Chemical Mechanical Planarization: Slurry Chemistry, Materials, and Mechanisms

Mahadevaiyer Krishnan,*† Jakub W. Nalaskowski,† and Lee M. Cook‡

IBM T.J. Watson Research Center, Yorktown Heights, New York 10598, and Dow Electronic Materials, Newark, Delaware 19713

Received April 30, 2009

Contents

1. Introduction 178
2. Planarization Process Application in Microelectronics 179
3. FEOL Applications: Device Level 180
   3.1. Shallow Trench Isolation (STI) CMP 180
   3.2. Replacement Metal Gate CMP 184
   3.3. Poly-Si CMP for FinFET Devices 186
4. MOL Applications: Contact Level 187
   4.1. Tungsten CMP 187
5. BEOL Applications: Multilevel Interconnects 188
   5.1. Copper Interconnect Technology 188
   5.2. CMP Challenges in Cu Interconnects 189
      5.2.1. Low-k and Ultralow-k Material Challenges 189
      5.2.2. Integration Challenges 190
      5.2.3. CMP Process Challenges 191
   5.3. Copper Planarization Process 191
   5.4. Ta/TaN Liner CMP Process 193
6. CMP Process-Induced Defects 194
   6.1. Defects in FEOL CMP 194
   6.2. Defects in MOL Tungsten CMP 195
   6.3. Defects in Cu BEOL CMP 195
      6.3.1. Corrosion of Copper 196
      6.3.2. Scratches 196
      6.3.3. Dishing, Erosion, and Trenching 196
      6.3.4. Mechanical Damage 197
      6.3.5. Other Defects 197
7. Models of CMP Processes 198
   7.1. Models Based on Contact Mechanics 198
   7.2. CMP Process Models 199
8. Alternative CMP Processes 200
9. Concluding Remarks 201
10. Acknowledgments 201
11. References 201

1. Introduction

The concept of chemical mechanical planarization (CMP) was invented in IBM in the early 1980s by Klaus D. Beyer in an attempt to create a highly planar surface and enable subsequent lithographic imaging without significant distortion.1 The first application involved filling the trenches with a dielectric such as silicon dioxide and removing the excess material by polishing it off. Since then the CMP process has been used to planarize a variety of materials including dielectrics, semiconductors, metals, polymers, and compos-
Lee M. Cook has been with Dow Electronic Materials since 1993. He has more than 30 years of experience in materials technology R&D and product development in optical, electro-optic, and semiconductor materials and processes. One of the pioneers in CMP research, he has held positions in R&D management and technology strategy during his tenure at Dow. Winner of four R&D 100 awards for product development, he is a frequent contributor to scientific journals, and holds more than 75 patents. He is a graduate of Grinnell College, Iowa.

Chemical Mechanical Planarization

2. Planarization Process Application in Microelectronics

Applications of CMP in microelectronics can be found in all three main areas of semiconductor device manufacturing. The transistor device part is called the front end of the line (FEOL), the contact metal (usually W) level is known as the middle of the line (MOL), and the interconnect part is called the back end of the line (BEOL). Figure 1a,b shows the cross-sectional images of the FEOL, MOL, and BEOL structures. An important FEOL CMP process is the shallow trench isolation (STI) CMP. MOL CMP is mainly tungsten contact level metal/liner polish and interlevel dielectric (ILD) polish processes. Interconnect CMP is exclusively for Cu, TaN/Ta and other liner removal processes. As gate lengths start approaching molecular dimensions, conventional simple planar transistor designs can no longer meet the performance requirements. Innovations such as high-k/metal gate structures have already been introduced in 45 and 32 nm technology nodes with gate first and gate last (replacement metal gate, RMG) integration schemes. For 22 nm and beyond, ultrathin body devices (ultrathin silicon-on-insulator (UTSOI), extremely thin SOI (ETSOI), and local SOI (LSOI)), back-gated devices, and multiple gate devices (FinFET, double gate, triple gate, and variations) are being proposed. Replacement metal gate based multigate field effect transistor (MuGFET) integration schemes require multiple CMP steps that are critical in enhancing device performance. Other technology areas such as 3D integration,
affected by both the CMP process and the post-CMP wet etch processes. Dishing, erosion, and pattern density dependent nitride thickness are the major contributing factors to nonplanarity in the CMP process. However, even if the STI CMP process achieves perfect planarity and perfectly uniform post-CMP nitride thickness across all pattern densities, removal of the sacrificial nitride will result in a nonplanar surface. Since both polish rates and wet etch rates are pattern density dependent, nonuniformities accumulate from all sources, resulting in large topography variation.

Pattern density effect is perhaps the most critical factor affecting the post-CMP planarity. Due to the close similarity of the oxide dielectric material used in ILD polish, early STI CMP processes employed the same consumables (pads, slurries) and process parameters used for ILD CMP. Since it is essential to remove all the oxide over the nitride in very large active areas as well as highly dense arrays, a certain amount of overpolishing is required. This results in the removal of the entire nitride layer in some features and erosion of trench oxide in other areas.\textsuperscript{14,15}

To solve these problems several different approaches have been attempted. Davari et al.\textsuperscript{16} have proposed a planarization technique using a combination of reactive ion etching (RIE) and CMP called reverse etch back. In this technique, areas of high density that have large initial topography are masked and etched away prior to CMP.\textsuperscript{16} While modest improvements were achieved by these methods, increase in process steps and complexity were key factors that prevented their implementation in manufacturing.

Another approach is to use dummy fill structures to eliminate the extremes in pattern density variations. In this approach, the layout is modified by addition of fill structures to increase the density of low-density regions to achieve pattern density equalization across the whole chip.\textsuperscript{17} Extensive dummy fill algorithms based on CMP process modeling of pattern density dependent rate variations\textsuperscript{18–20} and Monte Carlo simulations have been developed and integrated with chip design programs. This approach has phenomenally improved the STI topography to the extent that it is part of the design ground rules manuals and any exceptions are considered violations.\textsuperscript{21,22} A combination of fill shapes and reverse mask and etch have also been proposed.\textsuperscript{23}

Despite the great success of the dummy fill strategy, additional improvements in slurry chemistry and CMP process are also necessary to achieve better planarity. This is necessitated by the continuous drive to reduce device dimensions and the simultaneous reduction in the nitride thickness. Furthermore, the single step shallow trench planarization process gained importance due to the perceived simplicity and fewer process steps involved.\textsuperscript{24} This approach required significant improvements in the polish rate selectivity of oxide to nitride in the conventional fumed silica based slurries in order to achieve highly planar final topography.

In aqueous solutions, the oxide surface is hydrated due the diffusion of water under applied pressure. Water can exist in silicon dioxide as molecular water (H\textsubscript{2}O) and hydroxyl water (–SiOH), and they are in equilibrium as interstitial and substitutional impurities.\textsuperscript{25,26} The diffusion coefficient of water increases exponentially with increasing tensile stress.\textsuperscript{27} The chemical reactions between the siloxane (Si–O–Si) bonds and water determine the polishing behavior of oxide and control the overall polish rate.\textsuperscript{28}

\[
(SiO\textsubscript{2})_x + 2H\textsubscript{2}O \rightleftharpoons (SiO\textsubscript{2})_{x-1} + Si(OH)\textsubscript{4} \quad (1)
\]
The surface hydroxyl equilibrium can be described as

\[ \equiv \text{Si} - \text{OH}_2^+ \rightleftharpoons \equiv \text{Si} - \text{OH} \rightleftharpoons \equiv \text{Si} - \text{O}^- \quad (2) \]

At pH values around 10, equimolar surface concentrations of each species exist and the polish rate reaches its highest value. The properties of the surface layer can be altered by the adsorption of surface active molecules.

The polishing of the nitride surface involves the conversion of Si–N bonds into Si–O bonds through a surface hydrolysis reaction. Hu et al.\(^2\) observed that Si–O bond density on the surface increased after CMP. In addition, formation of silanol groups and amine species (NH\(_x\)) on the surface was also confirmed by Fourier transform infrared (FTIR) spectra. The ammonia formed during the hydrolysis was detected by Nessler’s color reaction. On the basis of the evidence, Hu et al. propose the following reactions for nitride hydrolysis that modify the surface layer

\[ \text{Si}_3\text{N} + \text{H}_2\text{O} \rightarrow \text{Si}_2\text{NH} + \text{Si} - \text{OH} \quad (3) \]

\[ \text{Si}_2\text{NH} + \text{H}_2\text{O} \rightarrow \text{SiNH}_2 + \text{Si} - \text{OH} \quad (4) \]

\[ \text{SiNH}_2 + \text{H}_2\text{O} \rightarrow \text{Si} - \text{OH} + \text{NH}_3^+ \quad (5) \]

\[ \text{SiNH}_2 + \text{Si} - \text{OH} \rightarrow \equiv \text{Si} - \text{O}^- \equiv + \text{NH}_3^+ \quad (6) \]

\[ \text{Si}_2\text{NH} + \text{H}_2\text{O} \rightarrow \equiv \text{Si} - \text{O}^- \equiv + \text{NH}_3^+ \quad (7) \]

The modified surface layer further reacts with water and forms silanol groups

\[ \equiv \text{Si} - \text{O}^- \equiv + \text{H}_2\text{O} \rightarrow \text{Si(OH)}_4 \quad (8) \]

These reactions are accelerated by friction and wear and hence can be considered tribochemical in nature. Nitride polish rates are dependent on the kinetics of the hydrolysis reaction. In addition, modification of the surface layer by adsorption can be used to suppress the nitride polish rates.

A simple modification of the slurry pH to improve oxide to nitride selectivity has been attempted by Prasad et al.\(^3\) Polishing of oxide surface requires the presence of OH\(^-\) ions,
and thus increasing the pH results in higher oxide polish rates. For nitride the opposite is true, at higher pH the rate of hydrolysis of silicon nitride is significantly reduced. Thus improvements in oxide to nitride selectivity can be achieved at higher pH. This strategy appears to have worked as reported by Prasad et al.\textsuperscript{35} in enhancing the oxide to nitride selectivity. However severe pitting of the oxide surface was observed under these conditions.

An alternative means is to add components to the conventional silica slurry to enhance the oxide polish rate or suppress the nitride polish rate. When used together substantially high selectivity can be achieved. Strong chelating agents for Si$^{4+}$ in combination with surfactants that form a film on the nitride surface have provided selectivity in the range of 200–300 times in the pH range 9–11.\textsuperscript{31–34} Use of chemisorbed self-assembled molecular layers that are formed on silica or silicon nitride surfaces to control polish rates was proposed by Moudgil et al.\textsuperscript{35} Addition of sodium dodecyl sulfate (SDS) to silica slurries at pH 2 decreased the polish rate of silicon nitride much more than the polish rate of oxide. This is attributed to the preferential higher adsorption of SDS on nitride due to electrostatic attraction compared with oxide. The SDS film acts as a lubricant on the nitride surface resulting in lower polish rates. This effect levels off at concentrations twice the critical micelle concentration (CMC) of SDS. Adsorption of SDS on the silica surface is not insignificant and the exact mechanism of the adsorption of an anionic surfactant on a negatively charged silica surface has not been clearly established. Chemical interactions such as hydrogen bonding\textsuperscript{35} and Na$^{+}$-mediated surfactant bonding\textsuperscript{36} have been invoked to explain this phenomenon. Similar arguments can be made for higher adsorption of cationic surfactant cetyl trimethyl ammonium ion (CTA) on oxide surfaces compared with nitride surfaces. Using FTIR/attenuated total reflection (ATR), contact angle and $\zeta$ potential measurements, Singh et al.\textsuperscript{37} have shown that CTA films contain randomly adsorbed molecules at low concentrations. Above a certain critical concentration, they form hemimicelles that are two-dimensional aggregates with strong hydrophobic association between surfactant tails. At even higher concentrations, the transition from hemimicelles to randomly oriented spherical aggregates occurs. These transitions can occur at concentrations below the critical bulk micelle concentration indicating the role of the surface in triggering the transition. These changes should be reflected in the oxide polish rates provided no significant adsorption occurs on the abrasive particles. In the event that the surfactants are strongly adsorbed on the abrasive particles, no significant polish rate for oxide or nitride can be observed due to surfactant-mediated lubrication. This is indeed the case for most silica slurries.\textsuperscript{37,38}

Canaperi et al.\textsuperscript{39} attempted to populate the layer adjacent to the nitride surface with molecules that mimic water but cannot participate in the hydrolysis reaction to suppress the nitride polish rates. Amino alcohols including mono-, di-, and triethanolamines appeared to accomplish this, and nitride rates as low as $<10$ Å/min could be achieved by this method while keeping the oxide polish rates around 1000–1200 Å/min. As often is the case, these polish rate experiments are normally carried out with individual blanket oxide and nitride films deposited on wafers and selectivity numbers are calculated from rates measured on separate oxide and nitride wafers. In patterned wafers, oxide and nitride surfaces exist next to each other and are strongly influenced by the presence of the other. Thus the local polish rates are weighted average of the individual oxide and nitride polish rates (blanket), the weight factor being the relative dimensions of the oxide and nitride surfaces. The consequence of this effect is that in order for the selectivity enhancement to work, large surface areas should be covered with the nitride surface and when this is the case high selectivity slurries achieve excellent planarity compared with the conventional slurries. However, in the case of STI patterns, this is not the situation. Thus the final topography did not show any significant improvement due to increased oxide to nitride selectivity.

