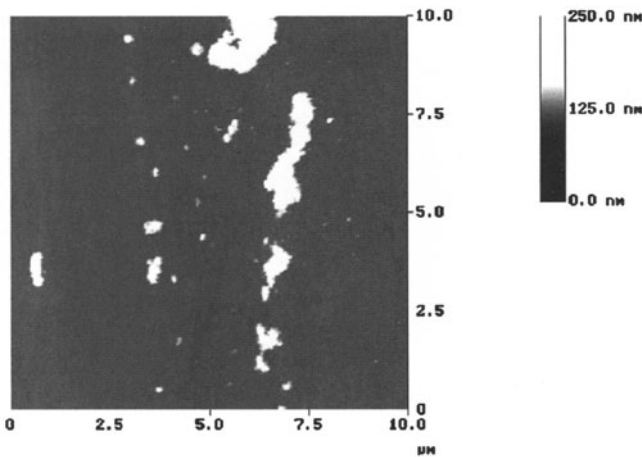
Fig. 8.2. Large-area Class B defects can be caused by agglomerated slurry particles strongly bonded to the wafer surface

Class C defects are ubiquitous to CMP. These defects are slurry particles loosely attached to the wafer surface. They are common to all polishing processes but are much reduced if a DIW buff is included. These particles come in a range of sizes since they are cause by agglomeration of particles. Class C defects, shown in Fig. 8.3 [23], are formed of individual slurry particles. SEM analysis has shown that these agglomerates are around 0.2 microns across and 0.1 to > 0.2 microns high. There can be $> 10^5$ of these particles on the wafer surface after polishing, but they are easily removed by standard cleaning techniques.

Class D defects are sufficiently small that they are often not detected by standard laser scattering techniques. Although AFM analysis indicates such defects may reach 0.5 microns in diameter, many are not detected using laser scattering because they are typically less than 600 Å high. The density of these
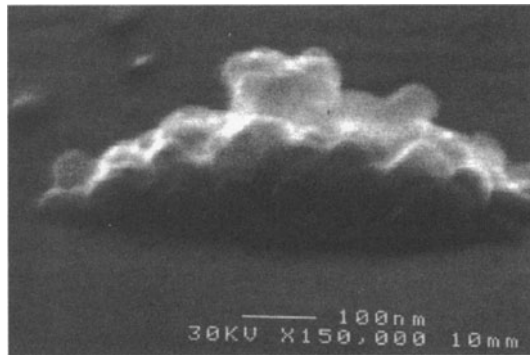
Fig. 8.3. Class C defects are caused by normal slurry agglomeration that is weakly bonded to the wafer surface. From [23]
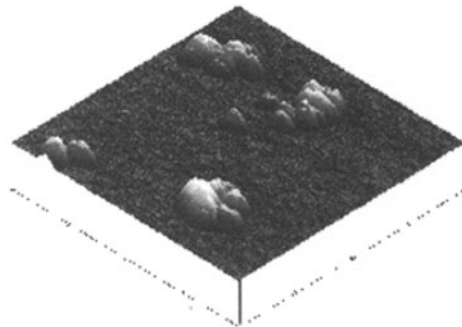
Fig. 8.4. Before cleaning, Class D defects are identified using AFM. The density of these clusters of slurry particles is dependent on polish parameters such as the rinse process
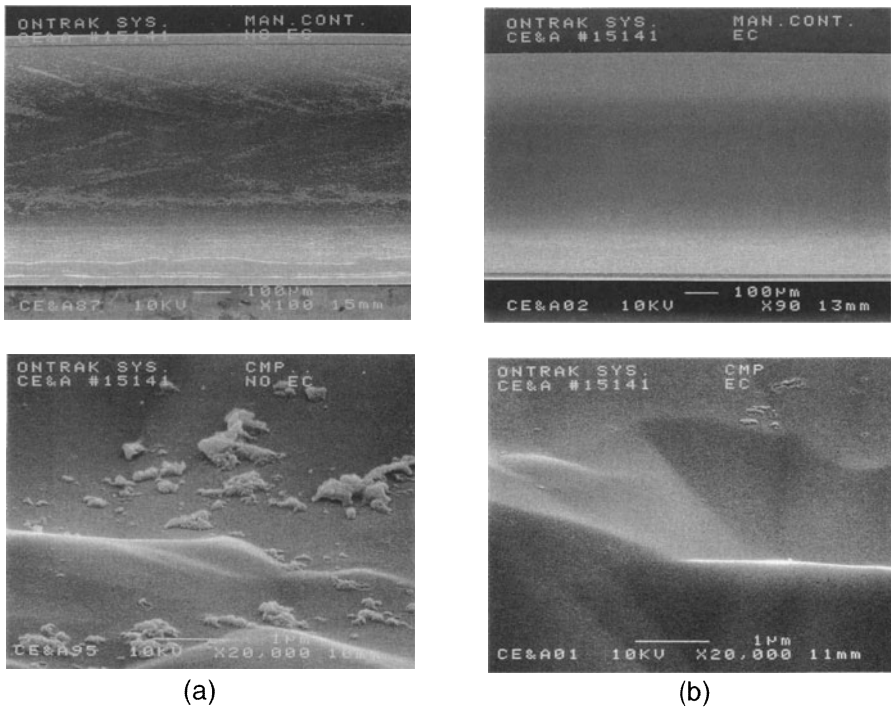
Fig. 8.5. A SEM comparison shows Class E slurry abrasive defects on the bevel edge of a wafer (**a**) before and (**b**) after cleaning

defects varies greatly, with estimates of total particle counts ranging from 10^3 to 10^9 defects/wafer. The lower densities are often found for DIW buff processes, while the higher concentrations can seen following a one step CMP process. The use of a water rinse on the hard pad can significantly increase the density of Class D defects if the rinse is not implemented correctly [24]. This type of defect differs greatly from the Class C variety: Class D defects are much smaller in height, can have an extremely high density, and are much more difficult to remove. AFM analyses indicate that these defects are composed of a small number of individual slurry particles bound together, typically in one or two layers, as illustrated in Fig. 8.4. Class D defects are poorly understood and because of their small size, their impact on device yield is not clear. These defects are strongly bonded to the dielectric surface but are readily removed with HF containing cleaning chemistries.

Class E defects cannot be observed using standard light scattering techniques because they are located on or near the bevel edge of the wafer. These defects are similar to Class C defects when viewed by SEM, yet they are not easily removed using standard brush scrubbing. The presence or absence is strongly dependent on the polishing equipment and process. This type of

contaminant can become dislodged during subsequent processing and then transfer to the front of the wafer where they can adversely affect device yield [10]. Figure 8.5 shows such defects before and after a scrubbing process with a special bevel edge cleaning hardware [25]. It is also possible to clean slurry contaminated bevel edges by rotating the wafer in a megasonic tank containing 2% NH_4OH and 2% H_2O_2 heated to $60°C$ when this process is followed by cleaning in a double sided brush scrubber [26]. Regardless which cleaning technology is employed, the amount of slurry adhering to the bevel edge and backside of the wafer can be minimized by hardware changes on the polisher carrier head [27].

8.2 Polishing and the Control of CMP Defects

8.2.1 Polishing Parameters and Polishing Defects

Although CMP cleaning technology is capable of reliably removing slurry residues, the polished film properties, including defect levels, can depend on the polishing system and operating conditions. For example, polishing can cause subsurface damage [2] altering the electrical characteristics of the polished dielectric. The degree of subsurface damage is dependent on the hardness of the polishing pad and the polishing pressure. Fourier-transform infrared spectroscopy has revealed chemical and structural modifications up to 0.2 microns below the polished surface [28]. Modifications in slurry chemistry and pad properties can produce dramatic improvements in light scattering defects [29]. Proper attention must be given to the rinse sequence following the polish and care must be taken to prevent slurry from drying on the wafer surface [30]. A change in defect surface morphology is observed for cases where platen rinsing follows the primary polish step [31]. With two step polishes, the chemical transition between the first platen, the DIW rinse, and the second platen needs to optimized to reduce defects [32].

Since the polishing process can produce light scattering defects that are not cleanable, such as scratches and pits, it is important to have a testing methodology for contaminating wafers independent of the polishing process. For this reason the "slurry dip" cleaning test was developed. As part of this process, bare wafers are dipped for several minutes in slurry of the same concentration as used on the polisher. They are then rinsed and run through the cleaning process. An optimized cleaning process will show no increase in defect counts when comparing particle counts before and after the slurry dip. This test provides a meaningful method to determine the first-order cleaning capability of a given cleaning technique. The slurry dip test can provide a baseline for additional cleaning tests following actual polishing. Care must be exercised when employing the slurry dip test if the polishing process changes the surface of the polished material from one condition to

another, such as transforming it from hydrophilic to hydrophobic. In this case, the slurry dip test can provide erroneous information.

8.2.2 To Buff or not to Buff

One of the standard planarization methods utilizes a two step polishing process. The primary polish on a hard pad which is used to remove the majority of the unwanted material is followed by a second step polish which is normally carried out on a soft pad. The purpose of this second step polish, or "buff," on a softer pad is to reduce surface defects including slurry and scratches. The use of a secondary platen buff with suitable chemical additives can further improve defect densities of a standard CMP cleaning process [33]. For W applications, and sometimes for dielectric applications, the second step buff is preformed with a silica slurry to remove scratches in the dielectric caused by the first step W removal with alumina slurry. Careful attention must be paid to the rinse step between the primary polish and the buff to minimize scratches and facilitate the removal of W and oxide slurry [32]. In Cu CMP, a second or third step is used for barrier film removal with an appropriate slurry.

With the development of advanced slurries, it can now be asked if buffing solely for the purpose of removing defects is really necessary [34]. Benefits derived from dispensing with the buff step include reduced complexity on the polisher and higher system throughput. Buffing is often a poorly controlled transition from polishing to cleaning [35]. Trace amounts of slurry transported on the wafers to a buffing pad may accumulate over time. In the case when buffing with DI Water is used as a pre-cleaning step, the buffing parameters, time, pressure and speed need to be optimized. The optimum buffing pressure for dielectric cleaning will depend on the pH of the buff [36].

8.3 Mechanical Brush Scrubbing for CMP Cleaning

Double sided mechanical brush scrubbers have been the mainstay for CMP cleaning since the inception of this planarization technology [37]. Brush scrubbing has the advantage of providing physical and chemical removal of surface contamination. Mechanical brush scrubbers are available in several configurations utilizing cylindrical, pancake, or pen style brushes. Regardless of brush configuration, the material of choice is invariably polyvinyl alcohol (PVA). The porous sponge-like PVA is compressed as it contacts the wafer surface during the cleaning process. These brush cleaning systems are often compatible with chemistries ranging from a pH of 2 up to 12. Mechanical brush scrubbing does have certain limitations in cleaning topography such as alignment marks or re-entrant holes in W plugs. These features are often found to be packed with slurry. Cleaning of these features is often improved with megasonic cleaning. Often scrubbing with dilute HF will aid in removal

of this type of topography related contamination. Brush scrubbers are often used in combination with non-contact cleaning, typically megasonics. The efficiency and benefits of these two techniques will often depend on the type of defect being removed [38].

8.3.1 Principals of Mechanical Brush Scrubbing

Current mechanical brush scrubbers are capable of reducing particle counts from >60,000 @ 0.2 microns to acceptable levels in the tens of particles per wafer. To accomplish this task, the scrubber and the process have been designed to remove the particles from the wafer surface, keep the particle entrained in the water, remove the contaminated water from the wafer surface, and finally dry the wafer with no water spot related residues. It is also imperative that the brush itself does not add particles to the wafer. If the brush is not properly maintained, a phenomenon commonly referred to as "brush loading" can lead to periodic and uncontrolled transfer of particles from the brush to the wafer surface. PVA brushes always contain particles within their pore network and effectively exchange particles with the wafer surface when brought into intimate contact during cleaning. Hydrophobic surfaces, such as HF stripped bare Si surfaces and as-grown or deposited oxide surfaces with high contact angles, have a strong tendency to pick up particles from the brush relative to hydrophilic surfaces. The hydrophobic surfaces may attract particles from the brush because of the multiple solid-liquid interfaces present on hydrophobic surfaces caused by the agitation of the water during scrubbing [39]. If the brush contamination is not severe, the brush can be recovered by running bare Si wafers through the scrubber. These Si wafers can accept particles from the brush and reduce the level of brush loading. If the brush cannot be recovered by this method, then the brush must be changed.

In order to accomplish the task of removing slurry abrasive with a mechanical brush scrubber, the mechanical, chemical, and electrical forces need to be controlled to optimize removal from various surfaces [13]. Understanding of the zeta potential concept provides a first order model for the differences in oxide and tungsten CMP cleaning [40]. According to this model, a particle immersed in water develops a charge at its surface [41]. The sign of the charge depends on the particle composition, the presence of molecular compounds attached to the surface, if any, the chemical makeup of the liquid, and the pH of the liquid. The zeta potential argument is based on the assumption that materials whose zeta potential have the same sign are repelled from each other while those materials with opposite signs are attracted. It is therefore obvious that it is beneficial for the polished surface to have the opposite zeta potential relative to the slurry abrasive. The zeta potential for the important materials in brush scrubbing are shown in Fig. 8.6. This figure shows that the PVA brush and a silicon oxide surface, and therefore the silica slurry abrasive, have approximately the same zeta potential over a wide range of pH. This

implies that silica slurry will be relatively easy to remove from the SiO_2 surface using DIW. Figure 8.6 indicates a significantly different zeta potential for alumina relative to silica. For example, at a neutral pH, a SiO_2 oxide surface and an alumina slurry particle have zeta potentials of opposite signs. This suggests that alumina slurry will be difficult to remove with DIW. Indeed, as will be seen in the section on W CMP, the use of a high pH cleaning chemistry is beneficial for removal of the alumina slurry. An additional method of overcoming zeta potential constraints is through the addition of a chemical that attaches to the surfaces and modifies the zeta potential. The ability to modify the zeta potential with this method allows cleaning to be carried out over a wider range of pH [42]. Figure 8.6 also suggests that alumina slurries can be readily removed at a pH less than 2, since in this regime the alumina and the oxide surface have positive zeta potentials.

The mechanical brush scrubbing process can be optimized by adjusting the speed of the brush and pressure of the brush on the substrate. Smaller brushes, such as pencil brushes, rotate at significantly higher speeds but have a smaller contact area relative to a slower rotating cylindrical brush. The relatively rigid brushes, typically nylon, that were use on brush scrubbers in the 1970's, could cause scratches if brought in contact with the wafer surface. This problem has been alleviated as PVA came to dominate scrubbing. This soft, flexible foam material is brought into direct contact with the wafer surface and is actually compressed several millimeters during the cleaning process. The PVA material has an open structure consisting of interconnecting cells that allow the brush to be constantly flushed with DIW and other chemicals during wafer cleaning. Brush life times of 20,000 wafers/brush are typical and over 100,000 wafers/brush have been recorded.
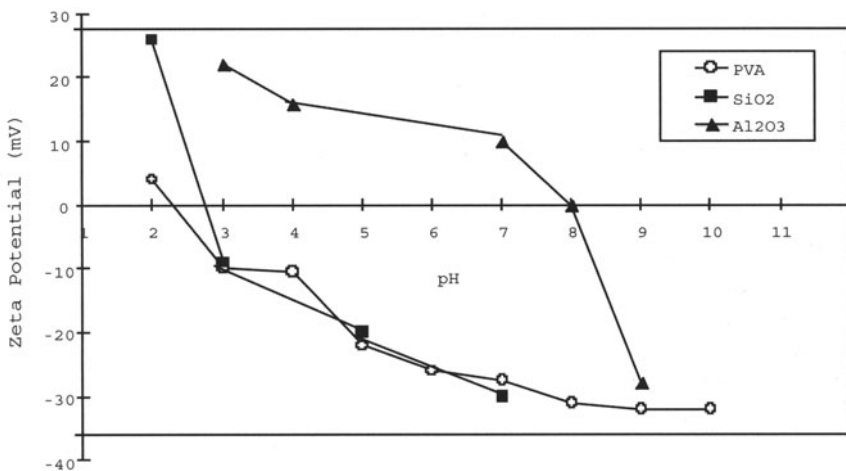
Fig. 8.6. Zeta potentials vs. pH are shown for PVA, SiO_2 and Al_2O_3

As the brush height is lowered and pressure on the wafer increased, the number of particles remaining on the surface reaches a minimum. This brush compression is usually found to be on the order of 2 mm for large cylindrical brushes. As the brush is lowered beyond the optimum point and there is excessive pressure on the wafer surface, the particle count will increase. This relationship between brush height and defect counts illustrates that mechanisms other than hydrodynamics are important in controlling defect removal. PVA brushes are compatible with cleaning chemistries in the 2 to 12 pH range, and NH_4OH is commonly used with this brush material for control of the zeta potential. Dilute HF in the 0.5% to 1% range can also be used with the PVA brush to remove trace metals and strongly adhered slurry. While the PVA is compatible with mild oxidants, such as a dilute SC-1, it is vulnerable to ozone. Care must be taken to isolate the cleaning system if the DI water source receives a periodic ozonation. UV light can also be used to break down incoming ozone it is present in the DI supply.

The removal of trace metals can also be accomplished on a mechanical brush scrubber [43]. This process used to remove metals will be discussed in detail in the section on Oxide CMP.

8.4 Non-Contact Processes for CMP Cleaning

There are several alternatives to the standard contact clean of the double sided mechanical brush scrubber including megasonic cleaning and spray processing. Each of these cleaning technologies has advantages and disadvantages depending on cost of ownership, fab integration issues, and process requirements. While there is some discussion that non-contact cleaning has less risk of damaging patterned wafers compared to contact cleaning, there is little evidence that this is the case. For example, mechanical brush scrubbing has been successfully utilized for the final clean in the manufacture of Si substrates [44]. Compared to the harsh chemical environment and large mechanical forces exerted on the wafer by polishing equipment, most cleaning technologies can be considered relatively benign.

8.4.1 Megasonic Cleaning

Optimization of megasonic cleaning has demonstrated the ability to clean polished wafers in batch [45] and single wafer processing [46]. Batch CMP cleaning processes based on megasonic energy can have significant throughput advantages over single wafer cleaning techniques, such as mechanical brush scrubbing. Batch megasonic cleaning approaches are well characterized since they are based on standard non-CMP cleaning technology. Extensive DOEs have been performed to optimize megasonic cleaning for CMP applications [47].

Single wafer cleaning using a proprietary megasonic source has demonstrated cleaning capabilities comparable to contact cleaning [48]. This method rotates the wafer under a quartz transducer held parallel to the wafer surface. This single wafer process is compatible with integration into a polisher for dry-in and dry-out cleaning. Optimizing the cleaning process with NH_4OH delivered to the frontside of the wafer and HF to the backside of the wafer allowed removal of slurry from the bevel edge of the wafer as well as the front- and backside. The cleaning process is optimized by adjusting process parameters such as the chemical composition of the cleaning liquid, liquid flow rate, the gap distance between the transducer and the wafer, and the megasonic power. The cleaning efficiency was also found to be dependent on the amount and type of gas dissolved in the cleaning fluid. De-gassing the liquid substantially reduced the cleaning efficiency.

8.4.2 Spray Processing

A study of spray processing technology with dilute concentrations of traditional cleaning solutions showed that this technique is capable of cleaning polished Si and TEOS surfaces [17]. The Si surfaces were successfully cleaned using repeated exposure to HF and SC1 chemistries. The process times were controlled such that the surface remained hydrophilic at all times. A comparison of different chemicals used on the spray processor indicate that ammonia peroxide mixtures provide better particle results compared to a dilute HF process. However, the HF last clean resulted in minimal levels of trace metal contamination.

Marathon testing of a non-contact spray process has been demonstrated to provide a reliable and effective means of wafer cleaning polished oxide surface with results comparable to those with brush scrubbing [49]. No statistically significant dependency was found relating cleaning efficiency to NH_4OH concentration or temperature. In a manufacturing environment, variations in cleaning results were related to post-polishing treatments such as buffing and rinsing limitations, not the performance of the cleaning process

8.5 Other Cleaning Technologies

Traditional brush scrubbing and wet immersion cleaning account for nearly all of CMP cleaning found in the manufacturing of semiconductor devices. However, in an effort to improve cleaning technology, other methods including aerosol cleaning, microcluster beams, and laser cleaning are being investigated. Aerosol cleaning has demonstrated cleaning capabilities on topography following CMP. This approach uses high pressure liquids, such as argon, nitrogen, or carbon dioxide, delivered through a properly sized nozzle to form a snow-like material. This solid material is used to transfer momentum to the wafer surface thereby removing surface contamination. These materials have

the advantage of complete evaporation from the wafer surface, ideally, leaving no residue. The use of CO_2 snow on polished substrates can reduce the defect level compared to mechanical brush scrubbing, particularly on wafers with topography [50]. Disk media for the hard drive industry often have considerable levels of topography present after polishing. The snow clean process appears to have an advantage in certain situations over brush technology. This technology is being developed to extend its capabilities to include back-side substrate cleaning as well as front-side. The ability to remove silica slurry residues with Ar aerosol cleaning has been demonstrated to be comparable to conventional cleaning processes [51]. However, there was a risk of pattern damage at high aerosol velocity and the cost per wafer was approximately three times higher.

8.6 Cleaning of Oxides, W, STI, Cu, and low k Materials

8.6.1 Oxide CMP Cleaning

As CMP became identified as an enabling technology, the planarization of oxides became the starting point for most new CMP installations. Therefore, much of the early studies on CMP technology are focused on oxides [17, 52, 53]. Most oxide CMP is performed with silica slurry. After polishing, the surface is grossly contaminated with Class C defects as described in the section on Metrology of CMP Contamination and Defect Identification. These defects are easily removed from the wafer surface using standard cleaning techniques. With these silica slurries, it is often possible to run DIW processes which have the advantage of being cost effective relative to those process that require the use of chemicals [54]. One major drawback with the DIW or NH_4OH process is their inability to remove trace levels of metallics, as
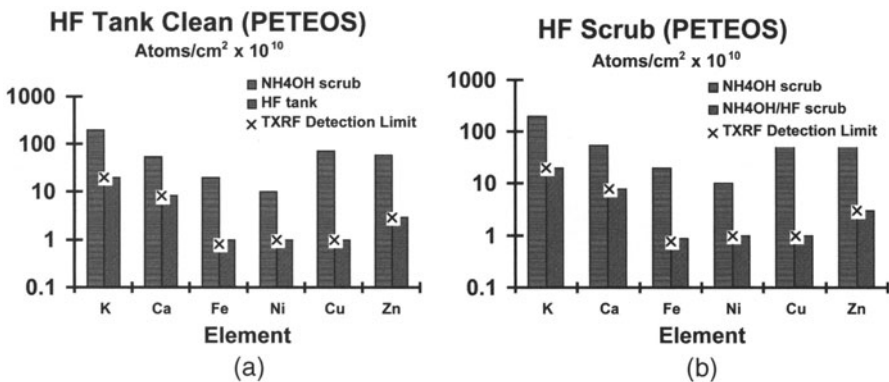
Fig. 8.7. Using HF as the final cleaning step, either (**a**) on the scrubber or (**b**) in a wet tank, reduces the trace metals concentrations to acceptable levels. From [3]
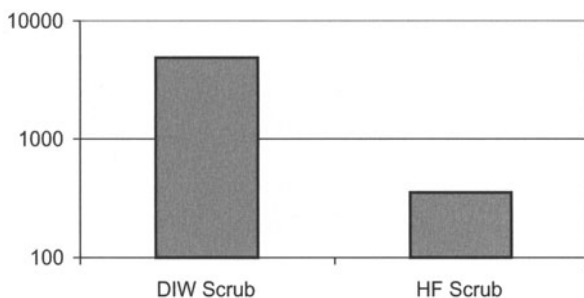
Fig. 8.8. Small Class D type defects > 0.16 microns visible on a laser particle counter are not removed with a DIW scrub but are removed with a HF process

indicated in Fig. 8.7. In this figure [3], TXRF analysis of polished oxides that were first given an NH_4OH scrub and then either a) dipped in an HF tank or b) given a subsequent HF scrub. This data shows that the NH_4OH scrub has a high level of metallic contamination while either HF process brings the metals level at or below detectability limits.

Another advantage of HF is the ability to remove small agglomerated slurry contamination described earlier as Class D defects. This type of defect, which can be observed with and SEM or AFM but are not always visible with laser scattering tools because of their minimal height. These defects are difficult to remove with DIW but readily removed with HF, as shown in Fig. 8.8.
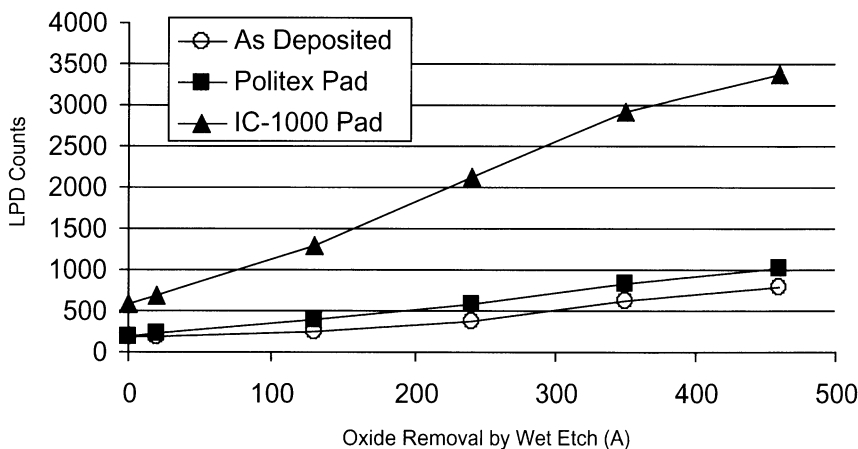
Fig. 8.9. Light point defects on oxide surfaces increase with increasing HF exposure. Surfaces polished with a harder pad exhibit higher defect densities. From [55]

Exposure of polished surfaces to HF does have potential disadvantages. Laser scattering studies of blanket polished oxide surfaces indicates that over-exposure to HF causes defect levels to increase dramatically. The degree of damage and the related defect level is dependent on the polishing process [2], as shown in Fig. 8.9, for defects > 0.16 microns [55]. The harder pad results in a higher level of defects compared to an unpolished surface and polishing with a softer pad. AFM analysis of polish defects before and after exposure to
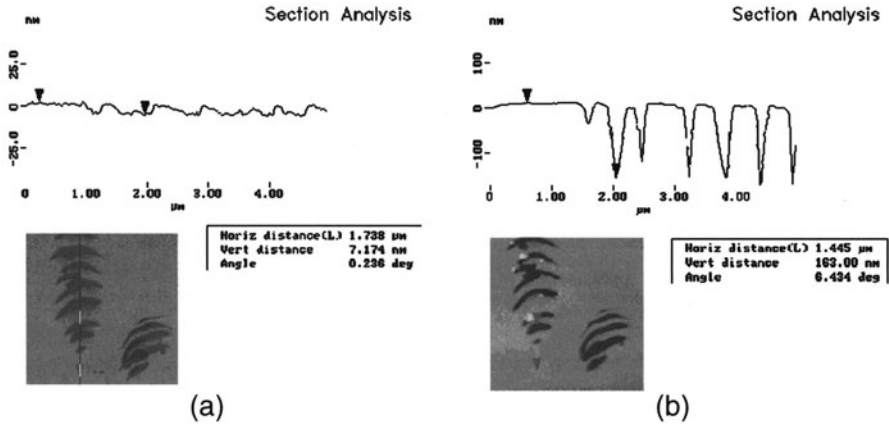
Fig. 8.10. AFM image of the same polish defect (**a**) before and (**b**) after exposure to HF shows accelerated etching of the damaged areas relative to the undamaged surface
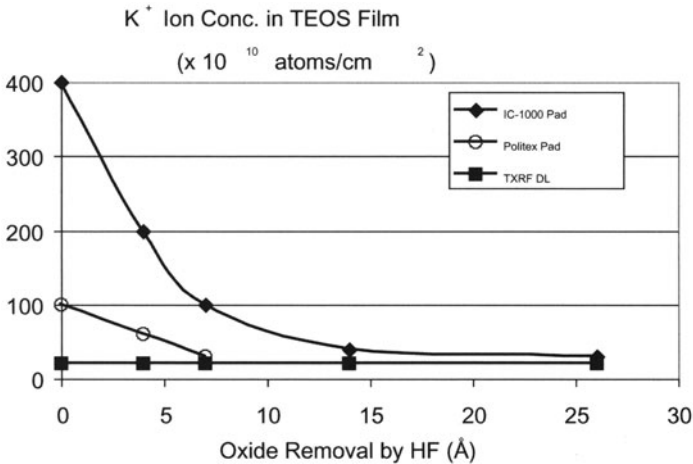
Fig. 8.11. TXRF measurements of K concentrations indicate that HF leaning can remove this material. The depth of penetration is dependent on the hardness of the pad

HF reveal that CMP related scratches increased in depth from 7 nm to over 160 nm when only 15 nm of blanket oxide was removed as shown in Fig. 8.10. TXRF analysis of the polished oxides revealed that the depth of penetration of the K below the surface is dependent on the hardness of the polishing pad, as shown in Fig. 8.11. A soft pad produced penetrations depths less than 10 Å compared to the 15 Å penetration of the harder pad.

8.6.2 W CMP Cleaning

W CMP has replaced traditional W etch-back for removing the over-burden material and defining the inter-metal contact. Processing W plugs with a RIE etch-back process is a notoriously dirty process which can reduce probe yields almost 15% compared to W CMP [56]. One slurry abrasive material commonly used for W CMP is alumina. While the alumina performs well for removing W, it often causes unacceptable scratching of the dielectric. Therefore, W CMP has been traditionally a two step process. A second step silica slurry was used to remove scratches in the dielectric caused by the alumina particles. Recent advances in alumina slurry technology has produced slurries which cause significantly less scratching of the oxide which can eliminate the need for the second step polish. Following W CMP, the surface of the wafer has exposed tungsten and barrier materials as well as the dielectric. Often the W plug is recessed after polishing. In addition, keyholes in the W plugs and recessed alignment marks are present. The combination of multiple materials with the topography can complicate the cleaning process.

As W CMP was developed, it was observed that the cleaning processes that worked well for oxide CMP were not effective. In the case of mechanical brush scrubbing with standard DIW processes, the brushes rapidly became loaded with the alumina slurry particles so that the efficiency of the clean was much reduced. Non-contact cleaning processes also did not produce acceptable results compared to oxide CMP cleaning. A first order explanation for this difference exists in the zeta potential differences with the alumina compared to silica. Figure 8.6 indicates that with a neutral pH DIW process, the alumina slurry particles have a positive potential relative to the PVA brush material and the oxide surface. Due to the zeta potential differences in these materials, alumina particles mechanically removed by the brushes would be attracted to the brush surface or redeposited on the wafer surface. This effect reduces the ability of the cleaning process to flush the slurry from the wafer surface in the DIW. It is well established that adjusting the pH to over 10 through the addition of NH_4OH to the scrubbing process enables removal of the alumina [57]. Particle detection analysis of patterned wafers cleaned with a) DIW and b) NH_4OH show the effectiveness of the high pH process for removing alumina slurry, as shown in Fig. 8.12. The high pH process reduces the defect count by two orders of magnitude.

Certain manufacturing process are sensitive to trace metals deposited on the dielectric surface following W CMP. Therefore, processes have been
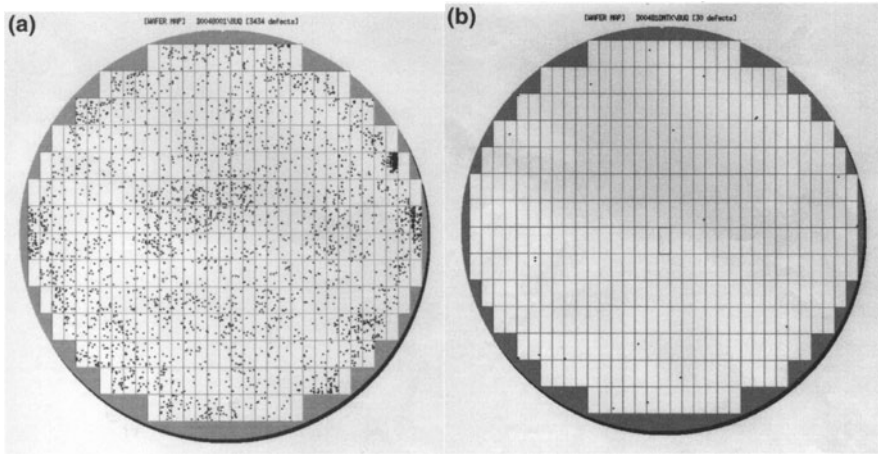
Fig. 8.12. Defect analysis of W CMP wafers show improved leaning for (**a**) a NH₄OH scrub compared to (**b**) a DIW scrub

Plug Size: 2μm
Scan Size: 1μm
Z Height: 50nm
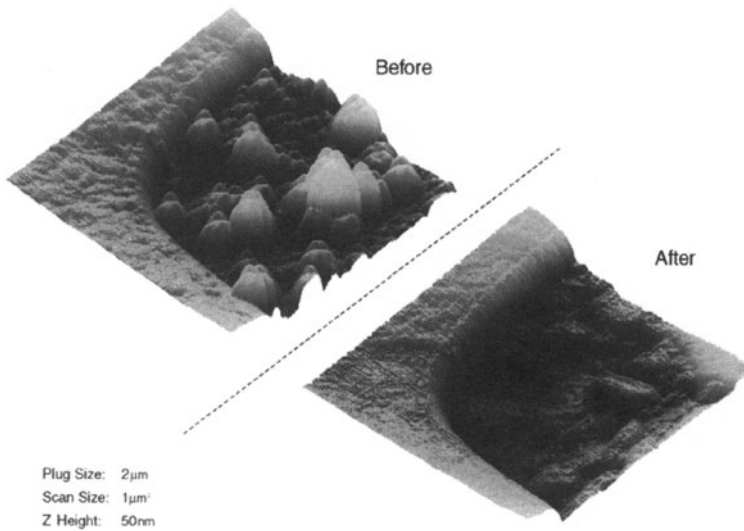
Fig. 8.13. A polished W plug befoe and after scrubbing with 0.5% HF does not show any damage to either the plug or the barrier material compared to plugs cleaned with NH₄OH. From [58]

developed to remove these contaminants with dilute HF as explained in the section on Oxide CMP. However, in the case of W CMP, the W plug as well as the barrier layer are exposed to the cleaning chemistry. Although W is known to be etched in HF, a judicious use of dilute HF will remove the contamination without damaging the plug or barrier layers. This can be seen
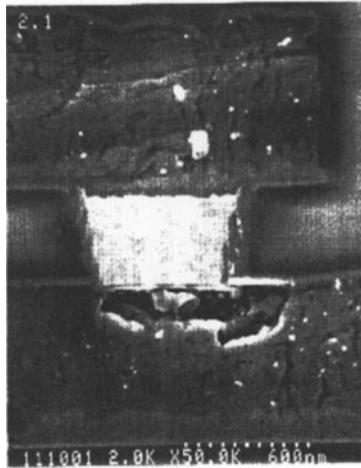
Fig. 8.14. The interconnect metal layers can be seriously damged if the W plug is overexposed to HF. From [61]

in Fig. 8.13, which compares a W plug after NH_4OH scrubbing and after 0.5% HF scrubbing [58]. As can be seen from these AFM images, the plug is slightly recessed following the polish and NH_4OH scrub. There is slightly less recess but no sign of damage to the plug after the HF process. The target for this clean is to remove approximately 100 Å from the dielectric surface. This slight etch is sufficient to remove metals plated on the surface of the oxide as well as mobile ions, particularly K, that have penetrated below the surface into the damaged subsurface layer. Although HF can readily remove trace metals, overuse of this chemistry can adversely impact the electrical properties of the polished oxide. Electrical leakage current studies of oxides polished with W slurries indicate that the leakage increases with increasing HF exposure [59]. Other studies show that a one step W polish followed with a dilute HF clean produced lower contact resistance for vias compared to a one step polish only and a W polish followed by an oxide buff [60]. These apparently contradictory results are likely due to differences in the primary polish conditions which would cause different levels of pitting and subsurface damage in the dielectric. However, it is clear that HF overexposure can seriously damage the tungsten plug and the underlying Al lines as seen in Fig. 8.14 [61].

8.6.3 Poly CMP Cleaning

Poly silicon continues to be an important material for such applications as trench isolation and local interconnects. Following the planarization of poly, the surface is predominantly hydrogen terminated and hydrophobic unless steps are taken on the polisher to alter the surface. For those applications

that do not require hydrophobic poly surfaces, most CMP cleaning process convert the surface to a hydrophilic state. This conversion is often performed on a megasonic tank using SC1 to grow a thin chemical oxide. An alternative method utilizes a mechanical brush scrubber with SC1 [62].

8.6.4 STI CMP Cleaning

Shallow trench isolation in one of the latest applications benefiting from the planarization capabilities of CMP. Since CMP is inherently sensitive to device pattern densities, integration problems are being attacked from the device side with the aid of reverse masks and dummy structures [63]. Slurry manufacturers are developing silica-based slurries as well as high selectivity ceria-based slurries in an effort to optimize dishing and erosion and minimize defects. Planarization with fumed silica slurries has shown acceptable particle residues and microscratches after cleaning. The polish process has the goal of planarizing the deposited oxide and stopping on the nitride layer with minimal erosion and dishing. Of equal importance to these polishing requirements is the clearing of the oxide from the nitride otherwise there will be a masking affect when the nitride is removed. The cleaning step must therefore have the capability of removing contamination from nitride and oxide surfaces.

8.6.5 Cu CMP Cleaning

Copper CMP poses a unique set of challenges to the CMP cleaning process [64]. Cu is much more susceptible to corrosion following planarization compared to other materials. Indeed, in some cases photo-activated corrosion can cause Cu lines to completely disappear. Contaminants from the polishing process must be removed to prevent corrosion of the Cu interconnects without damaging the conductor, while Cu must be removed to acceptable levels from the dielectric between the lines. In addition, since copper as a contaminant diffuses quickly in silicon and silicon dioxide, it must be removed from all wafer surfaces, (front, back, and bevel edge) in order to prevent an adverse effect on device performance.

During the copper CMP process, the copper layer is oxidized to form copper oxides and copper hydroxides, depending upon the slurry pH, electrochemical potential, and additives. In a basic or neutral pH cleaning on brush scrubbers, these copper oxides and hydroxides do not dissolve and may be easily transferred to the PVA brushes during brush scrubbing. If the brushes become contaminated, or are loaded by the copper oxides, they may transfer the copper contaminants onto subsequently processed wafers [65]. This brush loading effect would then cause severe copper cross-contamination. In the case of W CMP, this type of brush loading is prevented by delivering NH_4OH to the scrubbing process. However, while the ammonium hydroxide can prevent loading due to the alumina slurry, it cannot prevent loading due

to the copper oxides. Additionally, scrubbing with dilute NH_4OH can cause etching of the copper lines resulting in unacceptable surface roughening.

Brush loading from the alumina slurry and the copper oxides can be prevented without attacking the Cu lines by cleaning with low pH chemistries and the proper chemical additives. Several types of inorganic acids can be used to modify the zeta potential [66] such that the electrostatic charges on the surfaces are all positive during the cleaning process. As shown in Fig. 8.15, TEOS wafers dipped in an alumina based Cu slurry scrubbed with DIW result in typical brush loading. Those wafers scrubbed with the Cu cleaning chemistry containing inorganic acids to modify the zeta potential did not show evidence of brush loading. Cu cleaning chemistries often contain fluorinated species that are able to provide a slight etch of the dielectric to reduce trace metal contamination levels. Several alternative sources of Cu CMP cleaning chemical are available that show capabilities of removing alumina slurry particles and reducing trace levels of Cu to acceptable levels [67].

A non-contact Cu CMP clean has been demonstrated using both batch and single wafer processing [68]. Blanket Cu-TaN-FSG (fluorinated silica glass) films polished to clear the barrier metal using a two step polish without a buff cleaned using a scrubber or a megasonic bath showed equivalent results for defects greater than 0.2 microns. The cleaning chemistries consisted of caustic solutions followed by an organic fluoride mixture. A comparison of serpent-comb yield loss due to shorts was essentially equal for both the scrub and batch megasonic processes, indicating the effectiveness of both cleaning processes.
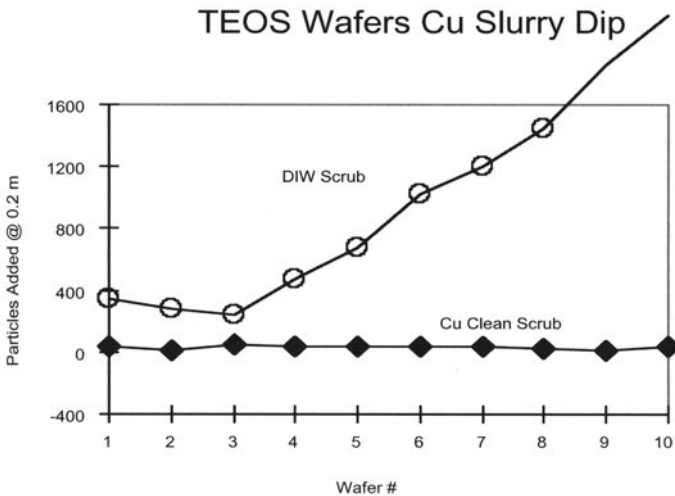
Fig. 8.15. Oxide wafers dipped Cu slurry and cleaned in a brush scrubber show expected brush loading when DIW is used. Adjusting the zeta potential of the chemical mixture results in clean wafers

Time-of-Flight Secondary Ion Mass Spectroscopy (TOF-SIMS) has been used to inspect copper patterned wafers for metal contamination on the dielectric between copper lines [69]. In the standard dynamic SIMS technique, the sample is sputtered away by a continuously operating primary ion beam. In contrast, TOF-SIMS instruments operate in or near the static SIMS regime where the primary ion gun operates in a pulsed mode and the secondary ions are representative of the immediate surface area. Typical analytical depths of TOF-SIMS are the top few monolayers of the sample. This technique is ideal for studying trace levels of Cu on the dielectric surface after CMP cleaning since it is also able to image the patterned Cu lines. Correlation studies between TOF-SIMS and TXRF have shown that these measurements are generally in agreement within a factor of two. TOF-SIMS was used to optimize a Cu CMP cleaning process to reduce the level of Cu between the Cu lines. As shown in Figure 8.16, a 60 micron square area is imaged using TOF-SIMS for CMP clean A and B. Clean B was the most effective on the 10 micron Cu lines with a 20 micron pitch, reducing the residual copper levels on the oxide areas by two orders of magnitude compared to clean A.

Cu CMP and cleaning present new problems relative to oxide and W processing since Cu is much more liable to corrode in both acid and basic environments. To reduce this type of problem, many slurry suppliers add corrosion inhibitors, such as benzotriazole (BTA) to their slurry to form a Cu-BTA

Clean A   Clean B
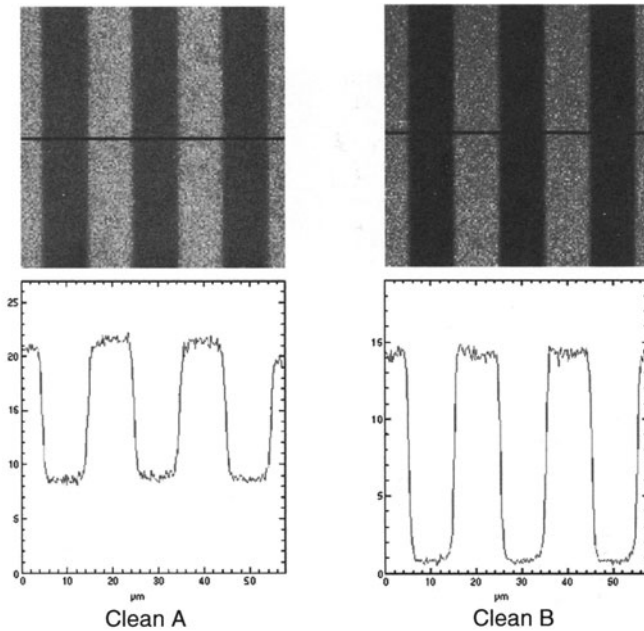
Fig. 8.16. Time of flight-SIMS is an effective technique to quantify trace levels of Cu on the dielectric between Cu lines

polymer on the surface. These Cu-BTA films can offer a significant degree of
protection against aggressive electrolytes as well as high humidity and ele-
vated temperatures [70]. While this passivating layer does reduce corrosion,
it makes the surface hydrophopbic and therefore more difficult to clean. Since
the presence of this organic polymer has also been linked to peeling of subse-
quent CVD films, it has limited use as a post-cleaning corrosion inhibitor. In
general, corrosion can be a more random event compared to particle removal
and it is therefore difficult, with some exceptions, to correlate corrosion to
a particular source. Corrosion is often seen associated with areas that can
trap chemical reactants such as interfaces, pits, and particles [71]. Typical
corrosion along the Cu/Barrier/Oxide area is shown in Fig. 8.17. This event
may have been caused by a defect in one of these layers that allowed the
slurry chemical to become trapped and not properly removed during the
CMP clean.

A particularly destructive corrosion process has been observed when
cleaning Cu CMP lines contacting active devices. This type of corrosion is
not observed when processing blanket Cu surfaces or patterned short loop
wafers where the Cu is not connected to the active devices. This type of cor-
rosion is attributed to current generated by exposure of the p-n junctions to
light [72]. This photo-assisted mechanism, illustrated in Fig. 8.18, results in
catastrophic displacement of Cu from the lines connected to the p-doped re-
gions to the lines connected to the n-doped regions. In this case, the cleaning
solution completes the circuit of the p/n junction and the junctions act like
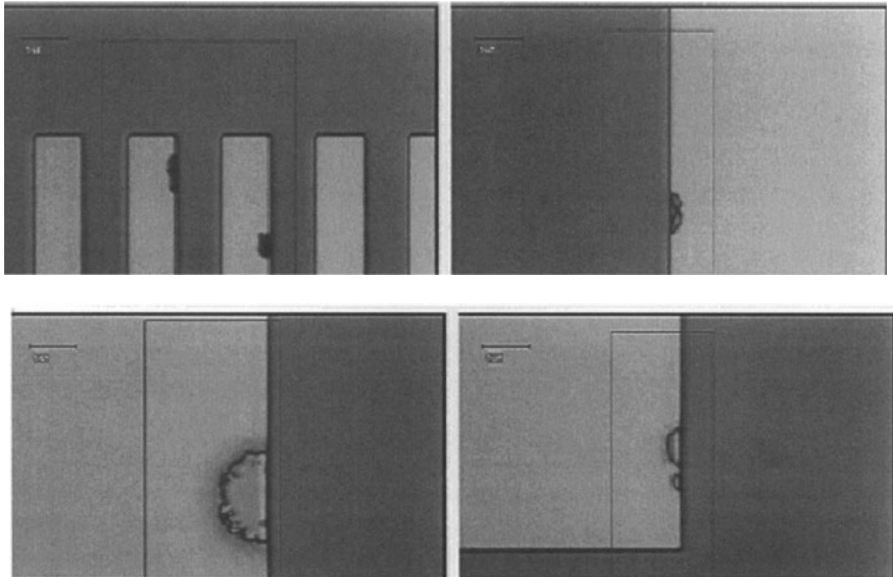
Fig. 8.17. The Cu/barrier/dielectric interfaces are prime areas for corrosion
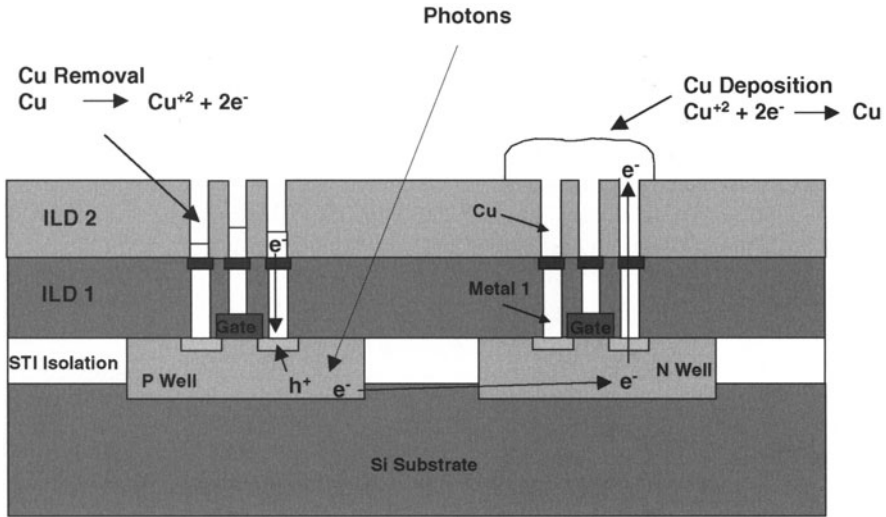
Fig. 8.18. Exposing p-n junctions to light can cause the connecting Cu lines to rapidly corrode

a solar cell. This deleterious phenomena can be reduced by adding corrosion inhibitors or reducing light impinging on the wafer surface during cleaning.

8.6.6 Low k CMP Cleaning

The path leading to the integration of low-k dielectric materials into semiconductor manufacturing is not yet clear. The many families of low-k materials, each with their own dielectric and mechanical properties, are competing in development labs and pilot lines with no obvious victor [73]. Novel organic and inorganic low-k materials include those dielectrics deposited by chemical vapor deposition (CVD) or spin-on coating. Of particular concern for CMP and cleaning is the ability of the low-k to withstand the pressures and chemistry of these wet processes. Often a two-step planarization process is required on these relatively soft, easily scratched films. Certain low-k films require a capping layer and will not be polished directly. In these cases, adhesion of the capping layer to the low-k is important just as is adhesion of the low-k to the underlying material. For those low-k layers that are polished directly, defect detection should not present any new issues although new defect classifications may be observed [74].

After planarization, low-k thin films are contaminated with slurry abrasive particles and trace metals in a situation similar to silica based dielectrics. However, there is an additional challenge with these newer materials as they are typically hydrophobic [75]. In the case of mechanical brush scrubbing, the commonly used cleaning chemistries such as NH_4OH and HF fail to wet

the polished low-k surface. This has led to the development of proprietary cleaning chemical mixtures that can be used in conjunction with Al or Cu damascene low-k applications. These mixtures were used in a cleaning study of a spin-on poly(arylene ether) organic low-k polished with an ZrO_2 slurry, an aromatic hydrocarbon polymer polished with an Al_2O_3 slurry, and a inorganic hydrogen silsequioxane polished with a silica slurry. Specific cleaning mixtures targeting Al and Cu metals were able to remove slurry particles without degrading the dielectric properties of the low-k.

8.7 Future Directions for CMP Cleaning

CMP and CMP cleaning will continue to be important technologies for the foreseeable future. The benefits of global and local planarization that are inherent with CMP will become increasingly important as device dimensions continue to shrink. The stand-alone cleaners that dominated the semiconductor industry through the mid 1990's are rapidly being replaced by cleaning systems integrated into polishers. With an accelerated merging of businesses in the semiconductor capital equipment industry, the number of companies producing CMP equipment is steadily decreasing. Most of these surviving companies have developed, or have purchased, their own CMP cleaners. The market for non-polishing companies to market their own cleaners will steadily decrease.

It appears unlikely that the established cleaning technologies, mechanical brush scrubbing and megasonic cleaning, or combinations of the two, will be displaced in the foreseeable future. These technologies have proven to be cost effective and able to deliver the required die yield. Alternatives such as snow cleaning, however, may be able to carve out a niche in certain specialty applications. Current CMP cleaning processes have seen been able to support CMP over a wide range of materials, starting from standard deposited dielectrics and extending to poly, tungsten, and copper. A potential hurdle that may be insurmountable to any aqueous cleaning process is the possible inability of very low k porous materials to handle exposure to water. Open cell materials may not tolerate exposure to CMP cleaning chemicals. They may require a different medium to provide momentum transfer to remove particles and a different solvent to remove trace metals.

8.7.1 CMP Cleaning at 300 mm and Beyond

CMP cleaning has made the transition from 100mm through all the wafer size changes up to 300mm. The cleaning process has not been pushed to its limits during this transition. In fact, although it is not well documented, blanket wafers with the smaller diameters are marginally more difficult to clean. Comparisons of particle data from double sided mechanical brush scrubbers

for 200mm and 300mm wafers often indicate that slightly lower defect densities are obtainable on the larger diameter wafers. Although this phenomena is not well understood, it maybe related to a better cleaning efficiency near the edge of a wafer compared to the center.

There are no fundamental limitations related to wafer size for the accepted methods of CMP cleaning. Brush scrubbers and immersion technologies will not be constrained by increasing wafer sizes. Wafer cleaning scales much easier than those technologies that rely on uniform distribution of gases and plasma sources. Planarization using conventional slurries with various abrasives is a well entrenched technology. There are however, many drawbacks with standard slurries such as consumable costs, delivery systems, and defects. These real or perceived limitations have led to the development of fixed abrasives as a replacement for traditional slurries. Polishing with these fixed abrasives has demonstrated good planarization capabilities [76]. As with standard slurry based polishing, scratches can also be an issue [77]. Although the fixed abrasive pads do not require a particle containing slurry, the pads are imbedded with a hard material that may be left on the surface of the wafer after polishing [78].

8.8 Conclusion

Wafer cleaning technology has demonstrated that it has the capability to support chemical mechanical planarization for dielectric, metal, and silicon applications. The user has a choice of cleaning technologies for the removal of the abrasive slurry particles and trace metal contamination. Contact cleaning, non-contact cleaning, and combinations of these processes are all found in manufacturing environments. These aqueous based cleaning technologies will likely dominate for the foreseeable future since they are cost effective and well understood. Alternative cleaning methods, laser or aerosol based, may have applications in niche areas where the planarized material cannot tolerate contact with DI water.

References

1. W. Kern and D. Puotinen, RCA Review, **31**, 187, 1970.
2. F. Kaufman, S. Cohen and M. Jaso, MRS Symposium Proceedings, **365**, 85, 1995.
3. M.A. Ravkin, D.L. Hetherington, J.M. de Larios, D.G. Gardner and W.C. Krusell, Proceedings 1996 CMP-MIC Conference, 177, IMIC, Tampa, 1996.
4. J. de Larios, J. Zhang, E. Zhao, T. Gockel and M. Ravkin, "Evaluating Chemical Mechanical Cleaning Technology for Post CMP Applications, MICRO, 61, May 1997.

5. J.M. Steigerwald, S.P. Murarka and R.J. Gutmann, "Chemical Mechanical Planarization of Microelectronic Materials," 1997, J. Wiley & Sons.