The allure of the simplicity involved in the one-step STI CMP process has prompted the development of ceria-based slurry systems.\textsuperscript{40} The average particle size of ceria varies from 50 to 200 nm with abrasive loading of 0.5–3% (W %). In addition, polyelectrolytes and other additives such as amines, amino alcohols, amino carboxylic acids, carboxylic acids, thiols, organic polyls, or amino acids\textsuperscript{41–45} are used in conjunction with ceria to enhance selectivity. However, the polyelectrolyte/ceria systems are the most widely used in practice and deserve to be discussed in detail. These systems typically contain poly(acrylic acid) and its ammonium salt or poly(acrylic acid) copolymers containing carboxylic acid and amide moieties as additives.\textsuperscript{46} These slurries are sometimes used as two part systems; the abrasive slurry and the solution containing the additives are supplied to the table by separate lines and are mixed on the table. The flow rates can be different and precision dispensing of the additive solution is required for polish rate control.

The role played by the poly(acrylic acid) and copolymers is multifaceted and complex. In general, the ceria nanoparticles in aqueous dispersions are stabilized by electrostatic interactions. Destabilization can occur from high-surface-to-volume ratio of the particles and from the strong reactivity of the surface sites to the physical or chemical changes occurring in the near vicinity.\textsuperscript{47} Dispersants such as poly(acrylic acid) with COOH groups form complexes with the surface hydroxyl groups and stabilize the colloids. Depending on the pH and the dispersant concentration dispersion—precipitation—redispersion is known to occur in these systems.\textsuperscript{47}

The higher polish rate selectivity of the ceria systems has been attributed to the preferential adsorption of poly(acrylic acid) on the silicon nitride layer.\textsuperscript{48,49} Significant change in the $\zeta$ potential values for the ceria in colloidal dispersions was observed by Zedwick et al.\textsuperscript{50} in presence of poly(acrylic acid) indicating change in the surface charge density due to adsorption (Figure 4). Adsorption of poly(acrylic acid) (PAA) on to SiO$_2$, Si$_3$N$_4$, and CeO$_2$ powders in suspensions has been observed by Kim et al.\textsuperscript{51} The adsorption isotherms (Figure 5) clearly indicate that PAA adsorption was nearly 10 times higher for Si$_3$N$_4$ than for SiO$_2$ under equilibrium conditions. The adsorbed amount was higher at pH 7 for nitride and nearly invariant with pH for oxide. The isoelectric point for nitride is around pH 6–7 and for oxide is around pH 2–3. Thus the poor adsorption on the oxide surface is attributed to electrostatic repulsion. ATR-FTIR spectra of oxide and nitride particles in suspension and contact angle measurements on oxide and nitride surfaces confirm that PAA is preferentially adsorbed on nitride. The passive layer formed on the nitride surface by the polyelectrolytes lowers the polish rate thus enhancing the selectivity to oxide.\textsuperscript{51,52}

The ceria/polyelectrolyte systems exhibit a self-stopping polish rate that is initially very fast and drops rapidly with
time (2–3 min) as polishing proceeds. Addition of water to the table at this point can restore the polish rate to the original levels. Similar behavior is observed in patterned wafers and as a result the planarization efficiency of ceria-based systems is high only in the initial stages of polish. At longer polish times (2–3 min), the selectivity advantages are lost. The polish rate variability for this process over extended periods (days) can be as much as 20–25% from the average rate. Because of the fluctuations in the polish rates, it is necessary to measure the thickness of the remaining oxide after polish for each wafer and ensure that the target thickness is achieved across the entire wafer. To ensure complete removal of oxide on top of nitride, additional touch up polish may also be required.

Ceria slurries with carboxylic acids and amino acids have also been proposed. L-Proline, arginine, lysine, alanine, glycine, and substituted glycines show enhancement of blanket rate oxide to nitride selectivity. The adsorption behavior of L-proline and L-arginine in the pH range 9–11 on oxide and nitride surfaces does not show significant differences in the amounts adsorbed. However, L-arginine lowers both oxide and nitride polish rates, while L-proline appears to have an effect on nitride rate and not on oxide polish rate. Thus the blanket polish rate behavior of ceria slurries containing amino acids cannot be explained by preferential adsorption on to nitride surfaces. More elaborate studies are needed to understand the mechanism of oxide and nitride polishing in the presence of amino acids.

For both ceria/polyelectrolyte and ceria/amino acid systems, the high polish rate selectivity observed in blanket wafers is not realized in patterned wafers. Thus the selectivity observed in patterned wafers is significantly lower than that predicted based on blanket wafer experiments. As mentioned earlier, the ratio of nitride covered areas to the oxide covered areas is a small fraction in STI patterns, and this is the primary reason for this behavior, and contrary to popular belief, the nitride layer does not seem to act as a global stop layer in STI CMP. For slurries with higher selectivity, higher nitride thickness has been observed in large structures that are completely covered with nitride compared with slurries with lower selectivity. This is strictly a local effect. High selectivity can be helpful to improve global planarity only when ratio of nitride areas vs oxide is large, as in the case of Ta liner vs. Cu in interconnect structures.

Application of fixed abrasive polishing to STI was introduced to improve the planarity as well as the uniformity of the conventional silica based slurry systems. These systems use a variety of abrasives incorporated in a specially constructed web matrix structure (sculptured pads, see Figure 6a) and additives in solution to achieve polishing. Applied Materials’ roll-to-roll web type polishers allow fresh matrix material to be introduced between wafers (Figure 6b).

Irrespective of the polish process (one, two, or three steps) and the slurry systems used, the observed nonplanarity is around 200–300 Å for most designs, with large STI features recessed with respect to the active area. The consequences of the STI recess are illustrated in Figure 7a,b. For 32 nm technology node and beyond, <100 Å topography is required (Figure 8) to achieve better static random access memory (SRAM) yields. Steigerwald has emphasized the importance of controlling the within-die (WID) thickness variations and
within-wafer (WIW) uniformity in STI polish for future devices. It may be difficult to meet these requirements by improved slurry chemistry alone. New integration schemes incorporating polish stops may also be necessary to meet the stringent planarity requirements of the future generation of devices.

3.2. Replacement Metal Gate CMP

The great success of the complementary metal–oxide–semiconductor (CMOS) technology can be attributed to the scalability of the transistor. For over a quarter of a century very little has changed in the basic transistor design except the dimensions. The scaling concept requires that all the physical dimensions (length, width, and thickness) be reduced simultaneously. As these physical dimensions start approaching molecular scales, it has become extremely difficult to achieve performance gain by simple device scaling alone.61 The high-$k$/metal gate approach is an attempt to extend the planar transistor design and preserve device scaling beyond $\sim 50$ nm gate length.

The SiO$_2$ gate oxide is a key material that has enabled scaling of CMOS devices to gain performance improvements. The physical thickness of gate oxide in transistors has been decreasing steadily with the decrease in gate length. For the 90 nm node, the thickness of gate oxide is 12 Å, and experimental transistors with 8 Å gate oxide thickness have been reported.62,63 However, continued gate oxide scaling is becoming extremely difficult because the gate oxide leakage in SiO$_2$ increases with decreasing physical thickness, and SiO$_2$ can no longer play its role effectively as the gate dielectric. High $k$ dielectric materials such as hafnium- and zirconium-based oxides reduce the leakage current significantly (by 100-fold) and can be used at higher thickness. However, replacing SiO$_2$ with high-$k$ materials in poly-Si gates leads to two major problems. Because of the defects that form at the gate dielectric/poly-Si gate electrode interface, the voltage at which the transistor switches (threshold voltage, $V_t$) becomes too high. Second, the electron mobility in the channel is severely degraded due to surface phonon scattering.64,65 Both these problems affect the transistor switching speeds. Chau et al.66 have demonstrated that use of metal gates with appropriate work functions can provide the right threshold voltages ($V_t$) and significantly reduce channel mobility degradation. With the combination of high-$k$ gate dielectrics and metal gates, the transistor performance can be significantly improved and scaling below $\sim 50$ nm becomes possible.62,63 However, finding a pair of n-type and a p-type metals with the right work function for metal gates proved to be a big challenge since most electrode materials reported in the literature have work functions that are close to midgap. Chau et al.64,66 have reported that they

Figure 6. SEM picture of fixed abrasive pad (a), schematic diagram of the fixed abrasive polishing tool (b). Reprinted with permission from ref 59. Copyright 2000 IMIC.

Figure 7. SEM pictures of 32 nm SRAM array showing (a) poor topography and (b) excess STI recess.

Figure 8. SEM picture of 32 nm SRAM array showing good STI topography.
have identified materials for n-type and p-type metal electrodes that have correct work functions on the high-$k$ dielectric for high-performance CMOS logic devices. The high-$k$ dielectric stacks have the equivalent oxide thickness of 10 Å with negligible gate oxide leakage, and channel mobilities that are close to the values obtained with SiO$_2$ gate dielectric.$^{64}$

The fabrication of high-$k$ metal gate devices proved to be another formidable challenge. In conventional CMOS processing, high-temperature dopant activation is used. Most metals with high work function are stable at these temperatures. However, the interface between the high-$k$ dielectric and work function metal is sensitive to high temperature processing and significant changes occur in the interfacial characteristics. This manifests as $V_t$ variation and decreased device reliability. Thus the materials used in the metal gate stack determine the device integration run path. A metal gate structure with stack materials that can withstand high-temperature processing can be integrated with gate-first approach and is compatible with conventional CMOS processing. Any metal gate structure with stack materials that cannot withstand high-temperature processing has to be built after all the front-end processing steps including high-temperature junction activation have been completed. This integration method involves fabrication of sacrificial poly-Si gates and replacing them with metal gates (hence the name replacement metal gate). The process flow includes several CMP steps and is known as the gate-last or damascene metal gate flow. Figure 9 shows the gate last process flow and the CMP steps involved.$^{67}$

The two CMP steps in the replacement metal gate (RMG) process flow are (i) the poly opening CMP and (ii) the metal gate CMP. Most of the integration challenges (Figure 10) in RMG CMP arise from poor thickness control during material deposition and CMP processes. Polish rate variation during the poly opening CMP results in gate height variation. Tall gates lead to higher aspect ratios and poor fill during metal gate material deposition. Unexposed poly-Si due to low oxide polish rates results in partially filled metal gates and device failure due to improper work function ($V_t$ shift) and higher gate resistance. High oxide polish rates can expose the source/drain regions that may be attacked during the poly-Si wet etch process. The metal gate polish problems are similar to the problems in interconnects such as shorting due to underpolish. The metal gate polish step must have sufficient overpolish window to overcome the topography created during the poly opening CMP. Excessive overpolish in any

![Figure 9. Replacement metal gate process flow. Reprinted with permission from ref 67. Copyright 2007 IEEE.](image)