6. M. Olim, J. Electrochem. Soc., **144**, 3657, 1997.

7. F. Zhang, A. Busnaina and G. Ahmadi, J. Electrochem. Soc., **146**, 2665, 1999.

8. K. Bahten, H. Liang, D. McMullen, E. Estragnat, T.G. Zhang and J. Lee, "A Study of the Mechanics of Brush Scrubbing and Particle Removal in Post-CMP Cleaning Applications," 5th International Symposium of Chemical Mechanical Polishing, August 13–16, 2000, Lake Placid New York.

9. F. Zhang, A. Busnaina, J. Feng and M. Fury, Proceedings 1999 CMP-MIC Conference, 61, IMIC, Tampa, 1999.

10. G. Banks, P. Carr, J. Farber and R. Kurjanski, "Reduction of Edge Defects in a High Volume Production Environment Using the Lam OnTrak Wafer Cleaner," SEMICON/West Seminar on Contamination Free Manufacturing, July, 2000, San Francisco, CA.

11. J. Shen, W.D. Costas, L.M. Cook and J. Farber, J. Electrochem. Soc., **145**, 4240, 1998.

12. Y. Gobil, M. Fayolle, F. Tardif, O. Demolliens, S. Deleonibus and F. Romagna, "Characterization and Cleaning of CMP Induced Defects," Continuing Education in Engineering, University Extension, University of California, Berkeley, October 4–6, 1994, Austin TX.

13. S. Roy, I. Ali, G. Shinn, N. Furusawa, R. Sha, S. Peterman, K. Witt and S. Eastman, J. Electrochem. Soc., **142**, 216, 1995.

14. W. Krusell, J.M. de Larios and J. Zhang, "Mechanical Brush Scrubbing for Post-CMP Cleaning," Solid State Technology, 109, June, 1995.

15. M. Oleson and B. Fraser, Proceedings 2000 CMP-MIC Conference, 530, IMIC, Tampa, 2000.

16. W. Fyen, R. Vos, I. Teerlinck, E. Vranckaen, J. Grillaert, M. Meuris, P. Mertens and Marc Heyns, Proceedings 2000 CMP-MIC Conference, 507, IMIC, Tampa, 2000.

17. M. Jolley, "Spray Cleaning Procedure for Silicon Wafers after Chemical/Mechanical Planarization Polishing (CMP)," MICRO '94 Proceedings, 546, 1994.

18. Y.-L. Wang, W.-T. Tseng and M.-S. Feng, Proceedings 1997 CMP-MIC Conference, 347, IMIC, Tampa, 1997.

19. R. Shah, S. Roy, G. Shinn and I. Ali, "Metrology: Key to Successful Cleaning Post Chemical Mechanical Polishing," Microcontamination Conference Proceedings, 185, 1994.

20. R. Howland and B. Trafas, "Wafer Inspection Techniques for CMP Applications," SEMI/Europa '96 Technical Program: Implementation and Integration of Chemical Mechanical Polishing, March 29, 1996, Geneva, Switzerland.

21. R. Hockett, "Ultratrace Impurity Analysis of Silicon Surfaces by SIMS & TRXF Methods," in Handbook of Semiconductor Wafer Cleaning Technology, ed. W. Kern, Noyes Publication, Park Ridge, N.J., 1993.

22. E. Tseng, Proceedings 2000 CMP-MIC Conference, 503, IMIC, Tampa, 2000.

23. Dale Hetherington, Private Communication.

24. M. Ravkin, D. Hetherington and D. Zhang, Proceedings 1997 CMP-MIC Conference, 423, IMIC, Tampa, 1997.

25. M. Moinpour, H. Nguyen, M. Salek, Y. Park, T. Bramblett, J. de Larios, L. Ryle, D. Anderson and W. Krusell, "Method and Apparatus for Cleaning Edges of Contaminated Substrates, US Patent 5,861,066.

26. J. Tang, B. Fishkin, A. Lerner. M. Sugarman, F. Redeker and B. Brown, Proceedings 2000 CMP-MIC Conference, 517, IMIC, Tampa, 2000.
27. K. Ikenouchi, T. Murakami and Y. Miyoshi, Proceedings 1999 CMP-MIC Conference, 271, IMIC, Tampa, 1999.
28. J. Trogolo and K. Rajan, Journal of Materials Science, **29**, 4554, 1994.
29. A. Francis, P. Feeney, G. Bogush and G. Khan, Proceedings 2000 CMP-MIC Conference, 391, IMIC, Tampa, 2000.
30. J. Schlueter, "Controlling Contamination During CMP Processes, Solid State Technology, 58, May 1996.
31. M. Ravkin, J. Zhang, A. Jensen, A. Pant, D. Hetherington, K. Achuthan and J. Doyle, Proceedings 1997 CMP-MIC Conference, 423, IMIC, Tampa, 1997.
32. K.-J. Chen, H.B. Lu and P.W. Yen, Proceedings 2000 CMP-MIC Conference, 511, IMIC, Tampa, 2000.
33. C. Huynk, M. Rutten, R. Cheek and H. Linde, "A Study of Post-Chemical Mechanical Polish Cleaning Strategies," Chemical Mechanical Planarization in Integrated Circuit Device Manufacturing, The Electrochem. Soc. Proceedings **96-22**, 17, 1996.
34. M. Ravkin and K. Mikhaylich, "Is Buffing Really Needed for Modern CMP?," SEMICON/West Technical Symposium, CMP Technology for ULSI Interconnection, July 10–14, San Francisco, CA.
35. M. Ravkin and K. Mikhaylich, "Transition from Polishing to Cleaning for Oxide CMP," 4th International Symposium on Chemical Mechanical Polishing, August 13–16, 1999, Lake Placid, New York.
36. I. Ali, S. Roy, G. Shinn, S. Raghavan, R. Shah and S. Peterman, "The Effect of Secondary Platen Downforce on Post-Chemical Mechanical Planarization Cleaning," MICRO '94 Conference, 196, 1994.
37. W. Krusell, "The Resurgence of Mechanical Brush Scrubbing Systems for Post-CMP Cleaning," Semicon West, 1994.
38. B. Albrecht, L. Fritz and R. Solis, "Reducing Defect Density Using an Optimized Wafer Scrubber," MICRO, 39, July 1997.
39. M. Ravkin, J. Farber, I. Malik, J. Zhang, A. Jensen, J. de Larios and W. Krusell, "Cleaning of SiO_2: Differences between Thermal and Deposited Oxides," Material Research Society Spring Meeting, April 17–21, 1995, San Francisco, CA.
40. W. Krusell, J.M. de Larios and J. Zhang, "Mechanical Brush Scrubbing for Post-CMP Cleaning," Solid State Technology, 109, June, 1995.
41. M. Ranade, "Adhesion and Removal of Fine Particles on Surfaces," Aerosol Science and Technology, **7**, 161, 1987.
42. L. Zhang, S. Raghavan, S. Meikle and G. Hudson, J. Electrochem. Soc., **146**, 1442, 1999.
43. A. Jha and P. Gopalan, "Particle Removal and Metallic Contamination Control on CMP Oxide Film Surface Using Auriga EC System," 7th International Symposium on Particles on Surface: Detection, Adhesion, and Removal, June 19–21, 2000, Newark, N.J.
44. D. Golland, P. Albrecht, W. Krusell and F. Puerto, "The Clean Module: Advanced Technology for Processing Silicon Wafers," Semiconductor International, 184, September, 1987.
45. B. Fraser, M. Olesen, T. Phan and B. Morrison, The Electrochem. Soc. Proceedings, **35**, 634, 1997.

46. C. Franklin, Y. Wu and T. Nicolosi, "Using Acoustic Parameters to Optimize Performance of Single Wafer Non-Contact Post CMP Cleaning," 5th International Symposium of Chemical Mechanical Polishing, August 13–16, 2000, Lake Placid, New York.
47. G. Gale and A. Busnaina, "Physical and Chemical Effects of Megasonics on Liquid Based-cleaning of Si Surfaces," Proceedings of the Adhesion Society, 520, 19th Annual Meeting, Blacksburg, VA, 1996.
48. M. Olesen, B. Fraser, C. Franklin and M. Bran, Chemical Mechanical Planarization in Integrated Circuit Device Manufacturing Symposium, The Electrochem. Soc. Proceedings **98–7**, 81, 1998.
49. K. Witt, M. Jolley and P. Burke, Proceedings 1997 CMP-MIC Conference, 351, IMIC, Tampa, 1997.
50. A. Hakanson, N. Garg, M. Borden and H. Chung, "Removing Post-CMP Residue through Carbon Dioxide Snow Cleaning," MICRO, 75, March, 2000.
51. M.H. Lee, K. Lee, Y.P. Han and S.R. Hah, "Study of Argon Aerosol Cleaning Method for Post-CMP Cleaning Application," 7th International Symposium on Particles on Surfaces: Detection, Adhesion, and Removal, June 19–21, 2000, Newark, N.J.
52. S. Cohen, M. Jaso and A. Bright, J. Electrochem. Soc., **139**, 3572, 1992.
53. D. Hetherington, P. Resnick, R. Timon, B. Draper, M. Ravkin, J. de Larios, W. Krusell and A. Madhani, Proceedings 1995 DUMIC Conference, 156, IMIC, Tampa, 1995.
54. M. Dax, "Contamination Control News: DI Water Process Shows Advantages," Semiconductor International, 50, March 1996.
55. E. Zhao, M. Ravkin and W. Krusell, Presented at the MRS Spring Meeting, San Francisco, CA, 1998. Symposium was not published.
56. V. Blaschke, L. Witters, S.-W. Hsia, D. Dornisch and K. Rafftesaeth, Proceedings 1997 CMP-MIC Conference, 219, IMIC, Tampa, 1997.
57. T. Myers, M. Fury and W. Krusell, "Post-tungsten CMP Cleaning: Issues and Solutions, Solid State Technology, 59, October 1995.
58. E. Zhao, R. Emami, I. Malik, K. Mishra, W. Krusell, J. de Larios and D. Hymes, MRS Spring Meeting, San Francisco, CA, 1997, MRS Symposium Proceedings, **477**, 1997.
59. J.Y. Kim, B.U. Yoon, I.K. Jeong, U.I. Chung, Y.B. Koh and M.Y. Lee, Proceedings 1997 CMP-MIC Conference, 433, IMIC, Tampa, 1997.
60. B.T. Lin, P.K. Nil, S.N. Peng and K.H. Wang, Proceedings 2000 CMP-MIC Conference, 139, IMIC, Tampa, 2000.
61. D. Hetherington, Private communication.
62. E. Zhao, D. Hymes, J. Zhang and W. Krusell, "Cleaning Process for Polysilicon CMP Applications," Cleanroom, January 1998.
63. J. Schlueter, "Trench Warfare: CMP and Shallow Trench Isolation," Semiconductor International, 123, October, 1999.
64. D. Hymes, H. Li, E. Zhao and J. de Larios, "The Challenges of the Copper Clean," Semiconductor International, 117, June 1998.
65. E. Zhao, L. Zhang, H. Li, D. Hymes, J. de Larios and W. Krusell, Proceedings 1998 CMP-MIC Conference, 359, IMIC, Tampa, 1998.
66. L. Zhang, S. Raghavan, S. Meikle and G. Hudson, J. Electrochemical Soc. Proceedings, **98–7**, 161, 1998.

67. M. Peterson, R. Small, G. Shaw III, Z. Chen and T. Truong, "Investigating CMP and post-CMP Cleaning Issues for Dual-Damascene Copper Technology, MICRO, 7, January, 1999.
68. B. Fraser, S. Rafie, M. Eissa and S. Joshi, "Wafer Cleaning: Noncontact Megasonics for Post-Cu CMP Cleaning," Solid State Technology, 105, July, 2000.
69. H. Li, D. Hymes, J. de Larios, I. Mowat and P. Lindley, "Using TOF-SIMS to Inspect Copper-Patterned Wafers for Metal Contamination, MICRO, 8, March 1999.
70. V. Brusic, M. Frisch, B. Eldridge, F. Novak, F. Kaufman, B. Rush and G. Frankel, J. Electrochem. Soc., **138**, 2253, 1991.
71. J. de Larios, "Cu and low k CMP Cleaning," SEMICON/West 99, SEMI Technical Program., CMP Technology for ULSI Interconnection, July 13, 1999, pg. H-1.
72. A. Beverina, H. Bernard, J. Palleu, J. Torres and F. Tardif, Electrochemical and Solid-State Letters, **3** (3), Mar 2000, p. 156.
73. M. Fury, "CMP Processing with low-k Dielectrics," Solid State Technology, 107, July 1999.
74. M. Fury and D. Towery, Journal of Electronic Materials, **27,** 1088,1998.
75. L. Jiang, M. Ravkin, D. Hymes, J. Zhang and J. de Larios, Proceedings 1999 CMP-MIC Conference, 278, IMIC, Tampa, 1999.
76. J. Gagliardi and D. Bange, SME 1st Int. Mach. & Grinding Conf., MR95-196, 1995.
77. A. Romer, T. Donohue, J. Gagliardi, F. Weimar, P. Thieme and M. Hollatz, Proceedings 2000 CMP-MIC Conference, 265, IMIC, Tampa, 2000.
78. K. Mikhaylich and M. Ravkin, Proceedings 2000 CMP-MIC Conference, 337, IMIC, Tampa, 2000.

9 Patterned Wafer Effects

D. Boning and D. Hetherington

9.1 Introduction

Chemical-mechanical polishing (CMP) is used in integrated circuit manufac-
turing to remove material from patterned wafer surfaces. The term "polish-
ing" or "planarization" is used to describe two types of processes: 1) topog-
raphy reduction of surface features that result from deposition, etch, or other
thin-film processes, and 2) removal ("polish back") of films leaving material
only in desired recessed regions. The later case is also called a "damascene"
or "polish-back" CMP process.

CMP was introduced in large part because of its ability to achieve better
planarity than other approaches that existed in semiconductor manufacturing
such as resist etchback and spin-on techniques [1, 2]. CMP has also been a key
enabler for the formation of damascene via and line structures. However,
CMP does not achieve ideally perfectly flat surfaces on the feature scale,
chip scale, or wafer scale.

In this chapter, we focus on how the "pattern" or initial topography on the
wafer surface (i.e., the layout pattern imprinted on the chip surface specific
to each product) impacts or interacts with CMP. We begin in Sect. 9.2 by
considering planarization terminology and the methods used to characterize
planarization processes in manufacturing. Next, Sect. 9.3 describes pattern
dependencies in oxide CMP emphasizing results obtained from experimental
studies published in the literature. This section also describes pattern depen-
dencies in shallow trench isolation (STI) polishing, an important polish-back
dielectric CMP process. STI CMP processes interact strongly with patterns
and are considered to be one of the more difficult challenges for integra-
tion into manufacturing. The polishing of metal films is then considered in
Sect. 9.4. First, we discuss pattern dependent polishing of tungsten to form
vias and local interconnects. Next, these issues are addressed for copper pol-
ishing, where dishing and erosion is a substantial yield and variation concern.

9.2 Planarization Terminology and Characterization

Planarization of topography is required for advanced integrated circuit manu-
facturing primarily due to photolithography depth of focus effects. Planariza-

tion minimizes surface topography that occurs, for example, when a thin film is deposited over patterned features such as a metal interconnect layer. The degree to which planarization relieves topography can range from local smoothing to complete global planarization as shown in Fig. 9.1 based on Wolf [3]. Complete global planarization is desired, however, there are no known processes that produce this effect over widely varying surface topographies and pattern layout densities. A smoothing process rounds off the topography features and is the least effective in planarizing. Local planarization generates a flat surface within a group or array of circuit features but does not significantly reduce topography at the edge of the array. Spin-on glass and high density plasma (HDP) chemical vapor deposition (CVD) techniques locally planarize [4].

Chemical-mechanical polishing processes fall into the near global planarization category, which is characterized by a high degree of local flatness with a reduction of the step occurring at the edges of large circuit arrays. CMP processes produce an effective planarity over millimeter distances. The degree of global planarization depends on many factors that will be discussed in this chapter.
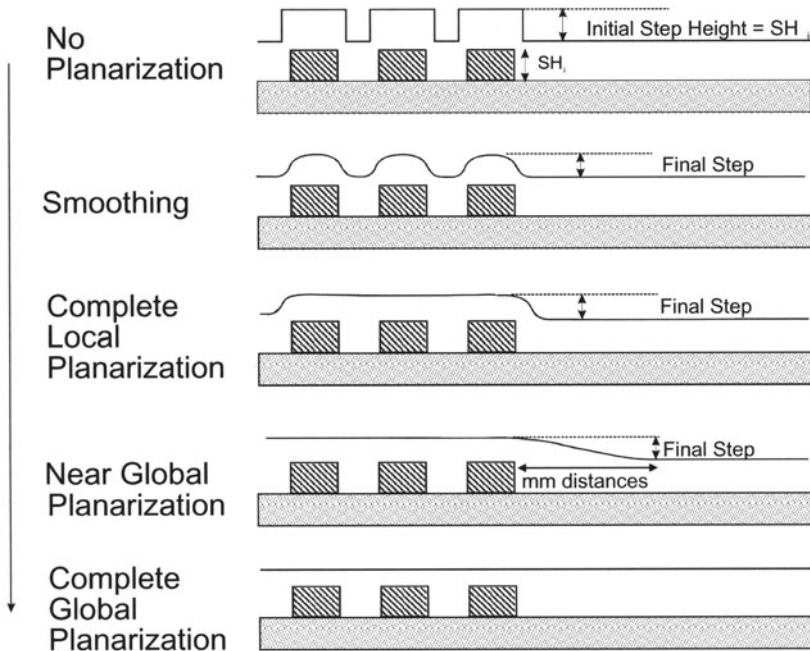
Fig. 9.1. Degrees of planarization that are possible in semiconductor processing technology. Chemical-mechanical polishing falls into the near global planarization category. After [3]

Characterizing planarization is time consuming and sometimes difficult to measure. Often the data gathered does not give a complete picture of the process capabilities but only results specific to a set of patterns. In this section we discuss the planarization terminology and review some of the methods used to characterize it.

9.2.1 Step Height Reduction

The term "step height" is defined as the vertical elevation of a surface asperity (or feature) as shown in Fig. 9.2. The step height of a film deposited over an isolated patterned feature approximates the thickness of the underlying patterned feature. An example of this is shown in Fig. 9.3. The cross-sectional scanning electron microscope (SEM) image shows silicon dioxide material deposited over a series of Al-Cu interconnect lines. The SiO_2 was deposited using a tetraethlyorthosilicate (TEOS) precursor in a plasma-enhanced CVD reactor. As shown in Fig. 9.3, the step height at the edge of the array approximates the thickness of the Al-Cu metal line ($0.8\,\mu m$). When the lines are closely spaced together, however, the conformal nature of the deposition causes the profile to overlap. In this case the step height of the oxide material reduces considerably from the underlying pattern thickness.

Step heights are measured using a stylus profilometer; however, for accuracy the tip must be able to penetrate and clear the space between the raised areas. In very small areas, atomic force microscopy (AFM) [5] or high resolution profilometry [6] can be used. Measuring step heights post-CMP across a die (global planarization category) is difficult. There is no abrupt step feature, but instead, a smooth and gradually changing surface over long-range distances (millimeters). Under these conditions, profilometer techniques can introduce large errors from wafer substrate and stage flatness.

Dielectric film thickness measurements can be correlated to step height. This is shown schematically in Fig. 9.2. Measurements are performed pre- and post-CMP using optical grating spectrometry and/or ellipsometry techniques. Dispersion models for each film stack are required that fit well to the experimentally measured data (goodness of fit > 0.9). The minimum meas-
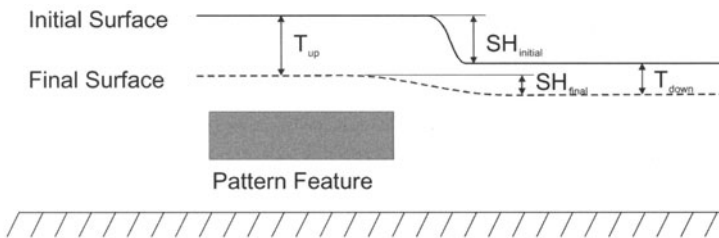
Fig. 9.2. Diagram of the initial and final surface above a patterned feature depicting changes in step height
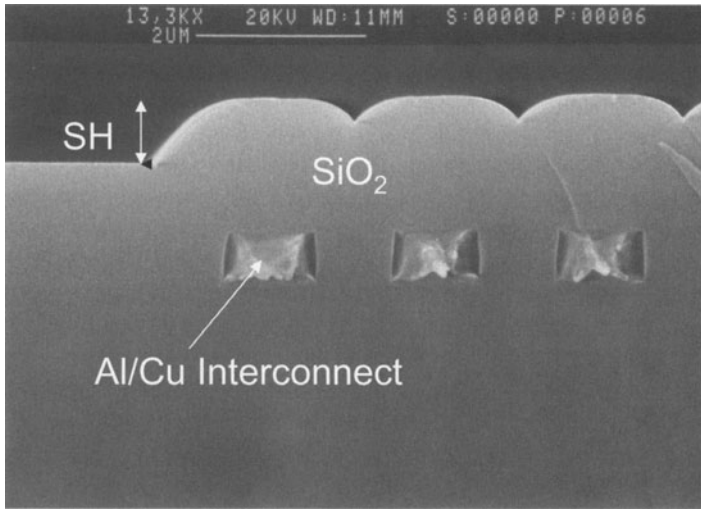
Fig. 9.3. Scanning electron microscope cross-section image showing silicon dioxide material deposited over an aluminum copper interconnect pattern. The step height between each feature differs from the step height at the edge

urement spot size is on the order of $10\,\mu\text{m}$, which prevents film thickness measurements in regions containing minimum size features. Measurement boxes are usually incorporated into the test chip layout area adjacent to the small feature size circuit region and are matched to the pattern density within the small feature size region.

Step heights (pre and post CMP) are related to material removed during planarization by

$$SH_{\text{final}} + T_{\text{up}} = SH_{\text{initial}} + T_{\text{down}}, \tag{9.1}$$

where T_{up} and T_{down} are the material thickness values *removed* in the "up" and "down" areas, respectively. SH_{final} is the final step height (after the planarization process) and SH_{initial} is the initial step height.

The step height reduction ratio ($SHRR$) is a measure of the degree of planarization and is given by

$$SHRR = 1 - \frac{SH_{\text{final}}}{SH_{\text{initial}}}. \tag{9.2}$$

Complete global planarization has a $SHRR$ value of 1.0. CMP processes can produce $SHRR$ values greater than 0.95 depending on the type of process, the consumable set utilized, and the geometric pattern layout.

9.2.2 Planarization Efficiency

Ideally, a CMP process would remove the material only over raised features producing perfect planarization. In actuality, material is removed from both

up and down areas. Referring to Fig. 9.2, planarization efficiency (EFF) is determined by

$$EFF = 1 - \frac{T_{\text{down}}}{T_{\text{up}}}.\qquad(9.3)$$

An efficiency of 1.0 implies that the step is removed without loss of material in the down areas. An efficiency of zero indicates that the down area polish rate is equivalent to the up area polish rate and the initial step height was not removed. Planarization efficiency described in 9.3 is equivalent to the SHRR given in 9.1 when T_{up} equals SH_i.

Other planarization efficiency metrics have been described by Borba et al. [7]. One term defines the time required to achieve the final step height (or to completely remove the local step). The other metric gives the average amount of film removed over the active area as well as the down area.

9.2.3 Pattern Characterization

Geometric Layout Parameters

The pattern dependent topography must be carefully considered when analyzing CMP results. Geometric terms such as pitch, linewidth and density are illustrated in Fig. 9.4 for a portion of an integrated circuit. The example shown in Fig. 9.4 is a plan view layout of a metal interconnect layer. Pitch
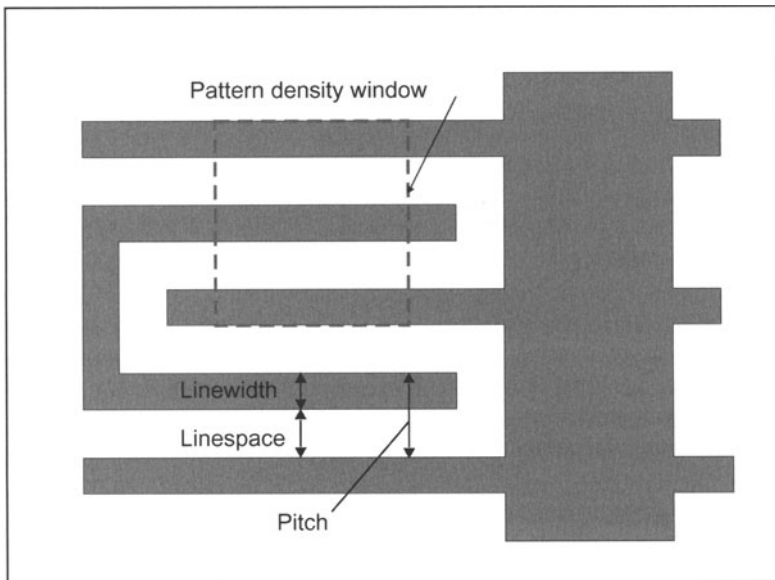
Fig. 9.4. Plan view example layout of a circuit pattern showing the concept of pitch, linewidth, linespace and pattern density

is defined as the length of a repeating edge of a metal line and equals the sum of the linewidth and linespace dimensions. Layout pattern density is the ratio of mask (chrome) layout area to the total specified area as shown in Fig. 9.4. Pattern density is sometimes referred to as pattern factor. The pattern density value is a ratio term that implies analysis over a fixed geometric dimension (pattern density window). The ratio of mask area to total area can produce different density results depending on the area over which density is computed. For the specific example given here, the pattern density value expressed as a percentage is 54.5%.

Film Deposition Pattern Dependencies

The deposition technique influences the asperity density of the initial patterned topography. Several techniques are used to deposit silicon dioxide including plasma- and thermally-assisted chemical vapor deposition (CVD) [8]. Thermal techniques include low pressure (LPCVD), atmospheric pressure (APCVD) and sub-atmospheric (SACVD). Plasma assisted methods include plasma enhanced (PECVD) and high density plasma (HDP). Spin-on glass technology is also a method utilized in many low-k material BEOL applications. For a review of dielectric deposition technology see Cotes, et al. [9]. The choice of deposition process is determined by the dielectric technology requirements, the ability to fill between patterned features, the allowable thermal budget of the process, local planarization characteristics, deposition rate (throughput), and equipment cost of ownership. Criteria such as electrical performance, reliability, stability, and mechanical integrity are key technology factors that must be considered in choosing a deposition method.

Figure 9.5 shows the profile of SiO_2 over patterned features using two different deposition techniques: (a) HDP and (b) SACVD [10]. The density

(a)

(b)
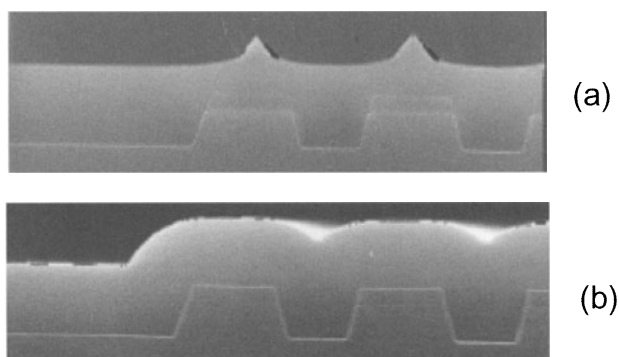
Fig. 9.5. Scanning electron microscope cross-sectional images of silicon dioxide material filled in a trench showing the differences between (**a**) high density plasma (HDP) and (**b**) sub-atmospheric chemical vapor deposition (SACVD) techniques. From [10]

of oxide asperities is much less for HDP compared to SACVD techniques. The HDP process fills SiO_2 into high aspect ratio gaps using a simultaneous deposition-etchback technique. This results in a triangular shaped asperity. High density areas tend to be locally planarized after deposition. In contrast, SACVD oxide conforms to the underlying feature. The asperity is rounded.

Deposition profiles must be taken into account when computing the pattern density. Normally the deposited film topography is correlated to the layout pattern density by applying the appropriate bias factor. If the deposition profile is conformal to the underlying features, then a positive bias is applied to the feature when computing the topography (asperity) density. HDP films, however, translate to a negative bias since the deposition technique produces an inward pyramidal shaped surface structure above each feature as shown in Fig. 9.5.

Dishing and Erosion

Dishing and erosion occur in damascene processes during the clearing and over-polishing stage. They are a result of polishing dissimilar materials that vary in pattern density. Shallow trench isolation processing is also considered a damascene process exhibiting dishing and erosion [11]. A typical damascene process sequence for the formation of metal interconnects was described in Chap. 5. First, a pattern is etched into the dielectric layer followed by metal deposition. The deposited metal films consist of a barrier layer and a primary interconnect layer, usually copper, that fills the trench. The polish step initially removes the primary metal layer that exists across the entire wafer. Eventually the underlying barrier layer is exposed followed by the oxide layer. Dishing occurs due to the differences in polishing rates of the two exposed materials, which in this case is oxide vs. metal. The oxide removal rate is generally much lower than the metal removal rate, which leads to dishing. Erosion is more dependent on the pattern layout. Pattern dependencies in copper damascene polishing processes are a significant issue.

Figure 9.6 is a cross-sectional diagram of a metal damascene structure (post-CMP) where dishing and erosion are defined. Dishing refers to the amount of material recessed in a local metal feature such as a trench, plug,
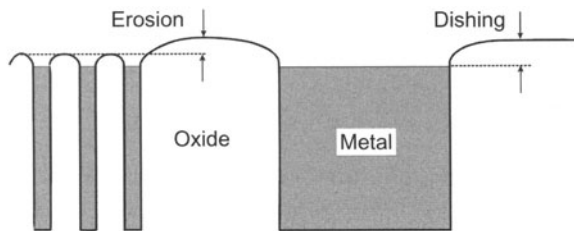
Fig. 9.6. Schematic of a damascene process. Dishing and erosion terms are shown

via, or interconnect. Dishing depends on the removal rate selectivity of the two films in question. Thinning is another term used in metal CMP similar to dishing. Erosion describes the thickness reduction of the oxide material (between the metal features) with respect to the oxide material at the edge of a group or array of features.

9.2.4 Characterizing Global Planarization

As described earlier, CMP processes generally fall into the near global planarization range. Step height measurements are difficult to measure globally (across the die). Other figures of merit can be used such as total thickness variation (TTV) and planarization length (PL).

Total Thickness Variation

Total thickness variation (TTV) within a die is a measure of the oxide thickness variation across a range of patterns. Figure 9.7 is a diagram depicting the metric TTV. Oxide thickness measurements above patterned features post-CMP are compared to the initial values. The TTV metric can quickly assess a particular CMP process or group of processes relative to a baseline process. However, the values are specific to a particular layout and do not necessarily translate to other die layouts.
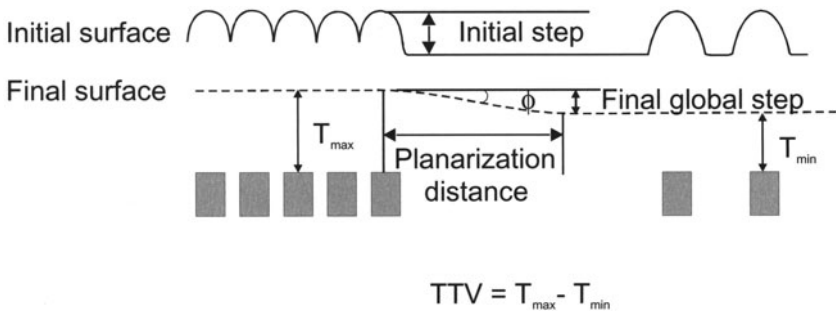
Fig. 9.7. Schematic cross-section of the initial and final surface over patterns of varying density. Total thickness variation (TTV) and planarization distance are shown

Planarization Length

The planarization relaxation distance is a metric used to compare the "characteristic planarization length" of a given CMP process. It is a parameter that is independent of the layout topography, but requires post processing

and modeling of measurement data in order to extract. Sivararm et al. [12, 13] originally defined the relaxation distance as the distance traveled over a step whereupon the original step height returns. Equivalently, an angle can also be used as the figure of merit. The definition was further modified to include the distance over which the height difference between the lowest and highest features becomes a non-changing fraction of the original step height [14]. Conceptually, the planarization length (PL) in CMP is the distance at which the CMP process variables (including the pad) no longer interact with step height and do not preferentially remove material from raised regions; instead, the entire surface continues to polish in unison. This is illustrated in Fig. 9.7.

Determination of the planarization length for a given CMP process can be accomplished by either direct measurement or indirect modeling of the post-CMP data. Although the direct measurement approach is by far the quickest, it is often impractical since the required test structures needed to characterize planarization length are large (mm distances). Instead, indirect modeling approaches are often used that correlate measurement data to a planarization model [15].

The procedures in determining planarization length are described in more detail by Stine and are reviewed in the next section [16]. Most oxide CMP processes have a planarization length on the order of 3 to 5 mm.

9.2.5 Planarization Test Masks

Background

Test patterns developed for characterizing planarization and published in the literature have varied considerably in layout dimensions. Most masks have been designed for testing specific effects rather than for generic use as CMP process characterization tools. Despite this, many of the main effects related to wafer polishing with patterns have been observed using relatively simple test patterns.

Renteln demonstrated that varying planarization rates and efficiencies could be characterized by using square wave test structures [17, 18]. The structures consisted of parallel trenches etched in silicon and coated with a deposited dielectric film. Patterns included regions of 1, 3, and 5 mm wide repeating structures. Each region was approximately square, several cm on a side. Profilometer scans and oxide thickness measurements were performed on wafers receiving different polishing times. From this simple set of test structures, Renteln showed that material removal rate was dependent on feature width. Small 1 mm wide features exhibited a faster planarization rate than larger 5 mm wide structures.

Siviram et al. developed a polishing pad deflection model based on a minimal test mask [12]. The test patterns consisted of a series of lines 5 mm in one dimension and varying width and space in the other dimension. The widest spacing was 2.5 mm and the range of feature widths was 1.0 μm to

40 μm. Based on results from this mask, Siviram reported that the CMP pad can be modeled with reasonable accuracy as a simple beam supported by the underlying features being polished. The pad planarization distance, which is the distance where polishing begins to occur in the down areas, varies as a function of remaining step height.

Pattern Factor Relationship to Polish Rate

Burke fabricated test patterns varying in size from 0.5 to 9000 μm and used these to characterize the feature dependencies on CMP removal rate [19]. The structures were anisotropically etched into LPCVD silicon dioxide to a depth of 0.6 μm. Polishing occurred for one minute at 5 psi with several different pads. Measurements were performed using standard optical film thickness tools. The polishing data derived from the characterization mask revealed that the oxide material in up areas polished faster than the down areas. The polish rate of large up features was equivalent to the polish rate of an un-patterned oxide layer. There was a dramatic effect on step height reduction versus the pattern density factor. The up area polish rate (PR_{up}) was experimentally determined to be related to the pattern density factor as

$$PR_{\mathrm{up}} = \frac{BPR}{PF},\tag{9.4}$$

where BPR is the blanket polish rate and PF is the pattern density factor. From this data, Burke developed a closed form empirical solution for step height reduction as a function of the down area removal rate and the pattern density factor. He showed that the step height reduction function is an exponential relationship in time and also developed a model that successfully predicted corner rounding of up features as well as down areas.

Warnock also used a test mask to develop a model for pattern dependent oxide polishing [20]. The structures ranged in dimension from 5 to 500 μm in linewidth and 5 to 40 μm in spacing and were etched into a thick layer of oxide deposited on a silicon wafer. The data generated from this mask was compared to a model that described a linear relationship of polish rate to local pressure within the array of features.

Hayashide et al. utilized two sets of test structures to characterize pattern sensitivity that was the basis of a model developed to analyze chip layout effects from CMP [21]. One test mask was used to determine the effects of pattern size and edge rounding. This mask included square and rectangular features varying from 0.2 mm to 5 mm in length. Another test mask consisted of square structures 2 mm in length formed on a 50 μm pitch that varied in pattern density from 10% to 90%. The test structures were prepared by etching the patterns into a 1 μm thick oxide film deposited onto a silicon wafer followed by an additional deposition of 1.5 μm thick silicon dioxide. A double stack polishing pad consisting of 0.8 mm thick polyurethane (top) and

1.2 mm thick coated felt (bottom) was used with silica slurry for polishing. The experimental results showed that removal was inversely proportional to pattern density and that stress concentration at a corner of a step caused rounding of the feature. The data also show that the bending of the polishing pad of adjacent "up" features is negligible if the distance between the features is less than 0.1 mm.

Grillaert et al. also characterized planarization using a set of test structures that included different pattern densities and pitch dimensions [22]. The density values ranged from 0 to 75% and the pitch dimensions ranged from 0.8 µm up to 1600 µm. Each test structure had an area greater than 2 mm². Using a polishing pad similar to Hayashide's experiments and a standard silica-based slurry, they found that the local removal rate is influenced by the pattern density and not the pitch dimension.

A systematic analysis of pattern dependencies in oxide CMP processes was performed by Stine et al. [16]. In this work, four characterization test masks were designed for rapid assessment of CMP planarization processes. As pictured in Fig. 9.8, each mask (1.2 cm by 1.2 cm in total area) emphasized one of four layout factors – pitch, density, area, and geometric aspect ratio. The structures in each mask were designed to fit in a 2 mm by 2 mm window. The pitch mask structures varied from 2 µm to 1000 µm while maintaining a constant density of 50 percent. The density mask varied systematically from 4 percent to 100 percent in density of chrome layout and the pitch was kept constant at 250 µm. A total of 25 density structures were incorporated onto the mask. The area mask varied the size of square regions from 100
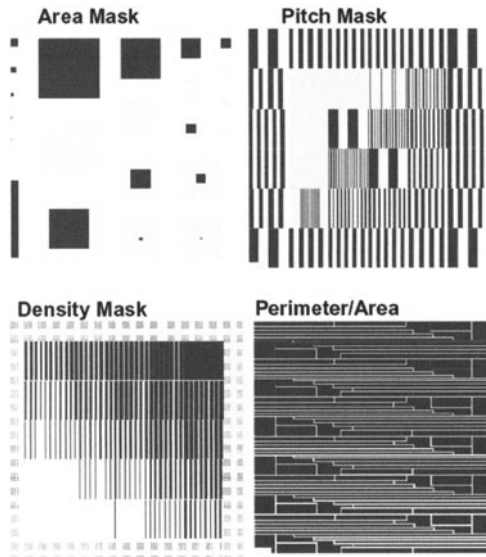
Fig. 9.8. Oxide CMP characterization mask set. From [16] (©2003 IEEE)

percent density to circuit-like density patterns across the die. Finally, the
aspect ratio mask explored equal area blocks with different perimeter areas.
Results from the polishing of patterned wafers using two different polishing
pads, an IC1000/SubaIV stack, and an experimental IC2000 pad, are shown
in Fig. 9.9.

The IC2000 pad has a harder top pad compared to the IC1000 pad; how-
ever, the underlying foam pad is similar in stiffness to the Suba IV. In this
plot, the clear linear dependence is seen between final oxide thickness in the
up areas and the effective layout density (averaged over a square region with
side length equal to the planarization length for the pad/process). In contrast,
all 50 percent regions in the pitch mask polish nearly the same, indicating
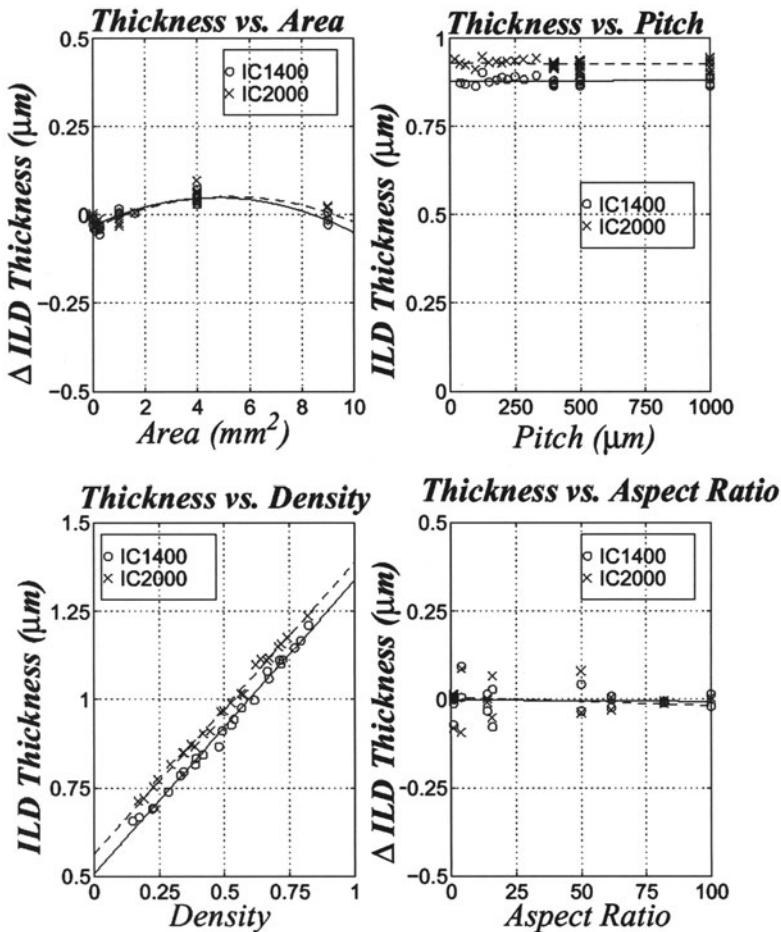
Fig. 9.9. Post-CMP oxide thickness vs. layout parameter for the four masks pic-
tured in Fig. 9.8. From [16] (©2003 IEEE)

again that density rather than pitch is the dominant factor for oxide polish-
ing. In the area and aspect ratio masks, the pattern density is first accounted
for (using the density mask data), and the remaining "delta" thickness as
a function of area or aspect ratio is plotted. Only minor contributions from
these two additional factors are seen. The up region polish rate dependence
on pattern density given in 9.4 was modified by Stine et al. to

$$PR_{\mathrm{up}} = \frac{BPR}{PF_{\mathrm{eff}}}, \tag{9.5}$$

where PF_{eff} is the effective pattern (density) factor determined by computing
a weighted average of the feature densities within a dimensional window. The
key difference between this result and the earlier study by Burke is the length
scale over which the pattern density factor is calculated. Based on this work,
Stine also developed a simple analytical model relating pattern density to
oxide CMP thickness [15, 16]. The simple relationship for total thickness
variation in the up areas (TTV_{up}) is given as

$$TTV_{\mathrm{up}} = SH_i * \Delta PF_{\mathrm{eff}}, \tag{9.6}$$

where ΔPF_{eff} is the total effective pattern density variation and (SH_i) is the
initial step height. This model assumes that no polishing occurs in the down
areas. Equation 9.6 can be used to estimate the tolerances of a particular
CMP process for a specific mask layout once the total effective density vari-
ation for the mask is determined. The dimension of the window size utilized
for density calculations is equal to the planarization length.

Comprehensive Dielectric Characterization Mask

Several refinements of Stine et al.'s mask and methodology were made culmi-
nating in a comprehensive mask for use in characterization of dielectric CMP
processes, including both oxide ILD and STI processes [23]. This single mask
pattern enables the study of both up and down area polish (where the down
area polish has been found to depend on the space between features), as well
as to understand the effect of deposition profiles.

The comprehensive dielectric mask is pictured in Fig. 9.10. The mask
dimension is 20 mm by 20 mm, consisting of a 5×5 array of 4 mm blocks.
The bottom two rows are gradual density regions where a constant pitch of
250 μm is filled with varying densities of line and space patterns. The middle
row is a "step density" structure, and the pattern density varies rapidly from
one block to the next in order to accentuate the transition between density
regions. The top two rows are constant 50 percent density regions, where
the line width and line spacing varies. Within this region (to the left side
of the second row), structures are added for profilometry or SEM extraction
of deposition profiles as a function of patterned feature size. In addition, the
fine feature blocks in all regions have a 20 μm square measurement site added
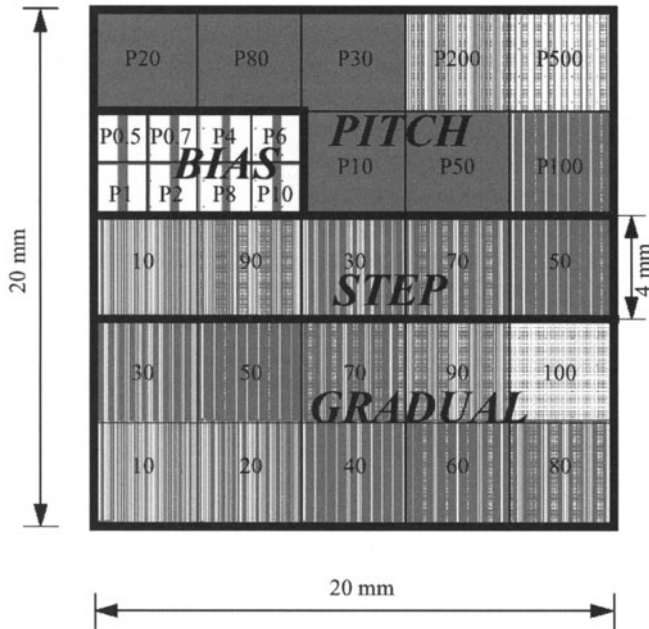
Fig. 9.10. Comprehensive dielectric CMP characterization test masks, integrating gradual and step density regions, pitch regions, and deposition characterization regions. From [23]

to the middle of the block, to enable optical film thickness measurements in these fine featured regions.

The structures included in this mask are effective in exploring the entire range of pattern densities and feature sizes, and give the widest possible characterization to a candidate CMP process. However, it should be noted that product wafers rarely cover the same density range (e.g. 4% to 100% pattern densities), and thus it is not expected that a given CMP process will perform well across the entire characterization mask. The test mask is also useful in determining the planarization length for a given polishing process. Variants of the test mask for STI characterization have also been reported that achieve pattern density using square features and include regions with realistic circuit patterns [24].

Copper CMP Test Masks

A number of test structures have also been presented for characterization of copper and other metal processes. Again, the entire layout of the chip is important, and reporting of performance on any single test structure (without information of surrounding area) can be misleading or difficult to interpret. A representative set of standard copper CMP test masks have been developed

by MIT in collaboration with SEMATECH and others, including the 931, 954, and 854 mask sets [25, 26, 27].

The first level of the MIT/SEMATECH 854 mask set is shown in Fig. 9.11. The primary test structure again consists of the arrays of lines and spaces, separated by unpatterned field regions.

Most of these structures can be electrically probed, in order to obtain electrical resistances of particular lines within the structure (implementing a spatial "sampling" near the edge of an array and across the width of the array). These same structures may also be measured using high resolution profilometry, or using emerging non-contact patterned copper optical mea-
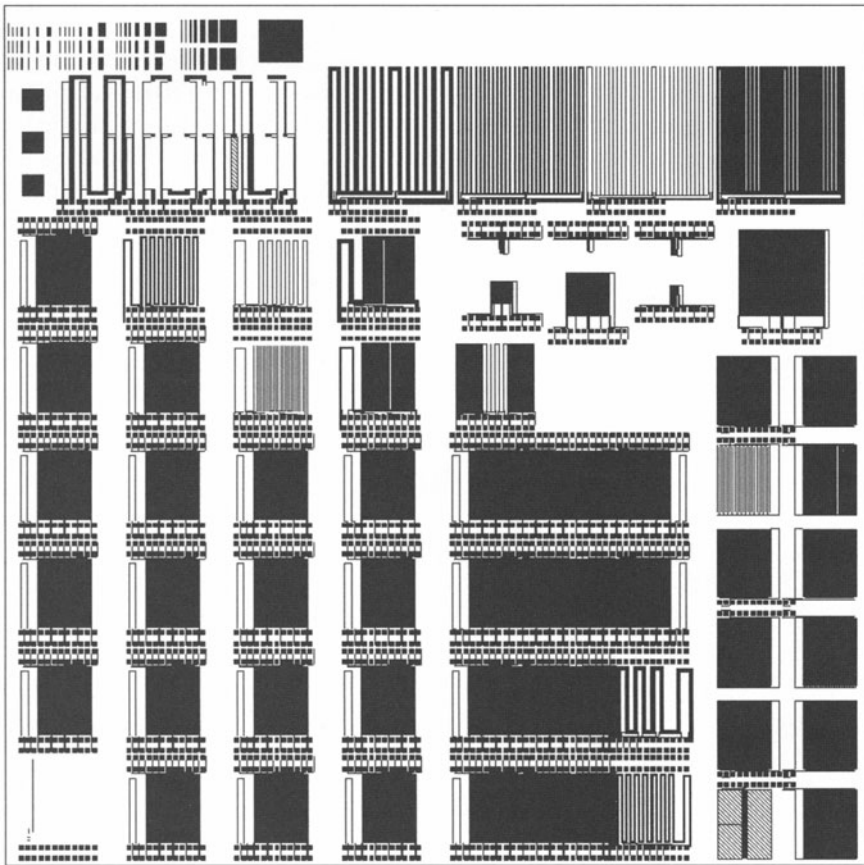
Fig. 9.11. Metal 1 mask in the MIT/SEMATECH 854 copper test pattern. The primary test structures consist of lines of arrays and spaces, with different pattern densities, line widths, and line spaces. Additional structures include via chains and area structures. From [27]

surement techniques [28]. Across the entire chip, these test structures again examine a very wide range of pattern density, line width, and line space parameters and their impact on dishing and erosion. Similar test masks have also been used to characterize pattern dependencies in copper electroplating processes [29].

Determining the Dielectric Planarization Length from Test Masks

Using a test mask that includes a large range of pattern densities, the planarization length can be extracted for a specific CMP process [16]. All of the local topography must be removed so that only global effects are present after polishing. The remaining oxide thickness is measured over each patterned feature location within the die. An example data set is shown in Fig 9.12. Notice that the 84% structure has the highest thickness value; the 100% structure is near to low density structures and thus has a lower effective pattern density. Next, a corresponding effective density value is computed at each site using a geometric shaped sampling window centered at the measurement site's coordinates. This is shown in Fig. 9.13. The initial value for the density window dimension (which varies iteratively in the model) is the as-drawn block dimension on the mask layout. After computing the effective density, a linear fit is performed on the thickness data versus effective pattern density. The
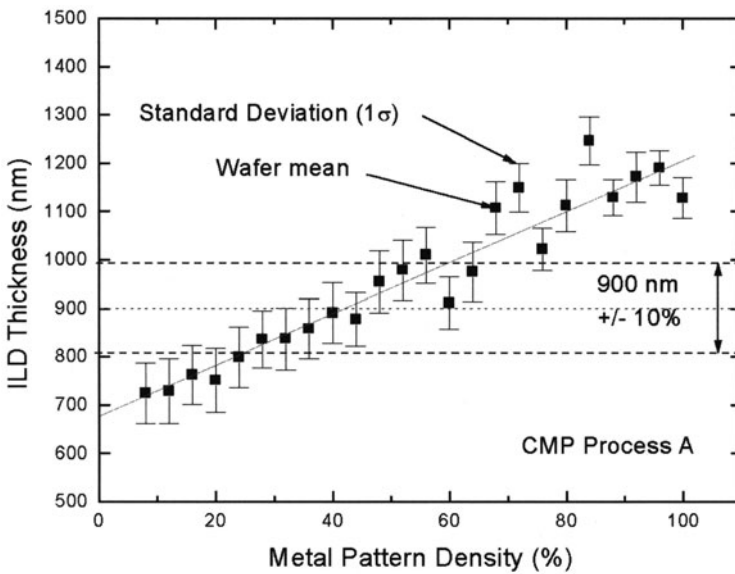
Fig. 9.12. Die-level oxide thickness versus metal pattern density using the mask shown in Fig. 9.8. The target thickness window of 900 nm +/− 10 nm is shown
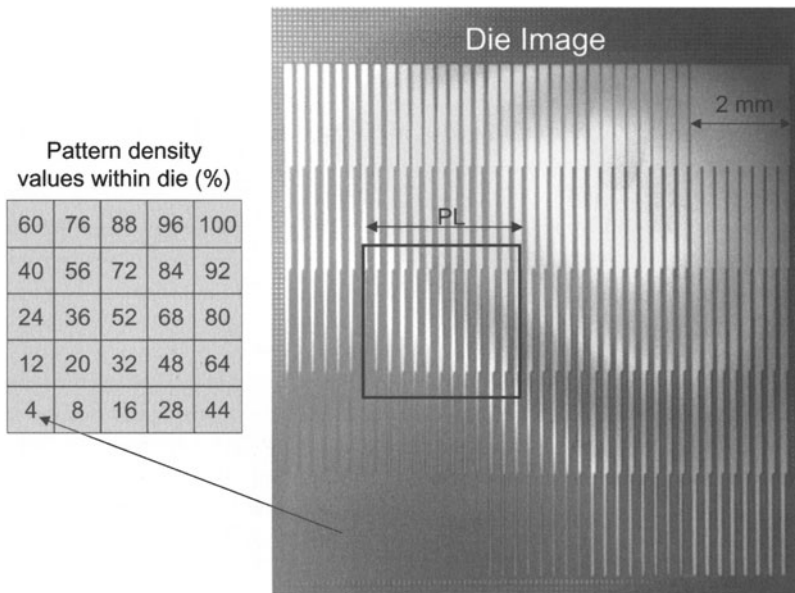
Fig. 9.13. Plan view micro-photograph of the die pattern density mask given in Fig. 9.8 following post-CMP. Pattern density values for each 2 mm structure are given at the left. The variation in contrast of the die image is due to the thickness variation of the silicon dioxide material over each pattern density feature. The measurement data is given in Fig. 9.12. Also shown is a drawn example of the effective density measurement window used in determining the planarization length. From [16] (©2003 IEEE)

window length dimension equals the planarization length when the slope of the linear fit equals the initial step height value.

9.3 Pattern Dependencies in Dielectric CMP

Dielectric CMP is utilized in both front-end (shallow trench isolation) and back-end (pre-metal and inter-metal dielectric) integrated circuit manufacturing processes. Planarization of patterned topography depends on many polishing parameters such as speed and pressure settings, pad and slurry types and the pattern layout itself. Regions of high pattern density planarize (polish) more slowly than regions of sparse topography.

We begin by briefly reviewing oxide CMP applications (Sect. 9.3.1) emphasizing the patterning effects that occur in both the back-end-of-line (BEOL) interconnect and the front-end shallow trench isolation (STI) modules. Section 9.3.2 presents the effect due to polishing pads; Sect. 9.3.3: process parameters such as pressure and velocity; Sect. 9.3.4: slurries. The final

Sect. 9.3.5 covers techniques used in manufacturing to alleviate intra-die non-uniformity from CMP.