<table>
<thead>
<tr>
<th>Integration Issues with RMG CMP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Issue</strong></td>
</tr>
<tr>
<td>Gate resistance variation</td>
</tr>
<tr>
<td>Poor gate fill</td>
</tr>
<tr>
<td>Unexposed poly Si gate</td>
</tr>
<tr>
<td>Residual material (underpolish)</td>
</tr>
<tr>
<td>Raised Si/D exposure</td>
</tr>
<tr>
<td>Contact etch window</td>
</tr>
<tr>
<td>Bavel edge redistribution of poly/metal</td>
</tr>
<tr>
<td>Structural damage to device layer</td>
</tr>
<tr>
<td>Strain/stress relaxation</td>
</tr>
</tbody>
</table>

![Figure 10. Integration issues in replacement metal gate process. Reprinted with permission from ref 60. Copyright 2008 IEEE.](image)
of these steps can cause thin gates with high gate resistance. Thus the two RMG CMP processes have very narrow process windows, and extremely tight process control efforts are necessary to have high yields. Slurry with high selectivity toward oxide is used to achieve excellent within-die planarity with process parameters optimized around the selected pad. Since the gate dimensions are very small, the metal gate device yields are sensitive to CMP-induced defects. Figure 11 shows the typical process-induced defects and their effect on the high-k metal gate device.60

In future metal gate devices, materials other than TiN and aluminum may be introduced, and the p-type metal–oxide–semiconductor (PMOS) and n-type metal–oxide–semiconductor (NMOS) gates may have different material sets as well. The n-type metal gate materials have work function between 3.9 to 4.2 eV and include hafnium, zirconium, tantalum, titanium, and aluminum and their alloys, as well as their carbides. The p-type gate materials have work function between 4.9 to 5.2 eV and include ruthenium, palladium, platinum, nickel, and cobalt and conducting oxides such as ruthenium oxide. The trench fill materials may be easily polishing materials such as tungsten, aluminum, titanium, and titanium nitride.68–70

The gate trenches may be half filled with gate material and the rest with trench fill metals. Slurries have to be formulated to handle multiple materials in contact with each other and electrochemical interactions between dissimilar metals have to be taken into account. Similar issues are also present in post-CMP cleaning. After metal gate CMP, the aluminum is exposed to the cleaning solutions and needs to be protected, especially in acid or alkaline environment.71

The replacement metal gate CMP is a new and emerging application, and most of the details regarding the CMP processes and slurries have not been disclosed at this time. As the RMG approach becomes widely adopted, many of the details regarding the CMP processes and slurry compositions may be published in the coming years.

### 3.3. Poly-Si CMP for FinFET Devices

Another way to overcome the difficulties associated with device scaling is to build multigate devices. These devices maintain scaling advantages by better control of the channel.61 As gate length becomes smaller and smaller, the close proximity of the source and drain reduces the ability of the gate electrode to control the potential distribution and current flow in the channel.72 Under these conditions, short channel effects (SCE) set in, limiting the ability to go below ~50 nm gate length. The short channel effects are attributed to two physical phenomena: modification of the threshold voltage ($V_t$) due to the shortening of the channel length and the limitations imposed on the electron transport (mobility) in the channel. The decrease of the threshold voltage ($V_t$) due to decreased gate length is called threshold voltage rolloff. The short channel effects can be minimized by reducing the gate oxide thickness, reducing the junction depth, and increasing the dopant concentrations in the channel along with the use of silicon-on-insulator (SOI) technology.73

These ideas have been already incorporated in modern transistor devices and have reached practical limitations as we approach 22 nm. In an effort to overcome short channel effects, transistors have evolved from planar single gate devices to three-dimensional devices with multigate (double, triple, or quadruple gate) structures (Figure 12). In reality, a double gate is actually a single gate electrode present on opposite sides of the device, and a triple gate is a single gate folded over three sides of the device.72

The process steps for FinFET fabrication are shown in Figure 13. The poly-Si planarization is one of the critical steps in achieving the right fin definition.74 The schematic diagram of the CMP process is shown in Figure 14. The requirements for poly-Si planarization are rather stringent. A highly planar post-CMP structure is required across dense pattern density fins and the gaps between the fins. The thickness of the remaining poly-Si after CMP should be less than as possible for gate height scaling, which affects the device performance. The within-wafer uniformity should be less than 100 Å for high device yields. In addition, the slurry should have high selectivity to poly-Si with very low polish rates for liner materials such as TiN and SiN. Most poly-Si slurries contain Si etchants such as quaternary ammonium salts to facilitate high removal rates. Thus tetramethylammonium hydroxide (TMAH), tetramethyl ammonium fluoride, tetra-ethyl ammonium chloride, dimethyl diethyl ammonium chloride, dimethyl diethyl ammonium fluoride, alkyl benzyl dimethyl ammonium hydroxide, diethylenetriamine, triethylenetriamine, and piperazine have been reported in the range of 0.01–5% (W %). In addition ammonium phosphate

### CMP Defect Modes

<table>
<thead>
<tr>
<th>Defect mode</th>
<th>Potential causes</th>
<th>Impact to device</th>
<th>Potential solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particles</td>
<td>Slurry/pad residue, Polish byproducts</td>
<td>Shorting/opens, Pattern distortion</td>
<td>Cleaner tooling, Clean Chemistries</td>
</tr>
<tr>
<td>Macro scratches</td>
<td>Large/hard foreign particles on polish pad</td>
<td>Pattern removal over multiple die</td>
<td>Pad conditioning, Pad cleaning Environment</td>
</tr>
<tr>
<td>Micro scratches</td>
<td>Slurry agglomeration, Pad asperities</td>
<td>Shorting/opens</td>
<td>Slurry filters, Pad/pad conditioning</td>
</tr>
<tr>
<td>Corrosion (metal CMP)</td>
<td>Slurry chemistry, Clean chemistry</td>
<td>Opens, Reliability</td>
<td>Passivating films, Chemistry optimization</td>
</tr>
<tr>
<td>Film delamination</td>
<td>Weak adhesion, CMP shear force</td>
<td>Shorting/Opens, Device parametrics</td>
<td>Improve adhesion, Low pressure CMP</td>
</tr>
<tr>
<td>Organic residue</td>
<td>Inadequate cleaning, residual slurry components</td>
<td>Shorting/opens, Disturbed patterning of next layer</td>
<td>Cleaner tooling, Slurry optimization, Clean Chemistries</td>
</tr>
</tbody>
</table>

Figure 11. CMP process-induced defects in replacement metal gate process. Reprinted with permission from ref 60. Copyright 2008 IEEE.
is used for buffering the slurry pH around 10. TMAH is widely used as the preferred component in many of these formulations. TMAH is highly toxic and is not easily removed from waste waters. Thus new slurry formulations that are nontoxic and are easily treated by existing waste treatment systems are needed.

Selective polishing of poly-Si for microelectromechanical systems (MEMS) applications was investigated by Veera et al. Much larger step heights on the order of 5 µm have to be planarized in MEMS structures and require high polish rates and good selectivity control to avoid dishing of large structures. Like the RMG CMP, poly-Si CMP is an evolving application. Multiple variations of the FinFET devices are being designed and developed at this time, and the details regarding the CMP processes and slurries may become available in the coming years.

4. MOL Applications: Contact Level

4.1. Tungsten CMP

Tungsten is widely used as the contact via material mainly due to its superior electromigration performance. It also acts as a good diffusion barrier and prevents the diffusion of Al and Cu into the device structures at the thickness used in contact level metallurgy. The processing steps include deposition of tungsten by LPCVD over patterned trenches in dielectric and removal of the overburden to create inlaid metal via. LPCVD tungsten deposits have been conformal and have had good fill characteristics up to 90 nm technology node. Starting from 65 nm node, because of the high aspect ratio vias, seams in tungsten deposits have been observed. So far this has not affected the use of LPCVD tungsten for contact level metallization even though efforts are underway to find alternatives. The W CMP process should provide excellent planarity since the first BEOL interconnect level will be adversely affected by the topography created by the contact via level polish. Thus dishing and erosion in W CMP could lead to serious shorting problems in the first interconnect metal level. This is exacerbated by the use of borophosphosilicate glass (BPSG) or borophosphorous tetraethyl orthosilicate (BPTEOS) as the dielectric material due to its softness compared with oxide. In addition, the FEOL topography incoming to contact level is also significant. It is necessary to reduce the incoming FEOL topography by using a CMP step after the BPSG dielectric deposition. This is known as the ILD polish. To improve the adhesion of tungsten to the dielectric, a Ti/TiN liner is used. This layer is also believed to act as a barrier layer for tungsten. The tungsten CMP process involves three steps: removal of tungsten and stopping in Ti/TiN liner, removal of the liner material, and stopping on the dielectric, followed by post-CMP cleaning.

The early W slurries consisted of alumina abrasive and ferric nitrate or potassium iodate as the oxidizers. Slurries with silica abrasives and hydrogen peroxide as oxidizer have been more commonly used in recent years. Slurries with alumina abrasives had better selectivity with respect to W vs Ti and had significantly less liner erosion after the first step. Silica-based slurries have poor selectivity, and as a result they are usually used as a single step process removing tungsten and liner at the same time. Sometimes a second buffing step may be used improve the planarity and lower defect counts. Jeong et al. have reported that with slurries containing H2O2 and silica abrasives, circular defects consisting of precipitated tungsten oxides were observed. These can be removed by using strong ammonia in the post-polish cleaning solutions.

Figure 12. Different types of multiple gates. Reprinted from ref 72 with the kind permission of Springer Science and Business Media.

Figure 13. Process steps in FinFET integration. Reprinted with permission from ref 74. Copyright 2008 IEEE.

Figure 14. Schematic diagram of the poly-Si CMP process for FinFET.
The tungsten surface in contact with strong oxidizers is covered with an oxide film in acidic pH range as indicated by the Pourbaix diagram.\textsuperscript{80} Furthermore, tungsten dissolution in static conditions (no polishing) is absent or extremely low in the presence of ferric nitrate and potassium iodate. The passivating oxide film also protects the final surface from corrosion due to any remaining oxidizer as well as from the atmosphere.\textsuperscript{81,82} Several studies have focused their attention on the surface passivation of W in the presence of oxidizers. The anodic potentiostatic polarization curves for W in presence of KIO$_3$ (3.8%) and of hydrogen peroxide (5% vol) at pH 4 have been compared by Tamboli et al.\textsuperscript{83} The measured dissolution potential of W in KIO$_3$ solutions is more anodic than that measured in H$_2$O$_2$ solutions. However, the current density in the passive region for H$_2$O$_2$ is one order of magnitude higher than that for KIO$_3$. Thus the W surface is much more strongly passivated in KIO$_3$ solutions than in H$_2$O$_2$ solutions. Using X-ray photoelectron spectroscopy (XPS) analysis of W surfaces with anodic films formed under polarization, Tamboli et al.\textsuperscript{83} conclude that in the case of KIO$_3$ solutions the passive film mainly consists of WO$_3$ and there was substantial increase in W–OH type bonds at higher potentials. The oxidizing action of H$_2$O$_2$ on the W surface was very different. XPS data show growth of peaks corresponding to substoichiometric oxides of W, such as WO$_3$ and WO, and peaks corresponding to WO$_3$ are absent. This coupled with higher current densities observed in the passive region lead them to postulate the formation of peroxy polytungstic acids due to the dissolution of tungsten oxides in H$_2$O$_2$.

Lim et al.\textsuperscript{84} have compared the oxidation of W in H$_2$O$_2$ and Fe(NO$_3$)$_3$ at pH 1.5 and concluded that a thin and dense oxide layer was formed in the slurry containing Fe(NO$_3$)$_3$, with a fairly high rate of WO$_3$ formation. With H$_2$O$_2$, a thick and porous layer was formed on the W surface. They believe that Fe(NO$_3$)$_3$ would be a better oxidizer for W CMP because of the high polish rate and better protection of recessed layers with a dense WO$_3$ passivation layer. In solutions containing H$_2$O$_2$, tungsten dissolves forming oxides such as WO$_2$ as shown:

\begin{equation}
W + 2H_2O_2 \rightarrow WO_2 + 2H_2O
\end{equation}

\begin{equation}
2WO_2 + 6H_2O_2 \rightarrow H_2W_2O_{11} + 5H_2O
\end{equation}

\begin{equation}
3H_2W_2O_{11} + 7H_2O \rightarrow 2H_2W_3O_{12} + 8H_2O
\end{equation}

Stein et al.\textsuperscript{85,86} have investigated the interaction of the W surface film and the colloid species in the slurries under polishing conditions. The polisher rate as a function of alumina concentration in the range 0–8 W % was measured at three different pressure–rotation rate settings. For all settings, at low alumina concentrations, the polisher rate increases rapidly with increase in alumina content. Beyond 1% and up to 7.5%, the polisher rate shows very little change with alumina concentration. At any given alumina concentration, higher polisher rate is observed at higher pressure–rotation rate settings. The polisher rate dependency on KIO$_3$ concentration also exhibits nonlinear behavior.

Stein et al. found that the process temperature was strongly dependent on the nature of the colloid species, their concentration, and the pH of the slurry but was independent of the oxidizer KIO$_3$ concentration. The polisher rates varied by 2 orders of magnitude (from ∼15 to ∼2000 Å/min) for slurries containing yttrium, cerium, zirconium, and aluminum oxide colloids under otherwise similar conditions such as abrasive loading, particle size, downforce, rotation rates, pad, and oxidizer concentration. They concluded that the nature and surface characteristics of the various colloids and their interaction with the oxide-covered tungsten surface are key factors in determining the polish rates. Understanding these interactions is vital in establishing the role of the various surface kinetic processes in the mechanism of the overall CMP process.

5. BEOL Applications: Multilevel Interconnects

5.1. Copper Interconnect Technology

On-chip interconnections (or interconnects) are local and global wiring that connect the various circuit elements and distribute power. They also function as the interface between the device and the package and require excellent control of mechanical properties. Thus both electrical and mechanical reliability are critical elements of a successful interconnect technology. Planar multilevel architecture was an important invention that enhanced the structural integrity of interconnects. The exceptional planarity provided by the CMP process enhanced the electrical and mechanical properties of the metal and dielectric films and in turn enhanced their reliability. In addition, enhanced planarity facilitated the continuous increase in circuit density, number of interconnect levels, and innovation of new design elements such as stacked vias.\textsuperscript{87}

The electrical performance of the interconnect is determined by the resistive and capacitive elements that contribute to the gate delay and hence affect the transistor performance.\textsuperscript{88} The gate delay has four components. The first is the intrinsic delay associated with charging the driver and receiver loads, the second is associated with the time needed to charge the conductor wires with finite driver currents, the third is the distributed delay of the interconnect load widely known as the RC delay. The fourth and last component is the time it takes to charge the receiver input via the wire resistance. The interconnection length at which the RC delay is half the total gate delay is defined as the critical interconnection length ($l_c$).\textsuperscript{88} For high performance interconnects, this value is 2–3 mm. For wires longer than the critical length, the performance is mainly affected by the resistance, and this can be minimized by increasing the cross-sectional area. For short wires, the main component to the delay is through the capacitive loading of the smallest devices, so their cross-section should be reduced to minimize their capacitance. To work around the long-line RC delay problem, two kinds of wiring circuits are used in high-performance microprocessor designs. The thin wire circuits have short average length, are scaled to the minimum pitch, and populate the first few wiring levels. The most effective way to increase their electrical performance is to lower their capacitance. For fat wires (which are also long, several millimeters), density is secondary to delay considerations. These are designed in such a way that their time for signal propagation is only a very small fraction of the CPU cycle time. They typically occupy the upper levels of the chip.\textsuperscript{89,90,91} Figure 15 shows the SEM cross-section of a modern multilevel interconnect structure with thin wire and fat wire levels.

Copper interconnects were introduced in 1997 and became industry standard soon afterward. Initially oxide ($k \approx 4.0$) was the dielectric material followed by carbon-doped oxide
5.2. CMP Challenges in Cu Interconnects

The continuous drive to improve chip performance has accelerated the introduction of new materials along with simultaneous reductions in minimum wiring dimensions and increase in the number of interconnect levels. The aggressive scaling of line width and the need to incorporate new materials has resulted in material, integration, and process challenges for the interconnect CMP processes. Furthermore, these challenges are often coupled, complex, and multifaceted. Thus changes in the dielectric material result in changes in integration schemes, and new processes need to be developed to enable the new integration scheme. In addition, changes in the upstream processes will affect the design and implementation of downstream processes. These are examined in detail in the following sections.

5.2.1. Low-k and Ultralow-k Material Challenges

The main purpose of using low-$k$ and ultralow-$k$ materials in the interconnect structure is to enhance chip performance by reducing propagation delays, cross-talk noise between adjacent metal lines, and power dissipation resulting from RC coupling interactions.92,93 The carbon-doped materials containing Si, C, O, and H are deposited by a PECVD process. For porous ultralow-$k$ films, an organic precursor porogen is mixed with the skeleton precursor for pure SiCOH during plasma deposition. The resulting film consists of a random, covalently bonded SiCOH skeleton to which the CH$_x$ fragments, created by the porogen precursor dissociation, are attached. The porogen fragments are less stable than the skeleton and are removed by subsequent curing, creating a porous structure. The skeleton is strong enough not to collapse during porogen removal; however some shrinkage does occur. The most commonly used skeleton precursors include tetramethyldicyclosiloxane (TMCTS), octamethylcyclotetrasiloxane (OMCTS), and diethoxymethylsilane (DEMS). Cyclopentadiene, butadiene monoxide, and cyclohexane are some of the common porogen precursors.94 Another way to make the porous SiCOH films is to add precursors that contain molecular pores. The deposition conditions are so optimized as to enhance the polymerization of the precursor and prevent its dissociation in the plasma. The unsaturated carbon bonds in the precursor molecule are opened, and new bonds to the skeleton are formed. The porosity in the film is created by the undissociated molecular pores as well as the arrangement of molecules. The substrate temperatures vary from $\sim$30 to $\sim$300 °C, and films with $k$ values close to 2 have been obtained. The stability of these films at $\sim$400 °C, required for successful integration, is not known. The precursors used in this method are molecules with unsaturated carbon bonds. Vinyltrimethylsilane, divinylidimethylsilane, tetravinylsilane, divinylsiloxane, trimethylvinylcyclosiloxane, and benzocyclobutene are typical examples. The first method is called subtractive fabrication, and the second method is known as structural fabrication.95

The composition of the SiCOH and porous SiCOH (pSiCOH) films are analyzed using Rutherford backscattering (RBS), forward recoil elastic scattering (FRES), and XPS techniques. Significant differences among the results obtained by the various methods exist and the relative concentrations of Si, C, O, and H vary by 5–10% depending on the analysis technique.95 Thus RBS data indicate that SiCOH ($k=3$) contains 22% Si, 11% C, 35% O, and 32% H, while pSiCOH ($k=2.4$) contains 28% Si, 28% C, 44% O (excluding H). For the same pSiCOH ($k=2.4$) film, XPS data show 33% Si, 18% C, and 49% O (excluding H). FTIR spectroscopy is used to establish the structure of these films. The spectra have similar features over a wide range of $k$ values including a main Si–O stretch band that appears around 970–1250 cm$^{-1}$ with a peak at 1047 cm$^{-1}$ and a shoulder at a higher wavelength. Deconvolution of this band shows that it consists of three peaks at 1023, 1063, and 1135 cm$^{-1}$ corresponding to Si–O stretch vibration in C–Si–O bonds, Si–O–Si stretch vibration with terminal Si–O bonds, and Si–O–Si stretch vibration in a caged structure. Since the intensities of C–Si–O and Si–O–Si peaks are approximately equal, the ring configuration of tetramethyldicyclosiloxane appears to be preserved in the film. In addition a Si–CH$_3$ stretch peak (1274 cm$^{-1}$), a small C–H stretch band (~2900 cm$^{-1}$), and a Si–H stretch band (2900 cm$^{-1}$) are also present. For pSiCOH films, an additional band at 2900 cm$^{-1}$ corresponding to C–H$_x$ ($x=1–3$) vibrations is present indicating large amount of incorporation of CH$_x$ moieties in the film. On annealing, this band intensity decreases indicating the loss of these

![Figure 15. SEM cross section of a multilevel interconnect showing thin wire and fat wire levels. Reprinted with permission from ref 67. Copyright 2007 IEEE.](Image)
species. Comparison of the spectra of SiCOH and pSiCOH films show that a significant amount of C–H bonds remains after annealing. The optical, electrical, and mechanical properties of the low-$k$ and porous low-$k$ materials are largely determined by the composition and structure of these materials.

The dielectric constant is a key component of the electrical properties of the low-$k$ and ultralow-$k$ materials. It depends on the deposition conditions and the nature and amount of porogen precursor used. Thus films with $k$ values as low as 2.0 have been reported. As mentioned earlier, the incorporation of a thermally unstable phase, which is removed by annealing, creates porosity in the film. If the geometrical dimensions are maintained during annealing, the resulting porous film will have a lower density and hence a lower dielectric constant. However, in practice, depending on the porogen molecules used, significant shrinkage occurs during annealing. In some cases, as the concentration of the precursor is increased, thickness collapse can occur instead of continuous increase in porosity. Thus the lowest value of $k$ that can be achieved for a given set of precursors becomes limited. For other molecules, the thickness reduction during annealing reaches a limiting value and further increase in porogen concentration results in higher porosity and lower dielectric constant. Thus a careful choice of precursor molecules and deposition conditions is required to achieve the desired dielectric constant. In addition to the increase in porosity, increase in cage structure can also help in achieving lower $k$ values. Leakage current is another important electrical property and a critical requirement for the low-$k$ and ultralow-$k$ films. Leakage current values below $2 \times 10^{-9}$ A/cm² at 1 MV/cm are required for successful integration. Typical values for SiCOH and pSiCOH are in the range $2 \times 10^{-10}$ A/cm, much lower than the requirements for integration of the dielectric in a chip.

The mechanical properties of the dielectric films are critical to the reliability of the interconnect structures. Cracks in the structures develop in thicker films when subjected to repeated mechanical and thermal stress, and this process is accelerated by exposure to an aqueous chemical environment. In general, PECVD films are more resistant to cracking, because of the random, covalently bonded three-dimensional structures, than spin-on films of similar $k$ values. The crack development velocity in water for SiCOH films is in the range $10^{-12}$ to $10^{-10}$ m/s, and for equivalent spin-on films, values are orders of magnitude higher and in some cases reach up to $10^{-3}$ m/s. The coefficient of thermal expansion of Cu and the dielectric materials should be close to each other to avoid thermal stresses and resulting structural damage during repeated thermal cycling. The values for SiCOH materials are around $(11.8–12.3) \times 10^{-6}$ K, comparable to $17 \times 10^{-6}$ K for Cu. The low-$k$ and ultralow-$k$ materials have to meet a significant number of integration requirements. Compared with SiO₂, low-$k$ materials are mechanically weak and contain high levels of porosity, which makes them susceptible to mechanical damage. This factor has to be taken into consideration in the design of the CMP slurries and processes for low-$k$ and ultralow-$k$ structures.