9.3.1 Impact of Pattern Density in Oxide CMP Processes

Figure 9.14 shows a portion of the process flow for a conventional BEOL multilevel interconnect technology. A silicon dioxide material is deposited on top of a patterned metal interconnect layer. Deposition of silicon dioxide is accomplished using plasma-enhanced chemical vapor deposition (PECVD) in combination with sputter etchback. The etchback is performed after partial deposition of the oxide material to ensure adequate fill of dielectric material in the gaps located between the metal patterns. Recently high-density plasma (HDP) deposition has been used as a method to achieve good gap-fill between sub-micron interconnects. After oxide deposition, a CMP process is performed to planarize the oxide topography and remove dielectric material to a desired thickness. The BEOL oxide CMP process is a timed process [30].

A FEOL-STI process flow is shown in Fig. 9.15. A thin pad oxide is grown on the silicon wafer surface and is followed by deposition of a LPCVD silicon nitride layer. Next, the nitride/oxide stack is patterned and etched to remove
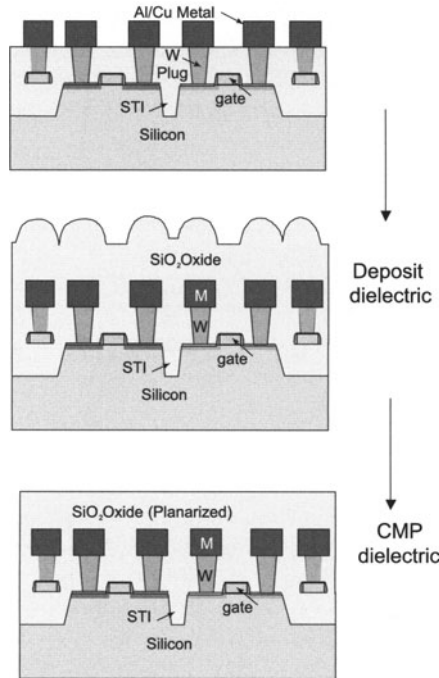
Fig. 9.14. Cross section schematic showing a portion of the process flow for a conventional back-end-of-line (BEOL) multilevel interconnect technology
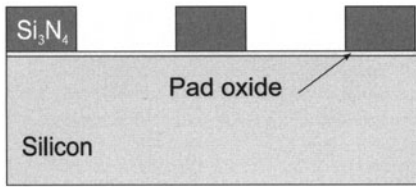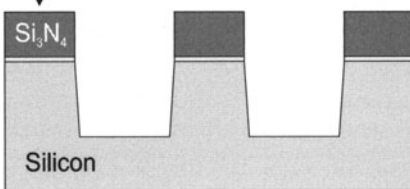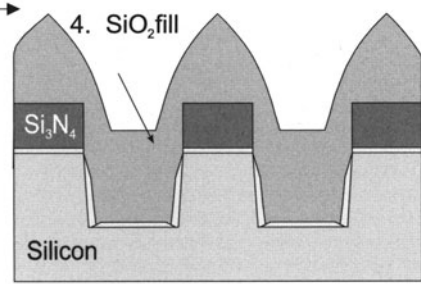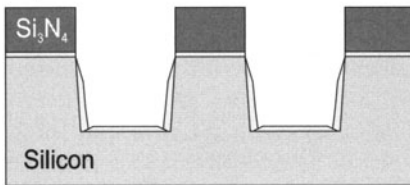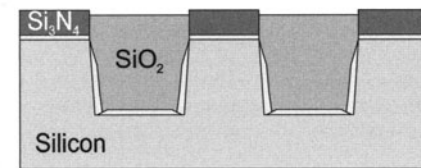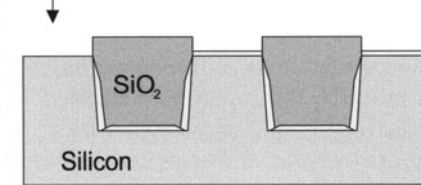
Fig. 9.15. Cross section schematic showing a portion of the process flow for a conventional front-end-of-line (FEOL) shallow trench isolation (STI) technology

the nitride material where the trench will be formed. The nitride layer remaining in the active area acts as a CMP stop layer during polishing. Once the nitride is opened, a trench is etched into the silicon substrate followed by a sidewall oxidation step in order to minimize reactive ion etching damage during the trench etch and assist in rounding of the top and bottom corners of the trench. After the oxidation step, a silicon dioxide insulator is deposited into the trench typically using a CVD or high density plasma (HDP) techniques. Next a CMP polishing step is performed to remove the silicon dioxide material and planarize the surface. The polishing process is complete when all of the silicon dioxide material has been removed above the silicon nitride layer. After CMP, the exposed nitride protective layer is etched away in a hot phosphoric acid mixture.

Pattern dependent oxide polishing can impact manufacturing yield loss, product performance and reliability. In the BEOL for example, thickness vari-

ations in the dielectric layer can cause over- or under-etching of contacts and vias. Mismatches in capacitance loading (due to dielectric thickness variations) can also impact circuit timing [31, 32]. As described earlier, pattern density is the most sensitive layout parameter to the CMP process compared to other layout parameters such as pitch dimension, etc. [16, 22]. Because the local pressure (and hence local removal rate) varies with pattern density, within-die oxide thickness uniformity can be equal to or greater than the wafer-level oxide thickness non-uniformity.

In the FEOL STI process, several conditions must be satisfied: (1) removal of all of the silicon dioxide material above the silicon nitride layer, (2) no exposure of the silicon material in active areas, and (3) minimal dishing of the trench oxide material located in the trench regions. Residual oxide that remains on top of the silicon nitride layer prevents the nitride film from etching away in the hot phosphoric acid step resulting in die loss. In addition, the degree to which the trench oxide material is over- or under-polished impacts transistor performance [33]. These requirements put a very tight constraint on the polish process window. The clearing and over-polish step is a critical part of the process. As expected, variations in pattern density make it difficult to maintain the above requirements.

Step Height Reduction vs. Pattern Density

Figure 9.16 shows a top-down microphotograph of a 256K SRAM integrated circuit metal interconnect pattern. Two locations that have different pattern densities are indicated – one near the middle of the dense SRAM area (high density region) and the other at the edge of the SRAM circuit (low density region).

Surface topography images and step height measurements were taken using an atomic force microscope (AFM) as a function of CMP planarization time [5]. The initial step height of the PECVD TEOS oxide topography was 0.8 μm. Figure 9.17 shows the AFM topography images for the two SRAM locations after 15 s, 45 s, and 90 s of polishing time. As shown in Fig. 9.17, the SRAM dense region located in the center of the circuit planarizes at a much slower rate compared to the edge of the SRAM. After 90 s of polishing time, the edge of the SRAM circuit is locally planarized whereas the center of the circuit has a remaining oxide step height greater than 2000 Å. Figure 9.18 quantifies these results showing a plot of step height versus polishing time for both locations of the SRAM. These results depict the main issue for CMP when applied to arbitrary die layouts. The planarization rate varies considerably within the circuit resulting in differences in final oxide thickness across the die.

Grillaert et al. also examined the step-height reduction versus time profile [34]. He showed that the exponential decay of step height versus time fit a model where the polishing pad is treated as a compressible material that deforms into the oxide features (at a transition step height). Initially the
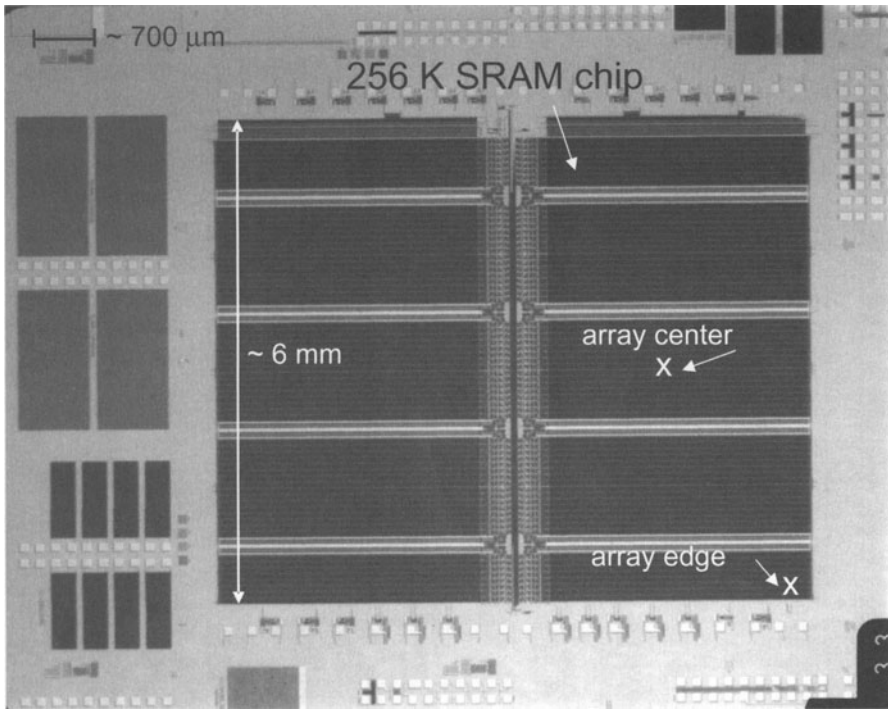
Fig. 9.16. Plan view micro-photograph of a 256K SRAM integrated circuit metal interconnect pattern. Two locations are shown – one near the middle of the dense SRAM area (array center) and the other at the edge of the SRAM circuit (array edge). From [5]

pad does not contact the down regions but only up the area where polishing/planarization occur. A transition occurs where the pad bends enough to contact the down region and the local pressure applied in the down area begins to remove material in the down region. The transition time is a function of initial step height and the mechanical properties of the polishing pad.

Polish Rate in Up and Down Areas

The polish rate on patterned topography also varies as a function of time and pattern density. In the up areas, the polish rate is initially higher than the blanket rate due to the enhanced local pressure applied to the features. As the step height reduces, the polish rate decreases until it approaches the blanket polish rate of the film. Likewise, the down area polish rate initially is very low since there is minimal contact of the pad and slurry against the material in the down areas. As the step height reduces, the contact area increases in the down areas and the polish rate increases.
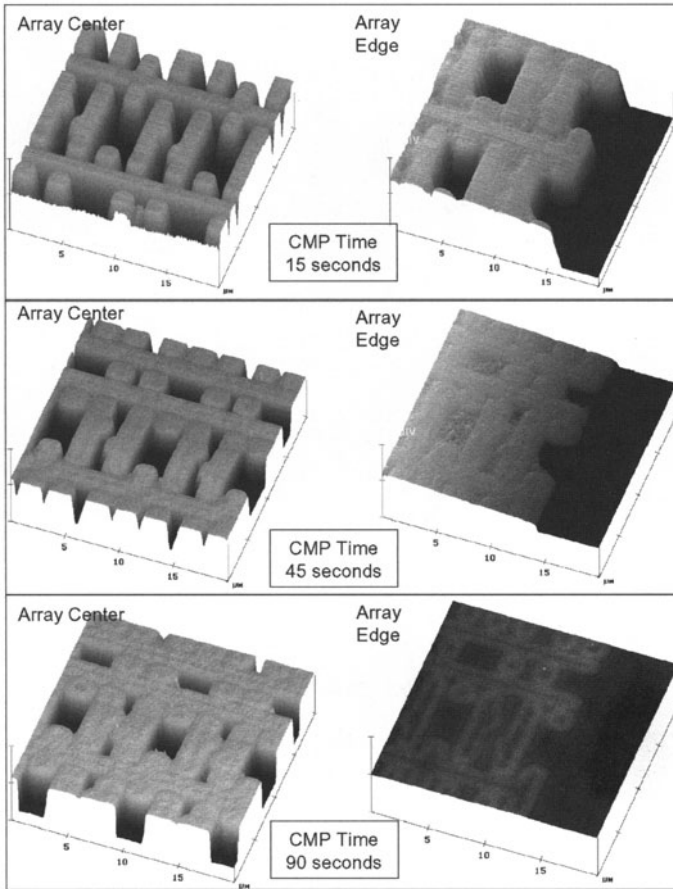
Fig. 9.17. Atomic force microscopy (AFM) images for the two SRAM pattern locations given in Fig. 9.15. The topography images show the time evolution of step height removal after 15 s, 45 s, and 90 s of polishing. From [5]

Figure 9.19 shows a plot of incremental polishing rate of up and down regions versus polishing time for different metal pattern layout densities corresponding to the mask of Fig. 9.13 [35]. The incremental polishing rate is defined as the incremental amount of SiO_2 removed during the polishing time interval divided by the time interval. The polishing was carried out using a conventional polyurethane stacked pad and a silica-based slurry. The carrier pressure setting was 9 psi and the rotation rate for both the carrier and platen was 30 rpm. A rotary CMP tool (IPEC model 472) was employed. The silicon dioxide material was deposited in a plasma-enhanced CVD reactor. The initial step height was 0.8 μm and the initial oxide deposition was 2 μm thick. The blanket oxide removal rate at this speed and pressure setting is approximately 3500 A/min.
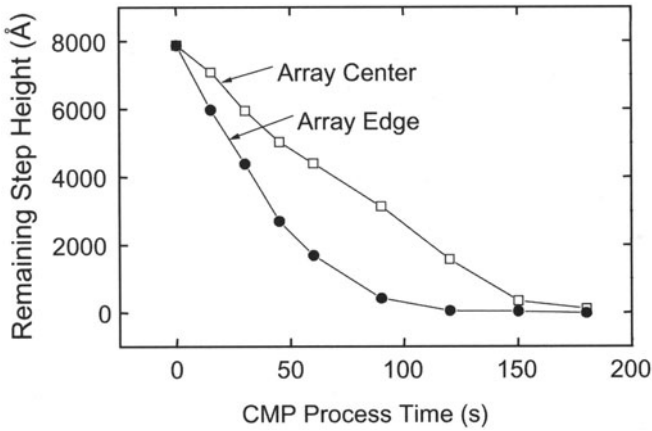
Fig. 9.18. Plot of remaining step height versus CMP time for the two SRAM pattern locations given in Fig. 9.15. From [5]
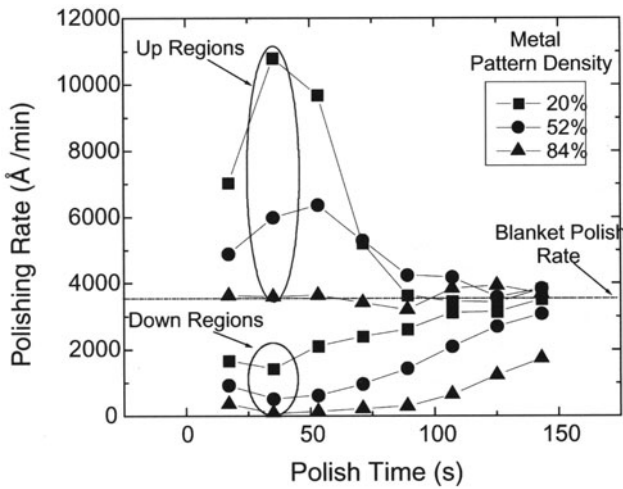
Fig. 9.19. Incremental polishing rate of up and down regions versus polishing time for different metal pattern layout densities. The incremental polishing rate is defined as the incremental amount of SiO_2 removed during the polishing time interval divided by the time interval. From [35]

As shown in Fig. 9.19 the incremental polish rate correlates with pattern density. The lowest density structure has the highest up area polish rate. The 84% structure has an up area removal rate very close to the blanket film polish rate. The effective density of this location (due to the surrounding density regions) is much closer to 100%.

All of the down areas have a finite removal rate indicating that local pressure (through the bending of the pad) is applied to these regions even during the initial stage of the polish cycle for these particular structure dimensions. The lowest density structure has the highest down area polish rate. The down area polish rates increase over time as the step height is removed until they eventually reach the blanket film polish rate.

Planarization (near global) is obtained when the up and down area polish rates converge to the blanket oxide removal rate. For the 20% and 52% structures, the time to planarization is approximately 110 s and 150 s respectively. However, the 84% density structure is not fully planarized since the down area polish rate does not reach the blanket polish rate as shown in Fig. 9.19. More polishing time is required to planarize this particular structure.

Step Height Reduction vs. Pattern Density in STI Processes

Step height removal in STI processes behave similarly to continuous layer polishing processes until the polish stop layer is uncovered. Figure 9.20 is a plot of step height versus time for various pattern densities [36]. In this particular study, the active region consists of a 1500 Å thick layer of silicon nitride on top of 100 Å of thermally grown oxide. The trench was etched to a depth of 5000 Å. A 0.85 μm thick silicon dioxide layer, deposited in a thermal CVD reactor, covers the active and trench regions. Conventional fumed silica-based slurry and a polyurethane stacked polishing pad were used. Polishing was performed in a rotary polisher (IPEC 472) with the platen and carrier rotation rates set at 30 rpm (matched). The carrier pressure was 7 psi.
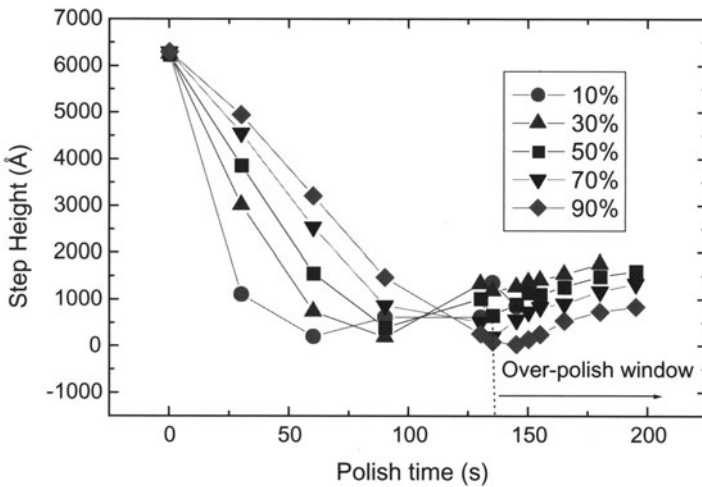
Fig. 9.20. Step height versus time for various STI pattern densities. The over-polish window is defined as the time following the exposure of the silicon nitride material located in the active region. From [36]

The plot shown in Fig. 9.20 reveals that the reduction of the step follows a similar trend to BEOL oxide planarization curve shown in Fig. 9.18 as a function of pattern density. Step height reduction rates for dense regions are slower than for sparse (low pattern density) regions. Once the step is removed in an STI process, however, the over-polishing phase will eventually cause an increase in the step height independent of pattern density [37]. This is due in part to the differences in removal rate of silicon dioxide to silicon nitride which is approximately 4:1 using conventional silica-based slurries.

Erosion and Dishing in STI Processes

Dishing and erosion in STI processes can occur in both the trench oxide and the active nitride regions. Figure 9.21 shows a plot of trench oxide dishing (or oxide loss) versus polishing time for various trench pattern densities. The wafer preparation and polishing conditions were described earlier in Sect. 9.3.1. Trench oxide loss occurs when the thickness value falls below the target value which, for this example, is 6500 Å. If the maximum acceptable value for dishing in an STI process is 500 Å, then the allowable variation in pattern density is approximately 20% as shown in Fig. 9.21.

Figure 9.22 is a plot of the silicon nitride erosion versus polish time for the same set of trench patterns described above. The initial thickness of the silicon nitride layer is 1500 Å. The polish time at which silicon nitride erosion increases from zero occurs when the oxide material above the nitride has been removed. The silicon nitride layer is completely removed when the erosion
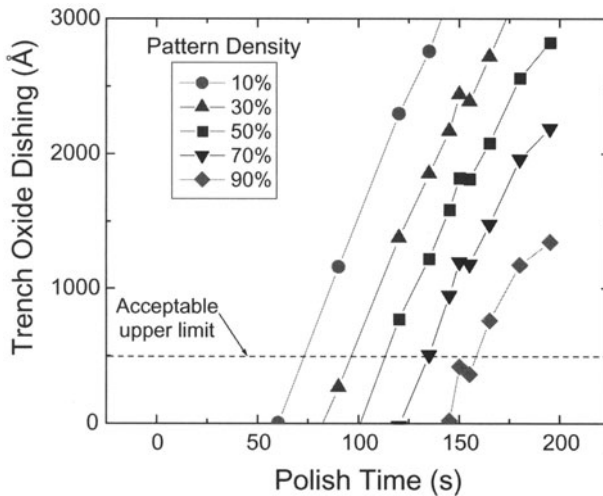
Fig. 9.21. Trench oxide dishing (or oxide loss) versus polishing time for various trench pattern densities. Also shown is an acceptable upper limit for dishing in a conventional STI process of 500 Å. From [36]
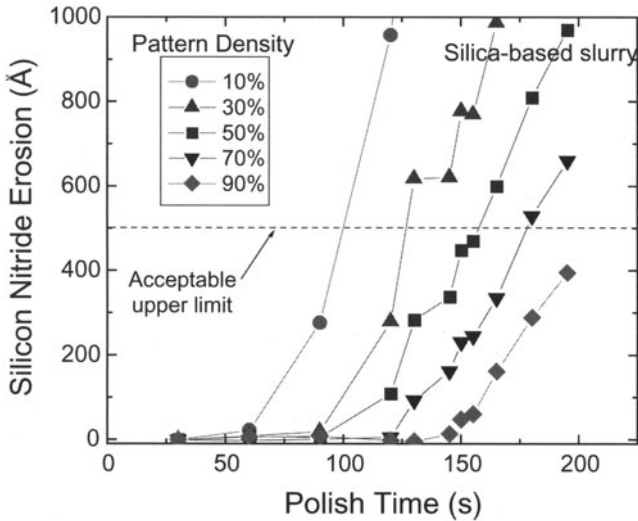
Fig. 9.22. Silicon nitride erosion versus polish time for various trench pattern densities. Also shown is an acceptable upper limit for loss of silicon nitride material that can occur during the over-polish step (500 Å). From [36]

value reaches 1500 Å. Data given in Fig. 9.22 show that the nitride layer for the 10% pattern density structure is completely removed in approximately 130 s while the highest pattern density (90%) begins to erode at a polish time of 150 s. It is clear from the data given in Figs. 9.21 and 9.22 that pattern density has a significant effect on the STI process window.

To summarize, the STI polishing process must ensure: (1) complete removal of residual silicon dioxide material over the active silicon nitride layer, (2) minimal loss of the silicon nitride layer after exposure to the process, and (3) minimal dishing in the deposited silicon dioxide material in the trench region. Acceptable values for nitride loss and trench oxide dishing in an STI process are in the 100 Å range. Variation in pattern density is a crucial aspect of the STI module due to the tight process constraints.

Polish Time Estimate for Patterned Topography

The local planarization polish rate of patterned dielectric topography assumes an inverse relationship of pattern density factor and the blanket polish rate of the oxide material such as the relationship given in 9.4. Calculating the polish time (PT) for patterned oxide topography to reach a specific target thickness is given by:

$$PT = PT_{LP} + PT_{OP}. \tag{9.7}$$

PT_{LP} is the polish time estimate to achieve local planarization, and PT_{OP} is the over-polish time estimate to reach the desired thickness once the step

height has been removed. When considering patterns of varying density, the average value of the polish time (PT_m) that achieves the desired thickness over a pattern is

$$PT_\mu = \frac{[SH_i\,(PF_{\text{eff}_u} - 1) + T_i - T_f]}{BPR}. \tag{9.8}$$

BPR is the blanket polish rate of the deposited film, SH_i is the initial topography step height, PF_{eff_u} is the mean value of the effective pattern density factor, T_i is the deposited film thickness, and T_f is the target post CMP oxide thickness. The total thickness variation (post-CMP) can be estimated using 9.6.

As an example, assume a reticle layout was analyzed for pattern density using a window size of 4 mm. The mean pattern density factor (PF_{eff_u}) was determined to be 0.65 and the maximum change in effective pattern density (ΔPF_{eff}) was 0.15. The blanket polish rate of the silicon dioxide material is 2500 Å/min; the initial step height is 0.5 μm; the initial deposition thickness is 2.0 μm, and the final target thickness is 1.0 μm. Using 9.6 and 9.8 the mean polish time PT_μ is 3.3 minutes to reach the target thickness over the 65% pattern density lines and the intra-die thickness variation for this layout is approximately 1500 Å.

9.3.2 Pattern Density Effects from Polishing Pads

The polishing pad invariably has a strong influence on the planarization performance. The planarization process is inherently dependent on the length over which the polishing pad remains rigid. For conventional silica-based slurries, the mechanical modulus properties of the pad show good correlation to the planarization length [12, 18, 19, 38, 39].

The most widely used pad for dielectric CMP is a composite (or stacked) polishing pad consisting of a hard polyurethane material placed over resilient soft foam as described in Chap. 8. The combination of hard/soft pad stack is intended to provide adequate wafer level polish uniformity while still producing a relatively high degree of die-level planarity [40]. One example of a composite pad is the IC1400 manufactured by Rodel, Inc. This pad consists of a 50 mil thick closed cell polyurethane material top layer and a 50 mil thick soft polyurethane foam sub-layer as shown in Fig. 9.23. Examples of mechanical behaviour of these pads are illustrated in Fig. 9.24.

Planarization tests performed using the CMP characterization test mask [16] described earlier in the chapter include patterns ranging in density from 4% to 100% [41]. Wafers coated with a 0.8 μm thick layer of sputtered Al/Cu metal were patterned and etched. Next, a 2 μm thick film of PECVD TEOS oxide was deposited over the metal features. Polishing was performed on an IPEC 472 rotary polisher using standard silica-based slurry (Cabot
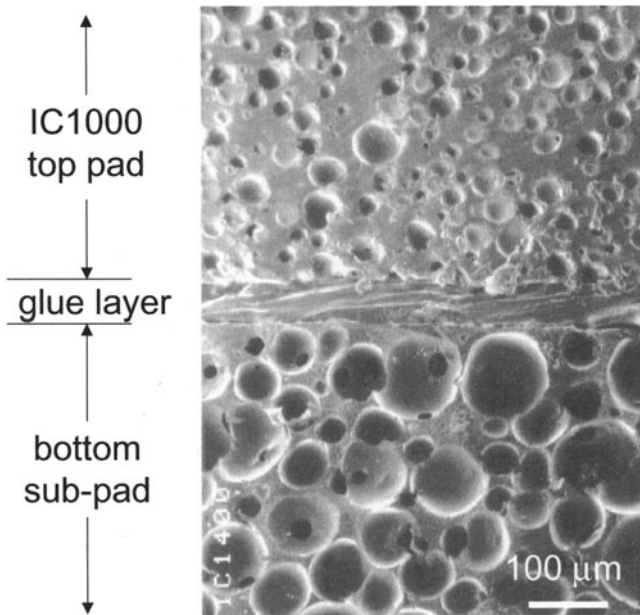
Fig. 9.23. Scanning electron microscope cross-sectional image of a polyurethane stacked polishing pad. Details of this pad technology are described in Chap. 8

SS12) for each of the composite pad types. Pressure and speed were also factors considered. The pressure settings were 3 psi and 9 psi. The polish rotation settings were 30 rpm for both the platen and carrier.

Figure 9.25 shows the average total thickness variation (TTV) post-CMP within the die for each processing condition. The average is based on a measurement of four sites across the wafer for each density structure. The stiff composite pad (fiberglass sub-layer) produces significantly better planarity compared to the foam-based sub-layer pad regardless of the polishing speed or pressure setting as shown in Fig. 9.25. Figure 9.26 shows the wafer-level post polish oxide uniformity results for each of the pad types. The oxide thickness uniformity is calculated from the ratio of the standard deviation (1σ) of the mean post polish thickness divided by the mean and expressed as a percentage. The foam sub-layer pad produces superior wafer level post oxide thickness uniformity compared to the hard pad consisting of a fiberglass sub-layer.

These results indicate that the sub-layer component of the polishing pad plays a crucial role in determining the wafer-level and die-level response. There is a clear tradeoff between wafer-level and die-level uniformity optimization and the type of polishing pad employed. The absence of the sub-layer pad creates a much stiffer composite pad and produces a better planarization response compared to a composite pad with a soft foam sub-layer. The
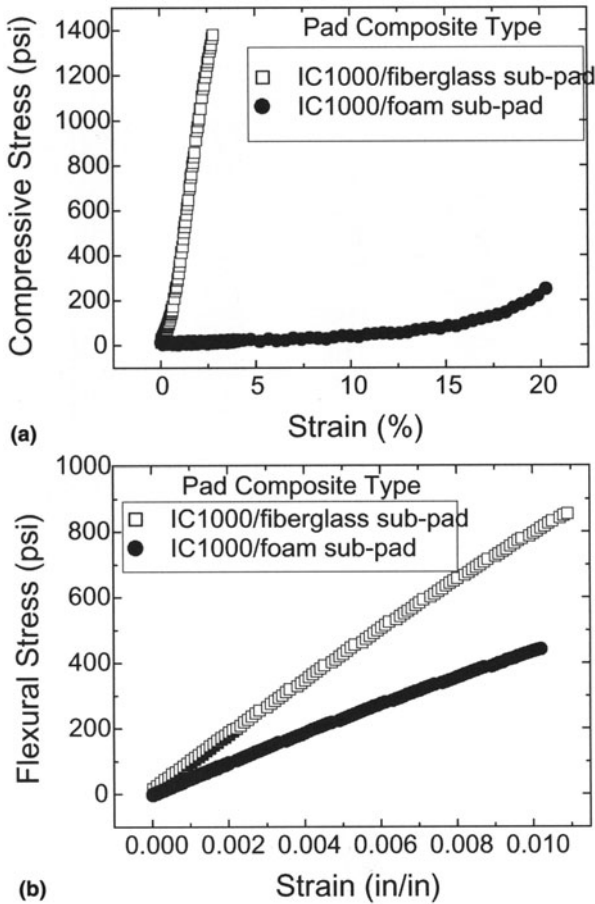
Fig. 9.24. (a) Compressive stress versus strain for two types of pad composites which vary only in the underlying layer material. The first type includes a 50 mil rigid fiber glass sub-layer material and the second type incorporates a soft foam sub-layer. The top layer on both pads is a 50 mil thick polyurethane material (IC1000). From [35]. (b) Flexural stress versus strain for the same pad types described in Fig. 9.23a. From [35]

soft foam sub-layer improves the wafer-level polishing thickness uniformity response; however, die-level planarity is compromised. A flexible pad such as the stacked pad with a soft foam sub-layer minimizes the mechanical alignment tolerance sensitivities and wafer-level non-uniformities associated with the polisher, wafer, and carrier design.
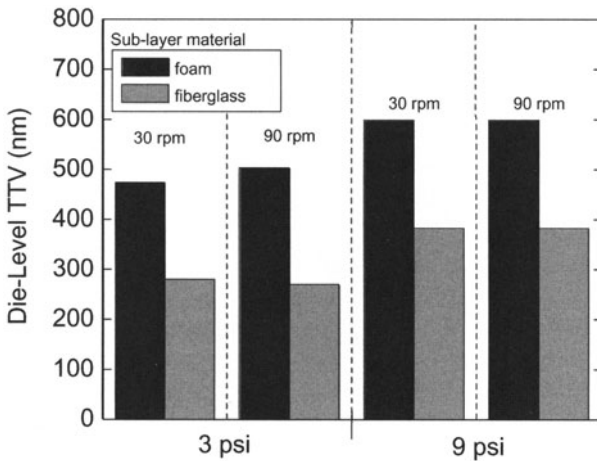
Fig. 9.25. Measured post CMP results for die-level total thickness variation (TTV) at different polishing pressure and rotation rates for the two types of pads described earlier in Fig. 9.23a. The density mask shown in Fig. 9.8 was used. From [35]
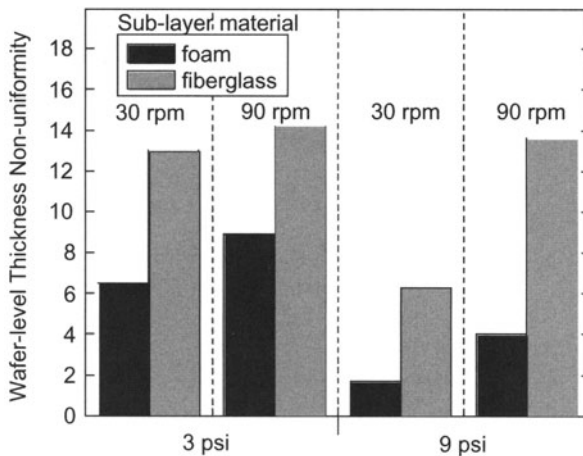
Fig. 9.26. Wafer-level post-CMP thickness non-uniformity at different polishing pressure and rotation rates for the two pads types described in Fig. 9.23a. The thickness non-uniformity is the standard deviation of the mean value expressed as a percent of the mean. From [35]

Effect of Pad Layer Thickness

Increasing the thickness of the top-layer hard polyurethane pad makes the overall pad stiffer and improves planarization performance. Figure 9.27 shows a plot of within-die non-uniformity (WIDNU) expressed as a standard deviation value versus polishing time for pads with different thicknesses [42]. In
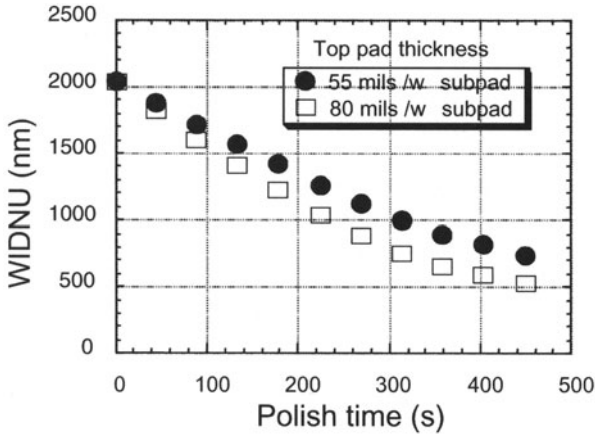
Fig. 9.27. Within-die non-uniformity versus polishing time for two different top pad thicknesses. The sub-pad layer is identical. The within-die values are the average values of the measured oxide thickness variation across each die. From [42]

this study, the WIDNU is defined as the level difference between the up area of the most dense structure (PD = 75%) and the down area of the most iso-lated structure (PD = 0%). The polishing pads (FX-9) were manufactured by Freudenburg Nonwovens Corporation. The top pad material is similar to the IC1000 pad described earlier and consists of polyurethane with a micro-filler. The sub-layer pad is a non-woven matrix. It was observed in this study that the WIDNU significantly improves by ~35% for the thicker top layer pad. However, the wafer-level non-uniformity (WIWNU) degrades as a function
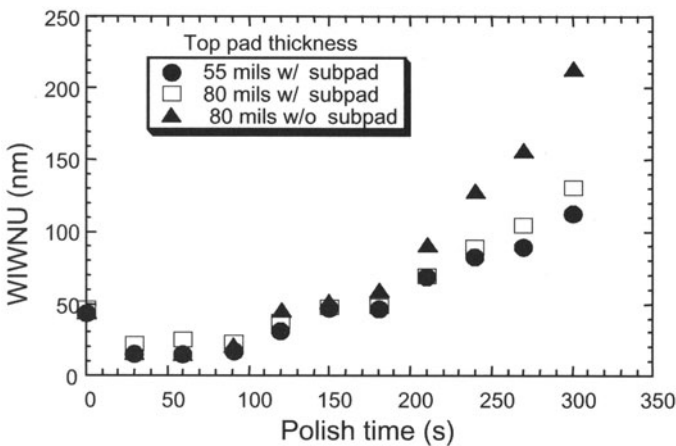
Fig. 9.28. Within-wafer thickness non-uniformity versus polish time for three different polishing pads that shows the effect of top layer thickness. From [42]

of patterned polishing time for the thicker pad as shown in Fig. 9.28. This is consistent with the earlier result given for the stiffer sub-layer pad material.

Pad Surface Conditioning

Surface pad conditioning also plays an important role in maintaining CMP process stability as discussed in Chap. 8. The conditioning process, which is a surface roughening technique, provides consistent transport of slurry through the surface layer asperities of the pad [43]. It was shown in earlier chapters that removal rate stability is a strong function of pad conditioning (Chap. 3); here we see that pad conditioning also affects planarization. Figure 9.29 shows a plot of within-die total thickness variation (WIDTTV) as a function of polishing time for conditioned and unconditioned polyurethane stacked pads [44]. The result shows that WIDTTV reduces by approximately 20% for the unconditioned pad. This is attributed to an increase in stiffness of the pad as a result of no conditioning [45], which is consistent with a measured decrease in pad surface bearing area ratio without conditioning [44].

In summary, there is a tradeoff in planarization length and wafer-level post-CMP oxide thickness uniformity. A stiffer pad increases the planarization length at the expense of increased oxide thickness variation across the wafer.
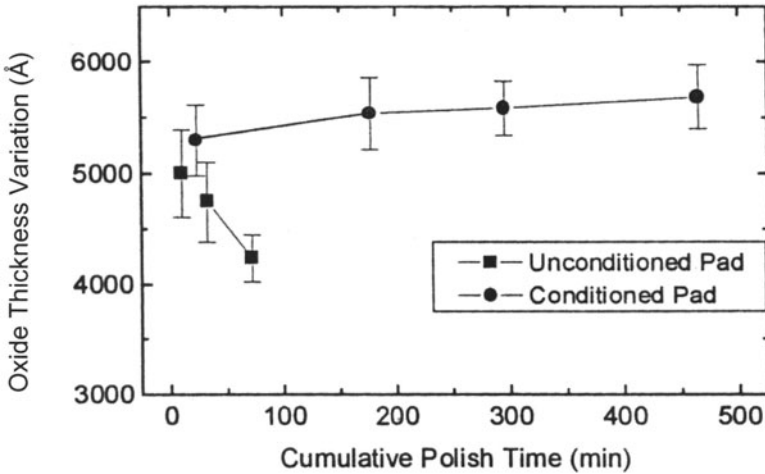
Fig. 9.29. Post CMP within-die oxide thickness variation (TTV) versus cumulative polish time for a pad that was conditioned concurrently during polishing and a pad that employed no surface conditioning. From [44]

Fixed Abrasive Pad

A fixed abrasive (FA) pad applied to semiconductor planarization was first described by Rutherford, et al. [46]. This pad differs significantly from standard polyurethane pads. The fixed abrasive pad is considered to be a "slurry less" pad since the abrasive particle is formed within the top polymer layer of the polishing pad and no additional particle-containing slurry is required to be dispensed during polishing. A dilute alkaline solution is normally required to lubricate the contact interface between the pad and wafer and allows the proper chemical reactions to occur during polishing. An example of a FA pad developed by 3M Corporation is shown in Fig. 9.30. The top abrasive layer is adhered to a series of backing layers. The top layer consists of a three dimensional structured abrasive material containing cerium oxide. It is well known that ceria behaves differently when planarizing topography compared to silica, even in slurry formulations [47, 48]. To date, there have been no published studies of fixed abrasive pads incorporating silica particles.

The shape and dimension of the surface cutting element has significant influence on polishing rate and stability [49]. Fixed abrasive pads are "tough" and require micro-topography on the wafer surface to activate the process and maintain the condition of the pad. Exposing fresh abrasive is critical to maintaining removal rate stability. The sub-pad layering stack also plays a critical role in determining the overall planarization capabilities similar to conventional polyurethane pads [50, 51].

Fixed abrasive pads can remove topography at a faster rate than conventional silica slurry and polyurethane pad systems. A three times increase in step height removal rate has been reported compared to conventional polyurethane pad and silica slurry due primarily to the ceria abrasive
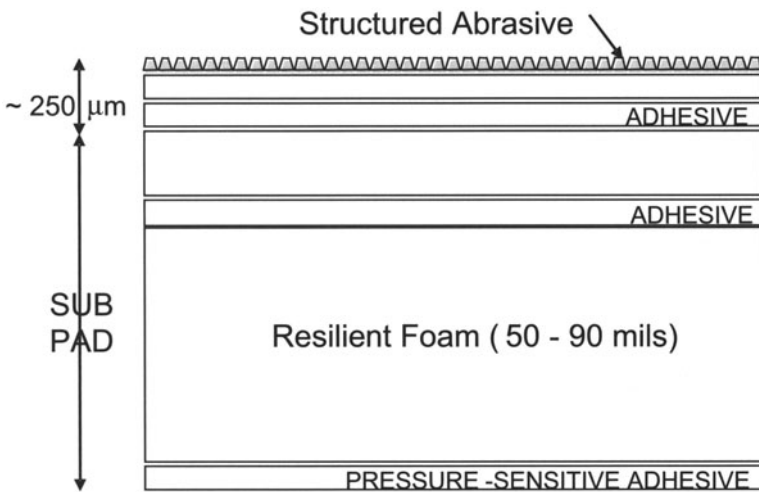
Fig. 9.30. Schematic cross-sectional diagram of a fixed abrasive pad. [Courtesy 3M]

employed in FA pads [51]. Cerium oxide also enables FA pads to exhibit a high selectivity to oxide topography. Fixed-abrasive pads are categorized as slow, medium, or fast depending on the removal rate of un-patterned blanket films [52]. The slow removal rate pad has a blanket oxide polish rate of $< 100\,\text{Å/min}$. The medium and high rate pads are designed to release more ceria during polish compared to the low rate pad. The blanket oxide polish rates for the medium and fast rate pads are $200\,\text{Å/min}$ and $2000\,\text{Å/min}$, respectively.

Fixed abrasive pads have been shown to significantly reduce within-die non-uniformity for oxide interconnect and STI CMP applications [49, 53, 54, 55, 56]. Schlueter et al. reported that oxide trench erosion for $100\,\mu\text{m}$ trenches filled with HDP silicon dioxide material [57] was lower when using a fixed-abrasive pad compared to silica or ceria-based slurries. Figure 9.31 shows a plot of trench oxide dishing versus pattern density - comparing the fixed abrasive pad to a conventional polyurethane pad and silica slurry process. As shown in Fig. 9.31, the trench oxide has minimal dishing when using a fixed abrasive pad compared to the standard pad and slurry. Figure 9.32 shows a plot of active area silicon nitride erosion as a function of pattern density. The fixed abrasive pad again erodes less than the standard pad/slurry combination.

For STI applications, careful tuning of the oxide deposition fill thickness value is required when using FA pads due to the high selectivity to oxide topography. The slow removal rate pad, for example, has a limited ability to
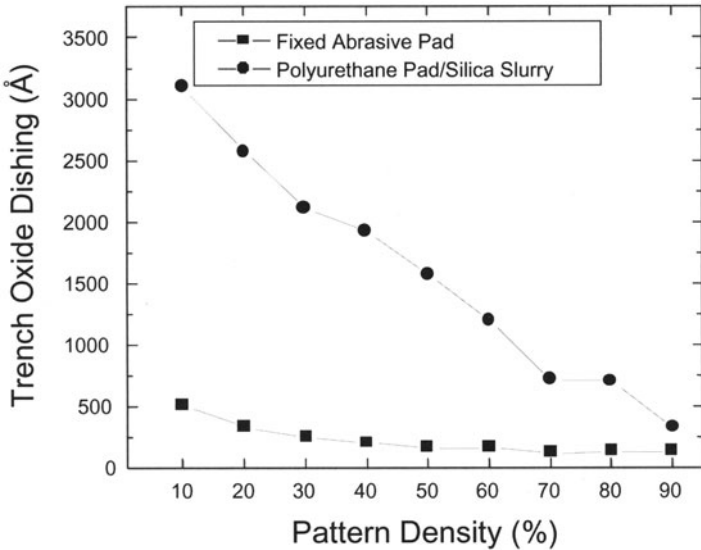
Fig. 9.31. Trench oxide dishing versus pattern density for a fixed abrasive pad and a conventional pad/silica slurry consumable set. [Courtesy 3M]
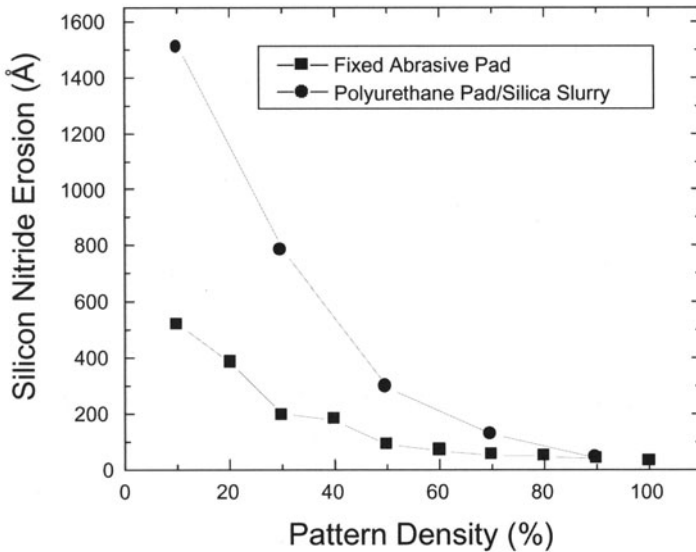
Fig. 9.32. Silicon nitride erosion versus pattern density for a fixed abrasive pad and a conventional pad/silica slurry consumable set. [Courtesy 3M]

remove oxide material over the active nitride layer once the structures are planarized. It is critical, therefore, to eliminate (or reduce significantly) the overfilling of the silicon dioxide into the trench. Values greater than 500 Å are impractical for manufacturing since the over-polish time is too long [52]. The slow rate pad also has no selectivity to silicon nitride, so its ability to self-stop is due solely to the reduction in surface topography [52].

9.3.3 Effect of Polishing Pressure and Velocity

In conventional polyurethane/foam stack pads and silica slurry systems, polish pressure has more of an influence on planarization length than does polish velocity [58, 59, 60]. Increasing the pressure will cause the pad to bend. The degree of bending varies as a function of pattern density.

Figure 9.33 shows the within-die total oxide thickness variation (WIDTTV) as a function of normalized polishing time for processes that used a standard polyurethane pad and silica-based slurry [59]. All wafers were polished on an IPEC 472 rotary polishing system. A blanket layer of PECVD TEOS oxide was deposited on the wafer followed by $0.8\,\mu m$ of sputtered Al/Cu metal. The wafers were patterned using the MIT pattern density mask described earlier [16]. After patterning and etching, the wafers were coated with a $2\,\mu m$ thick film of PECVD TEOS oxide. Standard polyurethane polishing pad (Rodel IC1400) and silica slurry (Cabot SS12) were used throughout.
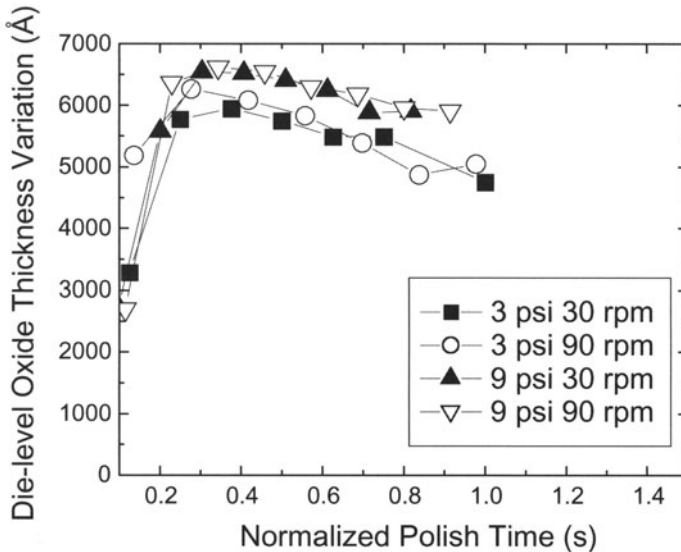
Fig. 9.33. Average value of die-level oxide thickness variation (TTV) versus normalized polish time for different combinations of polishing pressure and platen/carrier rotation rates. The TTV values were based on the density mask shown in Fig. 9.8. For a given process condition, each polish time is normalized to the time required to locally planarize of the 84% density structure and reach the target thickness. From [59]

All of the raw polish times for each process condition were normalized to the time required to achieve a post oxide thickness of $1.0\,\mu m$ over the 84% density structure. The TTV values are computed from the range of post CMP thickness values. Data show that die-level oxide thickness non-uniformity is more sensitive to polishing pressure compared to polish velocity for large pattern density variations. Polish velocity appears to have minimal impact on intra-die thickness variation. Higher applied polishing pressures produce larger variations in intra-die oxide thickness. The minimum WIDTTV condition was a low pressure, high velocity polish setting and the worst case WIDTTV was a high pressure, low velocity setting. This result is consistent with the earlier conclusion that the compressibility of the polishing pad significantly influences the planarization length.

9.3.4 Effect of Polishing Slurries

Silica-based Slurries

Silica abrasives are commonly used in oxide CMP slurries. They are normally dispersed in dilute solutions of KOH or NH_4OH chemistries to a pH of approximately 10–12 to ensure colloid stabilization. Silica particles are manufactured

primarily by two techniques: (a) liquid-based precipitation that forms spherically shaped particles or (b) high temperature oxidation. The latter forms a "fumed" silica particle. Figure 9.34 shows a transmission electron microscope (TEM) photograph for the two types of silica abrasives discussed (a) fumed silica manufactured in a high temperature furnace process (Courtesy Cabot) and (b) a spherical silica particle precipitated in a chemical solution (Courtesy Rodel).

Silicon dioxide removal rate, step height reduction rate (SHRR) and planarization length (PL) can potentially be affected by slurry parameters such as pH, weight percentage of abrasive (wt. % solids), and abrasive surface area. In an early study by Borba et al., increasing the pH from 9 to 11 in fumed silica slurry resulted in a 22% increase in removal rate but had no real impact on the planarization length [7]. When the wt. % solids increased from 12% to 18%, it resulted in a 13% increase in removal rate. However, the planarization length degraded when the solids content increased. This was ascribed to the fact that a greater number of slurry particles were available for abrasion in the lower regions which increased the down area polish rate. There was a strong interaction between wt. % solids and surface area. A summary of their findings is given in Table 9.1.

Evans et al. examined the effect of the cation (NH_4OH vs. KOH) in silica-based slurry on STI planarization performance [48]. The ammonia-based slurry was manufactured by Rodel, Inc. (ILD1300). The KOH stabilized silica slurry was also manufactured by Rodel, Inc (ILD1200). The particles were identical for both slurries. A standard polyurethane pad (IC1400) was used to polish PECVD silicon dioxide filled into trenches of various rectangular

(a)   (b)
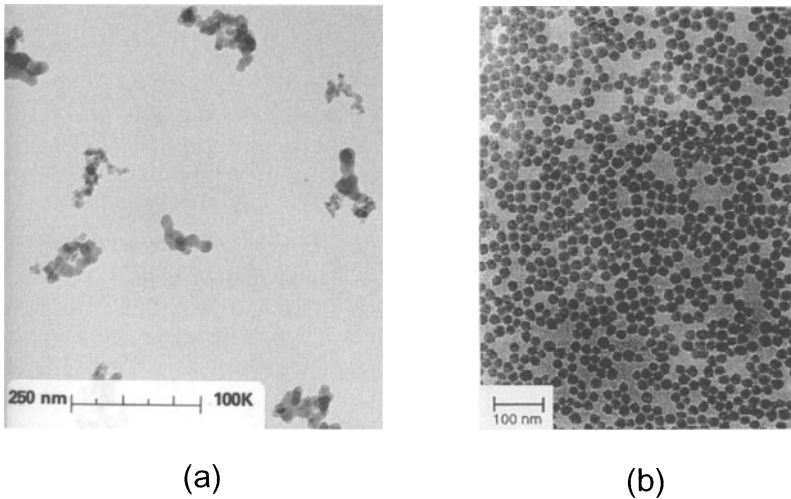
Fig. 9.34. Transmission electron microscope images of: (**a**) fumed silica slurry (courtesy Cabot), and (**b**) colloidal silica slurry (courtesy Rodel)

Table 9.1. The effect of fumed-silica slurry parameter properties on removal rate (RR), step-height reduction rate (SHRR) and P oxide. The definition of P oxide is the average amount removed in the up and down regions. From [7]

Response	Main effects		Interactions
	1	2	
RR	↑ pH (22%)	↑ wt. % solids (13%)	none
SHRR	↑ abrasive surface area (19%)	none	none
P oxide	↑ abrasive surface area (16%)	↓ wt. % solids (10%)	wt. % solids * abrasive surface area

shapes. They found that both slurries gave essentially identical dishing profiles. Figure 9.35 is a plot of trench oxide thickness versus pattern density for two different silica abrasives [36]. The fumed silica is KOH-based (Cabot SS-12) and the colloidal silica is NH$_4$OH-based (Klebesol 1501–50). Schlueter et al. also investigated the effect of fumed versus colloidal silica slurries on STI CMP processing and found slight differences in the trench oxide dish-
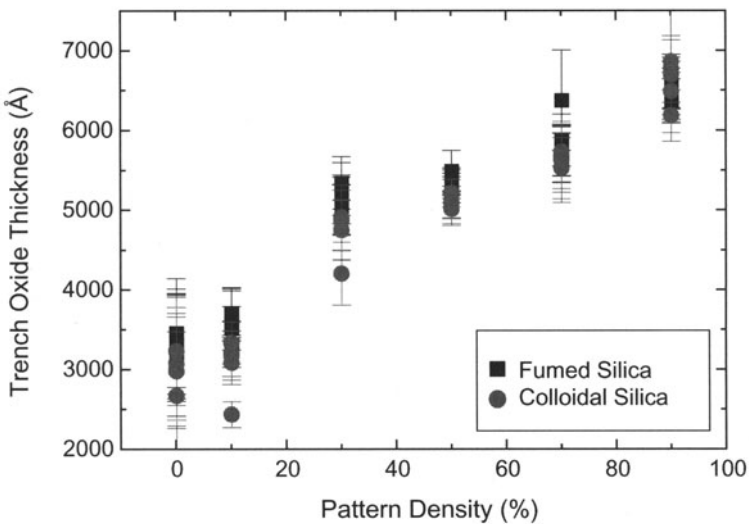
Fig. 9.35. Final trench oxide thickness versus pattern density for the two silica abrasive types shown in Fig. 9.34. From [36]. (Note: presentation material only – no paper published)

ing characteristics [57]. The fumed silica slurry showed slightly less dishing compared to the colloidal slurry.

Cerium Oxide

Cerium oxide slurries exhibit markedly different planarization characteristics (feature size, pattern density, etc.) compared to silica slurries [48, 61, 62, 63]. Nojo et al. demonstrated that ceria-based slurry formulated with a surfactant (2.5 weight %) produced self-stopping effects when the surface topography reduced [64]. The removal rate selectivity of oxide topography to a flat surface was 8:1. It was shown that this slurry would planarize areas as wide as 4 mm without dishing.

Additives to cerium oxide slurries can dramatically increase the selectivity ratio of oxide removal rate to nitride removal rate by a factor of 10 or more. High selectivity cerium oxide-based slurries (HSS) have been shown to improve planarization characteristics for STI applications compared to conventional silica-based slurries. [48, 65, 66]. An example is shown in Fig. 9.36 which is a plot of trench oxide dishing as a function of polishing time for a range of pattern density structures [67]. The standard silica slurry shows a fairly linear response as a function of polish time whereas the HSS ceria slurry exhibits non-linear behavior. The incremental trench oxide polish rate decreases over time for the ceria-based slurry as shown in Fig. 9.36. Figure 9.37 shows a plot of the nitride erosion versus polish time for the same
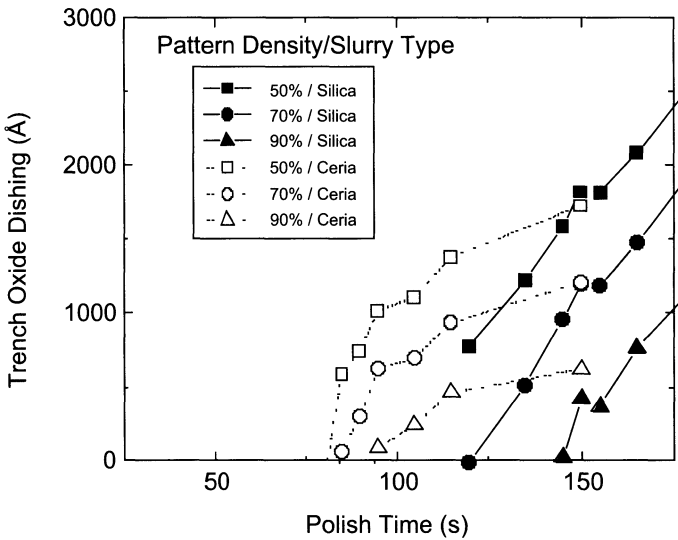
Fig. 9.36. Trench oxide dishing versus polishing time for silica and ceria abrasive types at different pattern densities. From [67]
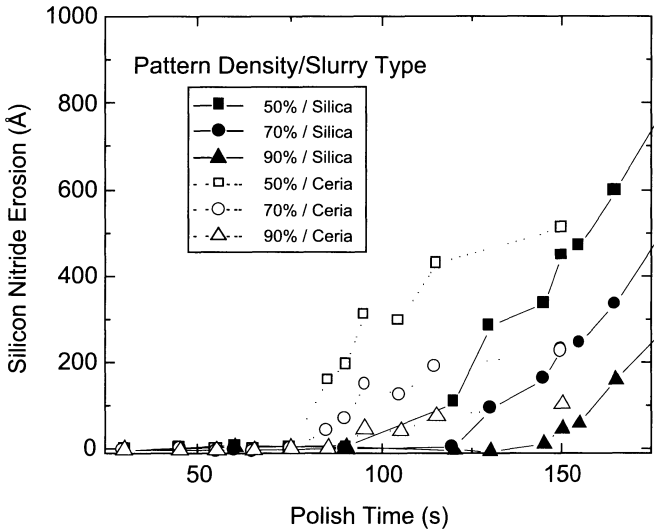
Fig. 9.37. Silicon nitride erosion versus polishing time for silica and ceria abrasive-based slurries. From [67]

two slurries. Again, the HSS ceria slurry exhibits more erosion limiting behavior over time compared to the standard silica-based slurry. Data suggest that that there is an improved planarization window when using the ceria-based slurry for the pattern density range (50%–90%) considered in this study.

9.3.5 Minimizing Pattern Density Effects

Layout Density Control Using Dummy Features

Variations encountered in CMP can be maintained within acceptable limits by controlling the layout pattern density. One method is to place dummy features into open spaces within the circuit layout. Several techniques have been reported [68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79]. Dummy features are normally rectangular or doughnut shaped polygons and are spaced accordingly in order to minimize capacitive coupling. They are constrained to specific areas based on physical and electrical design rules. Figure 9.38 is a microphotograph of an STI layout showing the placement of dummy structures between the circuit areas. As shown in Fig. 9.38, filling the sparse areas with dummy structures enables tighter pattern density control.

The optimum placement of dummy structures (also referred to as tiling) into an entire chip layout is an iterative procedure: (1) analyze pattern density, (2) choose dummy cell size, (3) insert cell into layout and (4) optimize for minimum density variation. Computer software can perform the complex analyses; however, the iterative nature of the density matching problem requires large memory space and long computing times.
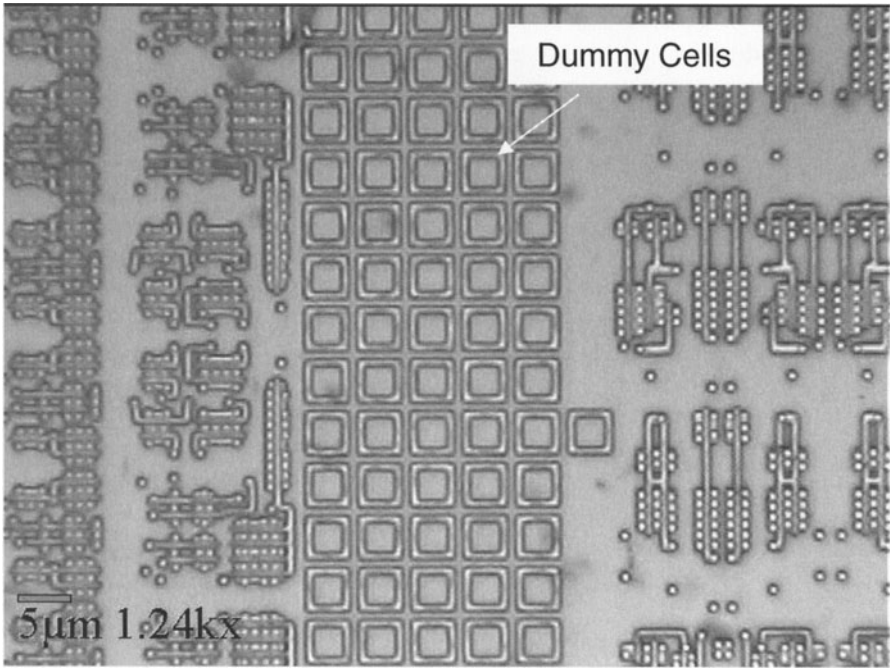
Fig. 9.38. Microscope image of a die pattern layout showing placement of dummy cells between active circuit areas
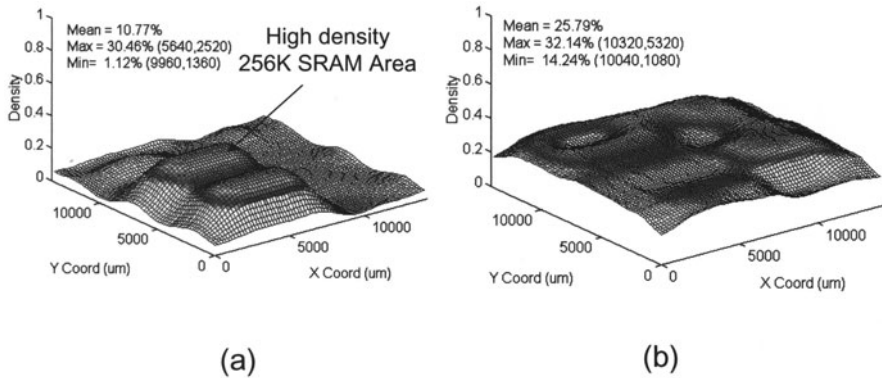
Fig. 9.39. Modeling results of density variation for a chip layout (**a**) before placement of dummy cells and (**b**) after placement of dummy cells. From [75]

Modeling pattern density effects allows for efficient and rapid analysis of die layouts [15, 80, 81, 82]. Die-level histograms and/or contour maps of pattern density variations can be generated for each level of CMP processing

(STI, Poly, Metal 1, Metal 2, etc.). Based on the results from the density analysis, a dummy cell is chosen with the appropriate size and pitch. Figure 9.39 is an example of a density contour map showing the original die layout (no dummy cells) and the final die layout after placement of dummy cells [75]. As shown in Fig. 9.39, the density variation range is considerably reduced from 29% to 18%.

Blockout Methods

The block technique is shown in Fig. 9.40. Following the dielectric film deposition, a pattern is printed in photoresist that masks off the trench areas (down regions). Next, a reactive ion etch process removes the majority of the oxide material in the raised areas. Finally a polish step is performed to planarize the remaining topography.

The advantage of this technique is that the polish time required for planarization significantly reduces since a considerable amount of material is already removed during the dry etch step. For large raised areas such as analog capacitors this method can be tailored to match the density of other parts of the chip [75]. The etch depth is a parameter that can be optimized for a given oxide deposition thickness. The disadvantages of this technique are that significant layout customizations are required in order to ensure success. Also, the final variation in pattern density is not necessarily reduced using the blockout method. In certain applications a blockout method combined with dummy fill is necessary in order to ensure good global planarization [83, 84].
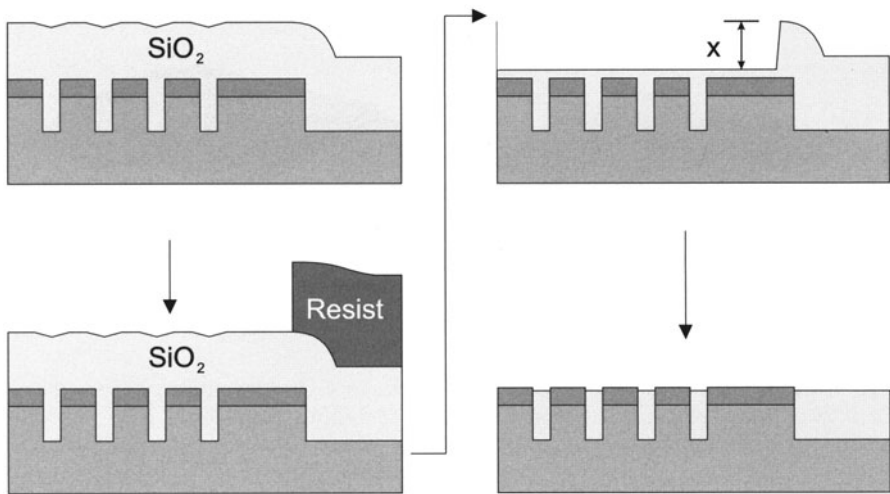
Fig. 9.40. Schematic process flow of the blockout technique to reduce pattern density

Other Integration Techniques

Several authors have studied the effects of incorporating a hard polish stop capping layer material on top of the primary dielectric film for improved planarity [85, 86, 87]. Figure 9.41 shows an approach for STI applications using silicon nitride as the capping layer. Following the trench oxide fill step, a thin layer of silicon nitride is deposited which acts as a polish stop material. This layer serves as an additional protection in the down regions during polishing. Depending on the layout, an additional mask layer is used to etch out the silicon nitride located in the up regions as shown in Fig. 9.40 [87]. This is particularly useful when addressing large structures such as analog capacitors or inductors [88].

The silicon nitride capping layer thickness is determined based on the trench fill thickness and the CMP polish rates of the oxide and nitride materials. Optimization is possible, although it does not work for all variations of pattern density [89].

Additional integration schemes include the incorporation of a polish stop layer within the dielectric film itself. Oliver et al. showed that planarization improved when a silicon-rich oxide film was inserted into the dielectric film stack at the desired final thickness level [90]. This integration scheme utilized a cerium oxide slurry formulation with highly selective polish rate properties
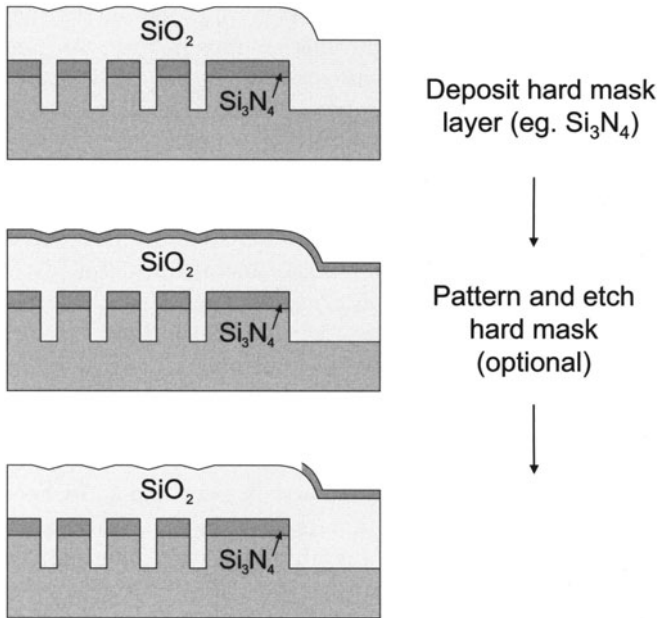
Fig. 9.41. Schematic process flow of the nitride hard mask approach for minimizing dishing effects due to pattern density variations. For improved results, the thin layer of nitride can be patterned and etched at the cost of an additional mask level

for standard SiO_2 compared to silicon-rich SiO_2. The polish rate selectivity was greater than 30:1.

9.4 Metal CMP Pattern Dependencies

Metal CMP processes, including tungsten and copper CMP, are also polish-back processes with many similarities to STI polish. In the case of tungsten CMP, one performs an etch of via or contact holes (or in some cases trenches for local interconnect lines), then deposits a metal stack usually consisting of thin Ti/TiN barrier layers followed by tungsten, resulting in excess metal both above the recessed regions and in "field" regions above the surrounding oxide. The metal stack must be polished back to leave a flat surface with tungsten in the desired recessed locations. The underlying oxide layer usually serves as the polish stop, having a much lower removal rate than the metal stack. Just as in STI polishing, the key pattern dependencies are dishing into individual tungsten via or line features due to the faster metal removal rates, and the erosion of oxide or dielectric spaces in array regions resulting from the inability of the polish stop to completely prevent polishing.

Copper CMP similarly enables the formation of inlaid lines or vias in either a single or dual damascene sequence, by polishing back excess copper over "field" or insulator regions to leave the deposited metal stack only where desired. In this sense, copper CMP is quite similar to the inlaid oxide trench structure in STI or inlaid vias or lines in tungsten, and many of the same pattern dependent concerns exist in copper CMP as in these processes. However, an additional complexity immediately presents itself in the case of copper CMP compared to STI CMP: rather than a dual material polish (oxide and nitride), copper CMP involves three materials often having dramatically different polish rates: copper, barrier metal, and dielectric. In addition, the implication of the dishing and erosion are different in the case of metal CMP, as the resulting thinning of metal lines and features induce not only topographical and yield concerns, but also impact the geometry and thus the resistance and capacitance of the patterned lines.

In Sect. 9.4.1, we first review the basic dishing and erosion dependencies in the formation of single level tungsten contacts or local interconnect. We next discuss, in Sect. 9.4.2, similar effects of dishing and erosion in single level copper CMP. The importance of the initial plating topography, and its interaction with copper CMP, is discussed in Sect. 9.4.3. In Sect. 9.4.4 we then consider another complexity not present in STI CMP arising due to the formation of multiple levels of metal, where the topography induced by one CMP step (e.g. at metal 1) can impact the polish of copper in subsequent layers (e.g. metal 2). Finally, in Sect. 9.4.5 we examine some of the approaches being investigated to reduce pattern dependencies or their impact in copper CMP. Of particular interest are dummy fill strategies, as well as new consumables.

9.4.1 Tungsten CMP

Several pattern dependent concerns arise with the use of CMP in the formation of tungsten vias and local interconnect. These include the total loss of line or contact thickness depending on pattern density and feature size, as well as the individual details of dishing and erosion in different regions of a chip. These can result in substantial yield and performance concerns.

Total Line or Contact Loss in Tungsten CMP

The earliest papers on CMP note concerns about pattern dependencies with tungsten CMP. Landis et al. [1] reported that tungsten CMP improves yield compared to the previous RIE etch back process by reducing center seams and edge etch effects, and by removing random defects from the surface of the wafer. When polishing patterned tungsten lines, however, a concern is the total line loss or normalized line thickness as a function of line width and pattern density. As shown in Fig. 9.42, wide lines are found to dish more strongly, and high metal densities (low oxide support densities) erode more strongly. An important point is that the extent of dishing and erosion is a strong function of the overpolish time. If one were able to stop at the nominal endpoint when the metal stack has "just cleared" in the raised field regions, then little dishing or erosion would be expected. If one must polish longer to ensure clearing across the entire chip due to longer range pattern dependencies or to ensure clearing across the full wafer due to wafer scale polish non-uniformities, then the degree of dishing and erosion will increase. Plots such as those of Fig. 9.42 can be used to set basic design rules on the layout. For example, if one wishes to guarantee that 90% of the line thickness will remain after polishing for the process of Fig. 9.42 with 25% overpolish, one might require metal pattern densities no greater than 30% and lines no wider than 10 microns.

Oxide Erosion in Tungsten CMP

Other publications have generally served to confirm and refine the early observations of Landis et al. [1]. Elbel et al. [91] report dishing and erosion in patterned tungsten arrays, and propose a model for dependencies on density and line widths. For conventional tungsten polish processes where large selectivity between tungsten and oxide exist, they found that erosion increases linearly with polish time and nonlinearly with pattern density, as illustrated in Fig. 9.43. The density of supporting oxide features inversely affects the polish rate of oxide in patterned regions in the same fashion as in ILD oxide CMP. This sets a density dependent rate for the oxide removal that is constant throughout the overpolish stage in CMP assuming that the metal polishes much faster than the oxide and thus has little effect on the erosion rate, resulting in a linearly increasing erosion depth with overpolish time. For
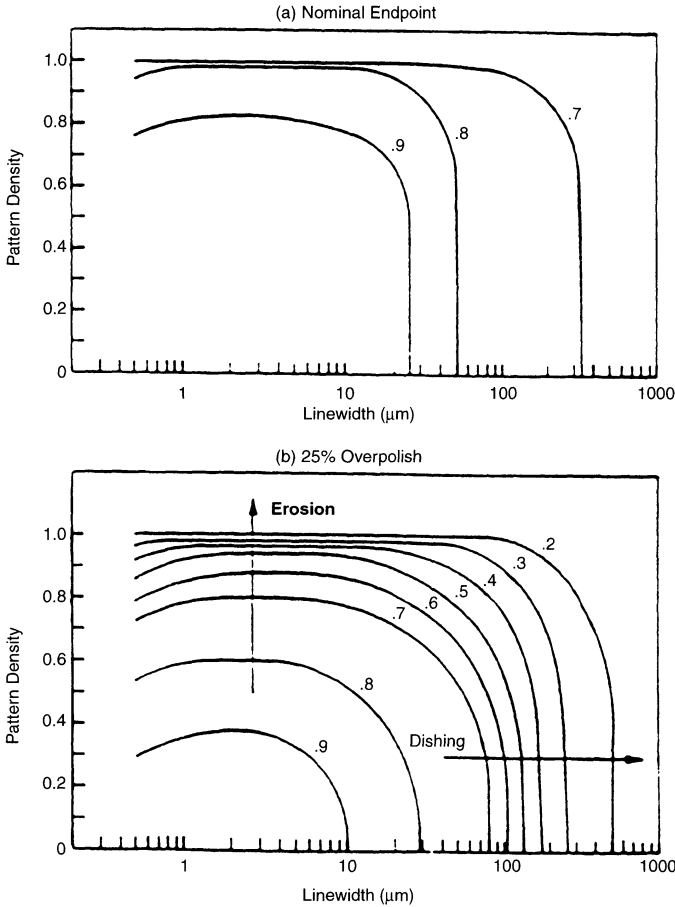
Fig. 9.42. Metal polish using a stiff pad with 40:1 metal to oxide selectivity, showing the normalized line thickness after CMP at (**a**) nominal endpoint, and (**b**) after 25% overpolish. The fraction notation on the contours indicates the portion of the line thickness remaining at the center of the lines. From [1]

the feature sizes considered (greater than $0.4\,\mu m$), the erosion was found to be independent of line width.

Tungsten Dishing

Elbel et al. also report dishing within arrays of contacts at 50% density [91]. They propose a theoretical maximum dishing parameter, d_{max}, which would be the depth of dishing (assuming zero polish rate for the surrounding oxide) at which the pressure on the recessed region vanishes and the dishing stops. This theoretical maximum is not reached; in their experiments, polishing
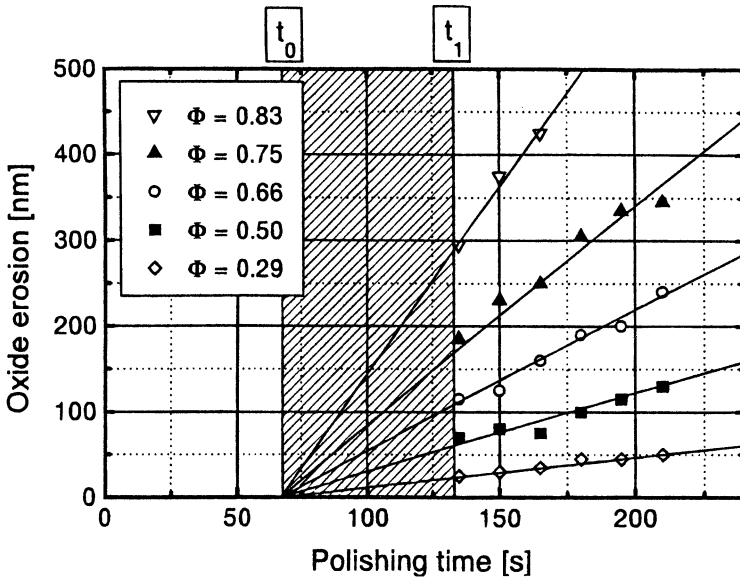
Fig. 9.43. Oxide erosion in an array of tungsten lines and spaces with given pattern density Φ, as a function of polishing time. Clearing of tungsten and barrier within the array occurs at time t_0; clearing of the Ti/TiN barriers in the field areas occurs at time t_1. From [91]

is performed long enough that a steady state condition is reached where an equilibrium is established between the polish rate in the recessed contacts and the rate of oxide erosion. Under these conditions, the actual depth of dishing is found to be a constant independent of the polishing time as illustrated in Fig. 9.44. However, the observed equilibrium dishing and theoretical dishing limit d_{max} do have a line width dependence as illustrated in Fig. 9.45. A spring model is used to predict the bending of the pad into small features and to simulate the dishing profile, with good matches to data.

The observation made by Elbel et al. of constant dishing independent of overpolish time is not always seen. The effect of overpolish time on dishing and erosion in tungsten lines has also been studied by van Kranenburg et al. [92]. As shown in the surface profiles of Fig. 9.46 and plot of Fig. 9.47, there is a small amount of dishing at nominal endpoint and it is seen to increase with overpolish time. These results are likely from a polish that is short enough that the "equilibrium" or steady state dishing condition described by Elbel is not yet reached. In practical CMP processes one often wishes to minimize the overpolish time, in which case the estimation of dishing will depend on time as well as the polish and pattern factors.
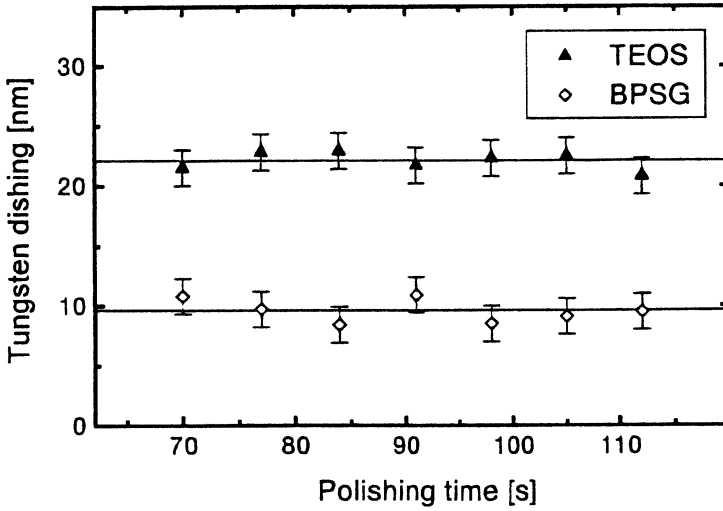
Fig. 9.44. Dishing in an array of 0.5 μm contacts, as a function of polishing time. The amount of dishing depends on the oxide material, but does not depend on time in this steady state dishing experiment. From [91]
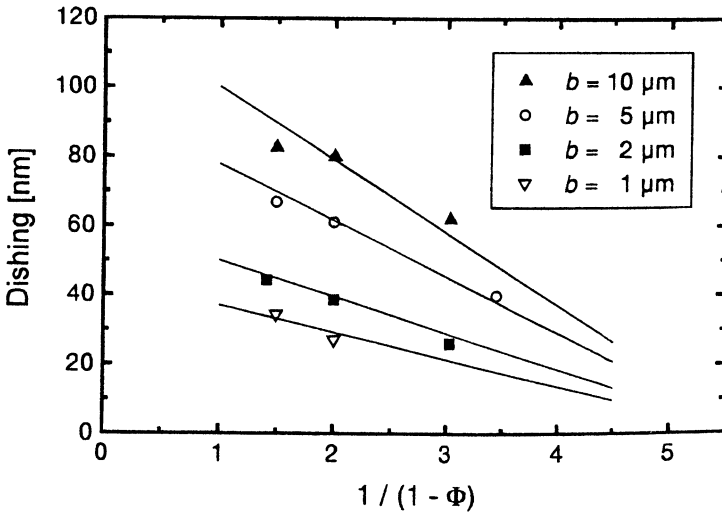
Fig. 9.45. Steady state dishing in an array of lines, as a function of pattern density Φ. The amount of dishing is seen to depend on the linewidth b. From [91]
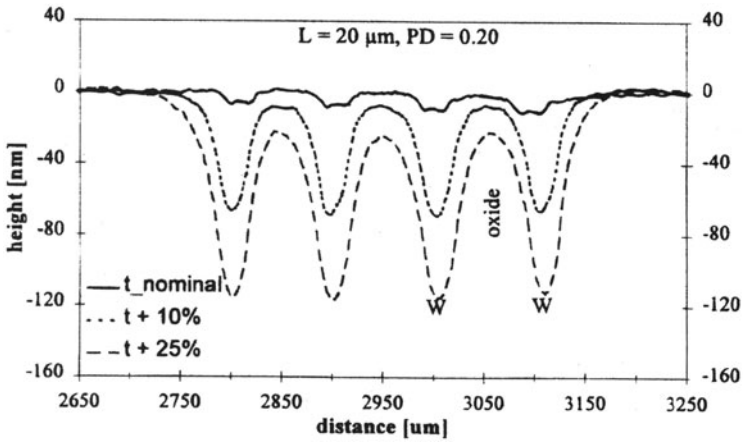
Fig. 9.46. Profilometry trace showing dishing in an array of tungstens lines (20 μm wide), with a line pattern density of 20%. As overpolish time progresses, the degree of dishing is also seen to increase. From [92]
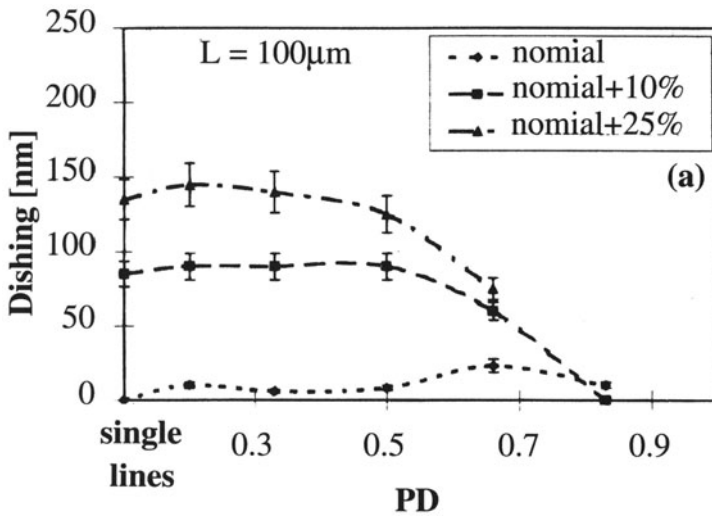
Fig. 9.47. Dishing as a function of pattern density, for different overpolish times. From [92]

Process Interaction and Yield Concerns in Tungsten CMP

The resultant topography created by dishing and erosion, and its interaction with other process steps, raises a number of yield concerns. While there has been little systematic study of these interaction effects on yield reported in the literature, yield concerns due to pattern-induced CMP variation are often raised in a qualitative fashion.

First, substantial dishing or plug recess may make the electrical contact of deposited aluminum metal stacks more difficult. With large dishing on small contacts, the deposited metal may be unable to effectively "fill" into the recess plug. Occasionally an oxide buff polish (also used to remove any residual metal or defects from the oxide surface) after the tungsten polish is designed so as to preferentially remove oxide and leave the top of the tungsten plug slightly raised. If the dishing is large, however, this could require a substantial polish and raise the process step cost.

Second, erosion can also give rise to yield concerns. The topography created by large eroded regions could impact depth of focus in subsequent photolithography, particularly if the erosion from tungsten polish is added to that arising from ILD oxide polish. Rutten et al. [93] also describe a via etch concern arising from the difficulty in completing a via etch in a large recessed region originally created due to tungsten polish induced erosion. The interaction between ILD CMP and tungsten CMP steps raises a similar concern. If the oxide polish results in substantial dielectric thickness variation over a given metal layer, then the via etch process may have difficulty in reaching through the thickest of these regions or may overetch laterally or vertically in the thinner oxide regions [94]. Finally, if a large recessed region exists due to either ILD polish or previous erosion profiles, the ability to completely clear the W/TiN/Ti stack in this recessed region may be a concern.

9.4.2 Single Level Copper CMP

In this section, we consider an idealized "single level" copper CMP process taking place on an initially flat substrate; in Sect. 9.4.4 we expand consideration to additional pattern dependent concern arising from multilevel interconnect formation. In the single level copper CMP case we are concerned with three stages of the process – overburden removal, barrier removal, and overpolish – and the topography CMP may create. In Sect. 9.4.3 we also consider the effect of initial topography variation arising from the copper plating process.

Pattern Dependencies in Copper Overburden Polish

The initial portion of a copper CMP process is focused on the removal of the copper overburden resulting from copper deposition and plating. Prior to reaching the underlying barrier metal, this polish appears similar to that seen

in a single material oxide polish, and the pattern density or area fraction of raised copper is the key pattern parameter affecting local polish rates across the chip.

The first concern relates to the removal of copper with different pattern densities in different regions across the chip. As in the oxide case, a planarization length parameter quantifies the spatial extent of nearby raised topography that affects a local polish rate. The characterization of partial polish in copper, however, is more difficult than in oxide; optical measurement of absolute copper thickness in both up and down areas where copper is still present in both is difficult [95]. Once measurements of thickness are made for a large number of structures on the chip, the planarization length is extracted by error minimization to account for the fact that any single structure's pattern density interacts with nearby structures.

An alternative approach has been used to overcome the difficulty of both measurement and extraction of planarization length. Lefevre et al. propose a wafer-scale pattern with trenches ranging from very small submicron features up to 20 mm in size, with large separations between these large structures [96]. These large area trenches fill nearly conformal. Upon polishing, one examines the amount of material removed in the center of these trenches as a function of the trench size. For small trenches, negligible copper is removed, while very large trenches will polish as if they were field regions. In between is the interesting behavior, as shown in Fig. 9.48. The notion of "minimum" and "maximum" planarization length is proposed to indicate, on a logarithmic plot, where down area polish just begins and where it reaches a maximum, respectively, with the "average planarization length" identified as the midpoint (geometric mean) on this plot. This large trench approach
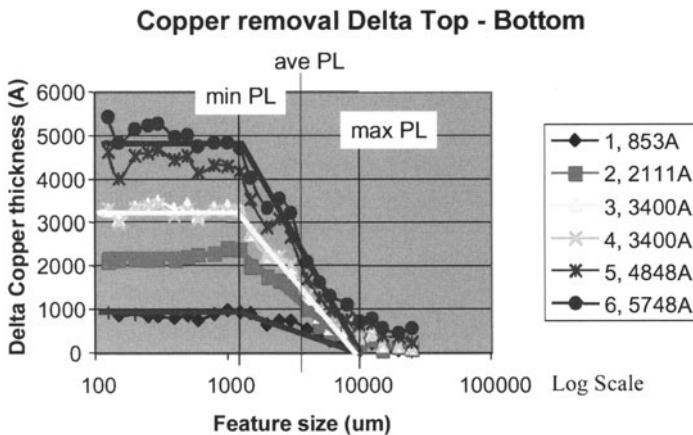
Fig. 9.48. Planarization length plot for copper polish using large trench test mask of indicated size. The delta copper thickness is the difference between the amount of up area (field region) polish and down area (trench region) polish. From [96]

closely matches the conceptual definition of planarization length, although more work is needed to unify this measurement with values used in copper CMP modeling [97].

Dishing and Erosion in Single Level Copper CMP

In one of the first examinations of pattern effects in copper CMP, Steigerwald et al. reported the amount of dishing as a function of the degree of overpolishing for patterned lines from $2\,\mu m$ to $200\,\mu m$, and densities of lines from 20% to 80% (using arrays of 20 lines and spaces) [98]. By polishing a wafer such that the barrier metal (titanium) had cleared in an outer ring of the wafer, dishing and erosion are measured at "endpoint." In addition, a 5% "overpolish" measurement is made by examination of regions on the wafer that receive approximately 5% more polish than points defined as endpoint (based on the amount of oxide thinning in isolated field regions).

Based on these experiments, Steigerwald finds that dishing depends primarily on the width of the line. Figure 9.49 shows a nearly linear amount of dishing for lines in the $2\,\mu m$ to $200\,\mu m$ range, both with and without 5% overpolish. In contrast to the line width dependence of dishing, Steigerwald concludes that erosion depends primarily on the pattern density. As shown in Fig. 9.50, the amount of oxide erosion is only a weak function of line width, but a strong function of pattern density. As the metal pattern density (percentage of area occupied by the copper trench) increases, the amount of erosion also increases. A model based on proportional loading of the pad on
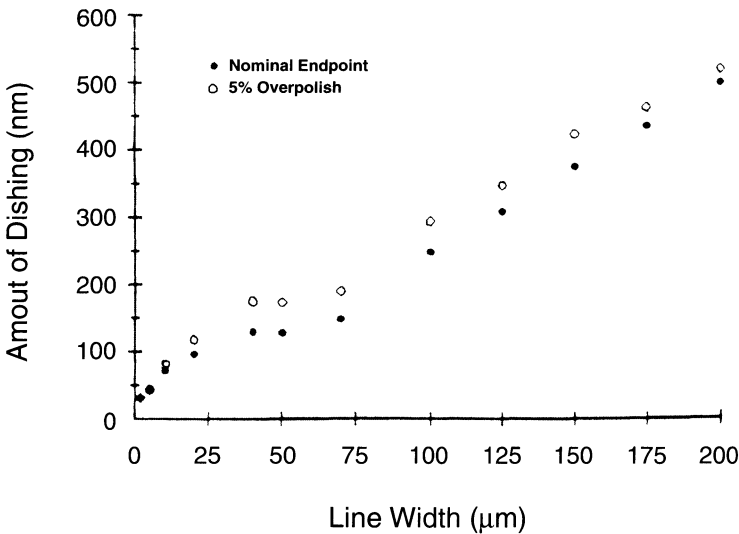
Fig. 9.49. Copper dishing as a function of line width. From [98]
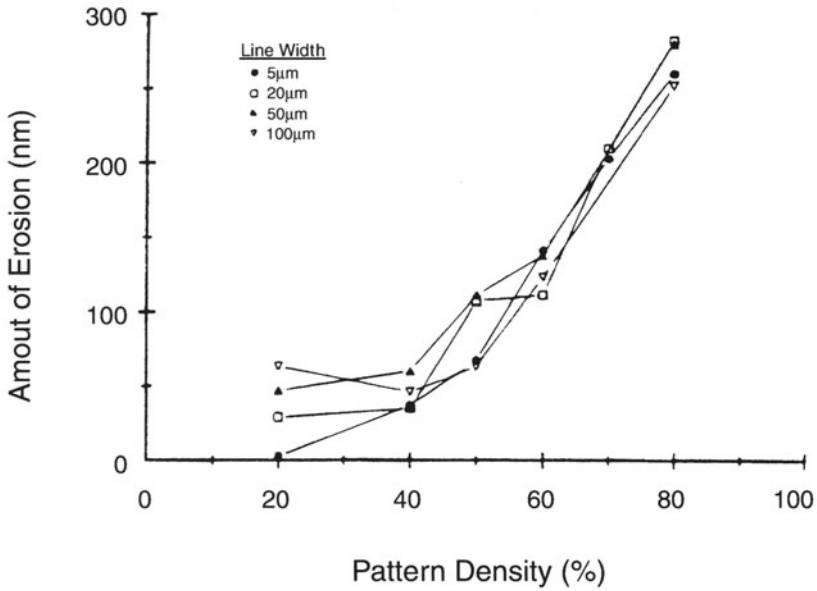
Fig. 9.50. Oxide erosion versus line width for several pattern densities. Erosion is seen to be a strong function of pattern density but not of line width. From [98]
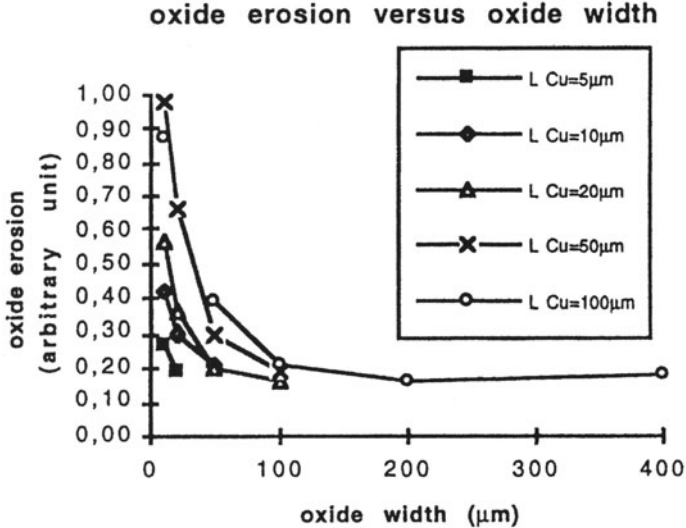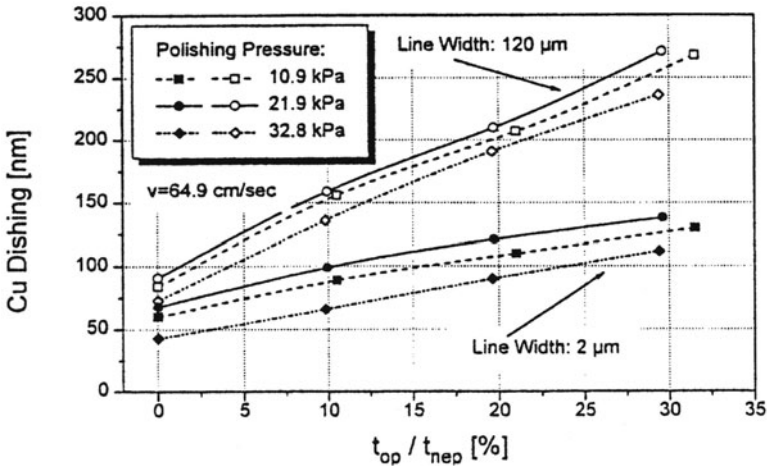

Fig. 9.51. Oxide erosion, as a function of oxide space, for different copper line widths. The oxide erosion is found to accelerate rapidly for thin oxide spaces. From [99]

the raised oxide is found to provide reasonable agreement with the observations.

Several other workers have also reported dishing and erosion results. Fayolle and Romagna [99] also found that when the oxide space is greater than 100 μm, oxide erosion is slight, but it dramatically increases with oxide space reduction as shown in Fig. 9.51. They found that for fine oxide spaces, erosion is dependent not only on oxide width but also on the copper line width of the pattern; this statement can also be interpreted as an additional density dependence.

Stavreva et al. report results for dishing and erosion as a function of over-polish time for a range of pattern density, line width, and line space values, as well as results for different polishing pressures and relative velocities [100]. As shown in Fig. 9.52, they observe dishing to continue to increase as a function of overpolishing time, and to show a small dependence on pressure. Oxide erosion also shows, in Fig. 9.53, only a small or negligible dependence on pressure.

The results from Steigerwald et al., Stavreva et al., and others require careful interpretation. Steigerwald, for example, shows erosion vs. pattern density for different line widths. Fayolle shows oxide erosion vs. oxide width for different copper line widths. Of course, these results are for different polishing processes (different pads, slurries, and tools, as well as different barrier metals and copper depositions). In addition, the sizes of the areas are not necessarily similar (and as we saw in the case of oxide polish, density interactions may depend upon relatively long length scales). Nevertheless, first order pattern dependencies in copper polishing are clear: dishing depends

Fig. 9.52. Copper dishing as a function of normalized overpolish time, for different polish processes. A strong dependence on line width is seen, but relatively weak dependence on pressure. From [100]
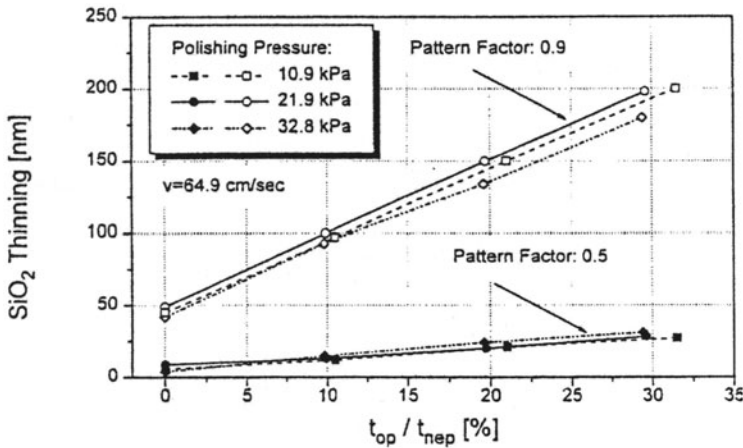
Fig. 9.53. Oxide erosion as a function of normalized overpolish time, for different polish processes. A strong dependence on pattern density is seen, but relatively weak dependence on pressure. From [100]

most strongly on copper line width, and erosion depends most strongly on pattern density. Additional effects, or the interactions between density, line width, and line space as a function of polishing time require careful examination.

In order to study the effect of fine line features as well as larger structures on dishing and erosion, Park et al. present test structures for both electrical and physical measurement [101, 102, 103]. Arrays of lines and spaces can be measured using surface profilometry (and using optical film thickness measurements for oxides regions 10–20 µm or larger), and the effective line dimensions can be extracted from resistance measurements. Profilometry traces are shown in Fig. 9.54 [104]. Here we see the dominance of erosion for narrow lines and spaces (each 1 µm), the dominance of erosion for very wide lines and spaces (each 50 µm), and the importance of both dishing and erosion in the case of intermediate lines and space (each 5 µm).

Tugbawa et al. propose models for copper pattern dependencies [97, 105]; these build on the modeling framework presented by Elbel et al. for tungsten CMP [91]. During the copper bulk removal stage, the pattern density evaluated over some planarization length determines the polish rate in local regions across the chip. This may impact the time at which the polish "touches down" on the underlying barrier metal in these different regions. During the second stage the barrier metal is removed; due to selectivity differences, some degree of copper dishing may occur during this time. For relatively small copper features, it is conjectured that the dishing is governed by the "surface compressibility" characteristics of the pad, a parameter very different from the planarization length, but consistent with a depth-dependent polish rate as
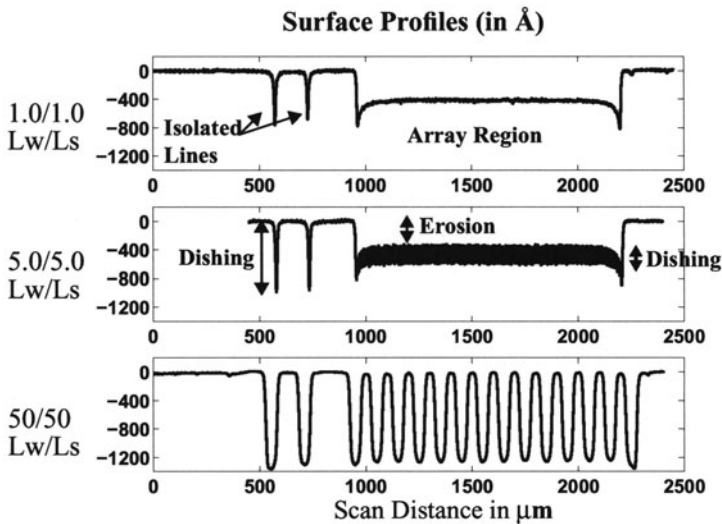
Fig. 9.54. Copper dishing and erosion across arrays with different line widths and line spaces (in microns). From [104]

in Elbel et al. The rate of polish within these "down" copper regions varies inversely with the induced dishing or step height. Once the barrier has been removed, both raised oxide and recessed copper now polish. The rate of raised oxide removal depends on the pattern density of the oxide features (with acceleration for very small oxide spaces), while the pad force is apportioned between the oxide and copper regions. Thus the degree of dishing and erosion depends fundamentally on the interplay between pattern density, line width, and oxide spacing.

Other Feature Level Effects in Copper CMP

In addition to the dishing and erosion, other feature scale effects also occur in copper CMP, as pictured in Fig. 9.55 [106]. The erosion "profile" across an array of lines and spaces has been previously discussed. At the scale of an individual oxide space, the rounding of the oxide spacer may also occur; similarly, the shape of the dishing within each trench or feature may vary. One effect that also merits discussion is the dishing "step" or offset from the edge of the oxide space to the start of the copper dishing profile. This offset is sometimes attributed to additional "chemical" etching of the slurry. Wrschka et al., for example, found a direct correlation between the copper line recess and etch rates for a variety of slurries used [106].
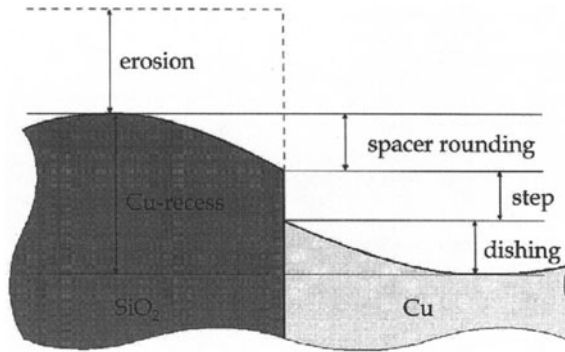
Fig. 9.55. Feature scale non-idealities in copper CMP. From [106]

9.4.3 Interaction with Copper Plating Pattern Dependencies

In addition to the pattern effects created by copper CMP, pattern depen-
dencies in other processes can interact with copper CMP. Among these, the
copper deposition profile is important in defining the initial topography at
the beginning of the polish. In relatively simple deposition processes, one of-
ten finds arrays of features are "recessed" relative to the surrounding field
region. In some "superfill" electroplating processes, on the other hand, the
addition of various leveling agents and other chemistries can result in dra-
matic excess plating thicknesses over individual small features, or result in
"bulge" of substantial regions on the chip. Park et al. present experimental
results for a variety of processes, in terms of the feature-level step height and
the array region recess, as a function of layout parameters [107]. The same
copper CMP test masks (as shown in Section 9.2) are utilized to characterize
the amount of array recess (or array bulge) and local feature step height, as
a function of the patterned line widths and line spaces; a typical superfill
plating result is shown in Fig. 9.56.

Clearly, variation in the thickness of the initial copper in different regions
on a chip can cause different clearing times for those different structures or
regions, with a corresponding variation in the degree of overpolishing these
various regions will see. Because dishing and erosion can be sensitive to the
degree of overpolish, the overall performance of the CMP process can be
strongly affected by the pattern dependencies of the initial plating profile.
Modeling and simulation of these effects are a challenge; a proposed approach
has been presented by Tugbawa et al. which combines a contact mechanics
evaluation of large-region pad bending and pressures, with a local pattern
density and step height model [105]. While improved copper plating technol-
ogy can be expected to decrease the overfill or underfill pattern dependent
problems, it is also clear that plating variation and nonuniformity will pose
a substantial challenge for the subsequent copper CMP process.
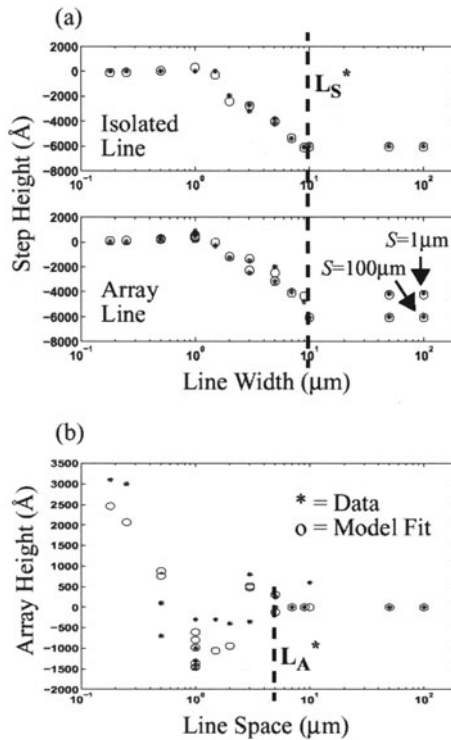
Fig. 9.56. Copper recess and step height in a superfill plating process, as a function of copper line width and line space. From [107]

9.4.4 Multilevel Copper CMP Effects

In advanced integrated circuits, multiple levels (rather than a single level) of metal interconnect are formed using copper CMP. Pattern dependencies create further topography during the polishing of second, third, and later metal layers. Not only does dishing and erosion occur as before on the upper metal level, but the initial topography existing prior to the upper metal copper polish may generate complicated surface heights and a diversity of copper line thickness, thus raising additional yield concerns.

Park et al. present test structures and test masks to study multilevel copper CMP effects [103]. Metal 1 patterns are first polished, resulting in dishing and erosion as shown schematically in Fig. 9.57. Subsequent oxide deposition is approximately conformal to this large-scale topography, resulting in an uneven surface for Metal 2 pattern and etch (assuming no additional oxide polish is performed). Metal 2 fill must then be polished back in Metal 2 CMP, and a substantial yield challenge is to completely clear the copper and barrier from the large recessed metal 1 erosion regions.
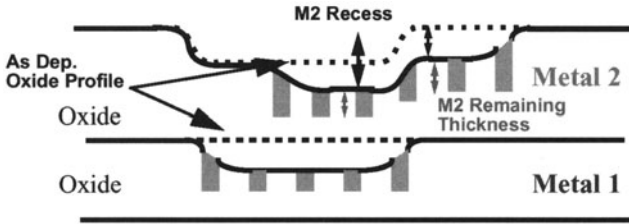
Fig. 9.57. Schematic illustration of erosion profile resulting from two level copper CMP. From [103]

The "half overlap" structure presented by Park, with an array of lines (as well as isolated lines) in metal is overlaid with another array of lines in metal 2. An example resulting surface profile is shown in Fig. 9.58, where 1 μm lines and spaces are patterned on metal 1, and 5 μm lines and spaces are patterned on metal 2. Looking from the left side, we see the dished isolated line from metal 1, the dished metal 2 isolated lines, then the metal 2 dishing/erosion profile sitting within the metal 1 erosion recess, and finally the new metal 2 dishing/erosion profile sitting over the field region from metal 1. Despite the metal 1 and metal 2 patterns being highly regular (isolated lines and fixed arrays of lines and spaces), the final metal 2 topography is quite complicated. Clearly, accurate estimation of the copper line thickness is difficult and the development of accurate models for copper interconnect resistance and capacitance is a challenge.
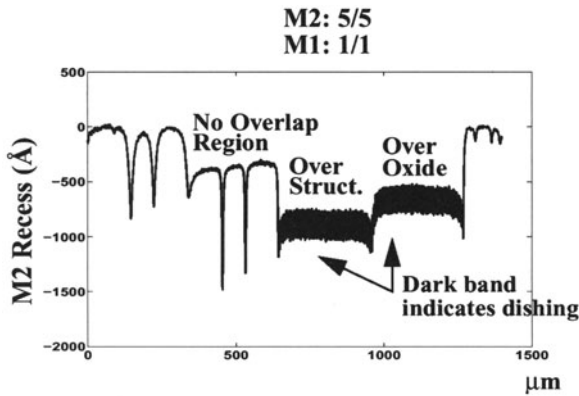
Fig. 9.58. Surface profile resulting from polishing of 5 μm lines and spaces (and isolated lines) over 1 μm lines and spaces. From [103]

9.4.5 Techniques to Address Copper CMP Pattern Dependencies

In this section, we summarize techniques to address or minimize the impact of copper CMP pattern dependencies, including dummy fill and slotting, process optimization, and new pad and slurry developments.

Dummy Fill and Slotting

In the case of oxide and STI polishing, process optimization has not been able to eliminate pattern dependencies, and as a result dummy fill strategies have been required and are widely used [108]. In an analogous fashion, dummy structures may be able to reduce copper pattern effects [109, 110]. The addition of dummy metal lines or structures can better equalize the pattern density across the chip, so that erosion is more similar across the entire chip. Note that the addition of dummy metal can only increase the copper line loss (by increasing erosion), but this may be preferable to unequal degrees of erosion. Alternatively, a dummy "slotting" approach can be considered, in which pillars or slots of oxide are inserted into wide trenches or large copper structures. These oxide structures conceptually "hold up" the polishing pad to reduce the degree of dishing. Thus a tradeoff exists between potential reduced dishing, and the inherent loss of cross-sectional area in large lines due to the presence of the oxide structures. In the case of pads or other large copper regions where dishing may be the largest concern while structure resistance is not critical, the insertion of different slotting patterns is often used. The benefits of oxide insertion in conducting lines are less certain.

Aside from the automatic insertion of dummy metal or dummy oxide, a related approach is to establish design rules that help to minimize or bound the pattern effects arising from copper polish. For example, design rules may specify that metal lines must have not only some minimum width (as is commonly done for electromigration reasons), but also must have some maximum width (to bound the degree of dishing in the line). Similarly, oxide space upper and lower bounds may be established, as well as pattern densities within some specified area.