The electrical and mechanical properties mentioned above pertain to as-deposited and well-annealed films. When exposed to various processing conditions such as high temperature, plasma (etch and deposition), chemical environment, and mechanical stress, one or more of these properties are altered. In addition, structural changes caused by a preceding process may influence the outcome of a succeeding process. A prime example of this is the effect of oxygen and CF₄ on the low-$k$ and ultralow-$k$ materials. Plasma etching processes are widely used for pattern transfer into dielectric films. Once the pattern is established, the photo-resist should be removed, and the trenches should be cleaned. For resist strip and etch residue cleaning, oxygen plasma has been the traditional choice for oxide dielectrics. With SiCOH materials, oxygen plasma causes severe oxidation and converts it into an oxide-like material. This oxidative conversion is accelerated due to the porosity of the pSiCOH, and the oxide-like layer is thicker. The top layer becomes hydrophilic due to the removal of Si–H and Si–CH₃ bonds, and the entire film becomes more porous. Film thickness is reduced due to significant loss of material. With CF₄ plasma, a fluorine-rich layer containing CFₓ and SiFₓ compounds is formed at the surface. While the top surface is rendered hydrophilic (low contact angle), the bulk remains hydrophobic as indicated by FTIR spectra. Low-pressure oxygen/nitrogen reactive ion etching (RIE) is also known to cause this type of surface damage. To avoid the exposure of the low-$k$ and ultralow-$k$ materials to plasma during etching, hard-mask layers consisting of one or more of SiO₂, SiC, SiCN, Si₃N₄, SiCH, and metals are used. Unfortunately the PECVD deposition of the hard-mask materials also induces plasma damage to the surface of the low-$k$ materials. The resulting hydrophilic surface enables the absorption of water, solution, and slurry components into the film during CMP when the hard mask is removed. The plasma damage can also weaken the interface between the hard mask and the dielectric causing delamination due to poor adhesion.

5.2.2. Integration Challenges

The ultimate performance of the CMP slurry and the polish process is heavily affected by issues that originate at other unit processes. The key concerns in low-$k$ and ultralow-$k$ PECVD deposition process are film thickness nonuniformity, film quality, thermal stability, water absorption, interfacial adhesion, and film integrity across the whole wafer. RIE processes may introduce side-wall and corner damage that result in poor liner and Cu seed layer coverage, resulting in defects in the plated Cu, which are exacerbated by the CMP chemistry. RIE selectivity to the hard-mask and liner materials determines the trench depth profile and the final spacing between adjacent lines. Trenches that are wider at the top will cause narrowing of the spacing between lines and require much longer overpolish to avoid shorting due to residual metal. The smaller spacing can also result in excessive current leakage. Such structures also alter the load distribution during CMP, and the higher localized stress may result in mechanical damage of the lines. The large topography variation associated with the copper electrodeposition process incoming to CMP requires longer overpolish and results in preferential dishing of large isolated structures. The interaction between upstream and downstream processes necessitates the simultaneous optimization of dielectric patterning, RIE and clean, liner/seed layer deposition, and Cu electrodeposition to create structures without weak interfaces and adhesion problems that are mechanically strong before and after CMP.
5.2.3. CMP Process Challenges

The performance metrics for the Cu CMP process include removal rates, selectivity, global and local planarity, surface topography, dishing and erosion, defectivity, and throughput. For low-κ and ultralow-κ interconnects, the primary concerns are weak interfaces, delamination, nonuniformity of the cap and dielectric films, and material compatibility with slurry chemistry. As the interconnect dimensions decrease further, many of the requirements and tolerances become extremely stringent. Post-CMP cleaning is a continuing and difficult challenge, especially for ultralow-κ interconnects. Thus for 22 nm node and beyond, the critical particle diameter for random surface particles is ~12 nm, and the target random particle/wafer pass is 23 defects/m². Innovations in slurry chemistry are essential to meet these requirements.

From a processing perspective, CMP is expected to accommodate the incoming variability associated with the upstream processes while meeting the exacting requirements specified by the downstream processes. Even more challenging is that with each technology node, the design ground rules, minimum line widths, number of interconnect levels, integration schemes, hard-mask and soft-mask materials, liner thickness, and dielectric materials change. However the same copper and liner slurries are expected to perform across all the technology nodes with perfection.

5.3. Copper Planarization Process

The major components of a Cu CMP slurry are oxidizer, abrasive, adsorbate (inhibitor), surfactants, chelating agents, and polyelectrolytes. Several oxidizers such as HNO₃, H₂O₂, Fe(NO₃)₃, KI, KMnO₄, K₂Cr₂O₇, and K₂S₂O₈ have been investigated. Most commonly used oxidizers in manufacturing are H₂O₂, K₂S₂O₈, and Fe(NO₃)₃. Most Cu slurries have alumina abrasive because of its high selectivity toward Cu and the ability to stop on Ta/TaN. Increasingly silica-based slurries are becoming more common, especially for the second step in a two-step copper polish. The silica abrasives have appreciable polish rates for Ta/TaN barrier and need additional components in the slurry to suppress liner erosion. Benzotriazole (BTA) is by far the most widely used adsorbate or inhibitor, even though other triazoles and several organic compounds that are known to adsorb on Cu have been investigated. Earlier slurries did not have strong chelating agents in the formulation because in acid pH highly soluble Cu⁺⁺ ions are the dominant species and achieving high rates is not a problem. Since the Cu polish rates are low in neutral and alkaline slurries, strong chelating agents were added to increase the rate. Addition of complexing agents is essential to avoid the precipitation of copper hydroxide above pH 6, keep the Cu ions in solution, and thus reduce surface particulate contamination. Ammonium ions, ethylenediamine, EDTA, and carboxylic acids including amino acids have been used to achieve this goal. Organic acids also form complexes with Cu via their carboxylate group. However, at low pH, they get protonated and thus cannot form stable complexes with Cu. At alkaline pH, they are unable to compete with hydroxyl groups and hence cannot stabilize Cu ions in solution. Thus many of the organic acid chelating agents can be used only in a narrow pH range. Ethylenediamine can be used over a wide range of pH and is found in many slurry formulations. Ethylenediamine exists in three different forms: H₂N–CH₂–CH₂–NH₃⁺, H₂N–CH₂–CH₂–NH₂, and H₂N–CH₂–CH₂–NH₃⁺ with equilibria between them. The dominant species below pH 6.8; H₂N–CH₂–CH₂–NH₂ is dominant above pH 9.9. H₂N–CH₂–CH₂–NH₂ exists between pH 6.8 and 9.9. The electrochemical behavior of Cu in solutions containing ethylenediamine indicates that due to its strong chelating ability ethylenediamine can stabilize copper ions in alkaline solutions and thus extend the Cu solubility range to higher pH. Since amino acids can also exist in three forms as cationic, anionic and zwitterionic species and form complexes with Cu, amino acids such as glycine have been investigated as possible candidates for chelating Cu with hydrogen peroxide as the oxidizer. Catalytic decomposition of hydrogen peroxide by transition metals and their complexes via generation of hydroxyl radicals is well-known. Cu–glycine complexes in the presence of hydrogen peroxide can generate high Cu dissolution rates and polish rates due to the enhanced generation of hydroxyl radicals. This reaction is also accelerated in the presence of Cu²⁺ ions, and addition of Cu²⁺ ions to glycine/H₂O₂/BTA solutions has been shown to increase the static etch rate. It is conceivable that the local concentration of Cu²⁺ ions can be substantially higher in high-density patterned areas and thus lead to much higher decomposition of peroxide in glycine/H₂O₂/BTA-based slurries. This can result in very high localized etch rates as well as polish rates. Significant dishing and corrosion can occur under these conditions. This aspect of the glycine/H₂O₂ systems needs to be carefully investigated further. It appears that the main role of the complexing agents is to accelerate the dissolution rate of copper. In general, these complexing agents do not have a strong affinity for copper surface or have the ability to compete with adsorbate molecules like BTA. Thus the passive films formed on the Cu surface in glycine–H₂O₂ systems in the absence of BTA consist of various oxides of copper such as CuO and Cu₂O.

Of all the chemicals in the Cu slurry, benzotriazole is perhaps the most important and critical component. It is known to form an adsorbed film on the copper surface that provides exceptional protection from all forms of corrosion. Numerous studies have focused their attention on characterizing the properties of BTA films formed under varying conditions (vacuum, solution, electrochemical, polishing) on a variety of Cu surfaces (PVD, CVD, electroplated, single crystal) probed by various surface microscopic (atomic force microscopy (AFM), electrochemical AFM (ECAFM)), ellipsometric and surface spectroscopic (infrared reflection/absorption spectroscopy (IRAS), surface-enhanced Raman spectroscopy (SERS), XPS) techniques. Several different species and structures have been proposed depending on the experimental conditions under which the films were formed and were investigated (in situ, ex situ, immersion/emersion). Film thickness values range from <10 to >4000 Å depending upon the surface and film preparation techniques used. Despite the number of such investigations, the answers to the most fundamental questions remain elusive and our understanding of the nature of the interactions at the interface incomplete.

Benzotriazole is known to exist in several forms in solution depending on the pH. At pH values around 2–3, the predominant species is the monoprotonated BTAH⁺. It forms diprotonated species with hydrogen on adjacent nitrogens (BTAH⁺⁺) at pH values close to zero. The anionic species BTA⁻ are formed in alkaline pH. The corresponding pKₐ-
values for the deprotonation steps are \( \sim 1.0 \) and \( \sim 8.2 \). The adsorption of BTA will also depend on the electrical surface charge density; thus different oxidizers and the presence of other molecules such as surfactants, anions, and cations can and do exert profound influence on the interfacial process.

Cohen et al.\(^{112}\) have studied BTA films grown under electrochemical conditions on 2 \( \mu \)m thick PVD Cu films using XPS (ex situ) and ellipsometry (in situ) at pH 2, 8.2, and 12 in the presence of \( \text{Fe(NO}_3\_\text{)}_3 \) and \( \text{K}_2\text{S}_2\text{O}_8 \). The results indicate that BTA adsorption occurs on all copper surfaces including \( \text{Cu}^0 \) (free of oxides), \( \text{Cu}/\text{CuO} \), and \( \text{Cu}/\text{Cu}_2\text{O} \). The composition and thickness of the films vary depending on the pH, electrode potential, solution composition, and amount of dissolved oxygen. When the Cu electrode is kept at cathodic potentials to ensure that the Cu surface is in a reduced state free of oxides and no oxidizer is present, BTA films are about 5 Å thick. At more anodic potentials where \( \text{Cu}_2\text{O} \) or \( \text{CuO} \) formation is allowed, the adsorbed film indicates that the surface layer has a significant amount of high solubility, and BTA exists in the protonated form \( \text{BTA}^+ \). The Cu surface is free of Cu oxides because of their \( \approx \) films are about 5 Å thick. At more anodic potentials where \( \text{Cu}_2\text{O} \) or \( \text{CuO} \) formation is allowed, the adsorbed film thickness is about 5–7 Å. This is consistent with the formation of a monolayer of BTA molecules adsorbed on the oxide-free or oxide-covered copper surface. This layer is formed first and believed to be present on any Cu surface that has been exposed to the BTA molecules in solution.

The situation is different in the presence of oxidizers where Cu dissolution is allowed. In acidic \( \text{Fe(NO}_3\_\text{)}_3 \) solutions (pH \( \approx 2 \)), the Cu surface is free of Cu oxides because of their high solubility, and BTA exists in the protonated form \( \text{BTAH} \). XPS analysis of films formed under these conditions indicates that the surface layer has a significant amount of \( \text{Cu}^+ \) in addition to BTA. The presence of a \( \text{Cu}^+-\text{BTA} \) complex is also confirmed by the absence of any peaks corresponding to \( \text{Cu}^{2+} \). Relatively thick films (80–100 Å) are formed under these conditions. For \( \text{K}_2\text{S}_2\text{O}_8 \) at pH 2, the thickness is about 10–40 Å.

In alkaline pH (\( \sim 12 \)) at potentials where \( \text{CuO} \) and \( \text{Cu}_2\text{O} \) formation and Cu dissolution are allowed, film thickness is about 10–35 Å. At pH 8.2 in borate buffer solutions, the thickness is slightly higher, 20–40 Å. Thus the BTA film consists of two layers, the first is a monolayer of adsorbed BTA molecules with strong interaction with the copper surface and the second is a much thicker layer of \( \text{Cu}^+-\text{BTA} \) molecules formed by the dissolution of Cu in the presence of oxidizers or at anodic potentials. Thicker layers are formed in acidic pH and in the presence of \( \text{Fe(NO}_3\_\text{)}_3 \), and much thinner films are formed in alkaline pH with no oxidizer present.

The nature of adsorption of benzotriazole was investigated by Weaver et al.\(^{113}\) using surface-enhanced Raman spectroscopy under controlled potential conditions. Vibrational frequencies and bandwidths corresponding to the functional groups directly attached to the metal surface will be affected by adsorbate/surface interactions. The relative intensities of the vibrational bands can shed light on the possible orientation of the adsorbed molecules. The Raman spectra for adsorbed benzotriazole in 0.01 M BTAH in 0.1 M \( \text{Na}_2\text{SO}_4 \) and 0.005 M \( \text{H}_2\text{SO}_4 \) at pH 2 on Cu electrodes kept at \(-0.7 \text{ V} \) (potential corresponding to adsorbed BTAH at pH 2 and not \( \text{Cu}^+-\text{BTA} \)) indicate that the benzene ring does not interact with the Cu surface. This observation is in agreement with interpretations based on photoemission, UV/X-ray photoelectron, and X-ray absorption studies. The attachment of benzotriazole molecule on the Cu surface occurs via the triazole rings with one and possibly two nitrogen atoms interacting with Cu, and the molecular plane is perpendicular to the surface. At more anodic potentials (\( \sim 0.3 \text{ V} \)), BTAH becomes deprotonated and formation of \( \text{Cu}^+-\text{BTA} \) occurs via the reaction

\[
\text{Cu} + \text{BTAH} \leftrightarrow \text{Cu}^+\text{−BTA} + \text{H}^+ + \text{e}^- \quad (13)
\]

Changes observed in the spectra are consistent with the Cu(I) coordination and the SER spectra of adsorbed and bulk \( \text{Cu}^+-\text{BTA} \) are closely similar indicating the formation of a phase film. This potential-induced transformation from adsorbed BTAH to a \( \text{Cu}^+-\text{BTA} \) phase film is also confirmed by the lack of any vibrational effect when the imino hydrogen is changed to deuterium. Comparison of the segment of the SER spectra for copper immersed in 10 mM BTAH and 0.1 M \( \text{H}_2\text{SO}_4 \) at \(-0.7 \text{ V} \) and the spectra of the same Cu surface obtained immediately after emersion into air revealed that spontaneous conversion of adsorbed BTAH to \( \text{Cu}^+-\text{BTA} \) occurs on exposure to air. This is explained by the following half-cell reduction reaction:

\[
\text{O}_2 + 4\text{H}^+ + 4\text{e}^- \rightarrow 2\text{H}_2\text{O} \quad (14)
\]

The electrons will be provided by the Cu surface resulting in the dissolution of Cu and the subsequent formation of \( \text{Cu}^+-\text{BTA} \). The overall reaction is written as follows:

\[
4\text{Cu} + 4\text{BTAH}^{\text{ads}} + \text{O}_2 \rightarrow 4\text{Cu}^+\text{−BTA} + 2\text{H}_2\text{O} \quad (15)
\]

Stewart et al.\(^{114,115}\) studied the effect of anions on benzotriazole films formed on copper surfaces in slurries containing silica abrasives (5 wt %) and \( \text{H}_2\text{O}_2 \) at pH 2.0 and 8.5. The Cu polish rate decreased by an order of magnitude when chloride was added to the slurry. AFM and ellipsometry measurements showed that thicker films were formed in acidic pH (2) compared with alkaline pH (8.5). In the presence of chloride and bromide, thicker films were observed at both pH values. BTA films grown in the presence of acetate, sulfate, and perchlorate were much thinner than the films in the presence of halides. SER spectral analysis exhibits no evidence of Cu−Cl interaction in the presence of BTA. Laser desorption ionization mass spectra (LDIMS) provide evidence for halide incorporation in the BTA films. A similar conclusion was also reached by secondary ion mass spectrometry (SIMS) analysis. Stewart et al. assert that the halide incorporation results in thicker films (>1000 Å) because the halide ions coordinate with Cu(I) in the \( \text{Cu}^+-\text{BTA} \) films stabilizing the Cu(I) and hence the entire film as well.

The picture that emerges from these studies is that at low pH (\( \sim 2.0 \)) benzotriazole exists in solution mainly as BTAH and at higher pH (\( \sim 8.5 \) and above) as BTA. The attachment to the Cu surface is via the triazole nitrogen atoms with the benzene ring perpendicular to the Cu surface. The BTA films on copper consists of two layers: the first thin (\( \sim 7 \text{ Å} \)) layer of chemisorbed BTAH or BTA− species and a second much thicker layer (up to \( \sim 700 \text{ Å} \)) containing Cu−BTA molecules. Spontaneous conversion of the chemisorbed BTAH or BTA− film to \( \text{Cu}^+-\text{BTA} \) phase film can occur upon initiation of Cu dissolution due to exposure to moist oxygen or change in electrode potential to more anodic values. Halide ions get incorporated into the BTA film because of the strong affinity toward Cu(I) in the \( \text{Cu}^+-\text{BTA} \) film, and thus very thick films (>1000 Å) are formed in strongly acidic solutions with oxidizers like \( \text{Fe(NO}_3\_\text{)}_3 \).
Surfactants have been used in Cu CMP slurries to modify and alter the BTA films or as better alternatives to BTA films. Anionic surfactants like sodium or ammonium dodecyl sulfate or cationic surfactants like cetyl trimethyl ammonium bromide (CTAB), dodecyl trimethyl ammonium bromide, and cetyl pyridinium chloride have been reported. The adsorption of these molecules can be classified as physical adsorption and often exhibits Langmuir isotherm behavior. Very often they also adsorb onto the abrasives and cause slurry instability. Thus addition of n-alkyl sulfates to alumina and addition of CTAB to silica slurries will cause immediate coagulation and precipitation. In some cases, redispersion can be accomplished by simple stirring, in other cases addition of dispersants and stabilizers may be needed to enhance the stability of these systems. It is also believed that these surfactants may cause aggregation of abrasive particles over time and lead to increased defectivity. Surfactants can also interact with pad materials and change the frictional forces involved in the polishing process. Depending on the nature of surfactants used they may increase or decrease the level of particulate contamination of the wafer. Thus surfactant interactions are highly complex and intricate, involving multiple components of the CMP process including abrasives, pad, metal, liner, or dielectric surfaces simultaneously.

5.4. Ta/TaN Liner CMP Process

Cu is extremely mobile in Si at high temperatures and when present creates trap levels in Si that adversely affect device operation. Hence, it is necessary to have an effective diffusion barrier that prevents Cu migration. One of the requirements for a diffusion barrier is that it is metallurgically stable with respect to Cu, and many refractory metals can meet this requirement since they do not react with Cu even at high temperature. Unfortunately, copper can penetrate through metal layers without reacting with it. Thus the design of a diffusion barrier for Cu posed some unique challenges.118

Hu et al. investigated the diffusion barrier properties of Ta, a refractory metal that does not react with Cu.119 They detected no diffusion of Cu through Ta layers deposited on oxidized Si up to 750 °C. The barrier properties of Ta can be further improved by incorporating impurities in the film. At high concentrations, these atoms in the Ta grain tend to segregate to the grain boundaries and block the pathways for copper diffusion. Holloway et al.118 demonstrated that nitrogen-alloyed Ta is a more effective barrier to copper diffusion. The diffusion barrier properties of Ta and its nitrides are enhanced as the nitrogen concentration in the film is increased.120 Cu interconnect structures typically have Ta/TaN layers as diffusion barriers; the thickness usually varies from 700 to <100 Å depending on the technology node.

Ta/TaN liner polish is a more challenging and complex CMP process than the Cu polish process. The Ta/TaN layers should be removed first, and then the hard-mask (or dielectric cap) materials should be removed to certain target thickness. For the direct polish scheme, the entire hard mask and about 300–400 Å of the dielectric should also be removed. From a slurry formulation point of view, this becomes extremely complex since the polish rate of Ta/TaN, Cu, the hard-mask materials, and dielectric are coupled. For example, the oxidizer concentration controls the Ta/TaN and Cu rates, and abrasive concentration controls Ta/TaN, Cu, hard-mask, and dielectric rates. As a result, independent control of the polish rates of different materials is nearly impossible. This is a critical problem in optimizing the slurry performance and fine-tuning the polish rates of different materials to achieve excellent planarity, electrical yield and minimum dishing/erosion simultaneously.

Typical Ta slurries contain oxidizers (H2O2, potassium persulfate, potassium iodate), abrasives, benzotriazole, Cu-complexing agents, surfactants, and polyelectrolytes. Surfactants and polyelectrolytes are generally used to eliminate water marks, particulate surface contamination, pad fibers, and stains. Polynonylpyrrolidone121–124 is used for the control of SiCOH-like hard-mask and SiCOH dielectric polish rates. Most liner slurries in practical use are based on colloidal monodispersed spherical silica particles. However polishing of Ta with alumina-based slurries over a wide range of pH has been observed.125–127

The variation of Ta polish rate over a wide range of pH has been investigated in the presence of oxidizers126,128,129 with alumina and silica abrasives. In most acid and alkaline noncomplexing solutions, the Ta surface is well-protected by the formation of tantalum pentoxide, Ta2O5, which is stable over a wide range of pH. However, partial dissolution of the oxide has been reported in the presence of complexing agents and at highly alkaline pH. Thus it dissolves slightly in sulfuric acid and more readily in oxalic acid and concentrated HF forming complexes. Thus the polish rate of Ta is strongly influenced by the kinetics of formation of the pentoxide film.

Oxidation of Ta occurs in aqueous solutions according to the reaction130

$$2Ta + 5H_2O \rightarrow Ta_2O_5 + 10H^+ + 10e^- \quad (16)$$

In acidic solutions, simultaneous reduction of hydrogen also occurs as shown in eq 17.

$$O_2 + 4H^+ + 4e^- \rightarrow 2H_2O \quad (17)$$

Since this reaction proceeds at a higher rate at lower pH, Ta oxidation rate can be expected to be higher at lower pH. In alkaline solutions (pH 12), a very high rate of oxidation is observed. Kuiry et al.129 attribute this to the rapid formation of Ta(OH)4+ species that decompose to form Ta2O5. It is also known that Ta2O5 dissolves in strongly alkaline solutions (pH 12) to form soluble hydroxotantalates [Ta(OH)]5−3+ and oxotantalates [TaO(OH)]3−2+. Under these conditions, in the presence of H2O2, the thickness of the tantalum oxide is governed by the relative rates of formation and dissolution of the tantalum oxide. For silica slurries, in general, higher Ta Polish rates are observed at pH ≈ 2–3 and 11–12. For alumina-based slurries, polish rate increases from pH 2 to 4 and remains constant in the range 4–8. A rapid increase (∼4-fold) in the polish rate is observed from pH 10 to 12, reaching a value of ~2500 Å/min. Jindal et al.126 observe that the oxide layer formed in alkaline solutions in presence of hydrogen peroxide is weak and highly porous129 compared with the oxide formed in acidic solutions. They suggest that the higher polish rate in alkaline pH can be attributed to the easy removal of the weak oxide film by abrasive action.

The thickness of Ta/TaN barrier has been steadily decreasing for each technology node, and for 45 and 32 nm interconnects, the total thickness is about 30–40 Å. Thus the barrier metal layers are removed in a few seconds, and
the rest of the liner polish is mainly concerned with the removal hard mask and 200–300 Å of the low-κ dielectric materials. The critical problem is that the removal of the hard-mask material is essential to avoid the increase in the effective dielectric constant in the final structure. On the other hand, complete removal of hard mask exposes the low-κ and ultralow-κ materials to the slurry chemistry and can adversely affect the electrical and mechanical properties of the low-κ and porous low-κ materials. Ishikawa et al. investigated the effect of CMP slurry and post-CMP cleaning chemistry on the degradation of spin-on porous low-κ materials. Upon exposure to the slurry and post-CMP cleaning solutions, the dielectric constant, leakage current, and refractive index of the porous low-κ films increased. The slurry and the alkaline cleaning solution also altered the hydrophobicity of the surface as indicated by the decrease in the contact angle. Rinsing with isopropyl alcohol did not restore the dielectric constant, leakage currents, and refractive index to the initial values. Heat treatment at 400 °C in 1,3,5,7-tetramethyl-cyclo-tetrasiloxane vapor for 4 h restored the values to the original levels, and the film surface also became hydrophobic after heat treatment. FTIR spectra showed that absorption intensities of OH and CH₃ increased after exposure. This is attributed to the permeation of water and other components such as surfactants into the film. Thermal desorption mass spectral analysis showed that part of the surfactant that had permeated into the film could be removed at 400 °C while other components could be removed only by heating to 600–700 °C. The elements desorbing at higher temperatures were also removed by isopropanol rinse. Clearly it is evident that components in the slurry can penetrate into the porous low-κ films and change the electrical and mechanical properties of the films. It is conceivable that bonds such as Si–CH₃ are broken and replaced with Si–OH bonds in alkaline solutions.