Process Optimization to Minimize Copper Pattern Dependencies

Customization of the process parameters and consumable set can reduce dishing and erosion. One approach is to use slurries and pads with different selectivity and pattern sensitivity for different stages of the process [111]. For example, a highly selective first step slurry is used to stop on the barrier metal, followed by a highly selective barrier step to clear the metal. Alternatively, a 1:1:1 (copper, barrier, and oxide) selectivity slurry is used in the final polish step to remove excess copper and barrier metal, as well as reduce any existing oxide topography. An extreme "sacrificial oxide" approach is used

where preferential oxide slurry removes the oxide profile arising from erosion; the tradeoff here is that copper line loss is substantial but the topography is reduced.

Consumable Options to Reduce Copper Dishing and Erosion

Additional consumable options are also used to reduce dishing and erosion. Just as in STI, fixed abrasive pads are of interest as a means to minimize "down area" dishing into copper lines. In the copper case, scratching of the surface is a substantial concern [112, 113]. Another approach is "abrasive free" polishing, in which chemical slurry without the usual alumina or silica particles is used with a conventional pad and tool. In work reported by Hitachi [114], very little barrier metal polishing (and thus small erosion) with reduced dishing occurs during a first CMP step, followed by either plasma etch or conventional barrier metal CMP to achieve less overall dishing and erosion. The removal rate of copper as a function of applied pressure for abrasive-free shows a non-linear dependence on pressure, as illustrated in Fig. 9.59 [115].
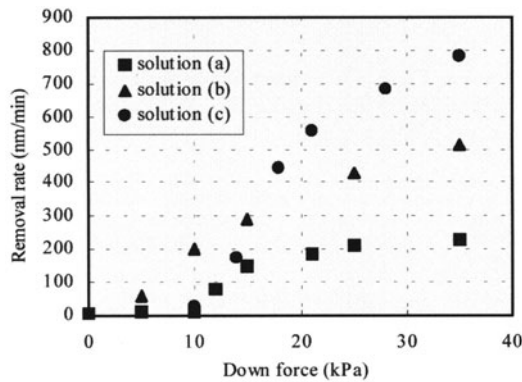
Fig. 9.59. Removal rate as a function of applied pressure for abrasive-free copper polishing slurry. From [115]

In this case, if a line is recessed somewhat from the surface, the pressure on the line may be decreased enough that the removal rate becomes negligible. Such effects can be integrated into pattern density/step-height or contact wear models [116].

References

1. H. Landis, P. Burke, W. Cote, C. Hoffman, C. Kaanta, C. Koburger, W. Lange, M. Leach and S. Luce, Thin Solid Films, **220**, 1, 1992.
2. W. Patrick, W. Guthrie, C. Standley and P. Schiable, J. Electrochem. Soc., **138** (6), 1778, 1991.
3. S. Wolf, Silicon Processing for the VLSI Era, Volume 2 - Process Integration, Lattice Press, Sunset Beach, CA, 1990.
4. T. Daubenspeck, J. DeBrosse, C. Kopburger, M. Armacast and J. Abernathey, J. Electrochem. Soc., **138** (2), 506, 1991.
5. D. Hetherington, A. Farino and Y. Strausser, Proceedings 1996 CMP–MIC Conference, 74, IMIC, Tampa, 1996.
6. J. Schneir, R. Jobe, V. Tsai, A. Samsavar, D. Hetherington, M. Moinpour, Y. Park, M. Maxim, J. Chu and W. Sze, in *Advanced Metallization and Interconnect Systems for ULSI Applications in 1996*, 555, Pittsburgh, Pa., Materials Research Society, 1997.
7. M. Borba, T. Myers, M. Stell, D. Scherber and M. Fury, Proceedings 1995 DUMIC Conference, 331, IMIC, Tampa, 1995.
8. S.V. Nguyen, D. Dobuzinsky, D. Harmon, R. Gleason and S. Fridmann, J. Electrochem. Soc., 2209, **137** (7), 1990.
9. D.R. Cote, S.V. Nguyen, W.J. Cote, S.L. Pennington, A.K. Stamper and D.V. Podlesnik, IBM Journal of Research and Development, 437, **39** (4), 437, 1995.
10. J.T. Pan, D. Ouma, P. Li, D. Boning, F. Redecker, J. Chung and J. Whitby, Proceedings 1998 VMIC Conference, 467, IMIC, Tampa, 1998.
11. C. Yu, P. Fazan, V. Mathews and T. Doan, Applied Physics Letters, 1344, **61** (11), 1992.
12. S. Sivaram, H. Bath, E. Lee, R. Leggett and R. Tolles, *Proceedings ULSI -VII Advanced Metallization for ULSI Applications,* Material Res. Soc., 1992.
13. S. Sivaram, R. Tolles, H. Bath, E. Lee and R. Leggett, Mat. Res. Soc. Symp. Proc., **53**, 260, 1992.
14. D. Boning, "Characterization Methods and Metrics for Patterned Wafer CMP," SEMI CMP Standards Meeting, SEMICON/WEST, San Francisco, CA, July 1999.
15. B. Stine, D. Ouma, R. Divecha, D. Boning, J. Chung, D. Hetherington, I. Ali, G. Shinn, J. Clark, O.S. Nakagawa and S.-Y. Oh, Proceedings 1997 CMP–MIC Conference, 266, IMIC, Tampa, 1997.
16. B. Stine, D. Ouma, R. Divecha, D. Boning, J. Chung, D. Hetherington, C.R. Harwood, O.S. Nakagawa and S.-Y. Oh, IEEE Trans. on Semi. Manuf., 129, **11** (1), 1998.
17. P. Renteln, M.E. Thomas and J.M. Pierce, Proceedings 1990 VMIC Conference, 57, IMIC, Tampa, 1990.
18. P. Renteln and J. Coniff, Mat. Res. Soc. Symp. Proc., 105, **337**, 1994.
19. P.A. Burke, Proceedings 1991 VMIC Conference, 379, IMIC, Tampa, 1991.
20. J. Warnock, J. Electrochem. Soc., 2398, **138** (8), 1991.
21. Y. Hayashide, M. Matsuura, M. Hirayama, T. Sasaki, S. Harada and H. Kotani, Proceedings 1995 VMIC Conference, 464, IMIC, Tampa, 1995.
22. J. Grillaert, H. Meynen, J. Waeterloos, B. Coenegrachts and L. van den Hove, in Advanced Metallization and Interconnect Systems for ULSI Applications in 1996, 525, Pittsburgh, Pa., Materials Research Society, 1997, 525.

23. D. Ouma, Ph.D. Thesis, Elect. Eng. and Comp. Sci. Dept., MIT, Nov. 1998.
24. D. Hetherington and D. Stein, Sandia National Laboratories, Dec. 1999.
25. T. Park, T. Tugbawa, J. Yoon, D. Boning, J. Chung, R. Muralidhar, S. Hymes, Y. Gotkis, S. Alamgir, R. Walesa, L. Shumway, G. Wu, F. Zhang, R. Kistler and J. Hawkins, Proceedings 1998 VMIC Conference, 437, IMIC, Tampa, 1998.
26. T. Park, T. Tugbawa, D. Boning, J. Chung, S. Hymes, R. Muralidhar, B. Wilks, K. Smekalin and G. Bersuker, Proceedings 1999 CMP–MIC Conference, 184, IMIC, Tampa, 1999.
27. T. Park, T. Tugbawa and D. Boning, Proceedings 2000 CMP–MIC Conference, 196, IMIC, Tampa 2000.
28. M. Gostein, M. Banet, M.A. Joffe, A.A. Maznev, R. Sacco, J.A. Rogers and K.A. Nelson, "Thin-Film Metrology Using Impulsive Stimulated Thermal Scattering (ISTS)," in Handbook of Silicon Semiconductor Metrology, Ed. A.C. Diebold, Marcel Dekker, NY, 2001.
29. T. Park, T.E. Tugbawa and D.S. Boning, Proceedings 2001 International Interconnect Technology Conference, 274, 2001.
30. J. McKinnis, S. Lantz and D. Sauer, Proceedings 2000 CMP–MIC Conference, 104, IMIC, Tampa, 2000.
31. B. Stine, V. Mehrotra, D. Boning, J. Chung and D. Ciplickas, IEDM Tech. Digest, 133, 1997.
32. V. Mehrotra, S. Nassif, D. Boning and J. Chung, IEDM Tech. Digest, 767, Dec. 1998.
33. D. Shum, J. Higman, M. Khazhinsky, K. Wu, S. Kao, J. Burnett and C. Swift, IEDM Tech. Digest, 665, Dec. 1997.
34. J. Grillaert, M. Meuris, N. Heylen, K. Devriendt, E. Vrancken and M. Heyns, Proceedings 1998 CMP–MIC Conference, 79, IMIC, Tampa, 1998.
35. D. Hetherington and D. Stein, "Dynamic mechanical response of polyurethane polishing pads to pad loads and its relationship to planarization," presented at Symposium P: Chemical-Mechanical Polishing – Fundamentals and Challenges, Materials Research Society Spring Meeting, San Francisco, CA, April 5–7, 1999.
36. D. Hetherington, D. Stein, D. Ouma and D. Boning, "Characterization of shallow trench isolation CMP," presented at 3rd Annual CMP Symposium, August 1998, Lake Placid, New York.
37. K. Smekalin, Solid State Technology, 187, July, 1997.
38. T. Murakami, M. Nishio and M. Hamanaka, Proceedings 1996 VMIC Conference, 413, IMIC, Tampa, 1996.
39. J. Grillaert, M. Meuris, E. Vrancken, K. Devriendt, W. Fyen and M. Heyns, "Modeling the Influence of Pad Bending on the Planarization Performance During CMP," Symposium P: Chemical-Mechanical Polishing - Fundamentals and Challenges, Materials Research Society Spring Meeting, San Francisco, CA, April 5–7, 1999.
40. T. Hyde and J. Roberts, U.S. Patent 5,257,478, November 1993.
41. D. Hetherington, D. Stein and M. Oliver, Proceedings 2000 CMP–MIC Conference, 399, IMIC, Tampa, 2000.
42. K. Devriendt, E. Vrancken, J. Grillaert, M. Mueris, N. Heylen and M. Heyns, Proceedings 1999 CMP–MIC Conference, 227, IMIC, Tampa, 1999.
43. L. Cook, J. Wang, D. James and A. Sethuraman, Semiconductor International, 141, Nov, 1995.

44. K. Achuthan, "Evaluation and characterization of polyurethane chemical-mechanical polishing pads," Ph.D. Thesis, Clarkson University, 1998.
45. W. Li, D. Shin, M. Tomozawa and S. Murarka, Thin Solid Films, **270**, 601, 1995.
46. D. Rutherford, D. Goetz, C. Thomas, R. Webb, W. Bruxvoort, J. Buhler and W. Hollywood, "Abrasive construction for semiconductor wafer modification," US Patent 5,692,950, Dec. 1997.
47. L.M. Cook, J. Non-Crystalline Solids, **120**, 152, 1990.
48. D. Evans, B. Ulrich and M. Oliver, Proceedings 1998 CMP–MIC Conference, 347, IMIC, Tampa, 1998.
49. J. Gagliardi, Proceedings 1999 VMIC Conference, 223, IMIC, Tampa, 1999.
50. D. Goetz, Proceedings 1999 CMP–MIC Conference, 234, IMIC, Tampa, 1999.
51. P. van der Velden, Microelectronic Engineering, **50**, 41, 2000.
52. L. Economikos, F. Jamin, A. Ticknor and A. Simpson, Proceedings 2001 CMP–MIC Conference, 553, IMIC, Tampa, 2001.
53. M. Fayolle, J. Lugand, F. Weimar and W. Bruxvoort, Proceedings 1998 CMP–MIC Conference, 128, IMIC, Tampa, 1998.
54. S. Kweon, J. Kang, B. Kwon, H.-J. Kim, J.-H. Lee and J. Lee, Proceedings 1999 CMP–MIC Conference, 242, IMIC, Tampa, 1999.
55. J. Gagliardi and T. Vo, Proceedings 2000 CMP–MIC Conference, 373, IMIC, Tampa, 2000.
56. A. Romer, T. Donohue, J. Gagliardi, F. Weimar, P. Thieme and M. Hollatz, Proceedings 2000 CMP–MIC Conference, 265, IMIC, Tampa, 2000.
57. J. Schlueter, I. Kim and F. Krupa, Proceedings 1999 CMP–MIC Conference, 336, IMIC, Tampa, 1999.
58. D. Hansen, H. Sun and R. Wall, Proceedings 1999 CMP–MIC Conference, 417, IMIC, Tampa, 1999.
59. D. Hetherington and D. Stein, Proceedings 2000 CMP–MIC Conference, 399, IMIC, Tampa, Mar. 2–3, 2000.
60. A. Francis, W. Fortino, P. Feeney, B. Mueller, G. Bogush, S. Ganeshkumar, S. Yontz, F. Khan and C. Baker, Proceedings 1999 VMIC Conference, 240, IMIC, Tampa, 1999.
61. K.-S. Choi, S.-I. Lee, C.-I. Kim, C.-W. Nam, S.-D. Kim and C.-T. Kim, Proceedings 1999 CMP–MIC Conference, 307, IMIC, Tampa, 1999.
62. S.-I. Lee, C.-I. Kim, H. Kim, J.-H. Kim, C.-W. Nam, S. Kim and C.-T. Kim, Proceedings 2001 CMP–MIC Conference, 218, IMIC, Tampa, 2001.
63. B.-H. Kwon, J.-H. Lee, S. Kweon, S.-Y. Lee, B.-C. Kim, I. Yoon, S.-I. Lee and J.-G. Lee, Proceedings 2000 CMP–MIC Conference, 163, IMIC, Tampa, 2000.
64. H. Nojo, M. Kodera and R. Nakata, IEDM Tech. Digest,349, Dec. 1996.
65. K. Goh, F. Chen, S. Balakumar, C. Chen, C. Lin, L. Chan and G. Higelin, Proceedings 1999 VMIC Conference 531, IMIC, Tampa, 1999.
66. T. Park, J. Kim, K. Park, H. Lee, H. Shin, Y. Kim, M. Park, H. Kang and M. Lee, Tech. Digest VLSI Symposium,159, 1999.
67. D. Hetherington, "Dielectric CMP processes," Short course presentation – CMP Planarization for ULSI Multilevel Interconnection, Santa Clara, CA, Feb. 1999.
68. L. Camilletti, IEEE/SEMI Advanced Semiconductor Manufacturing Conference, 2, 1995.
69. B. Stine, D. Boning, J. Chung, L. Camilletti, E. Equi, S. Prasad, W. Loh and A. Kapoor, Proceedings 1996 VMIC Conference, 421, IMIC, Tampa, 1996.

70. A. Kahng, G. Robins, A. Singh and A. Zelikovsky, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, **18** (4), 445, 1999.
71. Y. Chen, A. Kahng, G. Robins and A. Zelikovsky, Proceedings of the ASP-DAC Design Automation Conference 2000, Asia and South Pacific, 523, Jan. 25–28, 2000.
72. S.J. Kim, Y.W. Lee, S.P. Jung, S.Y. Kim and J.S. Choi, Proceedings 1999 CMP–MIC Conference, 84, IMIC, Tampa, 1999.
73. G. Liu, R. Zhang, K. Hsu and L. Camilletti, Proceedings 1999 CMP–MIC Conference, 120, IMIC, Tampa, 1999.
74. C. Gillot, E. DeBacker, J. Grillaert, N. Heyley, J. Vaca and G. Blavier, Proceedings 1999 CMP–MIC Conference, 413, IMIC, Tampa, 1999.
75. B. Lee, D. Boning, D. Hetherington and D. Stein, Proceedings 2000 CMP–MIC Conference, 255, IMIC, Tampa, 2000.
76. Y. Chen, A. Kahng, G. Robins and A. Zelikovsky, Proc. 37th Design Automation Conference, 671, June 5–9, 2000.
77. R. Tian, D.F. Wong and R. Boone, Proc. 37th Design Automation Conf., 667, June 5–9, 2000.
78. I.Y. Yoon, B.H. Kwon, Y.B. Park, H.H. Ryu and W.G. Lee, Proceedings 2001 CMP–MIC Conference, 69, IMIC, Tampa, 2001.
79. R. Tian, X. Tang and D.F. Wong, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, **21** (1), 63, 2002.
80. T.K. Yu, S. Cheda, J. Ko, M. Roberton, A. Dengi and E. Travis, IEDM Tech. Digest, 909, Dec. 1999.
81. K. Ryu, C. Ouyang, L. Milor, W. Maly, G. Hill and Y. Peng, International Symposium on Semiconductor Manufacturing Conference Proceedings, 221, 1999.
82. T. Ohta, T. Toda and H. Ueno, International Conference on Simulation of Semiconductor Processes and Devices, SISPAD '99, 195, 1999.
83. M. Nandakumar, A. Chatterjee, S. Sridhar, K. Joyner, M. Rodder and I.-C. Chen, IEDM Tech. Digest,133, Dec. 1998.
84. A. Chatterjee, I. Ali, K. Joyner, D. Mercer, J. Kuehne, M. Mason, A. Esquivel, D. Rogers, S. O'Brien, P. Mei, S. Murtaza, S.P. Kwok, K. Taylor, S. Nag and G. Hames, Journal Vacuum Science and Technology B: Microelectronics and Nanometer Structures, **15** (6), 1936, 1997.
85. J. Boyd and J. Ellul, J. Electrochemical Society, **143** (11), 371, 1996.
86. K. Ishimoto, S. Tanaka, M. Kishimoto and Y. Itoh, Proceedings IEEE International Symposium on Semiconductor Manufacturing Conference, F21, 1997.
87. J. Grillaert, N. Heylen, E. Vrancken, G. Badenes, R. Rooyackers, M. Meuris and M. Heyns, Proceedings 1998 CMP–MIC Conference, 313, IMIC, Tampa, 1998.
88. G. Badenes, R. Rooyackers, E. Augendre, E. Vandamme, C. Perelló, N. Heylen, J. Grillaert and L. Deferm, J. Electrochem. Soc., **147** (10), 3827, 2000.
89. K. Devriendt, "Shallow Trench Isolation: The Process and Its Integration Issues," in "Advances in CMP Technology" short course, SEMICON EUROPA 2000, April 4, 2000, Munich, Germany.
90. M. Oliver, S. Hosali, D. Evans, D. Hetherington, D. Stein and J. Stevens, Proceedings 1999 CMP–MIC Conference, 383, IMIC, Tampa, 1999.
91. N. Elbel, B. Neureither, B. Ebersberger and P. Lahnor, J. Electrochem. Soc., **145** (5), 1659, 1998.

92. H. van Kranenburg and P.H. Woerlee, J. Electrochem. Soc., **145** (4), 1285, 1998.
93. M. Rutten, P. Feeney, R. Cheek and W. Landers, Semiconductor International, 123, Sept. 1995.
94. B. Smith, S. Blackley, R. Carter, S. Cheda, P. Crabtree, D. Farber, M. Gall, R. Islam, D. Jawarani, C. King, D. Menke, R. Nelson, L. Pressley, D. Smith, T. Sparks, T. Stephens, E. Travis and S. Venkatesan, Proceedings 1999 International Interconnect Technology Conference, 106, 1999.
95. M. Gostein, M. Banet, M.A. Joffe, A.A. Maznev, R. Sacco, J.A. Rogers and K.A. Nelson, "Thin-Film Metrology Using Impulsive Stimulated Thermal Scattering (ISTS)," in Handbook of Silicon Semiconductor Metrology, Ed. A.C. Diebold, Marcel Dekker, NY, 2001.
96. P. Lefevre, A. Gonzales, T. Brown, G. Martin, T. Tugbawa, T. Park, D. Boning, M. Gostein and J. Nguyen, Materials Research Society Spring Meeting, San Francisco, CA, April 2001.
97. T. Tugbawa, T. Park, D. Boning, L. Camilletti, M. Brongo and P. Lefevre, Proceedings 2001 CMP–MIC Conference, 65, IMIC, Tampa, 2001.
98. J.M. Steigerwald, R. Zirpoli, S.P. Murarka and R.J. Gutmann, J. Electrochem. Soc., **142** (10) , 2841, 1994.
99. M. Fayolle and F. Romagna, Microelectr. Eng., **37/38**, 135, 1997.
100. Z. Stavreva, D. Zeidler, M. Plötner and K. Drescher, Microelectr. Eng., **37/38**, 143, 1997.
101. T. Park, T. Tugbawa, J. Yoon, D. Boning, J. Chung, R. Muralidhar, S. Hymes, Y. Gotkis, S. Alamgir, R. Walesa, L. Shumway, G. Wu, F. Zhang, R. Kistler and J. Hawkins, Proceedings 1998 VMIC Conference, IMIC, Tampa, 1998.
102. T. Park, T. Tugbawa, D. Boning, J. Chung, S. Hymes, R. Muralidhar, B. Wilks, K. Smekalin and G. Bersuker, Proceedings 1999 CMP–MIC Conference, IMIC, Tampa, 1999.
103. T. Park, T. Tugbawa, D. Boning, S. Hymes, T. Brown, K. Smekalin, and G. Schwartz, Proceedings of the Electrochem. Soc., **99-37**, 94, 1999.
104. T. Park, T. Tugbawa and D. Boning, Proceedings 2000 International CMP–MIC Conference, 196, IMIC, Tampa, March 2000.
105. T. Tugbawa, T. Park, B. Lee, D. Boning, P. Lefevre and L. Camilletti, "Modeling of Pattern Dependencies for Multi-Level Copper Chemical-Mechanical Polishing Processes," Materials Research Society (MRS) Spring Meeting, San Francisco, CA, April 2001.
106. P. Wrschka, J. Hernandez, G.S. Oehrlein and J. King, J. Electrochem. Soc., **147** (2), 706, 2000.
107. T. Park, T.E. Tugbawa and D.S. Boning, Proceedings International Interconnect Technology Conference, 274, June 2001.
108. R. Tian, D.F. Wong and R. Boone, IEEE Trans. on CAD, **20** (7), 902, July 2001.
109. J.T. Pan and P. Li, Proceedings 2000 VMIC Conference, 197, IMIC, Tampa, 2000.
110. R. Tian, X. Tang and D.F. Wong, Proceedings 2001 CMP–MIC Conference, 57, IMIC, Tampa, 2001.
111. K. Wijekoon, S. Mishra, S. Tsai, K. Puntambekar, M. Chandrachood, F. Redeker, R. Tolles, B. Sun, L. Chen. T. Pan, P. Li, S. Nanjangud, G. Amico, J. Hawkins, T. Myers, R. Kistler, V. Brusic, S. Wang, I. Cherian, L. Knowles,

C. Schmidt and C. Baker, Advanced Semiconductor Manufacturing Conference, 354, Oct. 1998.

112. V.N. Koinkar, R. Golzarian, M. Van Hanehem, Q. Luo, J. Shen, P. Burke, T. Fletcher, L.C. Hardy, J. Kollodge, J. Trice, T. Engfer and E. Funkenbusch, Proceedings 2000 CMP–MIC Conference, 58, IMIC, Tampa, 2000.
113. D.R. Evans, M.R. Oliver and M. Kulus, Proceedings of the Electrochem. Soc., 2000–26, 122, 2001.
114. S. Kondo, N. Sakuma, Y. Homma, Y. Goto, N. Ohashi, H. Yamaguchi and N. Owada, Proceedings International Interconnect Technology Conference, 253, June 2000.
115. N. Ohashi, Y. Yamada, N. Konishi, H. Maruyama, T. Oshima, H. Yamaguchi and A. Satoh, Proceedings 2001 International Interconnect Technology Conference, 140, June 2001.
116. D. Boning, B. Lee, T. Tugbawa and T. Park, "Modeling the Effect of Non-Prestonian Pressure on Pattern Dependencies in CMP," 6th International Symposium on CMP, Lake Placid, NY, August 2001.

10 Integration Issues of CMP

K.M. Robinson, K. DeVriendt, and D.R. Evans

When CMP was introduced, its first role was to planarize ILD films to enable multiple levels beyond two or three of metal interconnects. As with any other step in semiconductor processing, the presence of any CMP step comes with a significant number of integration issues, and their associated performance tradeoffs. This chapter addresses those issues and tradeoffs. Though there are a number of integration issues that are common to all CMP steps, such as defect reduction, in general, each type of CMP, such as tungsten CMP or STI CMP, has issues that are specific to that specific step. For that reason the chapter is grouped into the major CMP process types: oxide, tungsten, STI and copper. There are other CMP processes, but these four areas cover most of the issues with the more uncommon processes, such as poly-silicon CMP.

Again, many integration issues are similar among the various types of CMP. But, as an example, the dishing and erosion issues of copper CMP, in general, are much more demanding than those of tungsten. These integration requirements drive the specific CMP process requirements for each step. The goal of this chapter is to describe the integration issues, and how they impact the CMP process. As with all semiconductor processing, the specific process control limits tighten with each generation, but within a slowly evolving (from the point of view of adding new steps) process technology, the integration issues also evolve relatively slowly.

10.1 Oxide CMP Integration

10.1.1 Introduction

Pre-metal dielectric (PMD) and Interlayer dielectric (ILD) CMP are the most common and most studied of the CMP steps. Originally introduced to provide planarization for lithography, they have also come to be seen as enabling processes for etch and metallization. Unique amongst the CMP steps, PMD and ILD CMP are "stop-in-film" processes. There are no interfaces on which to stop the CMP process resulting in a process very dependent upon removal rate control, within die, wafer, lot and fab, for reproducibility. Collectively labeled "oxide" CMP, these two processes share many integration concerns

of deposition, planarity and defectivity, which will be covered first followed by more details of PMD and ILD specific issues.

PMD is meant to provide planarization between the front-end active devices and the back-end metallization. The reasons for the planarization are multiple, a) enabling contact lithography, b) enabling contact etch uniformity and c) enabling contact tungsten CMP. A new role for PMD CMP is the enablement of copper CMP in either single or dual damascene integration.

ILD CMP is meant to provide planarization between the increasing number of metal layers in the back-end. The reasons are two-fold, a) enabling via lithography and b) enabling via tungsten CMP. Although these two reasons appear similar to PMD CMP, the integration concerns are very different. It should be noted that although ILD is in the process of being replaced in the technology roadmap with copper CMP, it still represents a majority of CMP processing in manufacturing.

Integration of CMP began with ILD CMP. It was recognized that continuing device shrinkage dictates that the depth of focus for lithography tools would require true global flat surfaces. Depth of focus can simplistically be defined as the width of the optical interface with sufficient process margin to allow for within tolerance printing of mask features. It is highly dependent on the numerical aperture, which is lithography tool dependent, and the wavelength of the light [1].

Flat plate capacitor - substrate as bottom plate
No PMD CMP - Single layer metallization
LOCOS isolation

Gate structure

Stacked plate capacitor - patterned poly silicon bottom plate

PMD CMP - W etchback / W CMP interconnect
LOCOS isolation

Metal   isolation
contact

Vertical crown capacitor - Plug etchback / CMP capacitor interconnect

PMD CMP - W CMP interconnect
STI CMP isolation

capacitor

Fig. 10.1. Vertical growth of DRAM architecture as generation shrinks forces the capacitor structure from a flat plate to a three-dimensional structure

It is readily apparent that in order to pattern smaller features, the resolution of the lithography process must also improve and like depth of focus, resolution is also wavelength dependent [1]. As an example of the conflict inherent in shrinking dimensions, in a device such as a DRAM the minimum capacitance required for storage in an individual memory cell are determined by the cell area and the dielectric constant. The only way to maintain sufficient capacitance as the cell area shrinks is to go three-dimensional, as shown in Fig. 10.1 [2]. This increase in topography is counter-productive to the shrinking depth of focus and forces the need to planarize the topography before subsequent lithography steps. Similar arguments are used for logic devices. There are very good reviews on initial planarization process such as spin on glass and etchback and the reader is referred to those references for a historical perspective [3]. The focus of this section is to understand the integration of PMD and ILD CMP into present day chip manufacturing.

10.1.2 General Oxide CMP Integration Issues

PMD and ILD CMP are known collectively as oxide CMP. For both ILD and PMD, the slurry, pads, tools and ultimate purpose are essentially identical with some minor modifications based on proprietary needs of the particular devices. Subsequently many of the integration concepts are also interchangeable and will be covered with respect to planarization of a generic oxide film. More specific ILD and PMD concerns are covered separately.

Depth of Focus

Depth of Focus (DOF) is a description of the process margin for lithography patterning. In essence within the DOF, patterns maintain the same resolution across the reticle field. Variations in topography must be kept smaller than the DOF to assure design rule specifications on line widths and spacings are met. Local planarization techniques, such as spin-on glass and etchback, are adequate to maintain small dimensional consistency, over distances of 1–10 μm. However they fail for the larger dimensional needs, in the range of 10 μm–10mm, (cross-die and cross-reticle). At illumination wavelengths of 193 nm and 157 nm, an improved DOF is required to enable continuing technology critical dimension or CD shrinks [4].

Typical commercially available I-line (365 nm) or KrF (248 nm) based steppers have a DOF of $\sim$ 1 μm [1]. This is calculated on a simple formula:

$$\text{DOF} = \pm \text{K}\lambda/\text{NA}^2. \qquad (10.1)$$

K is an optical constant based on the stepper and resist, λ is the stepper wavelength and NA is the numerical aperture [3]. Typical values of NA range from 0.60 to 0.45 for I-line and KrF steppers. However as the limitations
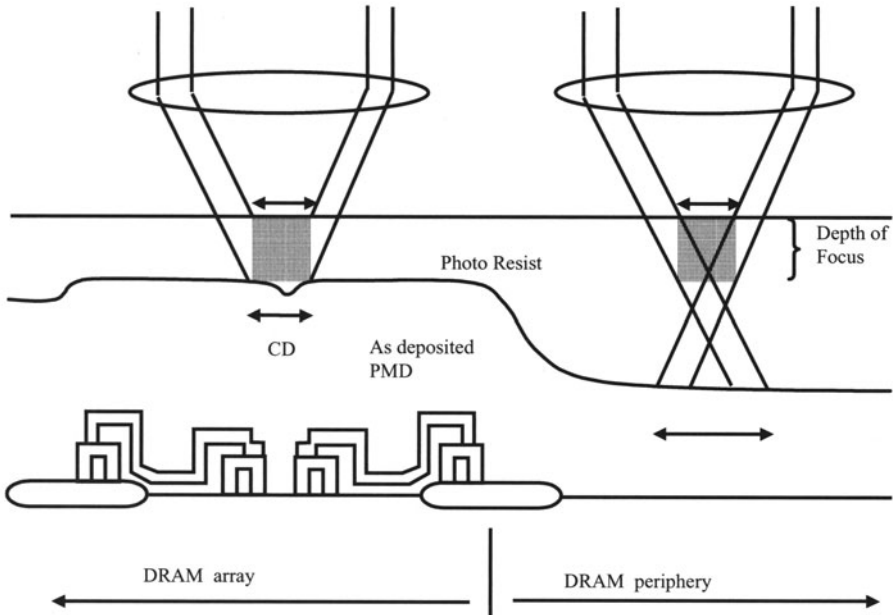
Fig. 10.2. Effect of topography on depth of focus

of KrF resists are pushed to enable 90 nm technology nodes, the DOF will shrink to 0.5 μm to allow for 120 nm line widths [5].

CMP enables the use of a reduced DOF system by reducing the topography height across the chip. In the DRAM example of Fig. 10.1, it is apparent that as the capacitor stack grows more vertical to accommodate a shrinking base, the step height induced at the edge of the array increases to greater than the DOF. Subsequent lithography process steps require making source/drain and gate contacts prior to metallization. As Fig. 10.2 depicts, without CMP the memory array contacts are out of focus with the peripheral contacts. Out of focus printing would result in underexposure of the resist. This leads to out-of-specification contact CDs or incomplete contact patterning that would block the subsequent etches. By reducing the topography, the array and peripheral contacts are within the DOF resulting in similar sized contacts or line widths across the die, as shown in Fig. 10.3.

Degree of Planarization

The basic purpose of oxide CMP is to planarize the dielectric over the underlying metal. By planarizing the dielectric, CMP improves interconnect reliability and yield [6]. The definition of planarization can be broken into three components, local, global and step height [3], which are described in Fig. 10.4. Local globalization can be considered on the same dimensions as

| 0 μm - planar | 0.8μm step | 1.9μm step | 3.0μm step |
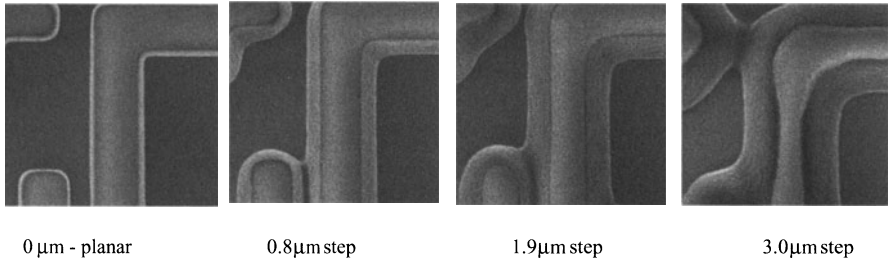
Fig. 10.3. Effect of step height on array, 0 μm, to periphery, 0.8 to 3.0 μm steps in photo resist CDs. The array is in focus while the identical structures in the periphery are out of focus. The pictures are shot using I-line steppers

the line widths or cell-to-cell structure in a DRAM. Global planarization can be considered on the dimensions of the full die or across a bank of devices such as the dome formed over the full memory array. Step height is similar to local planarization however it differs in that is a measured at the edge of a device density transition such as the edge of an array to the periphery. It is a resultant defect due to incomplete CMP or insufficient deposition. Based on a statistical approach, within die planarity variation is greater than wafer scale planarity variation [7] and any reported planarity data should dis-

Local planarization - due to conformity of deposited dielectric

Global planarization - due to pattern density changes across the device

Step Height - due to topography change at transition of high to low density patterns
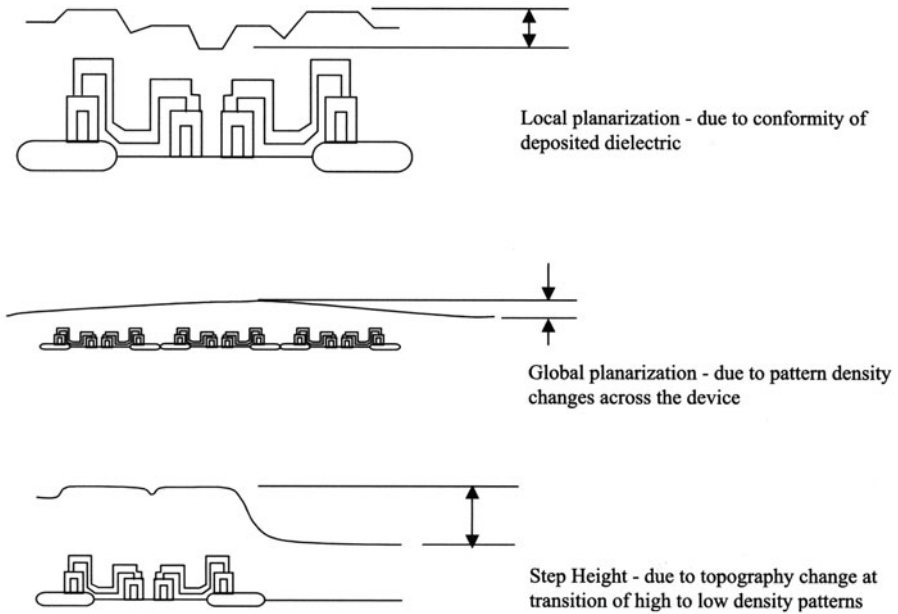
Fig. 10.4. Definitions of topography components, local, global and step height planarization

tinguish between these three components. As shown in Fig. 10.5, the total planarity of a flat surface with a small step height is identical to a domed surface with no step height, however the integration consequences are very different.

Oxide CMP is not necessary for local planarization. Processes such as SOG/etchback, HDP deposition and BPSG/reflow are adequate in reducing the small-scale topography within high-density patterns. Local planarization could be considered a by-product of the need for global planarization. One aspect of local planarization that is benefited specifically by oxide CMP is in the elimination of ILD voids, or bubbles, formed during deposition and reflow of doped oxides or from conformal PETEOS deposition processes [3]. In dense patterns, ILD deposition results in seams or bubbles in the dielectric film, as pictured in Fig. 10.6. In doped oxides, a resultant high temperature reflow can move these bubbles higher in the film or remove them completely. When high temperatures are not allowed due to metallization, initial depositions are followed by an Ar etchback to reduce the overhanging profile that leads to seams. A subsequent deposition is performed to fill the remaining gap. The result of both of these processes is an excess of ILD over the dense arrays. This excess film is polished back to remove the seam/bubble, which during subsequent metallization processes could lead to shorts caused by residual metal in the local topography of the seam.

Step height is a distinct subset of local planarization. It marks the transition between high and low density patterns within the die. Examples of high density patterns are a memory array in a DRAM or SRAM or a large MIM capacitor in an analog device. Examples of low density patterns are isolated NPN emitters in amplifiers or small test structures in the scribe lines. The size and density of the underlying devices define the initial step height. As discussed above, the actual step height may vary due to excess dielectric deposition on high pattern density structures. The initial CMP integration concern is the actual deposition thickness of the dielectric material. Sufficient dielectric must be deposited to result in planarization of the initial step height. A rough estimate for appropriate deposition thickness for complete
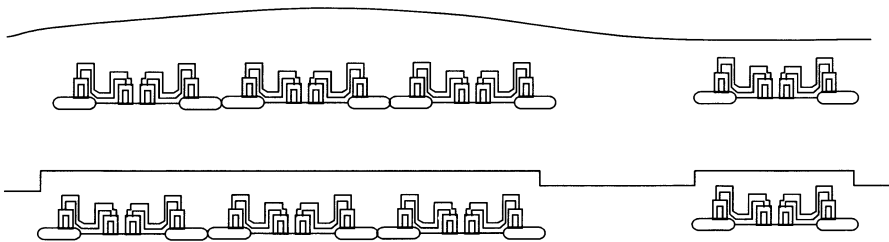
Fig. 10.5. Example of equivalent absolute topography in an IC device. Subsequent etch processes for the two devices would require very different integration schemes for the etch stop layer as the domed structure has more variability in the etch length
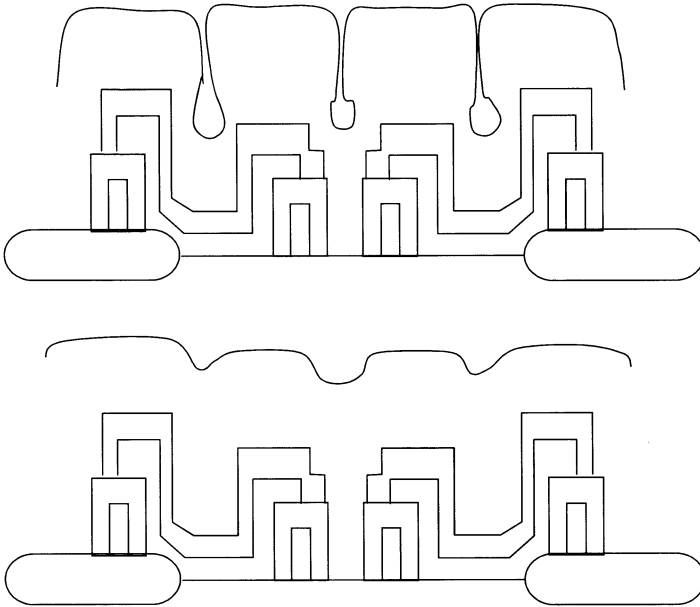
Fig. 10.6. Deposition of doped oxides creates voids in the ILD layer as the CD shrinks. High temperature anneals reflow doped-oxides to fill these voids but are insufficient to planarize the ILD over the topography. ILD layers in the backend are typically not reflowed due to the low temperature limitation of Al metallization

planarization of the step height during dielectric CMP is 1.5 times the initial step height in addition to the target thickness over the substrate [8], as drawn in Fig. 10.7.

Incomplete planarization of the step has several consequences. Metallization following oxide CMP is typically RIE etched to form metal lines. Topography in the blanket metal layers affects the RIE etch by leaving residual metal along the edges of the step, as seen in Fig. 10.8, known as stringers. Stringers are a result of the anisotropic RIE etch which is timed for minimal etch into the underlying oxide [9]. The stringers cause shorts between adjacent metal lines. They can be removed by increasing the RIE etch into the underlying oxide, however this increases the aspect ratio of the next oxide deposition. Higher aspect ratios are difficult to fill adequately and increase the probability of bubbles and seams. In addition to affecting the subsequent aspect ratios the increased etch requires increased thickness between metal layers and lower device layers as the over-etch could damage the underlying material. This increased thickness between layers increases the aspect ratio for subsequent via and contact etch and metallization resulting in a loss in interconnect reliability.

The most beneficial aspect of oxide CMP is measured by the improvement in global planarity. Apart from the DOF discussion above, an improvement
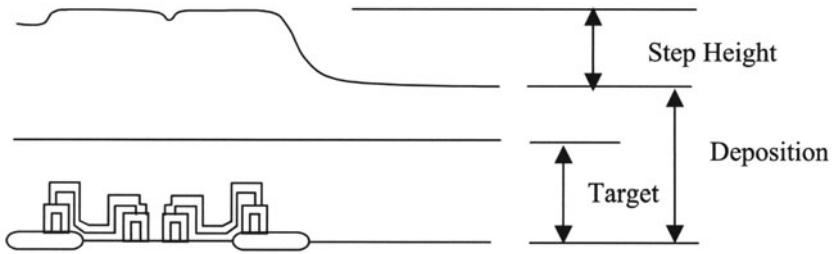
Deposition = Target + 1.5 x Step Height

Fig. 10.7. A rule of thumb for ILD deposition thickness is to add 1.5 times the step height to the final target thickness. The target is measured from the bottom of the topography not the thickness above the feature
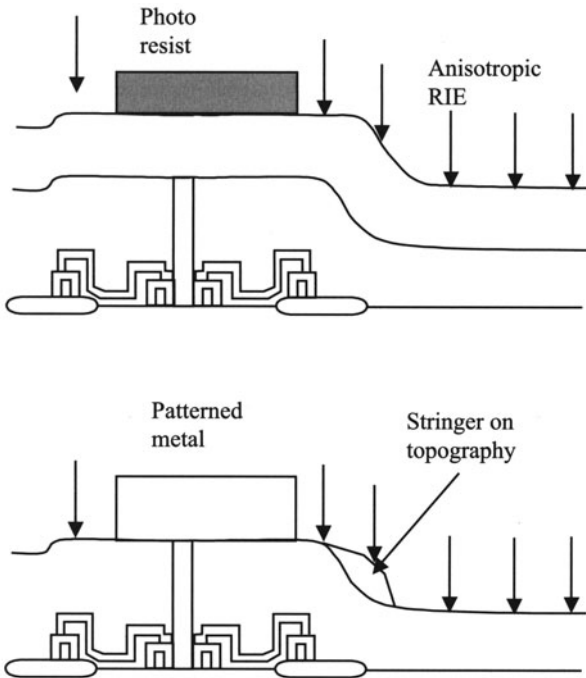
Fig. 10.8. Residual topography during RIE of metal lines leads to the formation of metal stringers that can short adjacent lines. The anisotropy of the RIE process, required for tight CD control, etches only in the vertical direction. Any topography produces an increased thickness in the metal at the step due to the conformal deposition of the metal layer

in the planarity of a die benefits subsequent etching of contacts and vias. One mechanism of improved etch is the reduction in resist thinning, as pictured in Fig. 10.9. When resist is spun onto the wafer, it self-planarizes over the topography. A long etch through the topography to the substrate requires more resist than the same etch in the periphery. This requires excess resist to be spun onto the wafer. CMP reduces the etch difference between array and periphery resulting in a reduction in resist usage.

Figure 10.10 depicts one risk of oxide CMP, the planarization of photolithography alignment marks. Steppers use alignment marks to coordinate or align the different mask levels that make up the device. During metallization patterning, the alignment marks are not visible. Instead, steppers align by the diffraction caused by HeNe laser light scattered from the topography of the top of the metal over the marks. A very efficient CMP process can remove the underlying topography and essentially blind the steppers. Typical robust marks are very long wafer scale features with extreme topography etched into the substrate so that the most aggressive CMP processes find them difficult to planarize. However as more and more oxide layers are added and as substrates are thinned, such as SOI, the initial topography of these marks is reduced. Fortuitously, Cu damascene integration, which is generally replacing AlCu/RIE metallization, does not require topography dependent alignment as the deposited metal is polished off before the next lithography step.

After an oxide CMP step, the global planarity is essentially the resultant ILD dome over large dimensional high-density features. There have been many studies of pattern density affects, both experimental [10, 11] and numerical [12, 13, 14, 15, 16, 17] as well as suggested process and integration methods to reduce pattern sensitivity. Process improvements are accomplished by appropriate experiments of the CMP process variables and improved con-

Minimum resist thickness for etch of long contact.

Excess resist for short contact.
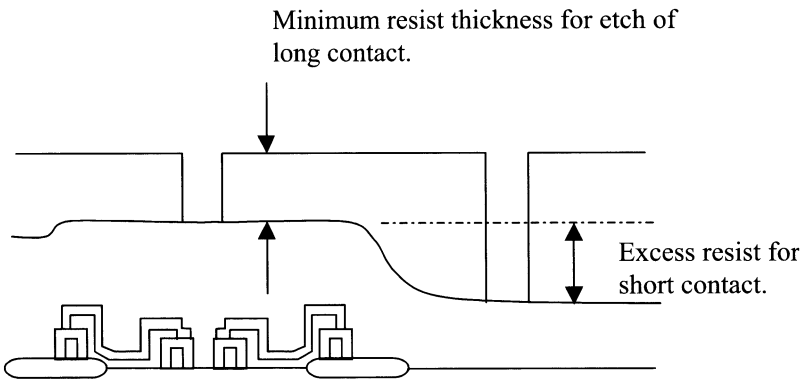
Fig. 10.9. The minimum resist thickness is determined at the top of the topography where etches are the longest. Resist in the low topography areas is excessive for the shorter etches
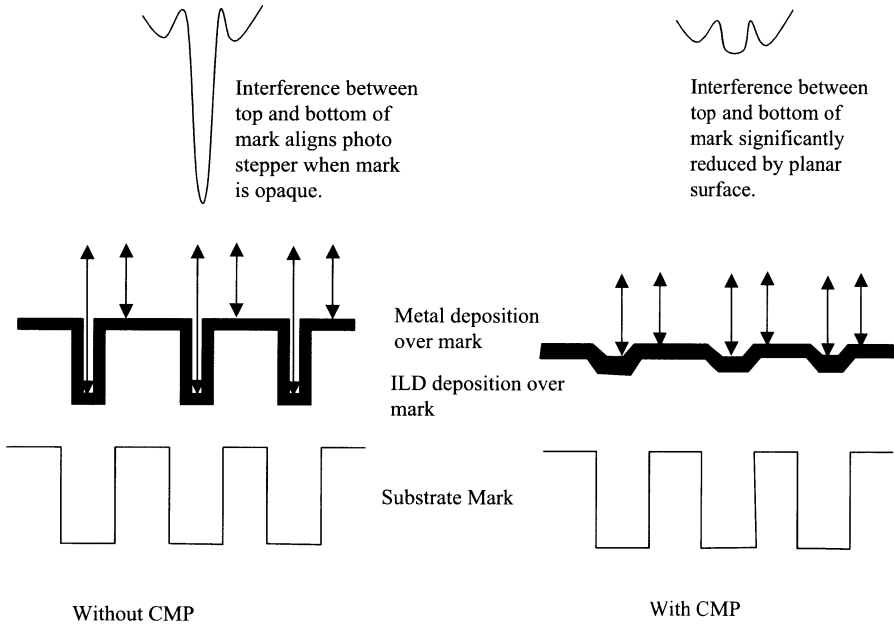
Fig. 10.10. Oxide CMP can blind the steppers to the alignment marks used in photolithography of the metallization layers. When the alignment marks are not visible through opaque layers, the steppers rely on the topography of the mark to align the metal layer to the underlying contact or via layer

sumables. Platen speeds, down force and carrier speeds have all been linked to improved planarity [11] and can be empirically optimized using common test masks. Changes in consumables that affect pad compression or the fluid dynamics also affect the global planarity [10, 14, 15, 18] although they are less well understood. The most common integration method for reducing global planarity is to incorporate dummy patterns into the periphery [3, 19]. Dummy patterns attempt to mimic the high-density patterns of the array devices but are not electrically active. However the complexity of the lithography process and the electrical concerns of a free-floating metal plate and how it can lead to MOSFET current mismatch [19] or capacitive coupling between the dummy pattern and the passive or active devices may lead to limitations on the dummy fill approach.

Uniformity

As previously mentioned, oxide CMP is unique as it does not involve a material interface, such as oxide/nitride or oxide/metal, to act as a stopping layer in the polish. Stopping layers can be used to improve CMP uniformity by using a consumable set designed to take advantage of the different removal rates of the two interfacial components. This lack of an interface makes

the ultimate uniformity of the wafer dependent solely on the removal rate control of the CMP process, tool and consumables combination. Incoming non-uniformity of the deposited material contributes to the post CMP thickness non-uniformity, as shown in Fig. 10.11 [20, 21]. Some typical deposition equipment non-uniformities averaged ~3% [20] and post CMP blanket wafers averaged ~5% [20, 21].

There are two components that determine oxide CMP uniformity: removal rate uniformity and film thickness uniformity. Most of the literature published on oxide CMP discusses removal rate uniformity [22, 23, 24, 25, 26] and is discussed in detail in another chapter. Initial data suggests that most non-uniformity is within the die and is pattern dependent [7]. However as wafer size moves from 150 mm to 200 mm to 300 mm in diameter, edge effects will also become a larger percentage of the non-uniformity [27]. For integration purposes, what is most relevant is the final film thickness uniformity, which is a combination of removal rate non-uniformity from the CMP process as well as residual non-uniformity of previous processes. It is this parameter that is ultimately integrated into the IC device structure. In fact it is possible to manipulate the CMP removal rate uniformity to compensate for non-uniformity in prior processing or for a known non-uniformity in a subsequent process. An example of this type of integration is controlling the center-to-edge capability of CMP to compensate for a measured center-fast, or center-slow, dry etch process. By using CMP in this manner, over-etch values can
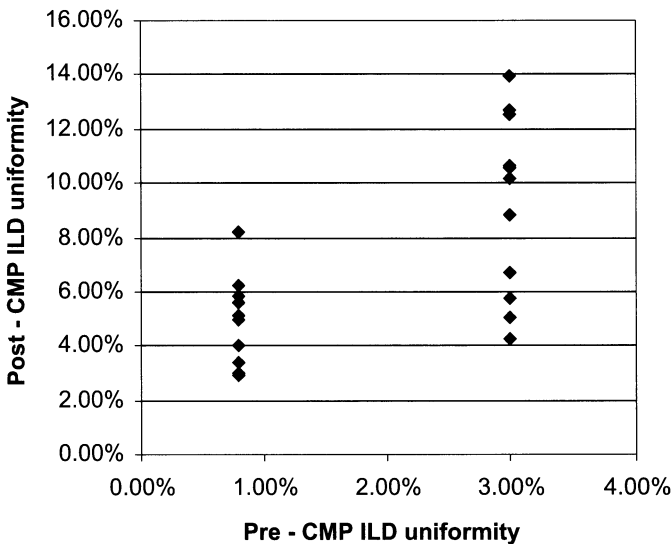
Fig. 10.11. The non-uniformity of two different oxide deposition processes affects the post oxide CMP non-uniformity. A typical process deposition of 3% 1-sigma leads to a range of post CMP non-uniformity of 4–14% 1-sigma. Improving the deposition to better than 1% 1-sigma improves the post CMP non-uniformity. From [20]
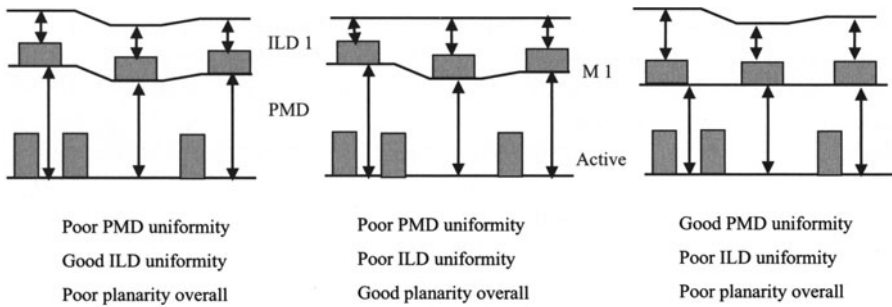
Fig. 10.12. Non-uniformity in the underlying substrates do not always imply continued non-uniformity in the subsequent polishes because the measurements are made relative to the substrate for PMD and to the metal layer, M1, for ILD

be reduced, which for example can reduce RIE over-etch and the resultant topography generated by that over-etch.

This film thickness variation must be less than the subsequent contact or via etch variation, not only across the wafer, but also, for small scale development, within the die, or for large scale production, within the entire product line. Monitoring of the CMP process, either during development or in production, is similar to all other processes that use a statistically determined parameter, C_{pk}. This is calculated by

$$C_{pk} = C_p(1 - (|x - \text{TGT}|/|\text{USL} - \text{LSL}|/2)), \qquad (10.2)$$

where

$$C_p = |\text{USL} - \text{LSL}|/6s. \qquad (10.3)$$

Here, LSL and USL are the lower and upper specification limits, TGT is the target value, s is the sample standard deviation and x is the sample mean value. For a well controlled manufacturing process a $C_{pk} > 1.5$ is desired. The value of x is dependent upon the oxide layer being polished. For example a PMD film is measured to the active silicon level while ILD is measured between metal layers, as seen in Fig. 10.12. It is apparent from Fig. 10.12 that the non-uniformity of the PMD level does not imply non-uniformity in subsequent ILD layers since the measurement locations are not identical. Uniformity is relevant to the uniformity of the underlying topography. However, from a process control point of view, it is best to be as uniform as possible at all levels.

Wafer/Film Properties

As discussed in general, oxide CMP has been shown to provide a platform for improvements in etch, lithography and metallization. However there are process considerations prior to oxide CMP that can have a profound effect on

improving the CMP process. Wafer profile and film homogeneity in particular stand out as having the most effect on CMP. Wafer profile is defined as the macroscopic warp or bow to the wafer disk. Film homogeneity assumes that the film properties, such as stoichiometry, dopant concentrations, moisture content and stress are consistent across the wafer surface.