The combined effect of stress and solution chemistry on the thermomechanical reliability of poly(methyl silsesquioxane) (PMSSQ) and SiOC nanoporous films during CMP has been investigated. In aqueous solutions, crack propagation occurs due to the breaking of Si–O–Si bonds and the formation of Si–OH and Si–O– bonds. Increasing pH accelerates crack growth rates. In the presence of nonionic surfactants, poloxymethylene ethers with hydrophobic tail and hydrophilic head, the crack growth is sensitive to hydrophilic ethylene oxide chain length as well as the number of CH₂ groups in the hydrocarbon chain. The magnitude of the sensitivity to crack growth varies with the pH of the solution. Acceleration in crack growth is observed with dimeric Gemini surfactants with two ethylene oxide chains. Thus a wide range of behavior varying from acceleration to suppression of crack growth can be observed depending on the surfactant used. These studies indicate that small changes in slurry chemistry such as pH, nature of surfactants, and additives can exert a large influence on the k values and crack growth rates. Significant degradation in the performance and reliability of ultralow-κ interconnects can occur due to these effects. Most of the studies found in the literature have focused their attention on spin-on low-κ/ultralow-κ dielectrics. A clear understanding of the effect of pH, molecular structure, diffusion properties of the surfactants/additives, and their influence on the electrical and thermomechanical properties of the PECVD porous low-κ materials is urgently needed to ensure the performance and reliability of the Cu–ultralow-κ interconnect structures.

6. CMP Process-Induced Defects

Defects are a major concern in any semiconductor unit process. It is more so for CMP since it involves removing excess material on top and exposing the structures, which enables visual inspection of the irregularities and imperfections lying underneath. Presence of defects is often directly associated with the presence or absence of one or more components in the slurry. Thus scratches, corrosion, and dishing/trenching are often attributed to the presence of abrasives and oxidizers and lack of adequate quantities of inhibitors and surfactants. A considerable amount of effort is devoted to reducing the number and severity of defects during slurry development and optimization. Even well-optimized and well-behaved slurries can occasionally cause high defect levels because of aggregate formation and other instabilities associated with colloids. Polish pads and conditioning processes are also other major contributors to high defect levels. Very often, there may be multiple sources for higher defect levels, and it may not be possible to determine the exact cause.

Many defects that are discovered after CMP originate at earlier process steps; however, they cannot be usually detected until CMP is completed. CMP processes often accentuate and exacerbate the preexisting defects, and thus many defects become detectable only after CMP. The CMP process itself can generate a variety of defects that are influenced by the problems in the preceding process steps such as liner deposition and CVD W and Cu deposition processes. Since defects often have a direct effect on the chip yield, lowering defect density is of paramount importance during CMP process development as well as during manufacturing. Typically the defect densities are high during the early stages of the life of a technology node or a product with a technology node. As the process technologies mature and successful efforts are undertaken to identify and eliminate the root causes of defects; defect densities become lower and yields become higher. The defect density requirements become more stringent as the device dimensions become smaller and smaller. The size of the defects and their density are continuously reduced for each technology node, posing enormous challenges for CMP consumables and process development efforts. In addition, the costs associated with the invention, development, implementation, and manufacturing become prohibitively high.

Some of the process-induced defects such as scratches, particulate contamination, surface residues, and residual material are common for all CMP processes, while others are specific for each CMP process. Even among the same defect class, some are “killer” defects and even when present in small numbers can cause huge yield loss, while others are cosmetic defects and may not result in significant yield loss. Even these cosmetic defects need to be eliminated because, if present in large numbers, they can overwhelm the defect inspection systems and prevent the detection of the small number of killer defects.

6.1. Defects in FEOL CMP

The family of defects generated in the STI, ILD, and poly-Si CMP processes are very similar in nature. The main differences are due to the variations in the type of oxides used and the corresponding differences in the type and severity of the defects. Most of these applications, with the exception of STI, involve polishing only one material, oxide,
boron- and phosphorus-doped silicate glass (BPSG), or poly-Si. BPSG is a softer material and more prone to defect generation than TEOS and HDP oxide.

Scratches are one of the most commonly observed defects in FEOL CMP processes. They are not usually easily detected immediately after CMP and are usually highlighted after etching in dilute HF. If the localized stress induced during the CMP process is high, then dislocation or fracture occurs on the silica surface. This may not be readily visible unless the near by material is removed by etching. These defects are called “chatter marks” and are shown in Figure 16. Small microscratches are also not easily detected after CMP and appear after exposure to dilute HF. The oxide scratches are usually attributed to the agglomeration of abrasive particles and the formation of large aggregates that cause scratch defects. Fumed silica has a higher tendency to form large particles compared with monodispersed, spherical colloidal silica.137 Figure 17 shows the comparison of defect densities for fumed silica and colloidal silica slurries. Similarly ceria is known to have a higher tendency to scratch oxide surface compared with silica even though the hardness values are comparable. Increasingly, slurries are being formulated to include surfactants and dispersants to improve the defectivity by reducing the abrasive—surface interactions. However, these additives might lower the polish rates at the concentrations needed to be effective and may affect the long-term stability of the colloidal system.

Another major contributor to scratching is the polishing pad. In general, hard pads exhibit higher scratch densities than softer pads. However, with certain abrasives, severe scratching can occur even with softer pads. Complex interactions exist between the pad, abrasives, and wafer surface during polishing depending on the pH, nature, and type of abrasive particles and the presence of surface active molecules in the slurry. For example, silica is known to cause pad glazing due to the formation of silica networks on the surface of the pad, and diamond dressing after each polish is necessary to eliminate pad glazing. When diamond dressing is inadequate, it can leave patches of glazed areas in the pad that can increase the localized stress and result in higher levels of scratching. Thus defect density reduction involves a series of complex optimizations involving several components such as slurry composition, pads, diamond dressing conditions, and polish process parameters for a given material set.

Complete removal of oxide on top of nitride is essential in the STI CMP process. This should be accomplished across all line widths, pattern densities and feature size. When thin oxide is left, it acts as a mask and prevents the removal of nitride during nitride strip. The residual nitride is one of the major yield detractors resulting from poor STI CMP process performance. A dilute HF etch step is used immediately after CMP to ensure removal of any residual oxide before nitride strip. Abrasive particles and pad debris or residue found on the wafer surface are the other commonly known defects in FEOL CMP indicating a poor post-CMP cleaning process.

6.2. Defects in MOL Tungsten CMP

Tungsten is relatively inert in most aqueous solutions containing oxidizers due to the rapid formation of protective oxides, and hence it is not normally prone to severe corrosion. However, few instances of corrosive attack have been observed in slurries containing hydrogen peroxide. When scratching is observed, it is usually in the BPSG dielectric and is very similar to the “chatter marks” seen in oxide polish. The most common defect unique to the W CMP process is the formation of voids in the W vias. The CVD tungsten deposition process is known to leave seams in the middle of high aspect ratio vias, and they appear as voids after CMP. This defect can be a major yield detractor if the electrical contact is lost due to the void. Improvements in via profiles and the CVD tungsten deposition process can alleviate this problem. Residual tungsten and liner due underpolishing is also a known concern. The post-CMP tungsten topography can introduce severe shorting in the first Cu interconnect level.

6.3. Defects in Cu BEOL CMP

From a defectivity perspective, Cu CMP is arguably the most challenging of all CMP processes. Cu is relatively soft and is easily vulnerable to corrosive attack in air and in solution. During CMP, it is in contact with more cathodic liner materials such as Ta and Ru. The interconnect integration schemes involve a number of organic, dielectric, and metal layers that are deposited by vacuum and wet chemical
methods creating several weak interfaces. To make matters worse, 10–12 interconnect levels have to be built on top of one another with varying thickness and multiple material stacks. This is perhaps the reason many felt in the early days that Cu CMP may not be a viable manufacturing process.

6.3.1. Corrosion of Copper

Corrosion can be defined as the unwanted dissolution of copper. Since Cu CMP slurries are formulated to remove the metal by dissolution, the possibility of corrosion is always present in the CMP process design. The aim is to control the kinetics of the dissolution and protective film formation processes so that the probability of unwanted dissolution or corrosion is made as small as possible. In patterned wafers, this need to be accomplished within a few nanometer spacing between the overburden areas and the recessed conductor lines. Thus the presence of corrosion defects indicates our inability to achieve this feat perfectly across every nanometer of the wafer surface at all times. In patterned wafers, this need to be accomplished within a few nanometer spacing between the overburden areas and the recessed conductor lines. Thus the presence of corrosion defects indicates our inability to achieve this feat perfectly across every nanometer of the wafer surface at all times. In properly formulated Cu slurries, the kinetics of the three fundamental processes, Cu dissolution, protective film formation, and film removal, are carefully balanced to minimize the probability of corrosion. Hence when corrosion occurs at a location, it indicates the presence of one or more of the following possibilities: the film formation is slow, the protective film is not formed at all, or the dissolution is accelerated. The root cause of many of the corrosion defects and its variations including pitting can be attributed to the above. Missing metal or hollow metal defects, in general, originate from processes other than CMP. When liner/seed layer coverage is poor, the resulting conductor lines have buried voids that are concealed by the overburden. These voids become exposed after CMP and are detected by the inspection systems. Similar defects can also be generated by poor trench filling performance of the plating process as well as the plating chemistries. Slit voids are also a form of corrosion related to the imperfections created by poor coverage by liner or seed layer. Examples of corrosion defects are shown in Figure 18.

6.3.2. Scratches

Long, deep polish scratches are a major yield detractor, and several different types of scratches are known to exist. Small, shallow microscratches are usually cosmetic and do not significantly affect the yield. However, they can swamp the defect inspection systems and may prevent the detection of other killer defects. For this reason, it is necessary to keep their counts as low as possible. Scratches are mainly attributed to abrasives and polish pads. Several factors can influence the scratching tendency of abrasives. Variability in manufacturing methods, impurities or contaminants, surfactants, additives, shelf-life problems, and slurry distribution systems can contribute to higher levels of scratching. Similarly for pads, inadequate break-in and diamond dressing are major sources of scratching. Examples of various types of polish scratch defects are shown in Figure 19.

6.3.3. Dishing, Erosion, and Trenching

Dishing, erosion, and trenching affect the local topography and are key components of performance metrics for CMP processes. Very often the slurry and process that causes minimum dishing, erosion, and trenching is preferred over others. Severe dishing of Cu occurs in large and isolated structures and is usually attributed to localized high polish rates as well as pad flexibility. Filling large Cu pads with oxide or dielectric fill shapes usually reduces dishing significantly. Liner erosion occurs during copper polish in large field areas adjacent to the edge of high density features. Liner erosion also can occur in the small space between wide lines. In general, significantly more erosion is observed for silica-based Cu slurries than alumina-based slurries. Dielectric erosion is observed after Ta liner CMP and is attributed to significantly higher polish rate for the dielectric compared with Cu. Trenching is the loss of dielectric around the large copper features, and the profilometric traces have characteristic features called “fangs” (Figure 20). This is caused by severe erosion of dielectric extending over several micrometers. The amount of trenching is usually dictated by pattern density and amount of filling. All three types of defects mentioned above result in residual metal at the next step.
level and cannot be eliminated without significant overpolishing at that level (Figure 21). This often leads to more dishing and erosion at that level causing residual metal at the next level.

6.3.4. Mechanical Damage

Low-\(k\) and ultralow-\(k\) interconnect structures are easily damaged by excessive localized mechanical stress, as well as the shear forces created by polishing actions. Cu lines may be deformed, smeared, and broken, and the dielectric is cracked under mechanical stress (Figure 22). Mechanical damage usually results in reliability failure and thus is a major concern. Excellent structural integrity and strong interfacial adhesion are prerequisites for a successful outcome in any CMP process. Since Cu interconnects contain more than 10 levels, it is necessary to ensure that the structures are robust enough to withstand repetitive mechanical and thermal stress.