The IC device manufacturing process includes deposition and removal of many different film types. Each of these films introduces new stress on the wafer substrates and results in various degrees of warpage in the macroscopic shape. As an example, silicon nitride can be deposited in a furnace near 900°C, a step known as thermal nitride, or near 300°C, known as PECVD [28]. Thermal nitride deposition is on both front and backside of the wafer. PECVD films deposit only on the front side of the wafer. Thermal nitride is tensile in nature. PECVD is compressive in nature although this can vary with deposition RF power. Since silicon nitride is used frequently in front-end and backend device manufacturing, these stresses can result in different wafer profiles at the different oxide CMP steps. Another example of stress induced CMP non-uniformity is a change in the wafer substrate. With the increase in SOI wafers in IC devices [29], changes in the stress also result in changes in uniformity; some representative results are pictured in Fig. 10.13. CMP tools are designed to partially compensate for these stress changes by process optimization of the tool parameters or wafer carrier profiles [30].
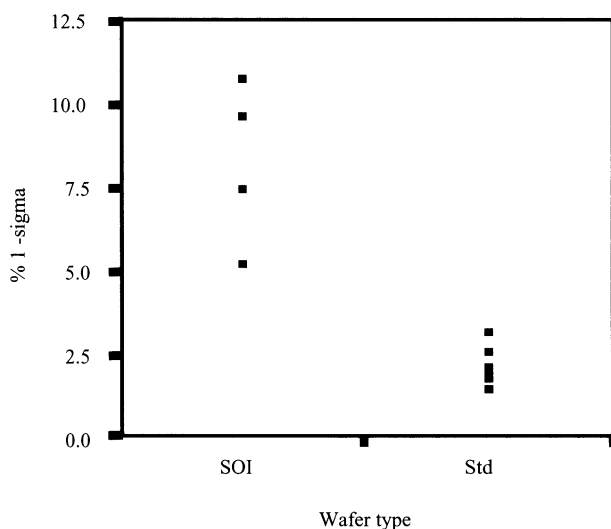
Fig. 10.13. Comparison of standard Si crystalline wafers and SOI wafers on the uniformity of oxide CMP. The non-uniformity is a result of both non-uniformity in the substrate surface and SOI induced stresses during CMP

Process steps prior to CMP can also be used to improve CMP. Figure 10.14 compares data from an oxide CMP polish with and without a wafer backside etch prior to CMP for a specific process sequence. It is apparent that relieving the stress on the wafer caused by the residual film on the back of the wafer improved the CMP uniformity. In addition to oxide CMP uniformity, removal rates are also stress dependent. In general, a tensile film results in higher removal rates and higher non-uniformity [31, 32]. The mechanism of increased removal rate with stress is not clearly understood although local film dislocations and fluid dynamics are suspected to play a role [33].

A larger concern for the integration of these film stresses is the consistency within the wafer. In particular, doped oxide films can have a large affect on CMP removal rates [20, 21, 34]. Not only do doped oxide films have higher rates than both PETEOS and thermal oxide films but subtle variations in dopant concentrations across the wafer can result in increased non-uniformity. Table 10.1 shows the removal rate increase for the common dopants of B, P and F used in PMD and early low-k dielectrics. It is apparent from Table 10.1 that consistency of the dopants, particularly P, plays a role in reducing non-uniformity across the wafer but also is important in improving wafer-to-wafer and lot-to-lot consistency. Without an understanding of the variation of the incoming film, each individual wafer lot needs to be treated as unique material. Continual film deposition variations, due to deposition equipment maintenance, substrate vendor changes or process integration changes, is one of the variables, along with consumable inconsistencies, that result in the need
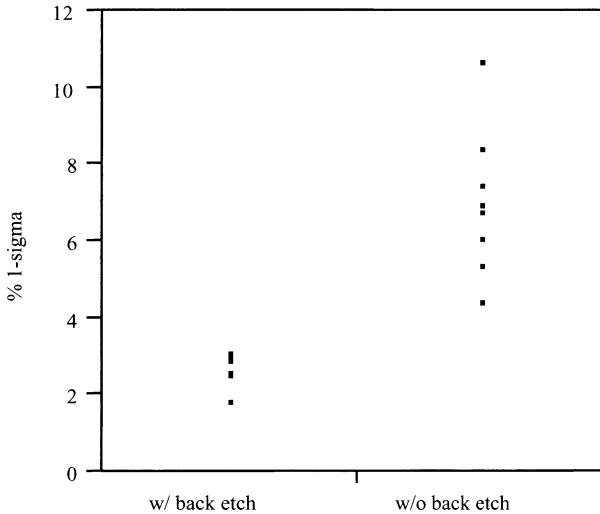
Fig. 10.14. Comparison of the non-uniformity post ILD CMP on wafers that previously had the backsides cleared of deposited high-stress films. Removing the induced warpage caused by the film stress reduces the non-uniformity created during CMP

Table 10.1. Comparisons of the effect of dopant concentration on normalized oxide CMP removal rate. P content has the largest impact on removal rate and is often limited to less than 6%. A small 0.1% change in P content leads to an approximate 5% variation in CMP rate

Dopant	Linear Component $(1 \; / \; \text{at} \; \%)$	Quadratic Component $(1 \; / \; \text{at} \; \%^2)$
F	0.054	–
B	0.0055	0.0017
P	0.0784	0.0021
B+P	–	0.0016

for individual lot targeting of the uniformity and removal rate. Targeting is a time consuming process of running lead wafers through the polish and measurement steps prior to committing the entire lot. This procedure results in a reduction in throughput of the oxide CMP process.

Defects

As with any process in IC manufacturing, oxide CMP creates defects that negatively impact yield. The most common defect associated with oxide CMP is scratching [21]. Whatever the source of scratching during CMP, such as the pad/wafer contact, particle/wafer contact or wafer handling error, the result is a crack in the dielectric surface that can lead to several failure mechanisms. The severity of the failure is dependent upon the dimensions of the scratch while the defect density affects the yield [35, 36]. A wide scratch can cause residual metal fill to short between two adjacent lines, as pictured in Fig. 10.15. Some circuit designs use redundancy in the layout to prevent a single defect from killing an entire line. However a single long scratch, on the order of several devices, can cross enough lines to negate the planned redundancy. A deep scratch provides a path for chemical cleans and contaminants to effect the underlying devices as shown in Fig. 10.15. Buff CMP processes are used to reduce scratch counts during the post CMP inspection, but do not necessarily remove the scratch. A buff process usually consists of a low stress polish on a soft conformal pad timed to remove a small amount of dielectric material and results in removing the sharp edges associated with a scratch. Buffing the scratches may improve the yield loss due to the length and width of the scratch, however they can do nothing to alleviate the loss due to the depth of the scratch. The buff process should remove sufficient material to reduce the scratch to a non-yield impacting defect, but not enough to become a significant main contributor to post CMP non-uniformity. Even with the use of a buff step, the best methodology to remove the scratch defect is to eliminate the source of the scratching from the bulk CMP process and consumables [37].
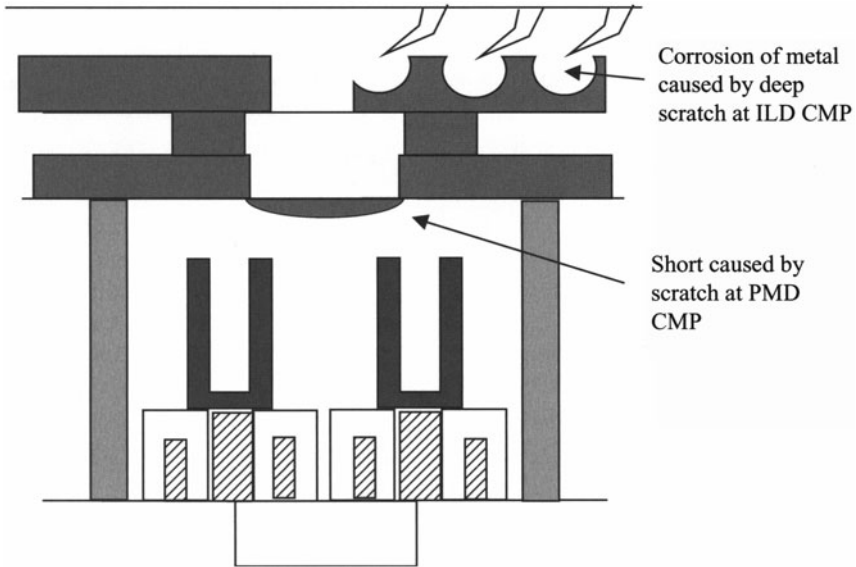
Fig. 10.15. Depiction of two electrical faults caused by CMP scratches at PMD and ILD. A PMD scratch has left residual metal, either from W CMP or Al RIE etch, that shorts two distinct DRAM cells. A deep ILD scratch has allowed cleans chemicals to corrode the underlying Al lines

Another typical source of CMP defects is residual contaminants left on the wafer surface such as particles and chemical contaminants. Post CMP cleanings are performed to remove these contaminants. The details of the cleaning processes are discussed in another chapter. Particles left on the wafer surface potentially can either be encapsulated by subsequent metal, photo resist or dielectric depositions [38] or can block subsequent etch processes. In-line metrology is often used to detect particles before the wafers are further processed when re-cleaning the surface is still possible. Redistribution of particles from the backside or the edge of the wafer, where in-line metrology is not effective, can also occur and be visible several process steps after the CMP process. The particle source is then often attributed to the wrong process step.

Chemical contamination post oxide CMP can potentially be more serious, especially considering the wet nature of the process and high mobility of the critical contaminants, primarily ions and metals. Gate oxide integrity is very susceptible to trace contaminants of mobile ions such as Na, K, Cl and Ca [39]. The presence of such contaminants can be source of much anxiety to any IC manufacturer. Most wet chemicals, including DI water, used in IC manufacturing have ppb level specifications of such impurities to protect the gate oxide. CMP slurries are no different. Although there is very little evidence that mobile ion content in oxide CMP processing actually affects

the gate oxide, the potential for cross-contamination requires the suppliers to continue to reduce the presence of these ions in the slurry. In particular as oxide CMP moves from the ILD to PMD level, closer to the gates and the high temperature anneals associated with active devices, residual ion contamination becomes more critical to yield and reliability. In addition to the consumable purity, metal cross-contamination is a concern due to sharing of CMP tools between ILD, PMD and STI. Although it is obvious that a Cu tool should not be used for STI polishing, sharing of an ILD tool can also lead to metals contamination. It is less of an issue of ILD over-polish exposing the underlying AlCu lines, but of backside and bevel residual metal contaminating both the pad and the wafer carrier, similar to residual particles as previously discussed. One integration solution to mobile ion contamination is the use of diffusion barriers between the gates and the CMP surfaces. Diffusion barriers are usually associated with Cu diffusion in advanced backend integration, however silicon nitride has been used as a diffusion barrier to ion mobility in AlCu backend processes. The silicon nitride layer, which is more often justified as an etch stop in high aspect ratio contacts, has the additional benefit of reducing ion mobility to the gates [40]. Similar to the scratching issue, the optimal response to mobile ions is to eliminate the source.

10.1.3 PMD CMP

Pre-metal dielectric CMP is a process that defines the interface between the frontend active devices and the backend metallization. As devices shrink, the need for a dielectric material to fill the narrow gaps between active devices has led to the wide use of doped oxides, PSG and BPSG, as the dielectrics of choice. PSG has been used widely prior to the advent of PMD CMP since P acts as a gettering agent for mobile alkali ions [21]. By addition of more dopants, the dielectric can be reflowed in a furnace anneal or RTA chamber at high temperature, greater than $850°C$ [20, 41]. The reflow accomplishes two goals: the first is to reflow the BPSG into the narrow gaps [42] and smooth out the local topography, and the second is to stabilize the BPSG film. As deposited BPSG films , in particular, absorb moisture from the atmosphere leading to crystallization of the dopants in the film. Longer anneals at high temperature result in smoother post-anneal topography. However, due to diffusion of the junction dopants at high temperature, long anneal times are not possible resulting in the need to planarize the remaining topography. Although not all process generations need the reflow capability of doped oxides, the need for planarization of the PMD remains.

Integration of PMD CMP enables new processes and concerns in addition to the general oxide issues already covered. The most obvious process enabled by PMD CMP is the use of W CMP for contact formation. W CMP requires a planarized oxide surface on which to stop otherwise the W CMP would require a severe over-polish to reduce the residual stringers formed around the topography, as pictured in Fig. 10.16a. Analogous to the W CMP process,

PMD enables Cu CMP of the first metal layer by providing a flat surface on which the Cu CMP stops. It also allows for uniform trench etches for the damascene process, as depicted in Fig. 10.16b.

In many integration schemes, CMP scratches created at PMD CMP are often coated with an additional dielectric layer of harder oxides, like TEOS or HDP silane. Reduction of defects at the PMD layer is important to minimize shorts between isolated W contacts or Cu lines. For W CMP, this is also advantageous as the oxide provides a harder layer on which to stop the W barrier polish. A harder layer reduces oxide erosion, caused by the low density of oxide in an array of W contacts, during W CMP. For Cu CMP, the harder layer is required, as it becomes the metallization dielectric in which the Cu is deposited as well as the layer on which the barrier polish stops. In addition to the improvements in W and Cu CMP, these layers also encapsulate the doped oxide and minimize the diffusion of moisture into the film. Along with the BPSG stabilization, minimal moisture content is desired to prevent oxidation of the W contact liners, which are Ti based.

In DRAM devices, PMD is not limited to the defining the first metallization layer, but also is used to define the poly Si contacts between the buried
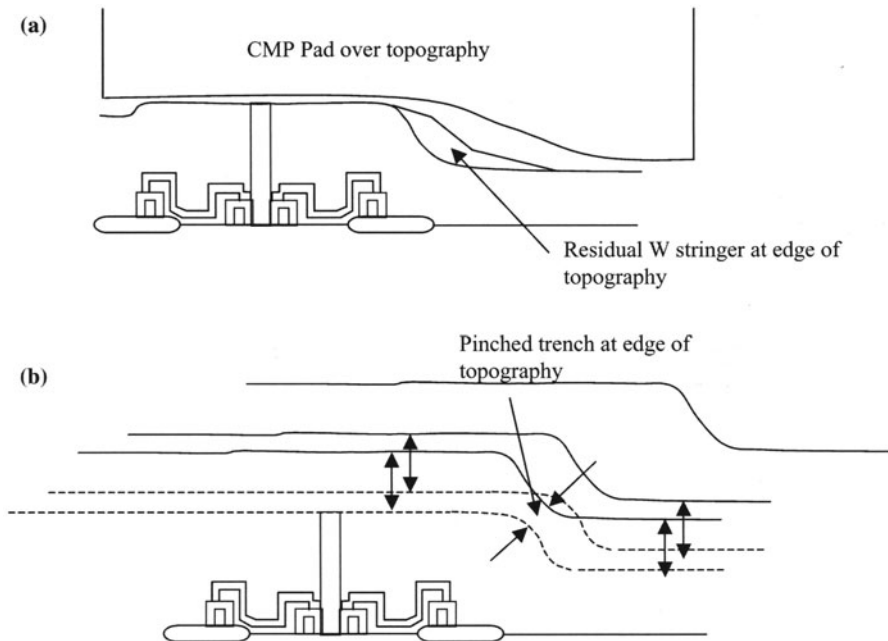
Fig. 10.16. A non-planarized PMD layer causes failure mechanisms in both W CMP (**a**) and Cu trench etch (**b**). The W failure is residual metal, or stringers, at the edges of the topography. The Cu failure is similar to RIE stringers, except that the result is a pinching of the trench at the edge of the topography. The resultant Cu CMP would also leave stringers similar to the W CMP failure

digit line and the source/drain junctions [43, 44]. Figure 10.17 illustrates two methodologies that exist for the poly contact formation: a selective approach analogous to the W contact formation and a non-selective approach in which the poly and BPSG are polished at similar rates [43]. The result is an isolated contact embedded in a planar dielectric on which the digit lines are then patterned. PMD CMP such as this is very defect sensitive since the following step is a poly-Si deposition. Poly-Si is a very conformal CVD process that can fill any voids or cracks left in the dielectric material leading to shorts between the poly-Si contacts. It is apparent that such a PMD polish also defines the uniformity of the vertical capacitors, stud or crowned style [2]. By providing a planarized uniform surface, test capacitors, for example, in the periphery have similar dimensions to those in the array which are built over the transistor lines. This enables test key devices to be more comparable in performance to the array device.

PMD CMP presents several unique challenges to IC manufacturing. The most obvious is the increased dependence on contact etch selectivity. Prior to PMD CMP, all contacts are similar in length. For example, a typical BiCMOS
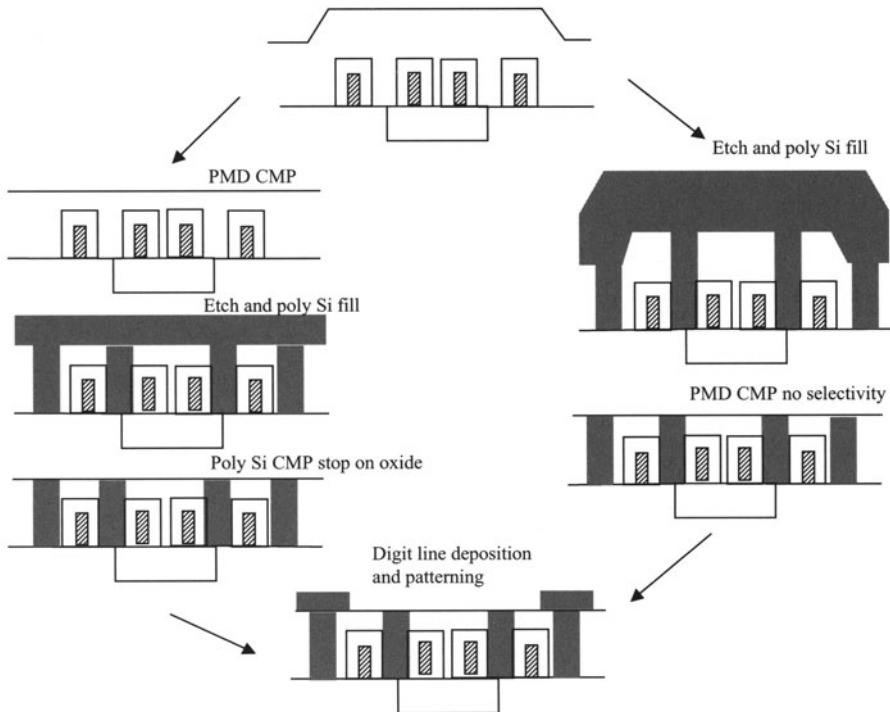
Fig. 10.17. PMD is used not only to planarize prior to metallization but also can be integrated into front-end fabrication. Two integration schemes depict DRAM poly Si contact formation for contact between source/drain and buried digit lines [43, 44]

device has contacts to the NPN emitter, the NPN base, the NPN collector, the CMOS gate and the CMOS source/drain, as shown in Fig. 10.18. Without a planarized PMD surface, all these contacts are similar in length allowing for a timed etch with minimal needs for etch rate selectivity between the dielectric and the contact bottom material. With a planarized PMD surface, this device now has five different contact lengths. A silicon nitride or silicon oxy-nitride layer is deposited on top of the active devices to serve as an etch stop layer to prevent damage to the junctions or the devices during contact etch by erosion of the underlying silicide. However a typical etch stop layer, which is minimized to reduce capacitive coupling between the first metal layer and the active devices, requires a 15–20:1 selectivity between the PMD dielectric etch rate and the stop layer etch rate. Although these selectivities are achievable in advanced etching systems, they rely heavily on the reproducibility of the PMD CMP uniformity, both within die and across wafer and on the strength of an endpoint signal.

PMD is rapidly becoming the driver for improvements in oxide CMP consumables. Improvements in the uniformity, planarity and defectivity of PMD CMP are important to the continued enablement of Cu CMP. However the success of PMD CMP also reduces the need for oxide CMP consumables as the enabled Cu CMP process replaces ILD CMP leaving PMD as the sole remaining stop-in-film oxide CMP process in advanced IC devices. This leaves open the need for advanced oxide consumables such as fixed abrasive, stop-on-planarity slurries and advanced colloidal particles.
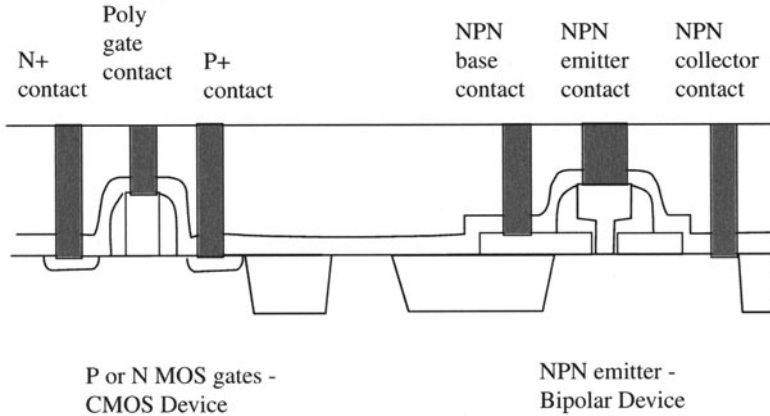
Fig. 10.18. BiCMOS contacts post PMD CMP have multiple lengths depending upon the active device being contacted. This requires the use of a etch-top layer with a high selectivity between the dielectric and the stopping layer to prevent damage to the underlying devices and erosion of the contact silicide

10.1.4 ILD CMP

Interlayer dielectrics are needed to isolate multiple layers of metallization in the backend of an IC device. Various forms of silicon oxide, such as HDP silane, PETEOS and SOG, have been used to provide this isolation. The thickness of the dielectric film is set based on a tradeoff between cross-capacitance coupling [7] and interconnects lengths which are frequently modeled to assist in the layout of the metallization layers [45, 25]. ILD CMP planarizes the dielectric layer deposited between metallization layers. By eliminating the need for metal lines to climb over topography generated by lower levels the yield on the wafer can improve [46]. The mechanism for improved yield may vary by part type. For example, a flat surface reduces the stress migration of the metal lines and the need for an increased etch of the oxide layer to overcome stringer formation [9].

There are many similarities between PMD and ILD, for example the enablement of W CMP for both contact and via formation. However there are several major differences that should be addressed. The first difference is that, post ILD CMP, via etches are of the same length. This makes the via etch process far more robust than the locally planarized backend, which is shown in Fig. 10.19. The uniform via length eliminates the need for nitride etch stop layers which can cause more capacitive coupling due to the high dielectric constant of most etch stop materials, such as silicon nitride. By reducing the need for etch stop layers and subsequent large over etch times in the via etch, a secondary benefit occurs with the reduction in plasma charging damage to the transistor gates. Charging damage occurs by an imbalance of plasma ions bombarding an exposed metal in contact with the gate oxide. The imbalance in the plasma induces a current to flow through the gate in an attempt to balance the charge build up. The result is a stored charge at the gate electrode that can result in premature breakdown of the gate oxide [47]. This is particularly important in backend metallization as the metal area exposed during etch is significantly higher, greater than 1000:1, than the gate oxide area. This large ratio means that the captured charge is multiplied by the 1000:1 ratio and is placed across the thin gate oxide. The resultant high electric field can, with sufficient charge, lead to gate oxide breakdown. Another source of charging is during the low temperature plasma enhanced CVD (PECVD) deposition of silicon nitride. Typically, charge traps formed during silicon nitride and silicon oxide depositions are annealed out with a high-temperature process. In the backend processing, unlike at PMD CMP, temperatures are typically limited to 450°C due to the AlCu metallization melting point. ILD CMP reduces the antenna charging by elimination of both the non-uniformity of the via etch and the need for the nitride etch stop deposition.

The post ILD CMP thickness is important in reducing capacitive coupling between metal layers. Capacitive coupling, or cross-talk, is a result of the proximity of the metal lines separated by a thin dielectric. The determin-
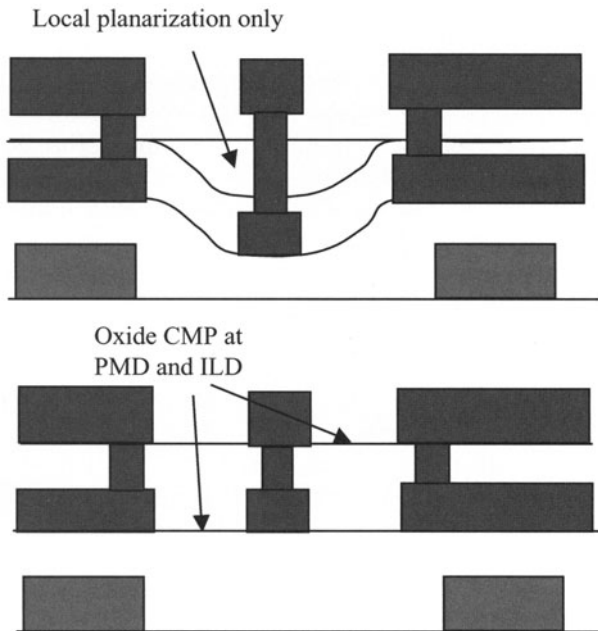
Fig. 10.19. Via lengths can vary due to use of local planarization on features with underlying topography. Depending upon the metal thickness and local planarization process, via lengths can vary more than 50%. ILD CMP makes all vias of equivalent length

ing factors in the magnitude of cross-talk are the signal voltage, separation distance and the dielectric constant of the dielectric. The operating voltage of the part typically fixes signal voltage, around 5V for older generation down to 1.5 V for new generation IC's. Separation distance is determined by the design rules for the generation of the part being manufactured. This leaves the dielectric constant as the remaining variable for reducing cross-talk.

Lowered dielectric constant materials are usually categorized in two groups, low-k and ultra low-k. Low-k groups, usually $k = 3 - 3.5$, are modified silicates, such as F doped TEOS or HDP silane (33). These films are simple replacements of the standard ILD and integrate readily into an AlCu/RIE etch backend metallization. They often result in an increase in CMP removal rate and require modifying the CMP process through appropriate design of experiments. However the standard consumables are generally capable of polishing these films. Ultra low-k films, usually $k < 3$, cover a wide range of organic and organo-silicate polymers, such as hydrogen silsequioxane, poly-imides, poly-quinolines and fluoropolymers [48]. These films are deposited by either CVD [49, 50] or spin-on techniques [48]. The CMP performance of these films is more frequently related to the films properties during Cu CMP as the ultra low-k film is the stopping layer for the polish [50]. Di-

rect replacement of the standard ILD with an ultra low-k film is also feasible, however as the films are no longer Si-O based, new consumables are required. In particular issues of different surface hydrophobicity, different iso-electric points and new chemical bonds varied the CMP removal rate [48]. In one study, ultra low-k films were polished only after a complete change in the slurry content [48]. In this process, zirconium oxide particles replaced the traditional silicon dioxide particles in the slurry, the pH was dropped to 4.6 from the traditional 10.5 range and the surface hydrophobicity ranged from high for methyl silsesquioxane to moderate for hydrogen silsesquioxane.

10.2 Tungsten CMP

10.2.1 Introduction

Tungsten CMP was the first metal CMP process to be implemented in IC manufacturing. It is a departure from the pre-existing oxide CMP process, not only technically but also in the manner in which CMP is perceived. Oxide CMP was seen as a "necessary evil" to enable continued advances in

Contacts:	Vias:
Variable lenght	Equivalent lenght
high aspect ratio	low - medium aspect ratio
Straight Wall	Tapered Wall
Stop on poly Si or cryst. Si silicides	Stop on metal/ARC stack
Small CD	Lager CD
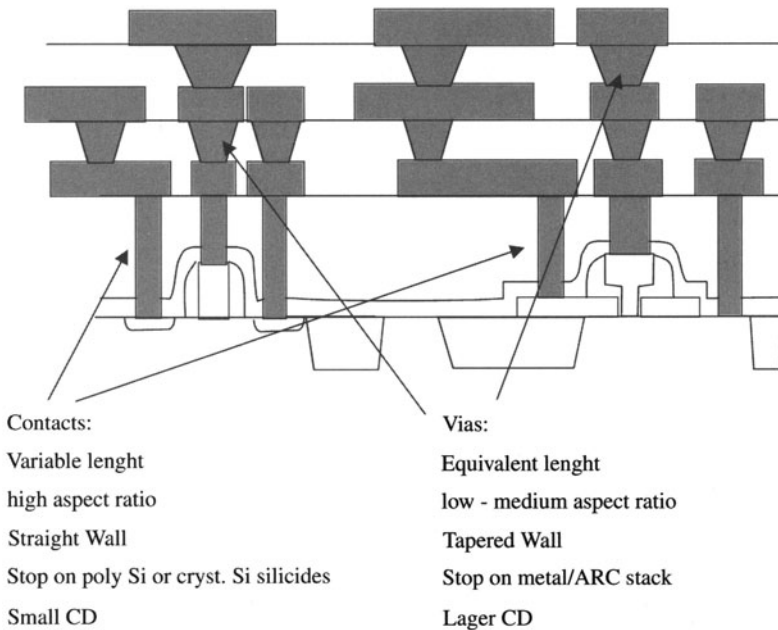
Fig. 10.20. Comparison of contacts versus vias. Contacts in typical IC devices are high aspect ratio interconnects between the front end active devices and the metallization layers. Vias in typical IC devices are lower aspect ratio interconnects between individual metal layers

photolithography and etch. If some other form of planarization was available oxide CMP may never have been implemented. Tungsten CMP, on the other hand, is a direct replacement of a pre-existing process, tungsten etchback that had reached the end of its viability in IC manufacturing. Tungsten CMP provided a reduction in etch-back particle formation and tungsten recess in the interconnect plug [51]. Tungsten CMP shifted the perception of the CMP process from "necessary evil" in the backend to that of a companion process for structure formation in the backend. The structures created by tungsten CMP are the metal interconnects.

Metal interconnects are divided into two categories, contacts, which connect the front-end device to the backend metallization, and vias, which connect metal layers. The purpose of W CMP regardless of the type of interconnect is identical, to leave electrically isolated plugs co-planar with the surrounding dielectric on which subsequent metal layers are deposited. The majority of the difference between the two types of interconnects involve the amount of W to be removed, the dimensions of W plugs and the film on which the W CMP stops, as shown in Fig. 10.20.

10.2.2 W Integration

Tungsten as interconnect metal replaced low-temperature PVD Al, which exhibited poor step coverage in deep contacts [51]. Attempts at improving the Al step coverage where partially successful, in particular high pressure force-fill [52]. Al plugs do have lower contact/via resistance than W, particularly as the dimensions of the contact are reduced, as shown by the data of Fig. 10.21. However as aspect ratio continued to increase, W deposition provided superior fill of the contact than Al deposition processes due to the interaction of Al with the Ti/TiN barrier forming high resistance alloys [53]. Table 10.2 lists the advantages and disadvantages of Al, W and Cu as interconnect metals. Although tungsten has many positive properties it is not a viable metal for wiring due its relatively high electrical resistance in comparison to Al or Cu. However, due to its deposition properties and reliability, it makes for an excellent local interconnect material, surpassed only by Cu dual damascene.

Tungsten contacts and vias are filled by deposition of a Ti/TiN barrier followed by WF_6 CVD. The Ti/TiN barrier is multipurpose. The TiN acts as a nucleation layer for the W-CVD reaction, however TiN has poor contact resistance. The Ti forms good contact resistance but reacts readily with WF_6. Therefore a Ti/TiN combined liner is used in the contact and vias. There are many references to the relative merits of the Ti/TiN deposition process [54], such as IMP Ti vs. collimated Ti [55], or CVD Ti/TiN [51] combination tools. However formed, the requirements of the barrier are high step coverage in the contact or via, no pinholes in the TiN film and low stress. Pinholes and stresses in the TiN film can lead to large defects due to the rapid growth of tungsten on the exposed Ti underlayer, called volcanoes, that are dislodged

Al vs. W plug resistance
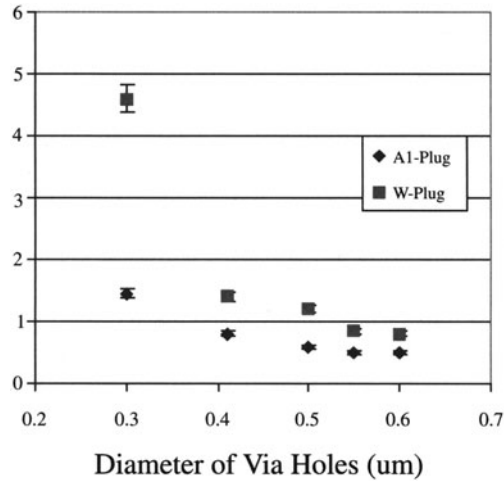
Fig. 10.21. Comparison of Al and W plug resistance. Al plugs typically have much lower resistance for a filled plug. From [52]

Table 10.2. A list of properties compares typical IC materials in the metallization layer. From [51]

Material	Resistivity ($\mu\Omega$ – cm)	Advantages	Disadvantages
Aluminum	$\sim$3	Universal Process PVD: High throughput Cheap process CVD: Good contact/via filling	Poor step-coverage
Tungsten	$\sim$10	High EM/SM resistance High temperature stability Excellent contact filling	High resistivity Volcano formation from WF_6/Ti
Copper	$\sim$2	Low resisitivity High EM/SM resistance	High diffusivity into Si/SiO_2 High corrosion susceptibility Difficulty of dry etch

during the CMP process. Since both Ti and TiN are high resistance material, it is desirable to keep these liners thin compared to the W to reduce the contact resistance. The appropriate thickness and chemical composition of the Ti/TiN barrier, which must be removed by the W CMP process [56], must account for the conflicting requirements of step coverage and defect density versus contact resistance.

Once the barrier has been deposited, the contact/via is filled with tungsten typically by means of a CVD process. The amount of tungsten deposited is balanced between sufficient fill of the contact/via hole and the need to remove the overburden by CMP in a reasonable time. In most advanced IC manufacturing, the contact/via layer is restricted to a single size or critical dimension, CD. There are several reasons for this but one benefit of this design rule is a uniform CVD W fill across the wafer [57]. Allowing variable contact/via size would produce excess overburden in smaller contacts/vias, in order to fill the largest structures, as is pictured in Fig. 10.22. In addition to contact/via size, the shape also is important for the fill of the structure. Figure 10.23 pictures how the contact/via etch process may produce straight walls, tapered walls and bowed walls, depending upon the ILD stack being etched [51]. For example a TEOS ILD that caps a doped oxide layer can lead to bowing of the contact if the dry etch being used has a higher etch rate in doped oxide than TEOS. Filling these holes is important to avoid voids and "keyholes" in the contact/via which lead to electrically resistive interconnects and electrical failures. Voids and "keyholes" are particularly susceptible to the oxidizing environment of W CMP and can provide a direct path for contaminants in the W slurry to the active devices underneath.

Insufficient W-Fill due to large CD via/contact

Small CDs are filled

Sufficient W-Fill due of large CD via/contact

Small CDs have huge W overfill for CMP removal
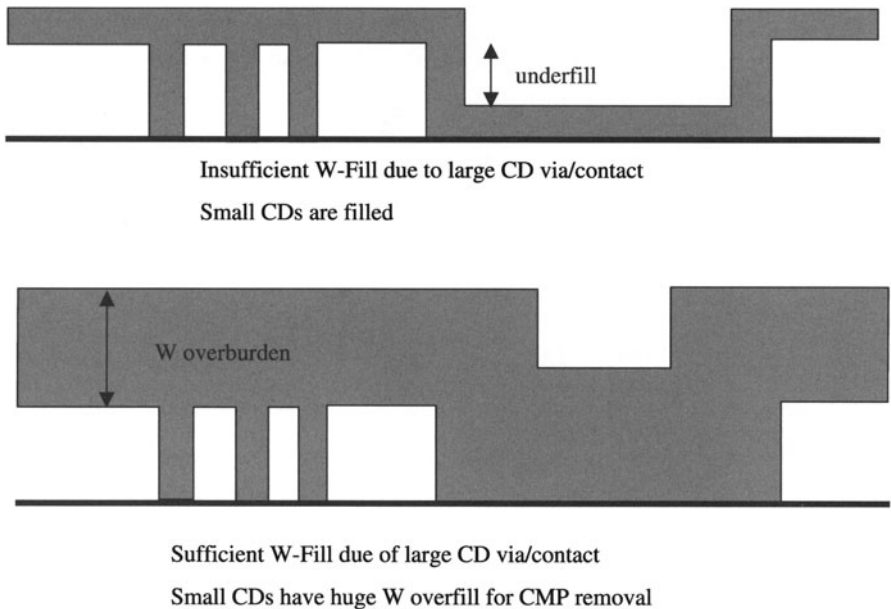
Fig. 10.22. Demonstration of underfill and overburden caused by variable CD contacts and vias
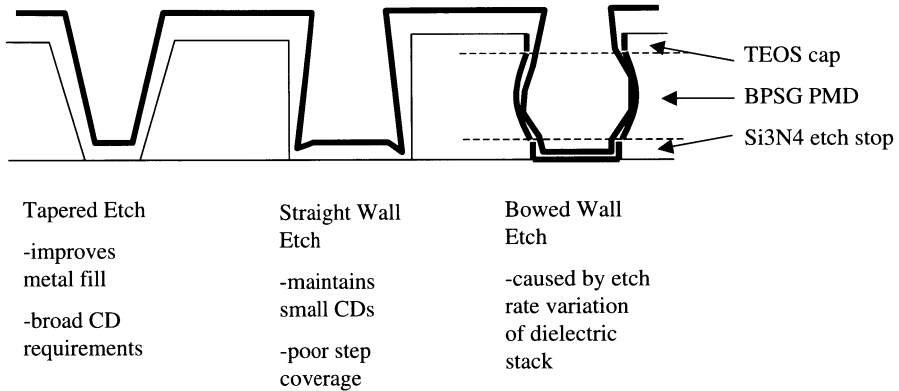
Fig. 10.23. Comparison of different via/contact etch profiles. The metal liner fill and CVD W fill is very dependent upon the profile and can lead to holes in the liner or voids in the plug

10.2.3 CMP Integration

The mechanism of W CMP is described in other chapters. Although the details of the process are not clear, the general needs of an oxidizing agent, abrasive particle and applied stress are understood. The W overburden is polished off by successive oxidation of exposed W metal and abrasion [54] of that oxidized layer. The removal rate of the W material is determined by the many factors similar to oxide CMP, such as relative velocity and applied pressure [58], in addition to the strength of the oxidizing agent [56, 59, 60, 61, 62]. There is usually no excessive topography associated with the W material as the pre-ILD CMP has planarized the surface on which the W material is deposited, with the exception of large openings that may be incorporated in test structures and heat conduction pipes. One advantage to this lack of topography is the removal rate that is calculated from blanket test wafers is applicable to determining the overall time needed to clear the W overburden on the product wafer.

Unlike ILD CMP that involves planarization and blanket removal with stopping in the film, W CMP is a blanket removal that has a Ti/TiN barrier interface as well as the oxide interface. The mechanism for Ti/TiN removal is similar to the W removal in that an oxidized layer is formed by the oxidizing agent and polished off by the abrasive component. The ILD layer is also slightly polished but this is due solely to the abrasive compound as there is little chemical attack by the slurry components. The amount of ILD removed is dependent upon the type of oxide under the barrier. Softer doped oxides polish more quickly than harder undoped TEOS films. As was noted above, the removal of the barrier is dependent upon the composition and treatment of the barrier [56]. With comparable W removal rates, different slurries can have very different Ti removal rates, including high selectivity between W

and Ti CMP. Also, similar selectivities can exist between the Ti removal rate and the oxide layer at the barrier – oxide interface. Figure 10.24 pictures how these different selectivity approaches allow for two different integration approaches to the full W CMP process. In the first approach, W CMP and Ti CMP have similar removal rates, or nearly 1:1 selectivity, while Ti CMP has a higher removal rate than oxide CMP. Here, the W CMP process stops selectively at the ILD interface. In the second approach, W CMP has a much higher rate than Ti CMP, or high selectivity, while Ti CMP has a moderate selectivity to oxide CMP. Here the W CMP process stops at the Ti interface and followed by a slower, more uniform Ti and ILD removal. Each approach has consequences on the final form of the contact/via post W CMP. Possible final forms are slightly protruding plugs, perfectly planar plugs or slightly recessed plugs [59].

The relative merits of protruding, planar and recessed plugs will vary depending upon the subsequent metallization layers. Protruding plugs are easier to make contact to but provide more surface area for charged particles to adhere. Recessed plugs create topography that the metallization layer must fill uniformly but provide the capability of performing an oxide buff process
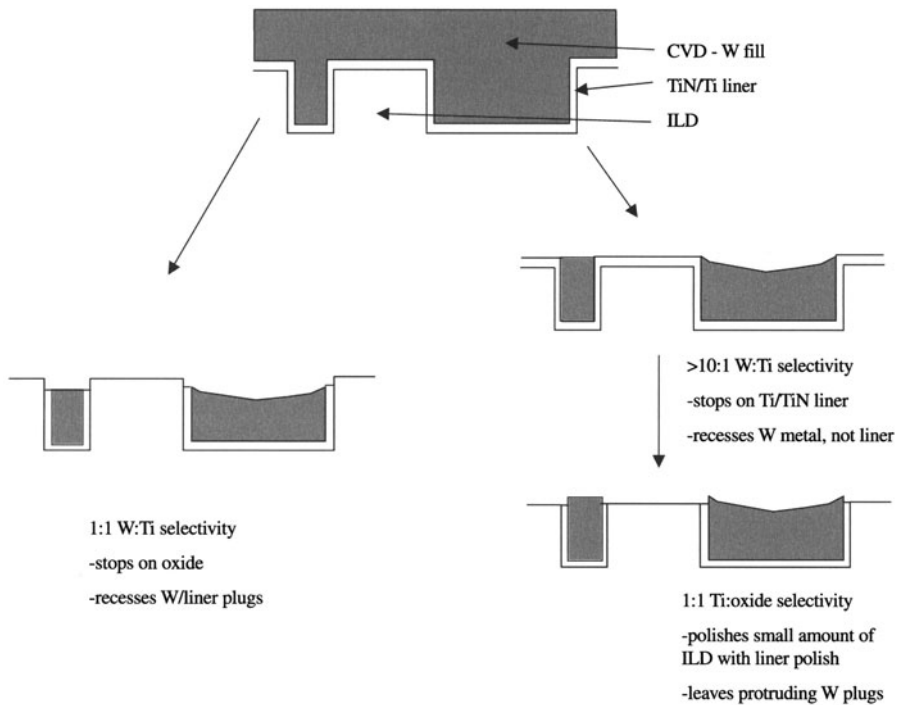
Fig. 10.24. Comparison of W CMP integration that leads to either recessed or protruding W contacts/vias

to remove ILD defects [59]. The degree to which these plugs shape changes depends not only on the integration approach but also the CMP pad being used in the approach [63, 64]. A hard pad tends to planarize better than a soft pad. Also, the chemical etch selectivity of the CMP cleans will also enhance or decrease the extent of the plug topography [64]. Regardless of the degree of this topography, both approaches present difficulties unique to a CMP process that relies upon interfacial selectivity.

Figure 10.25 depicts the two predominant interfacial concerns of W CMP, dishing and erosion. Dishing is the removal of W inside the contact/via below the ILD interface. Erosion is the removal of the supporting ILD after clearing the barrier [54]. Dishing is dependent upon the size of the plug [57], the W removal rate selectivity and the hardness of the pad. A large plug has no ILD support and is initially deposited with less tungsten. Higher removal rate selectivity to either Ti or silicon dioxide increases the rate at which the dishing occurs. Softer pads allow for the applied down force to reach into the plug [63, 64]. Dishing of the metal interconnects results in creation of topography that can be perpetuated, in extreme cases, through subsequent metallization. For example, Al lines could carry this topography to subsequent layers making stacked contact/vias difficult to manufacture. With increased dishing, the ILD etch of the next layer of vias, which is in a planarized ILD layer, would have to be increased resulting in overetch on isolated Al lines, as pictured in Fig. 10.26.

Erosion is dependent on the ILD polish rate, the Ti removal rate selectivity, the density of the plugs and the hardness of the pad. The ILD polish and Ti selectivity are related since the ILD rate used in the Ti selectivity calculation is oxide dependent. A high selectivity between Ti removal rate and oxide leads to reduced erosion, as shown in Fig. 10.27 [64]. However at contact W CMP, where the underlying ILD is often a doped oxide, the oxide has a relatively high removal rate. The incorporation of a thin TEOS layer as a W CMP stopping layer can reduce the erosion. High plug density puts more stress on the remaining exposed ILD film, resulting in a higher removal rate of ILD. As the ILD locally erodes, the W plugs also erode to maintain planarity with the ILD interface. Thus, increasing the density of the array of plugs results in an increase in the erosion [65, 66, 67]. The pad effect on

Planar plug   Dishing of W   Erosion of dielectric due to high contact
                   plug due to   density
                   wide CD
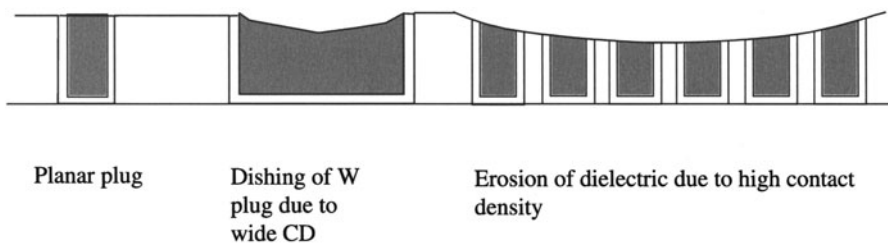
Fig. 10.25. Description of dishing and erosion in W CMP
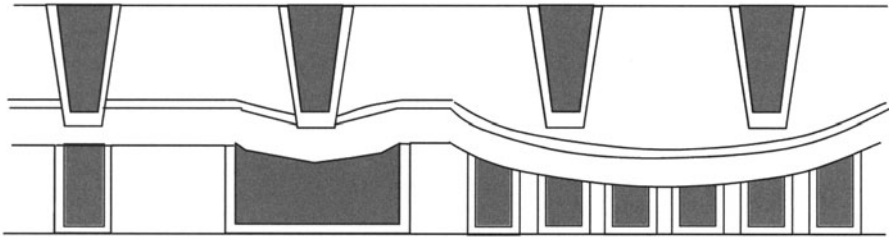
Equivalent etch on solitary planar plug over etches into the Al line.

Etch timed to make contact with Ti/TiN ARC on Al lines over dished W plugs.

Equivalent etch on eroded high density plugs end up short.

Fig. 10.26. Effect of dishing and erosion at contact W CMP on metallization deposition and via etch consistency
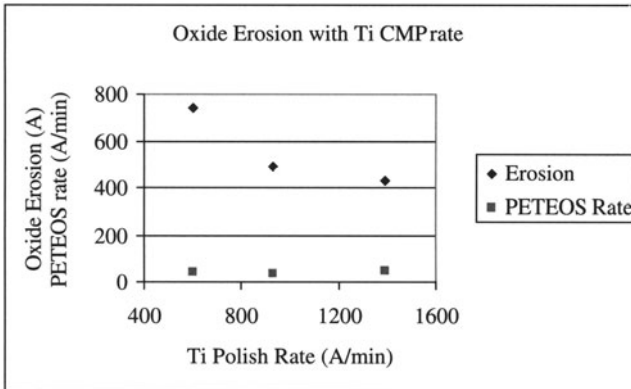
Fig. 10.27. Oxide erosion and blanket PETEOS polishing rate versus Ti polishing rate. Increasing the Ti polish rate and the selectivity to oxide CMP rate reduces the dishing. From [65]

oxide erosion is similar to that of W dishing. Softer pads enhance the erosion by applying higher force through the erosion-induced topography 66]. Similar to dishing, erosion creates topography that can be perpetuated through the subsequent layers. For example, an eroded dense array of contacts can leave excess ILD over the subsequent conformal Al lines. This again requires vias of substantially different lengths to be etched [67].

Both dishing and erosion can be minimized by appropriate selection of consumables and integration approach, as well as by design rules. CMP with endpoint detection can be used to assist in this reduction by sensing, either by changes in polish table current [68, 69] or optical signals [59, 68], the transition in the CMP process as it polishes through the interface, as described
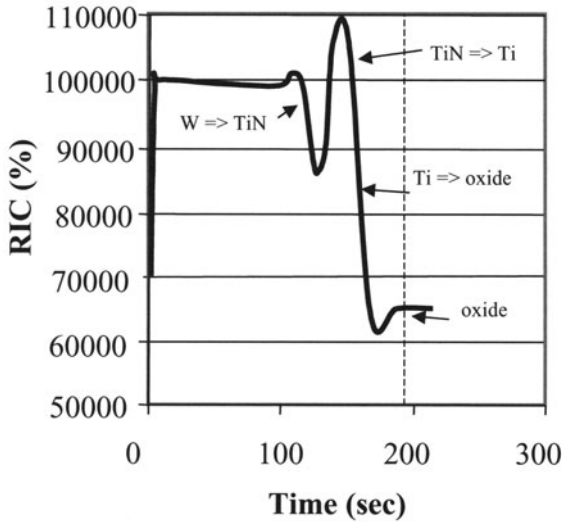
Fig. 10.28. Endpoint trace of the change in reflectivity through the W CMP interfaces. From [59]

in Fig. 10.28. Stopping the process at the correct time based on interfacial properties removes the obstacle of targeting polish time by lead wafer in a wafer lot, an issue inherent for CMP due to pad variations, slurry instability and deposition variations. Endpoint detection eliminates greatly reduces the need to over-polish to compensate for process variabilities, thus reducing both dishing and erosion. It also eliminates the risk of under-polishing the W layer, which leads to shorts between adjacent contact/vias from residual metal.

CMP defect levels in the ILD layers are also reduced by not over-polishing. The ILD is susceptible to the same chatter marks as in oxide CMP, however in W CMP the chemistries are not designed to hydrolyze, or soften, the ILD that might limit the extent of the damage from the particle impact. W CMP scratch marks can produce shorts between neighboring lines when the subsequent metal layer fills them during deposition. Buff processes can remove the ILD scratches but leave the plugs protruding from the ILD interface, as the buff chemistry does not etch W [59]. The buff process generally involves a soft pad that can affect the dishing and erosion. However, soft pads do have the ability to eliminate one source of defects associated with the W CVD process. W CVD deposits W metal to the edge of the wafer on to the bevel. During conventional hard pad W CMP and the subsequent cleaning, this metal can be dislodged and re-deposited onto the wafer surface leading to shorts [59]. The buff process can be used as an additional edge bead clean to remove the W on the bevel and eliminate the need for shadow rings in the
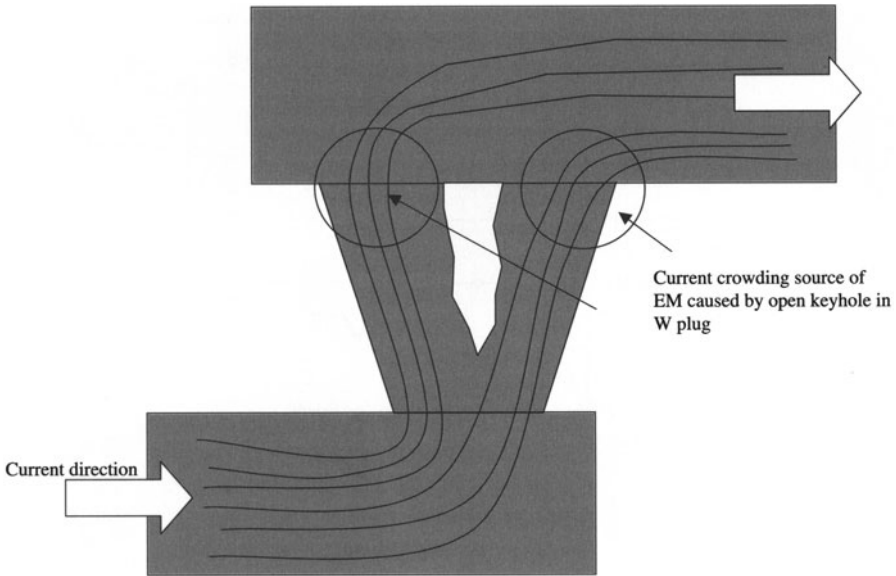
Fig. 10.29. Depiction of current crowding around a keyhole in a W plug. The current crowding is caused by the increase in current density at the edges of the vias. Higher current density increases electromigration. From [54]

CVD deposition process. Shadow rings are a potential source of particles in the deposition chambers.

Post CMP cleaning chemistry is primarily designed to remove particles from the wafer surface. The details are discussed in another chapter, however, certain cleans, such as HF, have the affect of etching the W plugs. This can eliminate the protrusion of the plug from the buff process. The concern with these types of chemistries, as well as the oxidative chemistries of the CMP slurry, is in opening up the W seam or keyhole formed during the CVD deposition. As previously discussed, keyholes not only increase the plug resistance but also provide a pathway for contaminants that can lead to corrosion of the metal or directly affect the active devices.

Keyholes also lead to an increase in current crowding in the vias [54], as illustrated in Fig. 10.29. Current crowding can lead to electromigration failures in the metallization.

10.2.4 Plug Reliability

Corrosion is the unwanted oxidation of the metal lines and plugs. The result of corrosion is an increase in line resistance and ultimately a completely open line. If water and Cl are present at the W or Al surface, they are sufficient to create corrosion. The Cl can come from the RIE etch of the Al lines. Usually, post-etch cleans are adequate in reducing the Cl content. The

moisture comes from the W CMP slurry and cleans, as well as some residual moisture absorbed in the ILD layer. When W and Al are exposed to both the Cl and moisture a galvanic cell is formed that facilitates Al corrosion [54]. Although the cleans are capable of removing the excess Cl from the etch, residual material left in opened keyholes during W CMP become a source for corrosion. Residual moisture in the keyholes acts as an electrolyte to further the corrosion reaction. Removing keyholes as a source of corrosion is one more assurance against metallization failures.

Electromigration is the result of transport of metal ions through the lines as a result of direct electrical current [51]. Typically this is studied as a line width issue; however the use of multiple materials in the plugs makes plug electromigration a concern. Poor step coverage and different materials cause electromigration in the plug. Keyholes in W plugs meet both conditions. First, the keyhole forces current to crowd the edges of the contact, shown in Fig. 10.29, and second the W-Al interface further crowds the current as flows around the corner from the plug to the metal line, described in Fig. 10.30. The second factor is alleviated by the use of TiN as a current distribution layer [70]. The first factor is alleviated only by not opening keyholes during W CMP.

Properly integrated W plug formation provides a highly reliable inter-connect [53, 71] metallization. Lifetime of these plug formations far exceeds standard Al forcefill and W etchback integration. In one report, time to break-down improved 26% for W CMP integration versus W etchback. When com-bined with the appropriate metal deposition layers, design rule changes, buff CMP steps, post W CMP cleans and consumable sets, W CMP produces consistent, long lifetime and reliable interconnects.
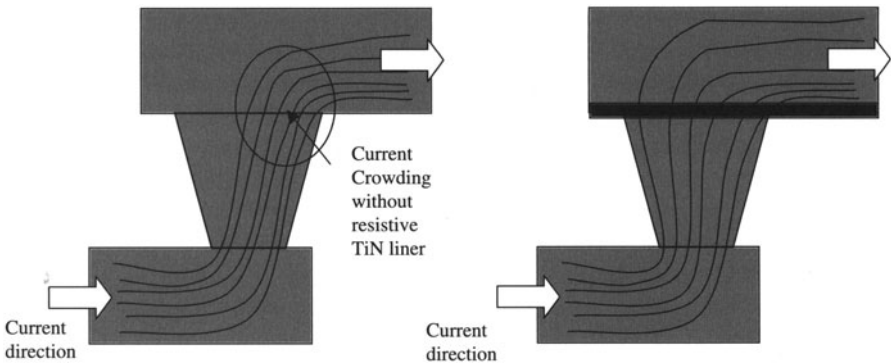
Fig. 10.30. Depiction of current crowding at the W/Al interface without the resistive TiN liner. The TiN liner reduces current crowding by distributing the current across the entire via interface [54]

10.3 STI Integration

The two main candidates for lateral isolation are LOCOS (LOCal Oxidation of Silicon)-based techniques and Shallow Trench Isolation (STI) with Chemical Mechanical Polishing (CMP). LOCOS isolation has been used for technologies down to $0.25\,\mu m$ because it is a very manufacturable and well understood process. Transistors are isolated by thermally growing a thick SiO_2 layer in the regions between them. A major disadvantage, however, is the large 'bird's beak' formation or the lateral enchroachment of the field oxide into the active area regions which is associated with the high temperature oxidation step. This 'bird's beak' of LOCOS isolation sharply limits its deep submicron scalability. In order to make it possible to downscale further to smaller isolation geometries, one has to switch to Shallow Trench Isolation with Chemical Mechanical Polishing. Here the field oxide is nicely embedded into the Si and is clearly distinguished from the active area regions. This allows for very small active area pitches and a higher device packing density. Due to the CMP step a superior planarity is created. This section discusses key considerations in the integration and implementation of the STI module in a conventional CMOS flow.

The basic STI process flow, i.e. without any additional modification to enhance the process performance, consists of the following process steps:

- Deposition of the isolation stack on the bare Si wafer. This stack consists of a thermally grown pad oxide layer with a furnace nitride layer on top.
- Next step is the active area patterning to define the regions of active area and field isolation and an etch through the Si_3N_4 and SiO_2, followed by further etching into the silicon substrate. The trench depth into the silicon depends on the technology used, but generally trenches become shallower for the smaller technologies (because of filling requirements).
- After etch and strip of the initial Si_3N_4 and SiO_2, a thermal oxide liner is grown along the etched trench sidewalls. Mostly, this step also includes a rounding of the trench corners, which is crucial for good electrical performance of the devices (see trench corner parasitic effect).
- The deposition of the trench filling oxide (thick enough to fill the trenches with a certain overfill) is one of the crucial steps in the STI process flow. The most common process is the HDP (High Density Plasma) CVD process because of its good gap-filling capacity (see trench refill).
- After trench refill comes an Oxide CMP step with polish stop on the nitride. Both low selective or high selective slurries for oxide to nitride can be used for this CMP step, dependent on the chosen approach (see CMP effects).

- The nitride on the active areas is removed in a hot phosphoric acid bath with high selectivity to the oxide. The thermal pad oxide is etched off in an HF solution.
- Then the gate oxide is regrown and the conventional transistor process fabrication flow continues.

Critical issues for STI Integration:
(A) Trench corner 'parasitic' effect.
A critical issue in the STI module is the formation of a 'parasitic' transistor at the active area edge. It is related to the local de-oxidation at the edge during the several wet etch steps (in HF) which are needed in an STI process flow. This local de-oxidation results in recess of the trench oxide and consequently thinning of the gate oxide in further transistor processing [72, 73, 74, 75].

A cross section of a wafer with STI isolation processed up to the transistor level is shown in Fig. 10.31. The polysilicon gate deposited over the active areas also extends over the oxide in the trenches. The critical place here is the edge of the active area. Local oxide thinning (creating a local high E-field region) and poly-gate wraparound at this recessed edge might result in the formation of a parasitic transistor with a lower threshold voltage (V_t) along the trench edge. This low V_t parasitic edge transistor provides a leakage path even before the real transistor is switched on.

This additional current manifests itself as a 'sub-threshold' hump in the I-V characteristics of the real transistor, as is shown in Fig. 10.32.

The main condition which should be fulfilled to avoid the formation of a parasitic edge device is that the active area edge should be rounded and
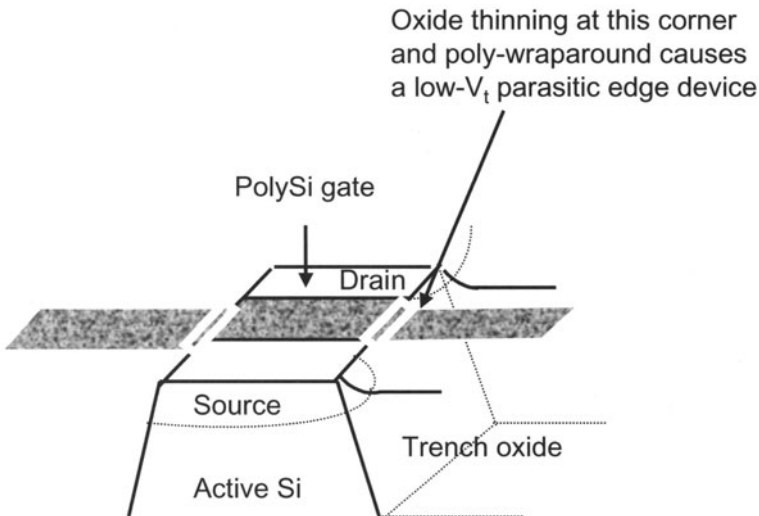
Fig. 10.31. The 'gate poly-Si wraparound' problem at the trench corner. From [72]
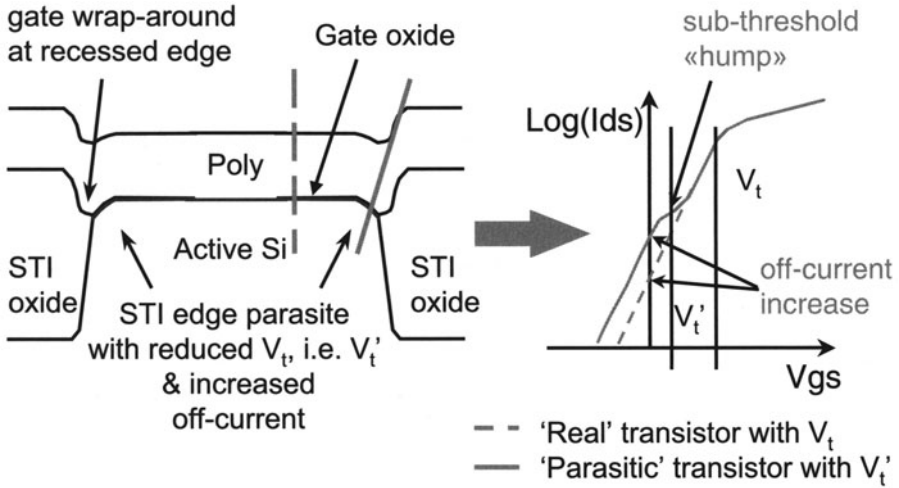
Fig. 10.32. STI edge effect on MOS devices

below the field dielectric. This is achieved by optimizing the trench etch, the liner oxidation, the trench-fill oxide deposition and the planarization (CMP) process. Specific issues for each of these steps are addressed below.

STI Trench Etch

The entire stack of nitride, pad oxide and silicon is etched to form trenches of typically between 0.3–0.5 μm, depending on the technology. The trench sidewall slope should be ideally ~80° to avoid sharp corners in the active silicon at the top of the trench. These sharp corners would lead to the creation of high field regions (as addressed previously) and consequently result in parasitic transistors or leaky diodes or gate-oxide integrity (GOI) problems. Typical etch profiles of 0.25 μm features with nicely sloped trench sidewalls are shown in the X-SEM pictures of Fig. 10.33 [76, 77].

Isolated 0.25 μm trench   Isolated 0.25 μm line   0.25 μm L/S
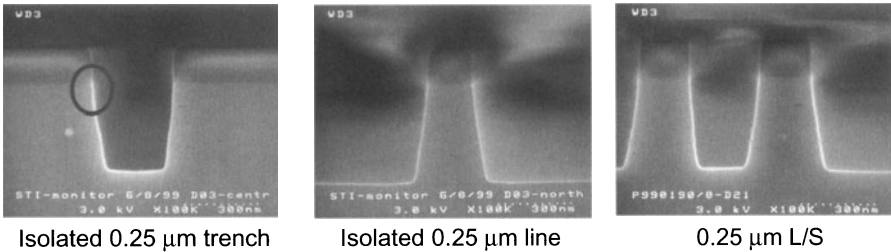
Fig. 10.33. STI trench profiles of 0.25 μm features

STI Liner Oxidation

The growing of the oxide liner on the trench sidewalls, which follows the trench etch step, is crucial to give the rounded corners and also allows for void-free and seamless gapfill during the trench-fill oxide CVD process. This oxide liner passivates the etched silicon surface and forms a barrier between the silicon edge and the deposited oxide. Among the several approaches to obtain rounded corners is the undercutting of the pad oxide beneath the active area protecting nitride mask by an isotropic wet etch before the oxide liner is grown, as is shown in Fig. 10.34a [73]. Another technique is to perform a high temperature oxidation of the STI dielectric, immediately after the CMP step, as is shown in Fig. 10.34b [73].

Figure 10.35 shows some TEM pictures of an STI structure with different degrees of corner rounding. Wafers with and without optimized corner rounding continued processing for full transistor fabrication using the standard process flow and were measured electrically [80].
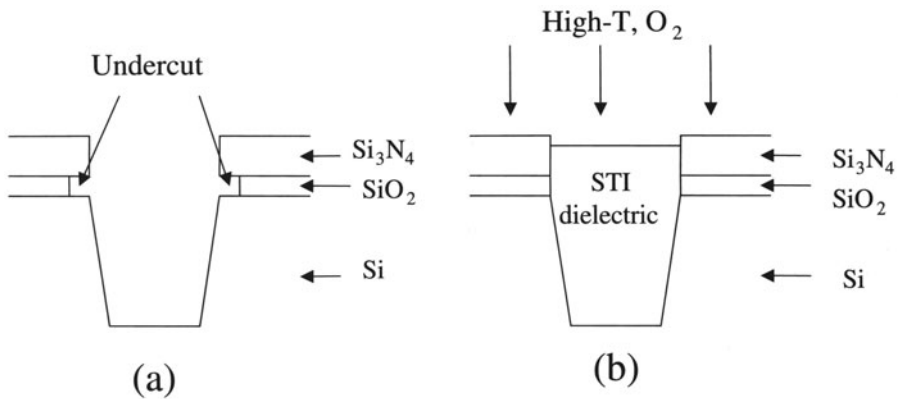
Fig. 10.34. Illustration of corner rounding technologies. From [73]
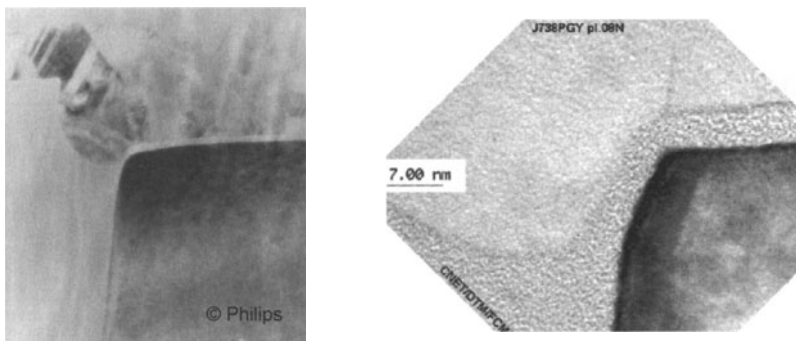
Fig. 10.35. TEM cross-section of a rounded active area edge
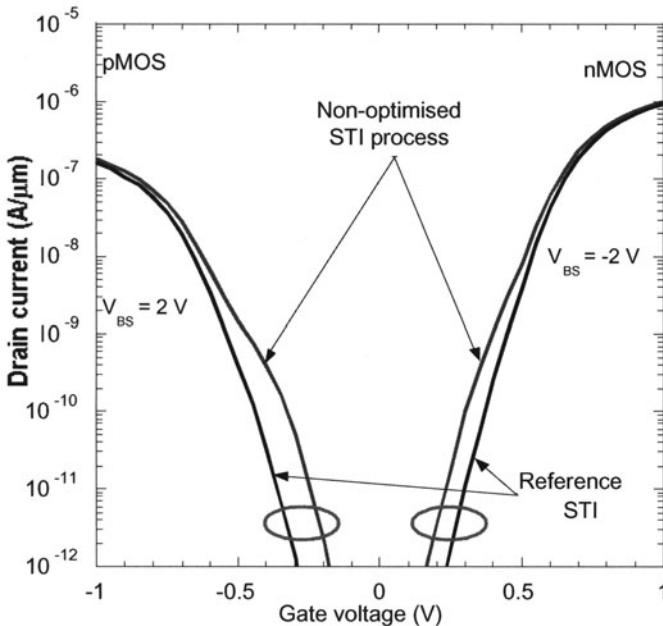
Fig. 10.36. I_d-V_g characteristics of pMOS and nMOS transistors fabricated with an optimized and non-optimized corner rounding STI approach

Figure 10.36 shows the drain current (I_d) versus gate voltage (V_g) characteristics of pMOS and nMOS transistors for the optimized (reference) and non-optimized corner rounding STI process. A substrate back bias is applied to make the double hump effect more prominent. This double hump shows up clearly for the non-optimized STI conditions. It is caused by the presence of a parasitic transistor with a lower threshold voltage (V_t) along the active area corner [81].

STI Trench Fill

The specifications for an STI-filling oxide are much more stringent compared to a normal deposited oxide used in the back-end. The STI oxide has to meet the following conditions:

- It should be thermally stable (and thus resist the high temperature steps such as source/drain anneal) and should have low shrinkage.
- It has to be resistant to wet etch (such as the multiple HF steps for oxide removal and the nitride etch). Otherwise the level of oxide in the trench could drop below that of the active area silicon, and thus could cause poly-wraparound and undesired electrical effects.
- It should not introduce mechanical stress into Si.

- It should allow for filling without voids or seams in high aspect ratio trenches. Voids above the active area level could be opened up during CMP, and could be filled with poly-Si during the gate formation leading to leakage paths.

The most commonly used STI-filling oxide is HDP-CVD oxide. It consists of a simultaneous ion sputtering during high density plasma (HDP) deposition and is pictured in Fig. 10.37 [82].

An important parameter to optimize in an HDP-CVD deposition process is the *Deposition-Sputter* (D/S) Ratio. In case of excessive sputtering damage to the trench corners might occur which would result in poly-wraparound and the formation of a parasitic transistor as discussed above. Another negative effect of excessive sputtering is transistor width reduction. The D/S ratio also influences the lateral distance of the filling oxide to the trench corners. This lateral distance should be maximized to avoid the detrimental corner effects. In Fig. 10.38, the SEM picture on the left shows the HDP oxide profile of an unoptimized process; after changing the D/S ratio the lateral distance to the trench corners is increased as is shown on the right SEM picture [83].

The oxide thickness on different pattern density active areas might be different for an HDP-CVD oxide process. In order to reduce the CMP pattern
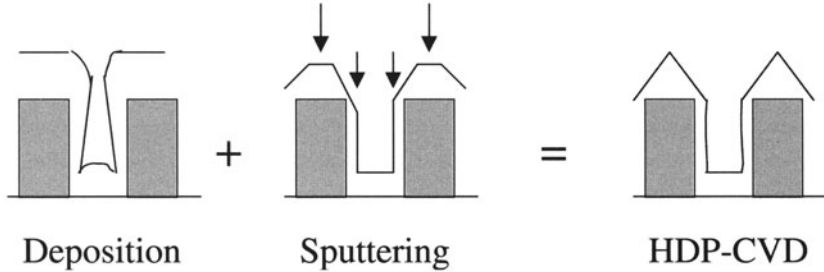
Fig. 10.37. Schematic of the HDP-CVD principle. From [82]
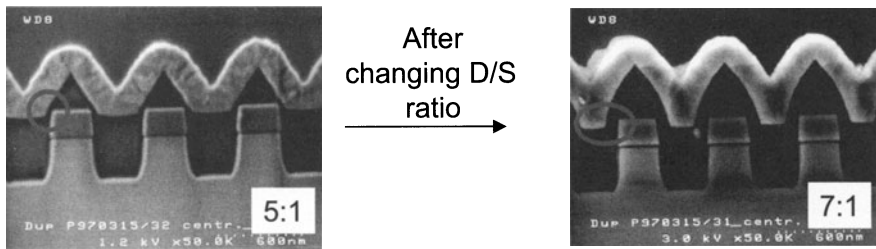
Fig. 10.38. Impact of the D/S ratio on the profile shape of HDP-CVD oxide. From [83]

TEOS-Ozone   HDP-CVD Oxide

Topography at the end of a dense structure

Topography over wide trenches
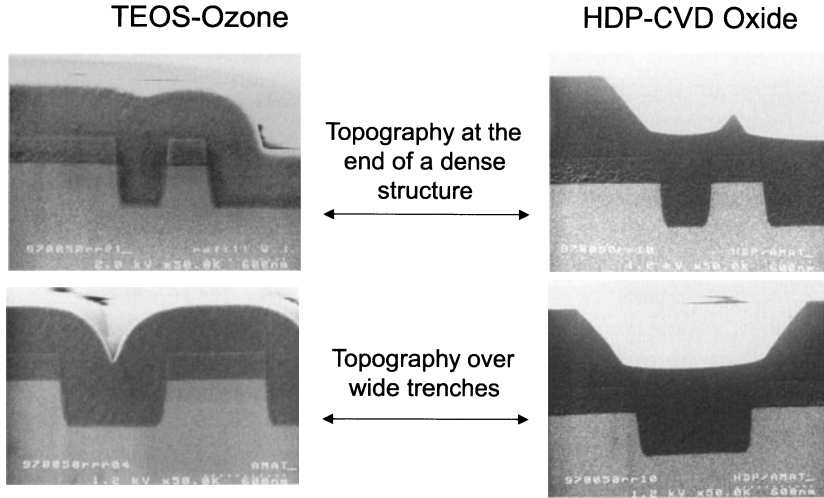
Fig. 10.39. Comparison of the profiles of a TEOS/O$_3$ oxide and an HDP-CVD oxide

density effects the D/S ratio should be so to have (at least) the same nominal oxide thickness on small, dense active areas as on wide ones.

Other trench filling processes use a TEOS/O$_3$ based chemistry CVD oxide. This process requires a high temperature anneal step to reduce the HF etch rate of the deposited oxide, and is more sensitive to voids. Also spin-on-glass materials can be used for STI trench fill. They have a very good gap-fill performance, but also need a high temperature anneal and can shrink up to 25% at high process temperatures, which might result in delamination of the oxide from the trench sidewall.

The choice of trench-filling oxide has a big impact on the subsequent process step, being the CMP. Final flatness after CMP depends very much on the topography before CMP. As can be seen in Fig. 10.39, the deposition profile at the end of a dense structure or over a wide trench is completely different for a TEOS/O$_3$ oxide compared to a HDP-CVD oxide. These different morphologies will lead to different dishing and erosion results after CMP.

Also, the deposition topography as a function of active area width differs significantly when comparing an TEOS/O$_3$ oxide to an HDP-CVD oxide as shown in Fig. 10.40. The TEOS/O$_3$ oxide has the same nominal oxide thickness, independent of the active area size, while the HDP-CVD oxide pyramid structure height increases with active area size. It will obtain its nominal oxide thickness on structures $> 0.5\,\mu m$.

TEOS-Ozone

+/- Same oxide thickness
everywhere

HDP-CVD Oxide

Different pyramids dependent
on active area width
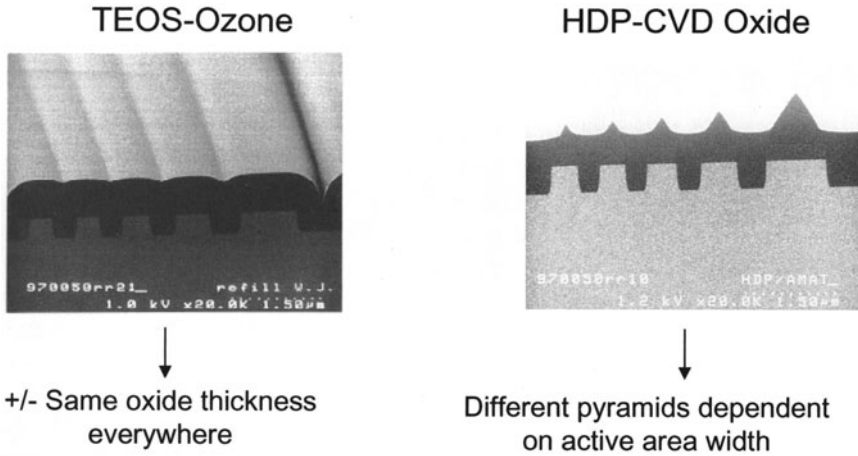
Fig. 10.40. Deposition topography for TEOS/O$_3$ oxide and HDP-CVD oxide as a function of active area width

Critical issues for STI Integration:
(B) Pattern layout effects in CMP.
The pattern dependency of the CMP process is one of the main reasons for the within-die nitride thickness variation between active area features of varying size and pattern density. It can be explained by the fact that the local removal rate in CMP is proportional to the local pressure on a feature, and inversely proportional to its pattern density. This within-die-non-uniformity (WIDNU) of the active nitride will result in variations in the active area to field step height after nitride strip, and thus in problems at the subsequent lithography steps.

Dishing in the large open field areas is another typical CMP related problem which of course negatively impacts the surface planarity. The critical places for CMP are schematically shown in Fig. 10.41.

Incomplete Polish or Erosion?
→ Dependent on fill and integration approach!

Over-polished
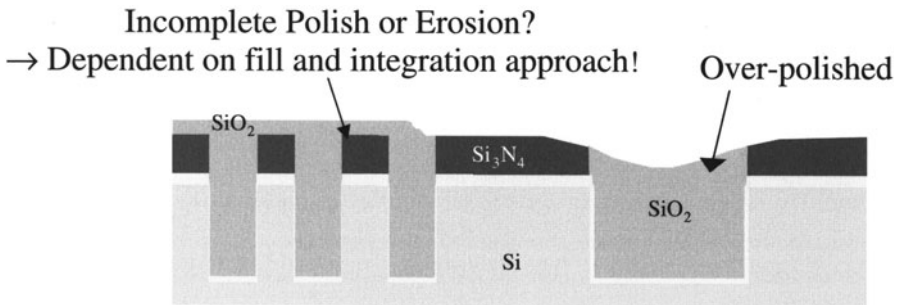
Fig. 10.41. CMP effects

Due to the pattern layout dependency of the CMP process the following thicknesses should be optimized for optimal performance:

The thickness of the active protecting nitride:

- It should be thick enough to leave a considerable amount of nitride in small isolated active areas (which are sensitive to overpolish!)
- A minimum nitride thickness is needed to have coverage of the trench corner by the fill oxide after the nitride strip and the subsequent HF steps (cfr. parasitic corner effect)
  - The nitride thickness determines the step between the active area and the trench-fill oxide; this step will be different depending on the active area pattern density!
  - The variation of this step over the die (WIDNU) and over the wafer (WIWNU) should be limited for lithography reasons; the following step is poly-gate patterning for which a maximum within-die topography of 100 nm is allowed!

The thickness of the trench-fill oxide:

- It should be thick enough to compensate for field dishing (not much field dishing allowed to keep a positive step!)

STI-CMP issues:
(A) Pad and slurry engineering

- The polishing pad:
  The standard pad being used in CMP consists of a stack of 'hard' polyurethane (typically 50 mils thickness) on a 'soft' bottom pad. The former is needed for good planarization and a low within-die thickness range; the latter gives a good within-wafer uniformity. Thicker pads generally give a lower WIDNU of the nitride on active but also might result in an increased WIWNU.
- The slurry:
  Standard Oxide CMP slurries with a selectivity of oxide to nitride of 4:1 might be used for STI-CMP although it highly depends on the chosen approach. Some STI integration approaches however need a high selectivity slurry; commercial slurries with a selectivity of oxide to nitride of over 100:1 are available on the market. This high selectivity limits the amount of nitride erosion, but might increase the amount of field dishing. In most cases however, the use of a high selectivity slurry widens the available process window. It should be noted that this selectivity is on blanket wafers; some studies indicate that this selectivity is pattern-dependent. Typically a selectivity loss occurs towards low active area density regions.

(B) Modifications to the process flow

1. Dummy structures

   The large field areas around isolated active areas can be filled up with *dummy* active areas to compensate for the limited within-die uniformity associated with STI-CMP. This yields a more uniform local density of active areas both over the die (WIDNU) and over the wafer (WIWNU) and seriously reduces the risk of overpolishing. It is the easiest way to solve the pattern density problem in CMP and is widely used for large scale production like memory applications [84, 85]. However, the approach offers a lot of design problems. In mixed-signal technologies it results in an increased capacitive coupling and noise (routing of the poly, disturb the functioning of poly-poly capacitors).

2. Oxide Reverse Etch (ORE) approach

   A very common STI planarization technique is Oxide Reverse Etch followed by CMP, as shown in Fig. 10.42. In short the process flow consists of the following steps [86]:

   - Pad oxide and nitride deposition
   - Trench patterning and etching
   - Sidewall oxidation and deposition of the trench filling oxide
   - *Removal of (part of) the oxide on large active areas (slowest polish rate!) with an additional litho and etch step. This step uses a certain oversize, i.e. it leaves a small border where no oxide is removed. In this way trench formation due to litho misalignment is avoided next to the actual active areas.*
   - Removal of the remaining oxide from the active areas with CMP
   - Removal of the nitride during trans etch step

   In summary, the most critical structures for the ORE approach are:

   - Small isolated active areas since these structures are very sensitive to overpolishing.
   - Dense active area structures (where there is no clear-out of the oxide!): Depending on the type of trench-filling oxide these structures will be prone to under- or overpolishing.
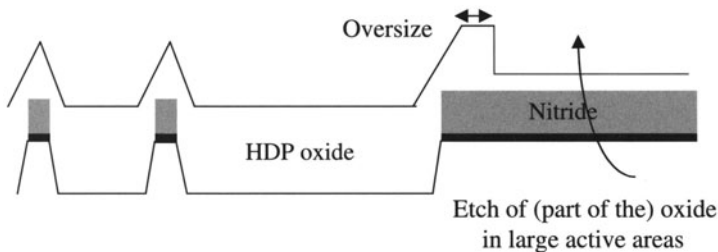
Fig. 10.42. Schematic of the Reverse-Oxide Etch approach

- Large field oxide areas: Severe dishing can be observed in wide field areas (up to 200 nm!). Enhanced dishing can be easily observed by the color variations of the oxide.

3. Dual Nitride (DN) approach

The dual nitride approach is an approach which completely avoids dishing in the wide field oxide areas. The process flow for the dual nitride approach consists of the following steps as depicted in Fig. 10.43 [87, 88]:

- Pad oxide and nitride deposition (referred to as nitride-1)
- Trench patterning and etching
- Sidewall oxidation and deposition of the trench filling oxide
- Deposition of a second nitride layer (referred to as nitride-2) and patterning of this layer in the field areas with a second litho and etch step. Again this litho step uses a certain oversize to avoid trench formation due to litho misalignment.
- CMP to planarize the surface (almost fully covered with nitride!)
- Etch off both nitride-1 and nitride-2 in one step

The main advantages of the dual nitride approach are:

- The polish rate of nitride-1 on small isolated active areas is reduced due to the presence of nitride-2 in neighboring field regions
- No dishing in field regions due to the protective nitride-2 layer

The approach also needs a high selectivity slurry to stop polishing in the field regions.

In summary, the most critical structures for the DN approach are:
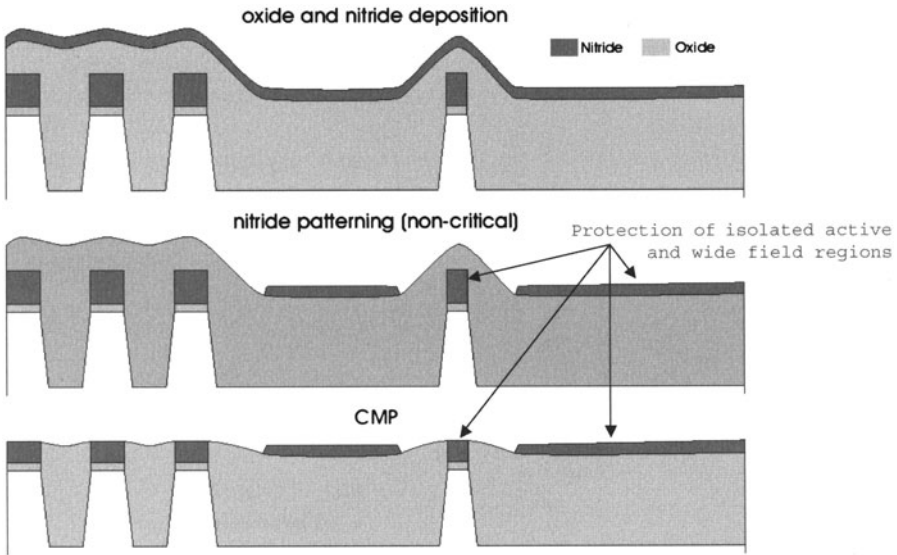
Fig. 10.43. Schematic of the Dual-Nitride approach
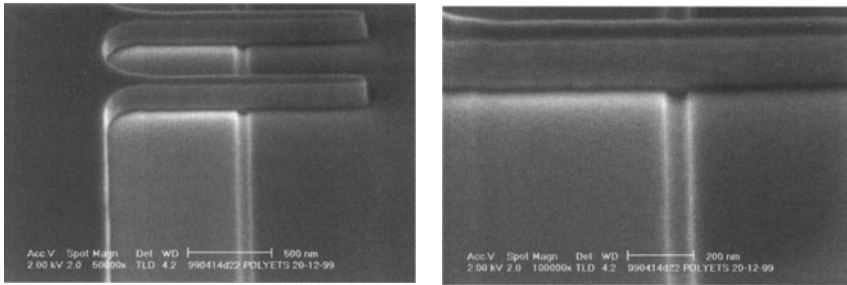
Fig. 10.44. Patterned poly-Si structure over an active area and extending into the field area. An almost flat transition from active to field regions is observed

- Large Capacitor-type structures: These structures need the longest (overpolish) time to clear the oxide. It can clearly be observed with a microscope if there is still some oxide left on these large structures. The remaining oxide after CMP will result in an incomplete nitride strip!
- Dense structures with small lines and spaces: These structures are not protected by the second nitride layer. Therefore, they are very sensitive to overpolishing into the silicon substrate.

The biggest advantage of the dual nitride approach however is the fact that dishing is completely eliminated, resulting in a very flat topography at gate level which is critical for the poly-Si gate patterning. This is shown in Fig. 10.44.

4. Resist Block approach
   In short the process flow of the resist block approach consists of the following steps:
   - Isolation patterning
   - Etching of the trenches and filling with oxide
   - Patterning of a block resist over the large field areas to bring the surface to about the same level as the oxide on the active areas. Hardening of the block resist
   - Spinning of a planarizing resist and cure
   - Reactive Ion Etch: with this process the flat resist surface is transformed into a flat oxide surface. The RIE is stopped before removing all oxide on the active nitride.
   - Removal of the oxide spikes with a short CMP step
   - Etch off the nitride on active

The resist block approach has an excellent planarization performance, but has as major disadvantages that an RIE process and a planarizing resist are required [89, 90].

5. Modifications without additional lithography steps

Additional lithography steps (as in the ORE or DN approach) signifi-cantly increase the cost of the STI module. Therefore, several alternative approaches have been proposed that do not need an extra litho step. Some examples are:

- The nitride overcoat approach [91, 92]:
    After trench refill with oxide, a thin nitride hard layer is deposited. Then follows a CMP step with a low selectivity slurry. The nitride overcoat is removed already after a short polish time in the dense active areas and on the small isolated ones. The large field areas how-ever are protected until the nitride overcoat above the dense areas is removed.
    This nitride overcoat approach may result in a very flat surface, but is very dependent on the design! A potential problem arises if very large active areas are present in the layout since the removal rate of nitride on these locations is comparable to the one in the large field areas, which makes no leveling possible.
- The use of an organic spin-on-glass as overcoat [93]:
    Here, the same remark is valid for this approach as for the nitride overcoat approach.
- Fixed abrasive technology (slurryless polishing):
    This new approach to STI-CMP has been developed by $3M^{TM}$ [94], where a combination of microreplication technology and coated abra-sive technology has resulted in a fixed abrasive matrix capable of CMP. The slurry particles are located inside the pad, and the fixed abrasive is designed so that the topology of the wafers conditions the abrasive composites to expose fresh mineral during polishing.

A key advantage of CMP using fixed abrasive is its selectivity to to-pography (100:1), while the selectivity between oxide and nitride is only $\sim 1:1$. This provides a high degree of planarity and a low susceptibility to dishing on overpolish.

10.4 Copper Damascene Integration

As stated elsewhere in this volume, copper polishing was originally devel-oped for fabrication of damascene interconnect [95]. This was made neces-sary by fundamental difficulties associated with conventional plasma etching of metallic copper, particularly the non-volatility of copper containing prod-uct compounds [96]. Within that context, some of the practical details of the copper polishing process itself were discussed, however, it is useful to consider the "larger picture" concerning the overall integration of copper metallization in fabrication of integrated circuits.

First of all, it cannot be overemphasized that manufacturers of integrated circuits are extremely conservative when it comes to the introduction of new

materials and processes. This is quite understandable in consideration of the risk involved and the economic impact of a mistake. Consequently, the introduction of any new material or process must provide some substantial and immediate benefit, which contributes ultimately to maintenance and improvement of profitability. Obviously, if this is not the case, there is no compelling reason to introduce any new technology. Indeed, the history of integrated circuit manufacturing is littered with unsuccessful attempts at introducing new materials and processes which did not meet this overall goal.

At this point it is useful to digress briefly to consider a historical perspective. Almost from the invention and initial commercialization of integrated circuits in the late 1950's and early 1960's, aluminum and aluminum alloys have been predominantly used for interconnect wiring. Even so, gold [97] was an early competitor to aluminum; however, in addition to its expense, gold is much more difficult to pattern and etch and is a particularly serious contaminant of semiconductor material in comparison to aluminum. Indeed, aluminum was found to be relatively easy to pattern and etch, first with acidic chemical solutions and later with chlorine containing plasma chemistry. Moreover, although aluminum is intrinsically quite chemically active, it forms an impervious, passivating surface oxide, which enhances its reliability in many applications; not the least important of which is integrated circuit wiring. Nevertheless, pure aluminum suffers from two serious drawbacks as an interconnect metallization. The first of these is its strong tendency to form an alloy when in contact with the silicon substrate. The result is the formation of contact "spikes" and collateral destruction of underlying *pn*-junctions. An early solution to this problem was the use of aluminum-silicon alloy instead of pure aluminum to reduce the thermodynamic driving force for the alloying reaction. Subsequently, refractory metals such as tungsten and titanium nitride have been used as thin film diffusion barriers definitely separating the materials. The second drawback is that pure aluminum recrystallizes at relatively low temperatures, e.g., 400–450°C, and forms large grains or "hillocks". Alloying a small amount of copper, typically less that 4%, with the aluminum during deposition has substantially eliminated this problem. In addition, the reliability aluminum-copper alloy wiring is superior to that of pure aluminum (98).

Clearly, aluminum alloy interconnect has been highly optimized over the last three decades. Even so, its electrical resistivity remains a fundamental limitation. By definition, resistivity is an intensive material property that determines the current carrying capacity of a conductive wire of given cross sectional area. To be specific, the smaller the resistivity of a particular metal, the larger the electrical current that can be accommodated in a wire of fixed cross section made of this metal in comparison to one of the same dimension made of some other conductive material. Indeed, excluding the possibility of room temperature superconductors, pure aluminum is the fourth most conductive material known having a bulk resistivity of 2.65 $\mu\Omega$ cm. Only

silver, copper, and gold are more conductive, having nominal resistivities at room temperature of 1.59, 1.68, and 2.27 $\mu\Omega$ cm, respectively [99]. Therefore, to reduce or merely to maintain the characteristic resistance of interconnect wiring, while at the same time maintaining or reducing its cross sectional area requires aluminum to be replaced with some more conductive metal.

This situation is further complicated because a typical integrated circuit is a dynamic device that operates by transforming electrical inputs into appropriate outputs during some prescribed time interval. Therefore, overall circuit speed is usually a paramount figure of merit, which has been steadily improving as device and wiring dimensions have shrunk. As a consequence, it is not only the resistance of a single wire that must be considered, but rather, the impedance of a complicated interconnect structure, which is made up of a complex array of wires separated by some insulating material. A pictorial representation of such a structure is shown in Fig. 10.45.

Here, R corresponds to the series resistance of the wire, C to the capacitance between adjacent wires on the same level of metal, and C$'$ to the capacitance between wiring levels. Within this context, both the characteristic resistance of the wiring and the characteristic capacitance of the insulator structure determine the aggregate "RC" time constant of the interconnect. In a broad sense, the reciprocal of this time constant corresponds to the time delay of an electrical signal propagating through the interconnect structure. Clearly, total interconnect performance can be improved by decreasing the dielectric constant of the insulator as well as the resistivity of the conductor metal. This will discussed in more detail subsequently.

Even so, none of this was of any consequence as long as semiconductor device dimensions were relatively large. In this case, circuit speed was dominated almost entirely by time constants associated with transistor performance and not by those of the interconnect structure. This situation changed dramatically as minimum device geometry fell below 250 nm and is the prevailing situation at present. This is illustrated in Fig. 10.46, which is obtained by consideration of a simple model [100].
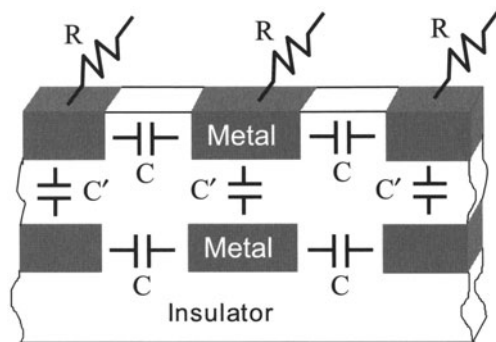
Fig. 10.45. Schematic of interconnect
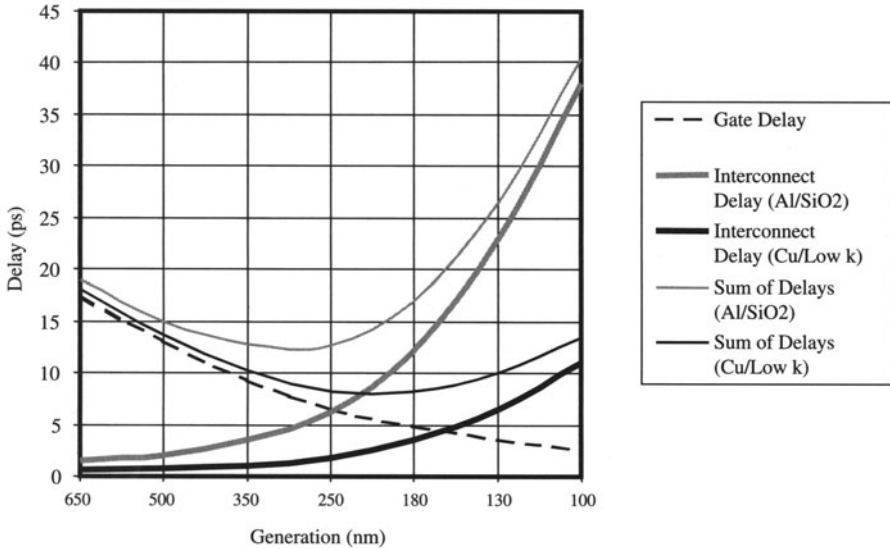
Fig. 10.46. Circuit delay vs. device generation. From [100]

Here, the heavy dashed curve denotes signal delay associated with fundamental semiconductor device structures, which decreases as transistor size and channel length decrease. In contrast, the light solid gray and black curves indicate pure interconnect delay for a typical aluminum alloy/silicon dioxide and copper/low-k insulator interconnect structures, respectively. (For the present purpose, the dielectric constant of the low-k insulator, i.e., k, is assumed to 3.) These delays increase due to the increased resistance of smaller dimension wiring and increased capacitance due to more closely spaced wires and thinner insulator. Thus, it is clear that the sum of these contributions, which represents signal delay for the overall circuit, becomes dominated by interconnect at small device geometry. This provides the fundamental motivation for the replacement of aluminum with a metal having lower resistivity. Moreover, of the three choices possible, in addition to the drawbacks mentioned previously, gold does not provide a substantial improvement over aluminum. Thus, the choice lies between silver and copper. Within this context, although the former has a slightly lower resistivity, it is more expensive and subject to chemical reactivity with sulfide and halide containing species that is thought to limit reliability. Hence, copper has been uniformly chosen as a suitable replacement for aluminum alloys.

An important issue in fabrication of copper interconnect is the decision whether to use a *single damascene* or a *dual damascene* process. If such a approach is applied to copper, each photolithography step in fabrication of the interconnect is accompanied by copper and barrier deposition followed

by CMP. Use of single damascene to fabricate a layer of interconnect wiring and vias is illustrated in Fig. 10.47.

Clearly, fabrication of tungsten plugs as described elsewhere is an also an example of a single damascene process. In contrast, two photolithography steps, one for wires and one for via holes are combined in a dual damascene process. This substantially reduces the number of metal deposition and polishing steps that are required to fabricate the interconnect structure. The dual damascene structure is pictured in Fig. 10.48.
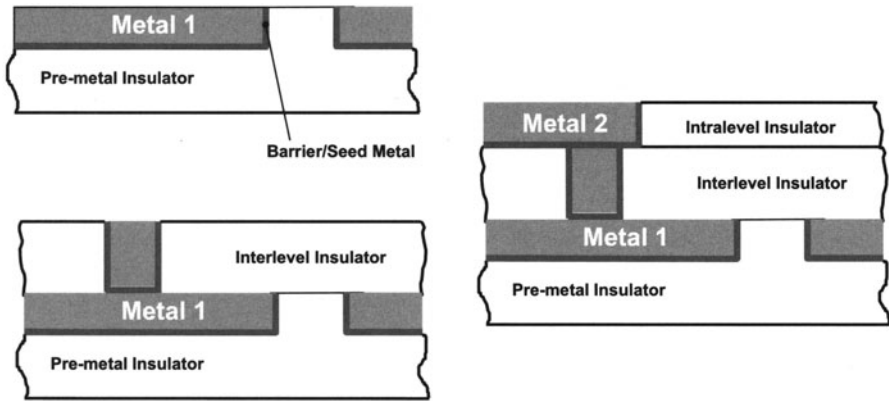
Fig. 10.47. Single Damascene: top: after lower level has been processed; middle: after via hole photolithography, etching, metal deposition, and CMP; bottom: after wiring photolithography, etching, metal deposition, and CMP
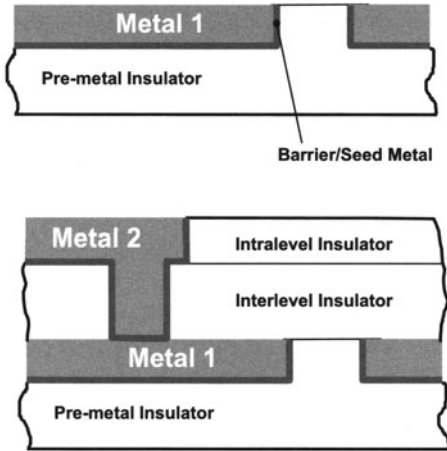
Fig. 10.48. Dual Damascene: top: after lower level has been processed; bottom: after via hole and wiring photolithography and etching steps, metal deposition, and CMP

Clearly, the only structural difference in the completed interconnect is the presence of an additional layer of barrier metal at the top of the via hole, which is unavoidable in the single damascene process. However, the series resistance due to this layer is generally negligible and, moreover, the additional barrier layer may improve electromigration resistance of via holes. To compare single and dual damascene processes, one observes that the dual damascene process eliminates about half of expensive metal deposition and polishing steps. However, this comes at the cost of making photolithography and to a lesser degree etching, more difficult. As mentioned previously, an additional choice of "via first" or "trench first" dual damascene must also be made [101]. Both of these have advantages and disadvantages, which are primarily determined by the details of the photolithography and etching processes. In contrast, single damascene is more conservative in the sense that photolithography is more standard. Of course, additional deposition and polishing is required.

Within this context, two significant issues emerge that should be considered in some detail. The first of these is choice of refractory metal for use as a diffusion barrier layer with copper metallization. Historically, titanium nitride has been used with conventional aluminum alloy wiring and tungsten plugs and has been thoroughly characterized. Moreover, titanium nitride can be deposited either by reactive PVD or by MOCVD and is easily etched by hydrogen peroxide solution or a chlorine containing plasma. Concomitantly, titanium-tungsten alloys have also been used as interconnect barrier metal; in particular with electroformed gold. Since, either titanium nitride or tungsten alloys (or nitrides) are easily removed using peroxide or some other mild oxidative chemistry; it would seem that they would be natural choices for use with copper damascene. However, for a variety of reasons, in the development of copper damascene, tantalum and tantalum nitride have become dominant as diffusion barrier materials. Unfortunately, neither tantalum or tantalum nitride are removed by a simple peroxide chemistry and, thus, require more aggressive chemical agents or abrasives. Moreover, as is made evident in previous discussion of damascene formation, use of these agents or abrasives must be compatible with metallic copper since the damascene structure is essentially completed prior to barrier metal removal.

An additional issue is that before removal of the diffusion barrier, copper and barrier metal are exposed simultaneously on the surface of the wafer and are in both direct electrical contact as well as contact through the chemical solution. This arrangement forms a short circuited electrochemical cell, which under severe conditions can lead to enhanced corrosion of one material or the other [102]. An extreme example of such phenomena is illustrated by the following micrograph, Fig. 10.49.

In this picture, copper has been completely removed from the field and the tantalum/tantalum nitride has been partially removed. What is striking is that no copper remains in bond pad areas connected to long leads, while
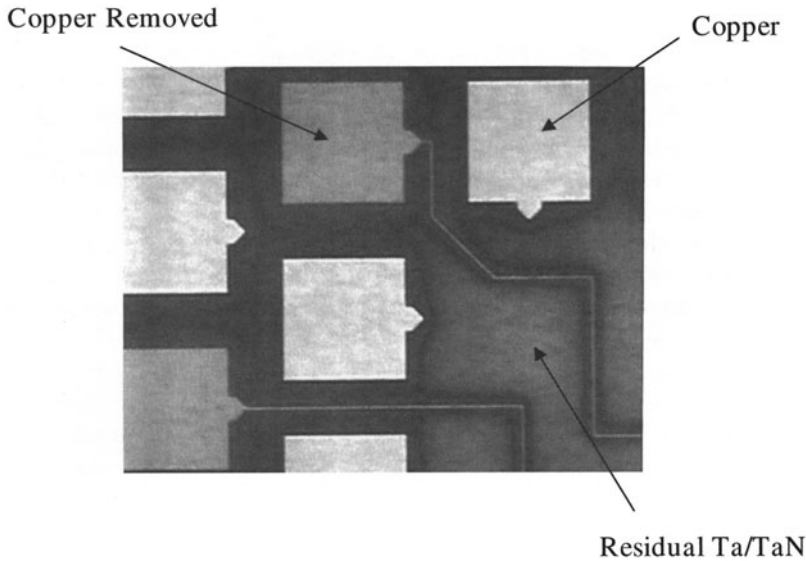
Fig. 10.49. Copper bonding pads with and without long leads

copper is present and the damascene structure remains intact in bond pads having no leads. (None of the leads or bond pads are electrically connected to the substrate.) Moreover, for this device, bond pads with and without leads were arranged around a larger central area in an alternating pattern. Clearly, this implies that excess copper removal from some bond pads cannot be attributed simply to dishing since this would affect all bond pads in exactly the same way. Thus, it seems likely that this behavior is due to the flow of current in the short-circuited cell. The difference between pads with and without leads arises because of the large difference in the perimeter to area ratio. Indeed, experiments have shown that a significant electrochemical potential can exist between tantalum or tantalum nitride and copper. In contrast, the potential between titanium nitride and copper tends to be much smaller. Moreover, since tantalum and tantalum nitride are generally removed with much more difficulty than either titanium nitride or tungsten alloys, any parasitic electrochemical cell formed can persist for a significant period of time during overpolishing, which exacerbates this problem.

A second significant issue facing anyone integrating copper interconnect into manufacturing is associated with completion of the damascene structure. Although, this has been touched upon elsewhere in this volume, it is worthwhile to consider it here in some detail. Of course, the simplest CMP process that can be envisioned for copper damascene formation is a "one step" process that removes copper and barrier metal in the field and stops selectively on the intermetal insulator. Unfortunately, this ideal has not been realized in practice. Although the issues can be quite complex, it is funda-

mentally a matter of process control. To understand this, one observes that because the copper overburden can be quite thick; a significant removal rate is required to achieve an acceptable processing time. Consequently, at the endpoint of any one step polishing process the margin between overpolishing and underpolishing becomes quite small and difficult to control. Obviously, this can be remedied by stopping short of endpoint and reducing the removal rate to increase the process margin. This is the so-called "soft landing" or "two step" approach. Even so, it would still seem that a process in which copper and barrier metal are removed selectively with respect to insulator would be desirable for the second step. However, this may not be the case. High selectivity tends to cause pattern topography to re-emerge during finish polishing and overpolishing. Therefore, even if, as discussed previously, the polished copper surface is rendered almost perfectly flat during intermediate stage polishing, the finished damascene structure may exhibit significant relief. Of course, the severity of topography re-emergence is highly process dependent, but in general, it is associated with high selectivity and high copper removal rate. In addition, pad hardness and conditioning can also affect topography re-emergence. Typically, use of a harder pad tends to reduce re-emergence, but increased conditioning tends to offset this (presumably by increasing density and dimension of pad asperities). Depending on circumstances, pattern relief in the finished damascene structure can be of the order of 100 nm. At first glance, this might not seem too severe, however, state-of-the-art integrated circuits now typically require more than five interconnect levels. If this amount of topography re-emergence occurs for each layer, then the aggregate relief after fabrication of several layers becomes unacceptably large.

An obvious method to preserve planarity is the use a nearly non-selective slurry for second step polishing, i.e., a so-called 1:1:1 process. (In practice, the selectivities required for best performance are within a factor of 3 or so of the 1:1:1 selectivity ratios.) In this case, copper, barrier metal, and insulator are all removed at the nearly the same rate and the planarity achieved during intermediate stage processing is transferred into the completed damascene structure. This is illustrated in Fig. 10.50.

However, as is almost always the case, there is no perfect solution to any engineering problem and trade-offs are required. Although a non-selective second step polishing process allows better planarity to be achieved for the finished damascene structure, variation in absolute conductor thickness and, hence, resistance, tends to be larger than for a selective process due to pattern density dependence. Naturally, this variation must be explicitly incorporated into metallization design rules. An alternative which allows the use of a selective second step process is the inclusion a subsequent insulator polishing step to remove residual topography. Obviously, this adds both cost and complexity to the overall manufacturing process and, as will become evident in what follows, may not be compatible with insulator materials having a low
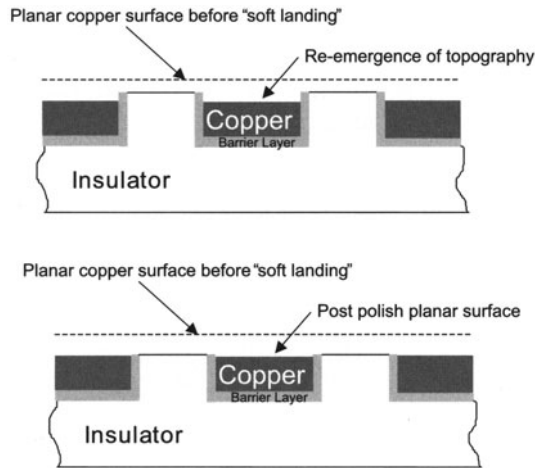
Fig. 10.50. Top: Selective second step process; Bottom: Non-selective second step process

dielectric constant. At this point, it suffices to observe that either selective or non-selective second step polishing processes can be used successfully for fabrication of copper damascene interconnect. However, it is of utmost importance to understand the associated trade-offs.

An additional, very important issue for integration of copper interconnect is the choice of intermetal insulator. As observed previously, just as there has been a strong motivation to use metals with lower resistivity, there has been an equally strong motivation to use insulators having lower dielectric constant. Unfortunately, in contrast to the selection of copper as a conductor metal, choice of insulator is much more difficult. To provide some background, it is useful to observe that vitreous silica (quartz glass) has a typical dielectric constant of about 4. By definition, a vacuum has a dielectric constant of exactly unity and air is only slightly higher than this value. This means that for an ideal parallel plate capacitor, the observed value of capacitance is four times higher if the gap between the plates is filled with vitreous silica rather than vacuum (or air).

One of the first approaches toward lowering the dielectric constant of the insulator was to add fluorine to vitreous silica during deposition to produce fluorosilicate glass (FSG). Early work indicated that a dielectric constant as low as 3.2 or 3.3 could be obtained using this material system. Unfortunately, fluorosilicate glass is highly moisture sensitive and FSG is unstable if it contains sufficient fluorine content to reduce the dielectric constant below about 3.5 or 3.6. Nevertheless, fluorosilicate glass has been used successfully and reliably with both conventional aluminum alloy and copper damascene metallization [103, 104].

Another material used with aluminum interconnect is polyimide, which is an organic polymer having a dielectric constant substantially lower than that of vitreous silica [105]. (The dielectric constant of polyimide is dependent on exact material formulation and processing conditions.) However, polyimide polymers have rarely been used with copper metallization [106]. The most likely reason for this is basic process incompatibility such as the use of oxygen plasmas to etch polyimide which result in severe oxidation of any exposed copper.

Subsequent to the development of these materials, which have been used with aluminum alloys, various other low-k insulators have been formulated and tested. Among these are amorphous boron nitride, amorphous fluorinated carbon [107], fluorinated polyarylene ether [108], silsesquioxanes [109], parylene-N, parylene-AF4, etc. These materials have been deposited by a variety of methods; however, CVD and spin-on processes are the most common [110]. While it has been possible to demonstrate dielectric constants of between 2 and 3 using these materials, insurmountable incompatibilities have been invariably encountered in any attempt to integrate these with copper.

At present, the most likely candidates for low-k dielectric are benzocyclobutene (BCB) polymers and related materials commonly sold under the trade name of SiLK^{TM} [111], and organosilicate glass (OSG) [112]. BCB and SiLK are pure organic materials, which are wet coated and thermally cured. According to the manufacturer, an effective dielectric constant as low as 2.5 can be practically achieved with these materials. In contrast, OSG is deposited by plasma enhanced CVD using and organosilane precursor, e.g., mono-, di-, tri-, or tetramethyl silane. Consequently, OSG can be viewed as similar to vitreous silica with organic functional groups substituted into the glass network structure. Generally, OSG has a dielectric constant of 2.5 to 3.

The major difficulty with any low-k dielectric material including BCB, SiLK^{TM}, and OSG, is that in order to obtain a low dielectric constant, other material properties are generally modified in an undesirable direction. For example, almost invariably mechanical strength and hardness of any low-κ insulator are much less than that of vitreous silica. This directly affects CMP because down forces and rotation rates must be lowered to avoid mechanical damage and scratching. This problem can be alleviated to some degree by the inclusion of a hard, thin stop or cap layer on top of the low-k insulator. However, any such hard material generally has a dielectric constant at least as large as that of vitreous silica. Hence, inclusion of such materials in any substantial amount immediately dilutes any beneficial effect of the low-k material. Another issue is layer-to-layer adhesion. Often, adhesion layers must be included between insulator layers to provide sufficient mechanical strength to prevent delamination during subsequent processing. (Ideally, stop and cap layers mentioned previously and adhesion layers should be the same; otherwise an undesirably complicated structure will be the result.) Finally, thermal stability of low-k materials is often poor in the temperature range necessary

for stabilization and recrystallization of the metal. This is particularly the case for the purely organic polymers. Taken together, it is clear that integration of copper damascene with an appropriate low-k insulator is not a trivial matter. Indeed, a conservative approach would be to use a low dielectric material only for the intralevel insulator to reduce C as illustrated in Fig. 10.45. A material having a higher dielectric constant, but better mechanical properties, e.g., vitreous silica, could be used as the interlevel dielectric constant, since, C' can be reduced simply by making the level-to-level spacing larger. (Of course, this requires the aspect ratio of via holes to increase, which naturally makes etching more difficult.) Obviously, the most simple integration scheme would be the use of only one kind of insulator; however, optimization of all mechanical and electrical properties for a single material is difficult.

All of these difficulties become even more severe if one contemplates lowering the dielectric constant below 2 as would seem to be required in the not too distant future. In this case, it is likely that solid materials cannot be used to achieve such a low effective dielectric constant and that porosity will have to be introduced into the material structure in some controlled way [113]. Obviously, this is likely to reduce mechanical strength still further. In addition, there will likely also be additional problems of moisture uptake, contamination during processing, sidewall coverage, etc.

In conclusion, some minor integration issues remain to be discussed. The first of these is that copper is easily oxidized in air, which over time results in a surface oxide layer, which appears as a conventional patina or tarnish. Moreover, the oxidation rate can be greatly increased by heating or chemical exposure. Therefore, care must be taken to avoid air oxidation during CMP or associated processing. Another integration issue is the possibility of electrochemical self-biasing because of photogenerated currents arising due to ambient light penetration into the semiconductor substrate. This effect can lead to selective, pattern repetitive corrosion of copper wiring and is illustrated in Fig. 10.51.

Of course, the substrate current arises from photogeneration of holes and electrons in high field regions associated with pn-junctions in the substrate. This current can flow up through the interconnect structure and out into the slurry or polishing solution during CMP. In the solution it may be carried by
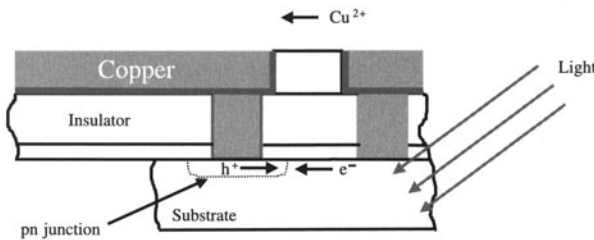
Fig. 10.51. Photogenerated corrosion current

aqueous copper ions formed by corrosion of the metallic copper surface. This problem is obviously eliminated by not allowing light to reach the substrate. Moreover, it is not such a problem for CMP itself, which by its nature tends to exclude light from the wafer surface during processing, but can become an important issue during wet handling and post CMP cleaning.

Naturally, it almost goes without saying that copper is an extremely undesirable contaminant for silicon, which severely degrades electrical performance of transistors. Therefore, it is of paramount importance to prevent copper contamination from propagating. In general, this is accomplished by implementation of rigorous handling procedures and scrupulous monitoring and cleaning of copper contamination, particularly on the backsides of wafers. Equipment manufacturers generally address these issues satisfactorily in the design of carriers, handling arms, etc. so that contamination is easily avoided by end users.

10.5 Other Applications of CMP

As is the case with other microfabrication process technologies, CMP was initially implemented for a particular, well-defined application, which specifically was interlevel dielectric planarization. Of course, there was also an initial skepticism and concerns about problems such as contamination and mechical damage. However, these have been largely overcome. As a consequence, CMP is now considered a full-fledged member of the complete suite of microfabrication processing techniques, which include photolithography, dry etching, chemical and physical vapor deposition, among others. As such, it has been implemented for other applications such as damascene copper interconnect, shallow trench isolation, etc. It is likely that this trend will continue unabated into the future with CMP being extended to different materials and new device structures.

However, before discussing new applications, it is of interest to consider an enhancement of an older application. In conventional STI (discussed earlier in this chapter), silicon nitride is generally used as the capping layer on the active areas of devices. The reason for this is partly historical and comes out of conventional LOCOS processing, which uses nitride as a mask for oxidation. Of course, once device islands are formed, the nitride must be stripped in conventional STI just as in LOCOS. In this case, the silicon surface is exposed and the gate oxide is formed by thermal oxidation. Because, no CMP process is ever perfect, gate oxide invariably must be grown over the edges or corners of device islands. Typically, this results in a reduced breakdown field for the oxide insulator at these points. In manufacturing, it is critical to control the morphology of the device edge in order to prevent loss of yield due to this falure mechanism. As an alternative, gate dielectric can be formed over the entire surface of the wafer and covered with polysilicon. Active areas are then patterned, etched, and oxide deposited. The oxide is polished back to expose

the polysilicon covering device areas, which is not removed as in the case for nitride in conventional STI, but is subsequently incorporated into the gate electrode. For conventional bulk CMOS processing, this scheme may provide only equivalent results. However, for fabrication of advanced devices using very thin epitaxial layers of silicon and, perhaps, silicon-germanium alloy, which would be disturbed during the nitride strip and gate oxide regrowth, "polysilicon-capped STI" provides an elegant solution. Indeed, this illustrates how even conventional CMP processes can be modified for use with advanced structures and materials [114].

It is unlikely that more compatible materials for electronic device fabrication than CVD polycrystalline silicon, thermally grown silicon dioxide, and single crystal silicon substrates could ever be found. Unfortunately, two fundamental limitations have now appeared that will probably require the use of different, less compatible materials in the future. Of course, continued transistor shrinkage will require gate capacitance per unit area to be made larger in order to obtain acceptable device characteristics. Therefore, if the gate insulator is limited to thermally grown silicon dioxide alone, this can only be achieved by reducing insulator thickness. Currently, for silicon dioxide the required thickness is rapidly approaching 1.5 nm or even less [115]. At such dimensions, even for a perfect, defect free silicon dioxide thin film, gate leakage current is unacceptably large due only to quantum mechanical tunneling. Clearly, this means that gate to channel leakage current can be reduced further only by replacing silicon dioxide with some different insulating material, having an intrinsically higher dielectric constant, which allows a larger physical thickness of the gate insulating layer to be used without a corresponding reduction in capacitance. At present, various heavy metal, e.g., hafnium and zirconium, oxides and silicates are under investigation as possible high-κ insulators. A second problem arises because, even if heavily doped, polysilicon remains a semiconductor and under conditions of high electric field, a carrier depletion layer is invariably present within the polysilicon at the interface of the gate insulator and electrode. This layer of depleted polysilicon acts essentially as additional insulator and, hence, makes a contribution that invariably reduces overall gate capacitance. Of course, this has always been the case, but in the past the effect of polysilicon depletion on overall gate capacitance has been insignificant in comparison to that of the gate insulator itself. However, this is no longer the case for projected values of gate insulator thickness required to scale transistor critical dimensions below 0.10 μm. Again, the required solution is the use of new materials to fabricate the gate electrode; in this case metals for which there are no significant carrier depletion effects.

Obviously, the practical details of process integration are critical to introduction of any new material to mainstream semiconductor device fabrication. Moreover, it has long been known that even trace amounts of most metallic elements seriously degrade the electrical properties of semiconductors. There-

fore, especially in semiconductor processing for which high temperatures are necessary, avoidance of metallic contamination of electrically active parts of the silicon substrate, i.e., transistor channel regions, source-drain junctions, etc., is critically important and requires rigorous cleanliness and scrupulous elimination of sources of metallic contamination. Of course, silicon cannot contaminate itself, which is, perhaps, the most fundamental advantage of conventional polysilicon gate CMOS device technology. Therefore, with very few exceptions, metals and metal oxides cannot be simply substituted for doped polysilicon and thermally grown oxide in gate electrode and insulator structures. Indeed, it is evident that such materials can only be introduced after all high temperature processing is effectively completed, that is to say, after dopants have been implanted and diffused and all junctions formed in the device. Unfortunately, if the gate electrode is not patterned until after implantation and diffusion are completed, no self-alignment between the gate electrode and source-drain junctions is possible. However, self-alignment of the gate/source-drain structure is precisely one of the most important enabling techniques that has allowed scaling of transistor devices to current deep submicron dimensions. Thus, a return to conventional photlithographic alignment would present severe, if not insurmountable, difficulties for device fabrication. In addition, even if the difficulties in patterning could be overcome, etching of a "new" gate electrode material and at the same time stopping on an also "new" and very thin dielectric layer would still be required in analogy to conventional polysilicon gate CMOS processing. Indeed, a great amount of effort has been previously expended on successful development of such etch processes for polysilicon and silicon dioxide but, there is no guarantee that similar results are achievable for metals and high-k insulators.

A possible approach that can preserve self-alignment of the gate electrode and source-drain after high temperature processing is the use of a "gate cast" or dummy gate, which is replaced by metal and high-k insulator later in the process [116]. This scheme is illustrated in Fig. 10.52.

In this case, polysilicon, silicon nitride, or some other non-contaminating material is used to fabricate a dummy gate electrode. Therefore, conventional self-aligned patterning and etching can be used to define the electrode structure. Implantation, diffusion, etc., then proceed as usual. Following this a blanket layer of silicon oxide is deposited over the structure and densified. This oxide layer is polished back to expose the top of the dummy gate using a CMP process quite similar to conventional STI. The dummy gate is removed by a selective chemical etch and the surface of the substrate is prepared for dielectric deposition. Next, the high-k insulator and metal are deposited and are then polished back to complete the structure. Of course, many detailed aspects of such an integration scheme remain to be determined and one can expect a number of associated difficulties. Among other things, deposition processes for both metal and insulators remain to be defined. (At present, atomic layer deposition, or ALD, appears attractive.) Moreover, it remains to
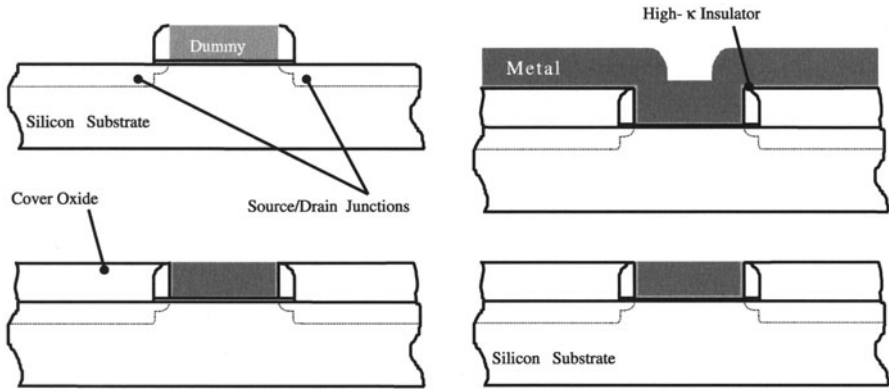
Fig. 10.52. Damascene gate fabrication

be seen if polishing and selective etching processes can be developed having sufficient process control. An added complication is that implementation of dual gate CMOS will require integration of two different metals having large and small work functions suitable for application to p-channel and n-channel devices, respectively. For n-channel devices, many different metals exist, e.g., titanium, titanium nitride, niobium, etc., which have sufficiently small work functions so as to be of possible use as the gate electrode. In contrast, only a few, generally noble, metals such as iridium and platinum have work functions large enough to be useful as gate electrodes for p-channel devices. Even so, CMP should provide the best method for optimized removal of all of these diverse materials and working devices have been demonstrated [117].

An application of CMP closely related to fabrication of damascene metal gate, high-k insulator transistors, is fabrication of so-called one transistor non-volatile ferroelectric memory [118]. Conceptually, these devices are similar to stacked gate flash memory, except that the upper dielectric layer is replaced by a ferroelectric layer and the floating gate is replaced by metal or is entirely absent altogether. To be more specific, in contrast to an ordinary dielectric, electrical polarization of a ferroelectric material exhibits hysteresis with respect to applied electric field. Therefore, dielectric polarization persists even in the absence of an applied field and the memory is non-volatile.

Ferroelectric materials are generally mixed heavy metal oxides having the perovskite structure, e.g., lead zirconium titanate (PZT), strontium bismuth tantalate (SBT), lead germanium oxide (PGO), etc. Moreover, noble metals, such as platinum and iridium are required for electrodes to prevent interfacial chemical reactions, which degrade retention properties of the ferroelectric. In addition, it is found that ion bombardment during dry etching can easily damage the ferroelectric layer, which also degrades its performance. Therefore, to avoid this kind of damage a damascene structure formed by CMP is an attractive alternative for the bottom electrode, the ferroelectric layer,

or both. The initial trench can be formed directly by dry etching or by variations of the dummy gate process as required by other integration issues. Typically, polishing characteristics of ferroelectric materials are similar to soft glasses. Even so, mixed oxide ferroelectric materials are quite complex chemically and considerable care must be taken to achieve acceptable results. Therefore, it becomes further evident that the flexibility inherent CMP allows removal processes for ferroelectric materials and/or noble metals to be highly optimized [119]. This is a clear advantage of CMP over more conventional material removal processes and is illustrated by fabrication of a small platinum bottom electrode directly on a transistor channel, which appears in Fig. 10.53.

This platinum electrode is self-aligned to source-drain diffusions in the direction parallel to the channel. In addition, it is self-aligned to the active area edge in the direction perpendicular to the channel.

Of course, in the future advanced CMP processes for metal interconnect will be necessary for integration of copper or even, perhaps, silver metallization with ultra low dielectric constant insulators. At present, solid materials having dielectric constants of between 2.5 and 3 are coming into use, but are much less structurally robust than conventional silicon oxide. Moreover, as mentioned previously, it seems likely that use of porous materials or even air gaps will be required to reduce the effective dielectric constant below 2. In this case, the mechanical strength and structural integrity of these materials presents a huge challenge for CMP and reduction of applied down force during damascene copper polishing is a priority for accommodation of fragile materials. However, this comes at the cost of reduction of throughput due to lower removal rates and more difficult control of thickness uniformity. Naturally, use of a more aggressive chemistry can increase removal rate, however this is likely to also be accompanied by increased corrosion. It seems that im-
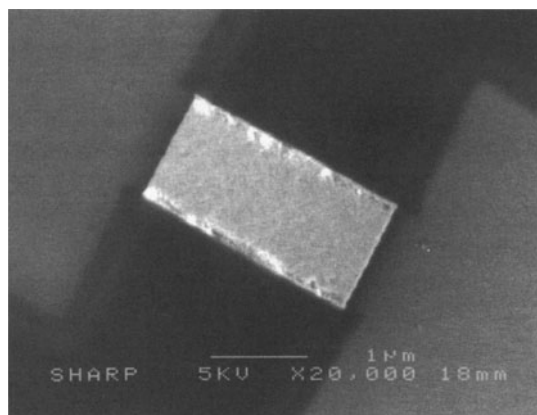
Fig. 10.53. Doubly self-aligned platinum electrode formed by CMP

provements in polishing pad characteristics combined with advanced machine control systems will become necessary. Indeed, "no contact" removal of copper by electropolishing has recently become an active area of interest. This is essentially the reversal of electroplating and it remains to be seen whether or not this will prove to be a viable alternative to more conventional CMP [120].

Finally, it is worthwhile to note that CMP finds ready application in fabrication of advanced substrate structures involving silicon-germanium alloys or silicon-on-insulator (SOI). At the very least CMP is useful to improve surface planarity following epitaxial deposition or wafer bonding [121]. Obviously, this usage is closely related to more ordinary, bulk silicon applications. In addition, CMP of polysilicon finds application in fabrication of passive components such as capacitors, as well as for fabrication of thin film transistors for displays and structural components of MEM's. Likewise, CMP of other materials, perhaps even organic polymers such as polyimide, may be useful in advanced packaging schemes.

References

1. K. Nakamura, "ULSI Technology", C.Y. Chang and S.M. Sze (eds.), McGraw–Hill, New York, 1996, pp 272–282.
2. C.Y. Lu and W.Y. Lee, "ULSI Technology", C.Y. Chang and S.M. Sze (eds.), McGraw–Hill, New York, 1996, pp 510–512.
3. R. Liu, "ULSI Technology", C.Y. Chang and S.M. Sze (eds.), McGraw–Hill, New York, 1996, pp 412–433.
4. A.E. Braun, Semiconductor International, **23** (12), 71, 2000.
5. L. Peters, Semiconductor International, **25** (2), 64, 2002.
6. P. Burggraaf, Semiconductor International, **18** (13), 74, 1995.
7. R.R. Divecha, B.E. Stine, D.O. Ouma, E.C. Chang, D.S. Boning, J.E. Chung, O.S. Nakamura, H. Aoki, G. Ray, D. Bradbury and S.Y. Oh, J. Electrochem. Soc., **145** (3), 1052, 1998.
8. D. Pramanik and M. Weiling, in Chemical Mechanical Planarization I, (eds.) I. Ali and S. Raghavan, ECS Proceedings, **96–22**, 47, 1997.
9. J.M. Steigerwald, S.P. Muraka and R.J. Gutmann, "Chemical Mechanical Planarization of Microelectronic Materials", John Wiley & Sons, Inc., New York, 1197, pp 173–174.
10. W.J. Patrick, W.L. Guthrie, C.L. Standley and P.M. Schiable, J. Electrochem. Soc., **138** (6), 1778, 1991
11. C. Oji, B. Lee, D. Ouma, T. Smith, J. Yoon, J. Chung and D. Boning, J. Electrochem. Soc., **147** (11), 4307, 2000.
12. B.E. Stine, D. Ouma, R.R. Divecha, D.S. Boning, J.E. Chung, D.L. Hetherington, C.R. Harwood, O. S Nakayama and S.Y. Oh, IEEE, **11** (1), 128, 1998.
13. D. Ouma, D. Boning, J. Chung, G. Shinn, L. Olsen and J. Clark, in Proceedings of IITC, pp 67–69, 1998.
14. C.H. Yao, D.L. Feke, K.M. Robinson and S. Meikle, J. Electrochem. Soc., **147** (4), 1502, 2000.
15. C.H. Yao, D.L. Feke, K.M. Robinson and S. Meikle, J. Electrochem. Soc., **147** (8), 3094, 2000.

16. P.A. Burke, Proceedings 1991 VMIC Conference, 379, IMIC, Tampa, 1991.
17. S.R. Runnels, J. Electrochem. Soc., **141**, 1900, 1994.
18. J. Warnock, J. Electrochem. Soc.,**138**, 2398, 1991.
19. H.P. Tuinlot and M. Vertregt, IEEE Trans. Semicond. Manufact., **14** (4), 302, 2001.
20. W.J. Schaffer, J.W. Westphal, H.W. Fry, P.J. Parikh and J.D. Lee, Proceedings 1996 CMP–MIC Conference, 299, IMIC, Tampa, 1996.
21. S.J. Fang, S. Garza, H. Guo, T.H. Smith, G.B. Shinn, J.E. Campbell and M.L. Hartsell, J. Electrochem. Soc., **147** (2), 682, 2000.
22. D. Castillo–Mejia, A. Perlov and S. Beaudoin, J. Electrochem. Soc., **147** (12), 4671, 2000.
23. D. Wang, J. Lee, K. Holland, T. Bibby, S. Beaudoin and T. Cale, J. Electrochem. Soc., **144**, 1121, 1997.
24. C. Srinivasa–Murthy, D. Wang, S.P. Baudoin, T. Bibby, K. Holland and T.S. Cale, J. Electrochem. Soc., 308–309, 533, 1997.
25. S.R. Runnels, I. Kim, J. Schleuter, C. Karlsru and M. Desai, IEEE Trans. Semicon. Manuf., **11** (3), 501, 1998.
26. R. Jairath, A. Pant, T. Mallon, B. Withers and W. Krussell, Solid State Technol., **39**, 107, 1996.
27. A.E. Braun, Semiconductor International, **24** (6), 70, 2001.
28. L.C. Klein, "Thin Film Processes II", 537, J.L. Vossen and W. Kern (eds.), Academic Press Inc., Boston, 1991.
29. K. Sakaguchi,T. Yonehara, Solid State Technology, **43** (6), 88, 2000.
30. J.F. Miner, W.Y-C. Lai and M. Hoffman, Proceedings 1995 VMIC Conference, 478, IMIC, Tampa, 1995.
31. Y. Zhang, P. Parikh, B. Stephenson, M. Bonsaver, J. Ling and M. Li, Proceedings 1996 VMIC Conference, 424, IMIC, Tampa, 1996.
32. W.T. Tseng, Y. H, Wang and J.H. Chin, J. Electrochem. Soc., **146** (11), 4273, 1999.
33. V. Sukharev J. Electrochem. Soc., **148** (3), G172, 2001.
34. W.T. Tseng, Y.T. Hsieh, C.F. Lin, M.S. Tsai and M.S. Feng, J. Electrochem. Soc., **144** (3), 1100, 1997.
35. L. Shi, L. Veltman, B. Zhang, U. Winkler, F. Verstraete and R. Schreutelkamp, Semiconductor International, **24** (8), 183, 2001.
36. T. Shadwick, B. Cote, W. Landers, D. Miura, B. Vollmer, K. Feldner, R. Cheek and M. Rutten, Proceedings 1995 VMIC Conference, 511, IMIC, Tampa, 1995.
37. K. Nicholes, R. Singh, D. Grant and M. Litchy, Semiconductor International, **24** (8), 201, 2001.
38. J. Teshima, Semiconductor International, **24** (8), 171, 2001.
39. C.Y. Yang and T.S. Chao, "ULSI Technology", C.Y. Chang and S.M. Sze (eds.), 60, McGraw–Hill, New York, 1996.
40. K.J. Cen, H.B. Lu and J.T. Lin, Proceeding 1997 VMIC Conference, 339, IMIC, Tampa, 1997.
41. W. Kroeninger, W. Redl, U. Hoeckele, M. Frank, M. Deshpande, W.F. Yau, W. Krogner and W. Rausch, Proceedings 1997 VMIC Conference, 606, IMIC, Tampa, 1997.
42. S.K. Tang, V.Y. Vassiliev, S. Mridha and L.H. Chan, Thin Solid Films, **352**, 77, 1999.
43. M.A. Jaso, J.P. Gambino, D.M. Dobuzinsky, M. Armacost and T. Ohiwa, Proceedings 1996 VMIC Conference, 407, IMIC, Tampa, 1996.

44. J.H. Han, D.U. Choi, H.S. Kim, B.H. Roh and J.W. Park, Proceedings 1997 VMIC Conference, 331, IMIC, Tampa, 1997.
45. Z. Lin, C.J. Sparos, L.S. Milon and Y.T. Lin, IEEE Trans. Semi. Manuf., **11** (4), 557, 1998.
46. P. Renteln and J. Coniff, Proceedings Mat. Res. Soc. Symp., **337**, 105, 1994.
47. Y.J.T. Li, "ULSI Technology", C.Y. Chang and S.M. Sze (eds.), 363, McGraw–Hill, New York, 1996.
48. W.C. Chen, S.C. Lin, B.T. Dai and M.S. Tsai, J. Electrochem. Soc, **146** (8) 3004, 1999.
49. H. Cui, I.B. Bhat, S.P. Muraka, H. Lu, W. Li, W.J. Hsia and W. Cataby, J. Electrochem. Soc., **147** (10), 3816, 2000.
50. C.L. Borst, D.G. Thakurta, W.N. Gill and R.J. Gutmann, J. Electrochem. Soc., **149** (2), G118, 2002.
51. Y.D. Kim, C.W. Nam, S.B. Kim, S.D. Kim and S.H. Yu, Conference Proceedings USI XIII, 747, MRS, Warrendale, PA, 1998.
52. A. Ishii, A. Ohsaki, Y. Takata, N. Morimoto, K. Maekawa, K. Mori, T. Tsutumi, Y. Mashiko, M. Hirayama and A. Inuishi, Proceedings 1997 VMIC Conference, 19, IMIC, Tampa, 1997.
53. E. Atakov, T.S. Sriram, D. Dunnel, S. Pizzanello, A. Ohsaki and K. Maekawa, Proceedings 1997 VMIC Conference, 473, IMIC, Tampa, 1997.
54. R. Liu, "ULSI Technology", C.Y. Chang and S.M. Sze (eds.), 371, McGraw–Hill, New York, 1996.
55. Z. Wang, W. Catabay, J. Yuan, J. Ku, N. Krishna, V. Pavate, A. Sudararajan, S. Saigal, B. Chang, M. Narashimhan, J. Egermeier and S. Ramswami, Proceedings 1997 VMIC Conference, 258, IMIC, Tampa, 1997.
56. G.F. Hudson and R.L. Elliott, Proceedings 1995 VMIC Conference, 514, IMIC, Tampa, 1995.
57. J.B. Choi, S. Hahn, J.W. Park and J.J. Kim, Proceedings 1997 VMIC Conference, 319, IMIC, Tampa, 1997.
58. C. Streinz, S. Frumbine, C. Yu, A. Zutshi, D. Schey and P. Meyers, Proceedings 1997 VMIC Conference, 313, IMIC, Tampa, 1997.
59. K. Wijekoon, R. Lin, S. Yang, F. Redeker, S. Nanjangud, M. Bakshi and S. Ghanayem, Proceedings 1998 VMIC Conference, 451, IMIC, Tampa, 1998.
60. D.J. Stein, D. Hetherington, R. Guilinger and J.L. Cecchi, J. Electrochem. Soc., **145** (9), 3190, 1998.
61. E.A. Kneer, C. Raghunath, V. Mathew, S. Raghavan and J.S. Jeon, J. Electrochem. Soc, **144** (9), 3041, 1997.
62. J. Zabasajja, R. Merchant, B. Ng, S. Banerjee, D. Green, S. Lawing and H. Kura, J. Electrochem. Soc., **148** (2), G73, 2001.
63. G.C. Lee, M. Weling, C. Drill and A. Hu, Proceedings 1997 VMIC Conference, 304, IMIC, Tampa, 1997.
64. E. Sicurani, M. Fayolle, Y. Gobil, Y. Morand and F. Tardif, Conference Proceedings USI XII, 561, MRS, Warrendale, PA, 1997.
65. C. Yu, T. Meyers and C. Streinz, Conference Proceedings USI XII, 519, MRS, Warrendale, PA, 1997.
66. J. Mendonca, C. Dang, C. Pettinato, J. Cope, H. Garcia, J. Saravia, J. Farkas, D. Watts and J. Klein, Proceedings of the 1998 IITC, 196, IEEE Electron Devices Society.
67. M. Rutten, P. Feeney, R. Cheek and W. Landers, Proceedings 1995 VMIC Conference, 491, IMIC, Tampa, 1995.

68. P. Holzapfel, K. Murella, J. Schlueter and C. Johnson, Proceedings 1997 VMIC Conference, 307, IMIC, Tampa, 1997.
69. S.H. Li, V. Bucha, B. Miller and K. Wooldridge, Proceedings 1998 VMIC Conference, 490, IMIC, Tampa, 1998.
70. J.T. Yue, " ULSI Technology", C.Y. Chang and S.M. Sze (eds.), 656, McGraw–Hill, New York, 1996.
71. S. Melosky, S. Li, J. Ling and C. Spinner, Proceedings 1997 VMIC Conference, 523, IMIC, Tampa, 1997.
72. S. Nag, A. Chatterjee, Solid State Technology, 129, Sept. 1997.
73. N. Balasubramanian, E. Johnson, I.V. Peidous, S. Ming-Jr and R. Sundarean, J. Vac. Sci. Technol. **B 18** (2), 2000.
74. Y.S. Chung, C.W. Jeon, J.H. Kim, S.K. Han, J.W. Hwang, S.Y. Kim, J.G. Lee, I.S. Hyun, J. Vac. Sci. Technol. **B 18** (1), 2000.
75. M. Nandakumar, A. Chatterjee, S. Sridhar, K. Joyner, M. Rodder, I.C. Chen, 133, IEDM, 1998.
76. D. Shaymirian, IMEC, Private Communication, 2000.
77. S. Lassig, C.S. Xu, A.J. Miller, S. Kamath, A. Romano and T. Kudo, Solid State Technology, 157, July 2000.
78. A. Chatterjee, I. Ali, K. Joyner, D. Mercer, J. Kuehne, M. Mason, A. Esquivel, D. Rogers, S. O'Brien, P. Mei, S. Murtaza, S.P. Kwok, K. Taylor, S. Nag and G. Hames, J. Vac. Sci. Technol. **B 15**, 1936, 1997.
79. C.P. Chang, C.S. Pai, F.H. Baumann, C.T. Liu, C.S. Rafferty, M.R. Pinto, E.J. Lloyd, M. Bude, F.P. Klemens, J.F. Miner, K.P. Cheung, J.I. Colonell, W.Y.C. Lai, H. Vaidya, S.J. Hillenius, R.C. Liu and J.T. Clemens, Tech. Dig. Int. Electron Devices Meeting, 661, 1997.
80. G. Badenes, IMEC, Private Communication, 2000.
81. S. Matsuda, T. Sato, H. Yoshimura, Y. Takegawa, A. Sudo, I. Mizushima, Y. Tsunashima and Y. Toyoshima, 137, IEDM 1998.
82. P. Van Cleemput, H.W. Fry, B. van Schravendijk and W. van den Hoek, Semiconductor International, 179, July 1997.
83. M. Schaekers and E. Sleeckx, IMEC, Private Communication, 2000.
84. J.T. Pan, P. Li, F. Redeker, J. Whitby, D. Ouma, D. Boning and J. Chung, Proceedings 1998 VMIC Conference, 467, IMIC, Tampa, 1998.
85. G.H. Koh, D.W. Ha, C.H. Cho, H.S. Jeong, G.T. Jeong, W.S. Yang, K.H. Lee, J.G. Lee, B.J. Park, J.K. Lee, J.S. Bae, J.H. Sim and K.N. Kim, Proceedings 1998 CMP–MIC Conference, 15, IMIC, Tampa, 1998.
86. K. Smekalin, Solid State Technology, 187, July 1997.
87. J. Grillaert, M. Meuris, N. Heylen, D. DeVriendt, E. Vrancken and M. Heyns, Proceedings 1998 CMP–MIC Conference, 79, IMIC, Tampa, 1998.
88. N. Heylen et al., Proceedings of 2nd Int. Symp. on CMP, ECS, 1998.
89. B. Davari, C.W. Koburger, R. Schultz, J.D. Warnock, T. Furukawa, M. Jost and Y. Taur, W.G. Schwittek, J.K. DeBrosse, M.L. Kerbaugh and J.L. Mauer, IEDM Tech. Digest, 61, 1989.
90. S.S. Cooperman, A.I. Nasr and G.J. Grula, J. Electrochem. Soc, **142**, 3180, 1995.
91. J.M. Boyd and J.P. Ellul, J. Electrochem. Soc, **143**, 3718, 1996.
92. J.M. Boyd and J.P. Ellul, J. Electrochem. Soc. **144**, 1838, 1997.
93. S.M. Yang, Y.H. Chen, C.L. Chang, C.H. Yu and J.Y. Chen, Proceedings 1997 CMP–MIC Conference, 186, IMIC, Tampa, 1997.

94. T. Vo, T. Buley and J. Gagliardi, Solid State Technology, June 2000, 123.

95. B. Luther, J.F. White, C. Uzoh, T. Cacouris, J. Hummel, W. Guthrie, N. Lustig, S. Greco, N. Greco, S. Zuhoski, P. Agnello, E. Colgan, S. Mathad, L. Saraf, E.J. Weitzman, C.K. Hu, F. Kaufman, M. Jaso, L.P. Buchwalter, S. Reynolds, C. Smart, D. Edelstein, E. Baran, S. Cohen, C.M. Knoedler, J. Malinowski, J. Horkans, H. Deligianni, J. Harper, P.C. Andricacos, J. Paraszczak, D.J. Pearson and M. Small, Proc. VMIC X, 15 (1993).

96. W. Lee, H. Yang and J. Lee, Proc. of the Electrochem. Soc., **2000–27**, 63, 2001.

97. D. Summers, Sol. State Technol., **26** (6), 137 (1983).

98. S. Wolf, *Silicon Processing for the VLSI Era – vol. 2*, 176, Lattice Press, Sunset Beach, CA, 1990.

99. R.C. Weast, Ed., *Handbook of Chemistry and Physics*, F-146, CRC Press, Boca Raton, FL, 1990.

100. M.T. Bohr, Proc. IEEE IEDM, 241, 1995.

101. C. Verove, B. Descouts, P. Gayer, M. Guillermet, E. Sabouret, P. Spinelli and E. Van der Vegt, Proceedings 2000 International Interconnect Technology Conference, 267, 2000.

102. David R. Evans, *AVS 1st Int. Conf. on Microelec. and Interfaces*, Santa Clara, CA, Feb. 9, 2000.

103. E.P. Barth, T.H. Ivers, P.S. McLaughlin, A. McDonald, E.N. Levine, S.E. Greco, J. Fitzsimmons, I. Melville, T. Spooner, C. DeWan, X. Chen, D. Manger, H. Nye, V. McGahay, G.A. Biery, R.D. Goldblatt and T.C. Chen, Proceedings 2000 International Interconnect Technology Conference, 219, 2000.

104. D. Carl, S. Schuchmann, M. Kilgore, R. Swope and W. van den Hoek, Proc. VMIC XII, 97, 1995.

105. R.V. Joshi, M. Jaso, H. Ng, L. Hsu, H. Dalal and P. Klymco, *Proc. IEEE VMIC VIII*, 75, 1991.

106. C.K. Hu, M.B. Small, F. Kaufmann and D.J. Pearson, "Tungsten and Other Advanced Metals for VLSI/ULSI Applications V", 357, MRS, Pittsburg, PA, 2000.

107. T.W. Mountsier and D. Kumar, Mat. Res. Soc. Proc. **443**, 41, 1996.

108. N.H. Hendricks, B. Wan and A.R. Smith, Proceedings 1995 DUMIC Conference, 283, IMIC, Tampa, 1995.

109. R.A. Donaton, F. Iacopi, M.R. Baklanov, D. Shamiryan, B. Coenegrachts, H. Struyf, M. Lepage, M. Meuris, M. Van Hove, W.D. Gray, H. Meynen, D. De Roest, S. Vanhaelemeersch and K. Maex, Proceedings International Interconnect Technology Conference, 93, 2000.

110. D. Pramanik, J.V. Tietz and K. Schiebert, Proc. VMIC X, 329 (1993).

111. G. Passemard, J.C. Maisonobe, C. Maddalon, A. Achen, M. Assous, C. Lacour, N. Lardon, R. Blanc and O. Demolliens, "Advanced Metallization Conference 1999", pg. 357, MRS, Pittsburgh, PA, 2000.

112. M.J. Loboda, "Advanced Metallization Conference 1999", pg. 371, MRS, Pittsburgh, PA, 2000.

113. H. Hanata, M. Miyamoto, T. Kamata, S. Matsumo and T. Tanabe, Proceedings International Interconnect Technology Conference, 61, 2000.

114. David R. Evans, *2000 Clarkson Univ. Int. CMP Symp.*, Lake Placid, NY, Aug. 13–17, 2000.

115. *International Technology Roadmap for Semiconductors*, 2001 Edition, http://public.itrs.net/Files/2001ITRS/Home.htm.
116. D. Evans and S.T. Hsu, U.S. Patent 6,133,106, 2000.
117. Y. Ma, D.R. Evans, T. Nguyen, Y. Ono and S.T. Hsu, IEEE Elec. Dev. Lett., **20** (5), 254, 1999.
118. T.K. Li, S.T. Hsu, B.D. Ulrich, L. Stecker, D.R. Evans and J.J. Lee, IEEE Elec. Dev. Lett., **23** (6), 339, 2002.
119. D. Evans, U.S. Patent 6,290,736, 2001.
120. D.H. Wang, M. Afnan and S.S. Chiao, Semicond. Fabtech, **13**, 255 (2001).
121. M.T. Currie, S.B. Samavedam, T.A. Langdo, C.W. Leitz and E.A. Fitzgerald, Appl. Phys. Lett. **72**, 1718, 1998.

Appendix: Pourbaix Diagrams

Oxidation-reduction chemistry of a given metal (or other material) in an aqueous medium is usually not dominated by a single process, but typically involves several different pH dependent half-reactions. In addition, solid oxides or hydroxides as well as dissolved species may be important and the chemistry further complicated by associated homogeneous and heterogeneous equilibria. A *Pourbaix diagram* includes all important half-reactions and equilibria and, thus, provides a convenient method for summarizing overall oxidation-reduction chemistry [1].

By convention, on a Pourbaix diagram pH appears on the horizontal axis and potential appears on the vertical axis. (Alternatively, thermodynamic free energy could be associated with the vertical axis.) Within this context, one may consider the potential as "applied" to the chemical system of interest. As a practical matter, this can be done directly by external application of an electrical potential to a metallic electrode or chemically by addition of a suitable oxidizing agent, e.g., hydrogen peroxide. Now, for a sufficiently negative value of potential, i.e., a reducing environment, the metal remains unoxidized, i.e., in an oxidation state of zero. However, if at some specified pH the value of the potential is increased arbitrarily, then one finds that there is some potential at which the metal is converted to an oxidized form. Usually, this oxidized form corresponds to the lowest commonly occurring oxidation state of the metal, but this is not always the case. Of course, if the potential is raised still further, more highly oxidized forms can be expected to appear. This continues until the highest commonly occurring oxidation state is reached; above which there is no further oxidation. The conversion potentials for these various oxidized forms are obtained using the Nernst equation and are plotted versus pH on the corresponding Pourbaix diagram as diagonal or horizontal line segments. In addition, equilbria between different species for which the metal is in the same oxidation state appear as vertical line segments [2, 3].

Naturally, actual Pourbaix diagrams are specific for a given material and can be relatively simple or quite complicated. However, as an illustrative example, a Pourbaix diagram for a hypothetical metallic element, M, that typically forms divalent compounds, i.e., in which M is in an oxidation state of

2, and that to a much lesser degree also forms tri- and tetravalent compounds (oxidation states of 3 and 4), might appear as in Fig. A1.

By definition, M denotes unoxidized metal, M^{2+} represents an aqueous divalent metal ion, and MO, M_2O_3, and MO_2 represent di-, tri-, and tetravalent oxides, respectively. (Generally, M^{2+} corresponds structurally to a so-called *aquo cation* in which water molecules are directly bonded to a central metal ion.) Clearly, the six different diagonal or horizontal line segments that appear are obtained from the Nernst equation and are formally identified with six corresponding half reactions:

$$M^{2+} + 2e^- \to M$$
$$MO + 2H^+ + 2e^- \to M + H_2O$$
$$M_2O_3 + 2H^+ + 2e^- \to 2MO + H_2O$$
$$2MO_2 + 2H^+ + 2e^- \to M_2O_3 + H_2O$$
$$M_2O_3 + 6H^+ + 2e^- \to 2M^{2+} + 3H_2O$$
$$MO_2 + 4H^+ + 2e^- \to M^{2+} + 2H_2O.$$
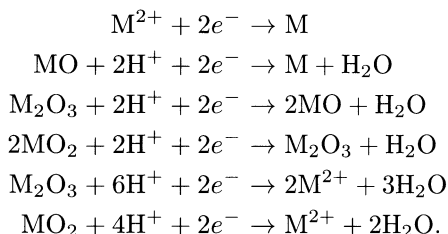
Naturally, the horizontal line segment corresponds to the half reaction in which hydrogen ions do not explicitly appear, i.e., n_H vanishes. The remaining vertical line segment corresponds to the hydrolysis equilibrium:

$$MO + 2H^+ \rightleftarrows M^{2+} + H_2O.$$
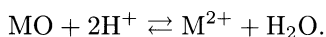
The equilibrium constant has the usual form and is easily obtained from published thermodynamic data. Obviously, this equilibrium depends on pH and is independent of applied potential.
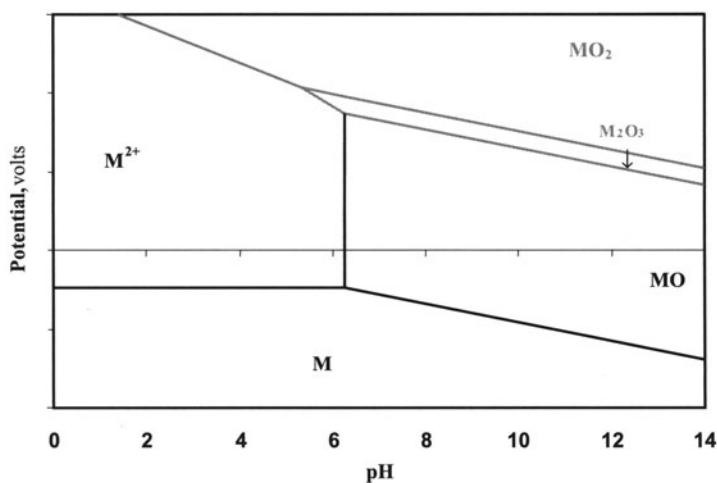
Fig. A1. Pourbaix diagram for hypothetical element M in water ($[M^{2+}]$ unit molar)

It is clear from the Pourbaix diagram that in the absence of an oxidizing agent or for a sufficiently negative applied potential, M remains in unoxidized metallic form. Now, in a strongly alkaline environment, if either the potential is increased or a suitable oxidizing agent is introduced, then M is oxidized and formation of one (or more) of the solid oxides, MO, M_2O_3, and MO_2, is to be expected. Of course, just which oxide will be formed depends on the value of the applied potential or the strength of the oxidizing agent. Furthermore, in an aqueous environment, metal oxides may be hydrated, therefore MO, M_2O_3, and MO_2, are, perhaps, more correctly considered as hydroxides or more generally just as designations for the appropriate oxidation state. (Indeed, if one formally replaces oxides with corresponding hydroxides, one finds relatively minor changes to the Pourbaix diagram in most cases.) However, irrespective of the exact chemical identity of the oxidized material, formation of solid oxides (or hydroxides) may be expected to suppress further oxidation and, thus passivate the metal surface. In contrast, oxidation of M in a strongly acidic environment results in immediate formation of the soluble aqueous ion, M^{2+}. Clearly, in an acidic environment one expects the metal, M, to dissolve readily under oxidizing conditions.

Application of Pourbaix diagrams to corrosion phenomena is fairly obvious. Indeed, it is clear that for the hypothetical metal, M, corrosion can be expected in an acidic environment, but at least some degree of passivity is expected in an alkaline environment. Now, if one accepts the Kaufmann model (or some suitable modification thereof) and in the absence of other factors, then one may conclude that successful CMP of M can be carried out in a neutral to alkaline environment. This is not to say that this is a strict requirement for acceptable results to be obtained, but this condition does serve to suppress direct dissolution of the metal surface, i.e., etching, and hence, to promote formation of a passivated surface. Furthermore, a cursory examination of the Pourbaix diagram suggests that in the absence of abrasion, the metal surface passivates over a fairly wide pH range. This is desirable for CMP since it implies that pure chemical etching can be easily avoided and that the process window should be reasonably large.

However, the situation is not really so simple as this. Formation and/or dissolution of an oxidized surface layer are strongly dependent on the hydrolysis equilibrium and collateral dependence of the equilibrium on the concentration of dissolved metallic species, e.g., M^{2+}. If instead of standard unit molar conditions, the concentration of M^{2+} is reduced to 10^{-6} molar, the resulting Pourbaix diagram (Fig. A2) shows a shift in the equilibrium to more alkaline values by three pH units.

Of course, if the concentration is lowered further, the equilibrium shifts to still higher pH values at a rate of one pH unit for every two orders of magnitude reduction in M^{2+} concentration. Therefore, it is evident that the concentration of dissolved metallic species directly affects the modified surface layer and, hence, can directly affect CMP characteristics.
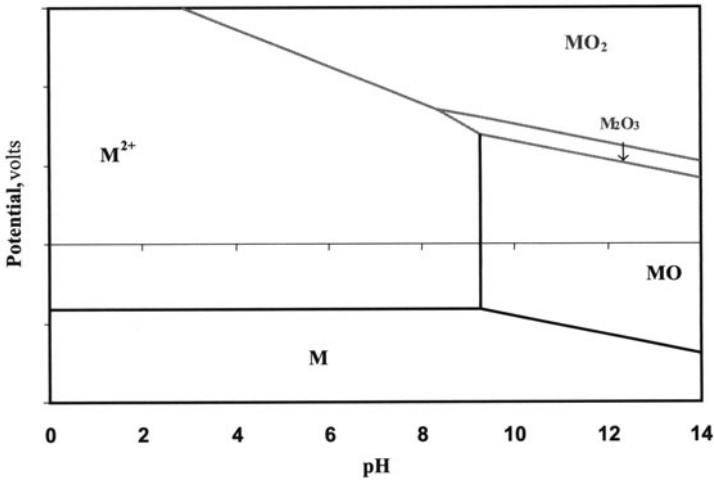
Fig. A2. Pourbaix diagram for hypothetical element M in water ($[M^{2+}]$ 10^{-6} molar)
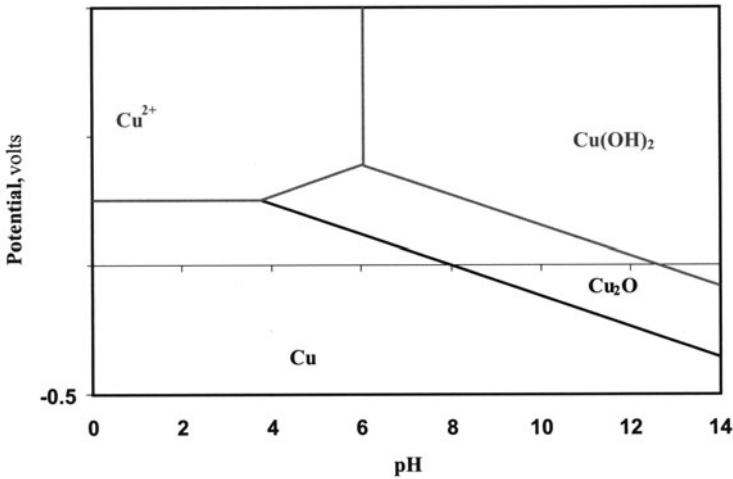
Fig. A3. Pourbaix diagram for copper in water

As a practical matter, copper presents an example of the previously described behavior. This is illustrated in Fig. A3. Clearly, surface oxides are formed in alkaline media and direct dissoultion occurs in acidic media. In contrast, the Pourbaix diagram for tungsten shows different behavior. Although all of the preceding discussion applies in a general sense, tungsten
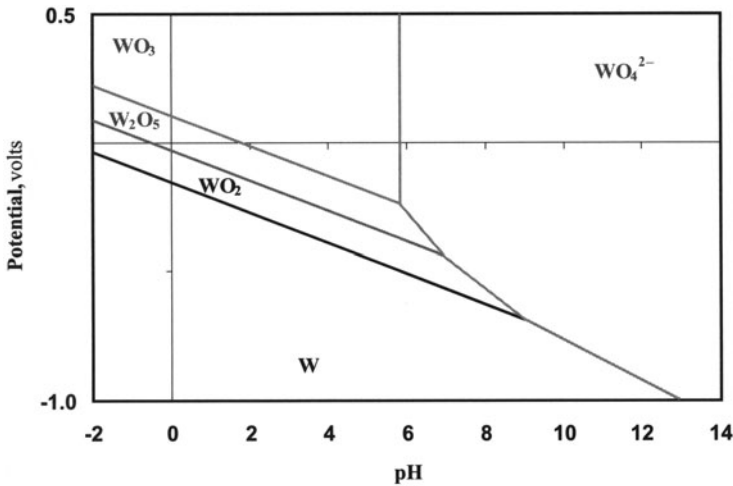
Fig. A4. Pourbaix diagram for tungsten in water

forms surface oxides in acidic media. In alkaline media, soluble tungstate ions are formed. (Of course, tungstate is an example of an *oxo anion* in which the central metal atom, although in a positive oxidation state, is strongly bound to oxygen atoms, which can be regarded as doubly negative "oxide ions". Hence, the aggregate species is an anion.) A Pourbaix diagram for tungsten appears as Fig. A4. Moreover, unless otherwise specified, Pourbaix diagrams address aqueous chemistry only and do not include interferences due to surfactants or complexation. These can directly affect surface dissolution reactions as well as effective concentrations of aqueous species. Indeed, as mentioned elsewhere, surfactants and complexing agents are often included in slurry formulations for just these reasons. For completeness, Table A1 lists various electrochemical half cell reactions [4].

Table A1. Electrochemical half-cell reactions and the related potential for typical CMP film and slurry components

Process	Role	Reaction	Potential: V (SHE)
W CMP	Oxidizer	$H_2O_2 + 2H^+ + 2e^- \rightarrow 2H_2O$	1.776
		$IO_3^- + 3H_2O + 6\ e^- \rightarrow I^- + 6OH^-$	0.257
		$2IO_3^- + 12H^+ + 10e^- \rightarrow I_2 + 6H_2O$	1.195
		$MnO_4^- + 4H^+ + 3e^- \rightarrow MnO_2 + 2H_2O$	1.70
		$Fe^{+3} + e^- \rightarrow Fe^{+2}$	0.771
		$Fe(CN)_6^{-3} + e^- \rightarrow Fe(CN)_6^{-4}$	0.358
	Film	$WO_2 + 4H^+ + 4e^- \rightarrow W + 2H_2O$	−0.119
		$WO_3 + 2H^+ + 2e^- \rightarrow WO_2 + H_2O$	0.036
		$WO_4^{-2} + 8H^+ + 6e^- \rightarrow W + 4H_2O$	0.053
	Liner	$Ti^{+2} + 2e^- \rightarrow Ti$	−1.63
		$TiO_2 + 4H^+ + 2e^- \rightarrow Ti^{+2} + 2H_2O$	−0.502
		$TiO_2 + 4H^+ + 4e^- \rightarrow Ti + 2H_2O$	−1.07
		$TiO_2 + \_N_2 + 4H^+ + 4e^- \rightarrow TiN + 2H_2O$	−0.272
Cu CMP	Oxidizer	$O_3 + 2H^+ + 2e^- \rightarrow H_2 + O_2$	2.07
		$NO_3^- + 4H^+ + 3e^- \rightarrow HNO_2 + H_2O$	0.94
	Film	$Cu^{+2} + 2e^- \rightarrow Cu$	0.3419
		$Cu_2O + H_2O + 2e^- \rightarrow 2Cu + 2OH^-$	−0.222
	Barrier	$Ta_2O_5 + 10H^+ + 10e^- \rightarrow 2Ta + 5H_2O$	−0.750
		$Ta_2O_5 + N_2 + 10H^+ + 10e^- \rightarrow 2TaN + 5H_2O$	−0.286
Al CMP	Film	$Al^{+3} + 3e^- \rightarrow Al$	−1.66
		$H_2AlO_3^- + H_2O + 3e^- \rightarrow Al + 4OH^-$	−2.33
Pt CMP	Oxidizer	$ClO_4^- + 8H^+ + 7e^- \rightarrow \_Cl_2 + 4H_2O$	1.39
		$Cl_2 + 2e^- \rightarrow 2Cl^-$	1.3600
		$Br_2 + 2e^- \rightarrow 2Br^-$	1.0763
		$I_2 + 2e^- \rightarrow 2I^-$	0.5361
	Film	$Pt^{+2} + 2e^- \rightarrow Pt$	1.118
		$Pt(OH)_2 + 2e^- \rightarrow Pt + 2OH^-$	0.14
		$[PtCl_6]^{-2} + 2e^- \rightarrow Pt + 4Cl^-$	0.755
Ru CMP	Film	$RuO_2 + 4H^+ + 2e^- \rightarrow Ru^{+2} + 2H_2O$	1.120
Si CMP	Film	$SiF_6^{-2} + 4e^- \rightarrow Si + 6F^-$	−1.24
		$SiO_3^- + 3H_2O + 4e^- \rightarrow Si + 6OH^-$	−1.697
		$SiO_2 + 4H^+ + 4e^- \rightarrow Si + 2H_2O$	0.857
		$F_2 + 2e^- \rightarrow 2F^-$	2.9178

References

1. W.M. Latimer, *Oxidation Potentials*, Prentice-Hall, NJ, 1952.
2. M. Pourbaix, *Atlas of Electrochemical Equilibria in Aqueous Solutions*, Pergamon, New York, 1966.
3. E.D. Verink, "Simplified Procedure for Constructing Pourbaix Diagrams", in *Uhlig's Corrosion Handbook*, Second Edition, Ed. R.W. Revie, Jon Wiley, New York, 2000.
4. More extensive data can be found in the *Handbook of Chemistry and Physics*, Eighty-first Edition, Ed. D.R. Lide, CRC Press, Boca Raton, FL, 2001; as well as in previous editions of this work.

Springer Series in
MATERIALS SCIENCE

Editors: R. Hull   R. M. Osgood, Jr.   J. Parisi   H. Warlimont

Springer Series in
MATERIALS SCIENCE

Editors: R. Hull   R. M. Osgood, Jr.   J. Parisi   H. Warlimont