6.3.5. Other Defects

Abrasive particle contamination (Figure 23), organic and polish residue (Figure 24), embedded contamination, and stains are other critical defects that are commonly observed in Cu CMP. Modern slurries contain polyelectrolytes and other additives that reduce the attractive interaction between the abrasives and the copper and dielectric surfaces. The main source of organic residues is pad material and pad debris. A
robust post-CMP cleaning process is necessary to eliminate these defects.

### 7. Models of CMP Processes

Several excellent reviews have appeared over the past decade on various aspects of the mechanics and mechanisms of polishing processes. Many of the early theories pertaining to material removal mechanisms viewed polishing mainly as a mechanical process. The earliest one was by Issac Newton who postulated, "The smaller the particles of the substance, the smaller will be the scratches by which they continually fret and wear away the glass until it is polished." The classic equation for polishing is the Preston equation that describes the process as a two-body wear problem. Abrasion, flow hypothesis, chemical hypothesis, and frictional wear hypothesis are the four main removal hypotheses that emerged from a large body of work that was done in the early part of 20th century. Since each one of these hypotheses failed to explain some aspect of polishing, combinations incorporating one or more of the above hypotheses were proposed. In general, CMP models fall under two main categories: contact mechanics based models that focus on the mechanical interactions that occur at the pad/abrasive–wafer interface and CMP process models that focus on the physicochemical interactions that occur at metal/dielectric–slurry/solution interface. They are discussed in detail in the following sections.

#### 7.1. Models Based on Contact Mechanics

The main goal of models based on contact mechanics is to develop an understanding of the mechanical aspects of the polishing process. According to the contact mechanics models, the polish rate is a function of the pressure distribution between the features and the polish pad. The chemistry aspects are usually represented empirically. These models provide predictions regarding polishability of various patterns in the mask design. Wafer scale, die scale, and feature scale models attempt to develop an understanding of the various phenomena occurring at different length scales ranging from millimeters to nanometers. Wafer scale models develop correlations between the pressure distribution between the wafer and the pad and the uniformity of polish rates across the wafer. Pad compressibility and slurry distribution are taken into account to model the slurry hydrodynamics. Tichy et al. consider the CMP process as a contact mechanics problem of the wafer on a pad lubricated by a thin film of slurry. The thickness of the film at the interface is controlled by the asperities in the pad, and the model is successful at predicting experimental pressure profiles. A statistical model describing the interaction between pad asperities and wafer has also been proposed by Yu et al. Yet another approach is to model abrasive
particle, pad asperity, and wafer surface interactions incorporating wear mechanisms such as adhesive wear, abrasive wear, erosive wear, and corrosive wear to predict the variation of removal rates with applied pressure.\textsuperscript{144–146}

Boning et al.\textsuperscript{147,148} have developed an integrated contact mechanics and density-step-height model that views the CMP process as a chemically enhanced mechanical process. Contact mechanics is used to compute the local pressure at various points on a three-dimensional envelope function that captures the long-range topography at chip level. The pattern density—step height formulation is then used to find the removal rates in the patterns for the computed envelope pressures at different points. To accomplish this, the polish rate vs pressure relationship is first determined for a given slurry system. Then using Hook's law, the variation of feature-scale pressure with pattern density and step height is established. The two relationships are then combined to compute the removal rate as a function of pattern density and step height. Comparison\textsuperscript{149} of simulated data with the experimental dishing and erosion data indicates that the trends can be accurately predicted by the model with in errors of 100–500 Å. They have also investigated the variation of chip level planarity across the whole wafer.\textsuperscript{150} Choi and Dornfeld\textsuperscript{151–153} postulated that the pattern density dependent topography variation is induced by the nonuniformity of the local contact pressure over the chip area. Elasticity theory was then used to calculate the topography-dependent local contact pressure. A chip scale model incorporating the pad asperity height distribution, pattern density dependent topography, and process parameters was developed. The chip scale model predictions are extremely useful in developing circuit pattern layouts that minimize dishing and erosion. However, as pointed out by Vlassak,\textsuperscript{141} they do not provide any insight regarding the physicochemical processes that occur during polishing.

Feature scale models attempt to describe and predict how individual features evolve during polishing. In the two-dimensional model proposed by Warnock,\textsuperscript{154} the elastic deformation of the pad sets the horizontal length scale, and the pad roughness sets the vertical length scale. The model uses pad roughness and compressibility as fitting parameters. This model provides quantitative predictions of topography in 5–500 μm wide features. A feature scale erosion model incorporating hydrodynamic fluid flow in the wafer—pad interface was proposed by Runnels,\textsuperscript{155,156} The shear stresses introduced by the slurry flow on the surface of the features are used in erosion models and microfracturing, and chemical effects are represented empirically. The profiles generated by the model are in agreement with experimental profiles in 10–500 μm features.

The general consensus is that as of now these models are not reliable enough to be used for verification of current processes or optimization of future processes. The main requirement is that the models should be able to address wafer, die, and feature scale issues in an integrated fashion and provide feedback to designers so that circuits can be designed for easy manufacturability and high yield. These models should incorporate linkage to preceding (Cu deposition) and succeeding (lithography at next level) processes so that interactions between them can be taken into consideration in CMP process optimization. Last but not least is the need for defectivity models that link pattern density effects and chip yield. This might require particle scale models that address the role of components such as abrasive particles, slurry chemistry, pad materials, and pad surface characteristics and their interaction with feature sizes, density, and materials in the patterns.\textsuperscript{157}

Philipposian et al.\textsuperscript{158} have developed a tribological approach in which analysis of Strubeck—Gumbel curves is used to identify the dominant tribological mechanism present during polishing. These curves describe the variation of the coefficient of friction with the Sommerfeld number (a dimensionless parameter dependent on the slurry viscosity, relative pad wafer velocity, applied pressure, and effective fluid film thickness) for the given CMP process. The effects of slurry surfactant, abrasive concentration, particle size, and applied pressure have been investigated using a variety of polish pads for ILD and Cu CMP.\textsuperscript{159} This analysis indicates the presence of a universal relationship between the average coefficient of friction and removal rates for various types of pads for oxide CMP.\textsuperscript{158} Philipposian et al. have extended this tribological model to Cu CMP with and without abrasives.\textsuperscript{159–160}

### 7.2. CMP Process Models

Many of the theories and concepts in CMP process modeling originated from glass polishing. In early models, Preston coefficient was identified with chemical processes involved in polishing. Grebenshchikov discovered that a silica gel layer is formed on the glass surface by chemical reaction with water and this layer is removed by polishing action.\textsuperscript{162} Brown and Cook\textsuperscript{163} first proposed the concept of “chemical tooth” to explain the removal of materials during glass polishing. Since interactions between abrasives and the glass surface occur at a molecular level during polishing, material removal is on the scale of chemical processes such as dissolution and not on the scale of mechanical processes such as abrasion. The surface layer is modified by diffusion of water into the glass surface. The subsequent protonation of the hydroxyl groups generates Si—O— moieties on the surface of the glass. The abrasive particles interact with the surface Si—O— species and form Si—O—A linkages. Under optimum conditions, the mechanical stress detaches the silica, and the hydrated silica is stabilized in solution. Surface dissolution described above occurs between abrasive particle impacts. Brown and Cook suggested that abrasives such as ceria and zirconia possess a “chemical tooth” that accelerates the bond formation at the surface and removal of silica from the surface faster than the re-deposition. In the case of silicon nitride, hydrolysis of nitride precedes the formation of surface silanol groups. For TaN and TiN, surface hydrolysis results in the formation of the respective oxides.

As discussed in section 4.1, the tungsten surface in contact with strong oxidizers is covered with an oxide film in acidic solutions. In addition, no significant dissolution of tungsten is observed under static conditions in the same solutions slurries. Similarly, the presence of a strongly adsorbed film on the Cu surface (Cu oxides, BTA) has been found to be necessary for achieving good planarity in Cu CMP. Very little planarization occurs in the absence of such surface films. Modification of these films by anions and surfactants affects the polish rates, as mentioned in section 5.3. These observations lead to the conclusion that polishing of metals involves continuous cycles of passivation, removal of the surface films, and repassivation of the exposed metal.\textsuperscript{82}

The general mechanism of the CMP process can be described as a sequence of the following steps: surface modification and formation of a surface layer, removal of the surface layer by abrasive action, and dissolution of the
material followed by reformation of the surface layer. For oxide surfaces modification occurs due to the reversible depolymerization reaction in the presence of water and the formation of Si(OH)₄.¹⁶³,¹⁶⁴ For metals, the surface layer could be a film of metal oxide or a film containing adsorbed organic molecules or a complex of the metal ion. Repetitive cycles of surface film formation, film removal, and metal dissolution result in eventual planarization of the structure. The overall polish rate then depends on kinetics of the three rate processes and the way they are coupled. The mechanical component is incorporated into the model as a parameter that influences the film removal rate. The film formation, removal, and dissolution rates can be expressed as a function of concentrations of the slurry components, the applied pressure, and the rotation rates as appropriate. This approach is an elegant way to relate the overall polish rate to the various kinetic processes occurring at the surface modulated by the mechanical parameters.

Paul et al.¹⁶⁵ have proposed a steady-state surface kinetics model where mechanical removal is represented as step in the mechanism. As mentioned above, the removal rate is expressed as a function of the concentrations of the oxidizer, abrasive, inhibitors, and the mechanical variables, pressure and rotation rates. The surface film formation during CMP due to the reaction between the metal surface M and a chemical component C in the slurry can be written as

$$M + C \rightleftharpoons MC^*$$  \hspace{1cm} (18)

where MC* is a surface complex in the film. The film can be removed in one or more of several possible ways: it can be mechanically removed by abrasive action or simply dissolved in solution (etching). In general, the etching rate is slow compared with the removal by abrasion. The etching and mechanical steps can be represented as

$$M_{\text{bulk}} + MC^* \rightarrow MC_{(aq)} + M$$  \hspace{1cm} (19)

$$M_{\text{bulk}} + MC^* + AB \rightarrow MC - AB + M$$  \hspace{1cm} (20)

where MC(aq) is the dissolved species MC, AB is the abrasive, MC − AB is the material removed from the surface, and M is the freshly exposed metal surface. The corresponding rates can be written as

$$R_{fR} = k_{fr}N_M[C] \quad \text{and} \quad R_{fr} = k_{fr}N_{MC^*}$$  \hspace{1cm} (21)

$$R_D = k_{D}N_{MC^*}$$  \hspace{1cm} (22)

$$R_M = k_{M}N_{MC^*}(N_d/A)$$  \hspace{1cm} (23)

where $k_{fr}$, $k_{fr}$, $k_{D}$, and $k_{M}$ are rate constants for the reactions described above, $N_M$ and $N_{MC^*}$ are the number of M and MC* sites on the surface, $N_d$ is the number of effective abrasive particles per unit area ($A$) of the surface and $[C]$ is the concentration of the chemical species in the slurry. At steady state, the film formation and removal rates are balanced, and hence the rate of change of $N_{MC^*}$ becomes equal to zero:

$$\frac{dN_{MC^*}}{dt} = k_{fr}N_M[C] - k_{fr}N_{MC^*} - k_{D}N_{MC^*} - k_{M}N_{MC^*}(N_d/A) = 0$$  \hspace{1cm} (24)

The total number of surface sites is given by $A/d_{\text{s}}^2 = N_M + N_{MC^*}$, and let $\theta = N_d/N_{MC^*}$. Solving the equations for $N_{MC^*}$ and using the definition for $\theta$, gives the expression for $\theta$ in terms of the various rate constants and the concentration of the chemical species and the abrasive.

The mechanical abrasion process is modeled in a similar fashion, and the mechanical removal rate, $R_M$, is expressed in terms of the abrasive surface density, $c_P$, and the fraction of the pad covered with the abrasive particles, $\theta_A$.

$$R_M = k_MN_{MC^*}(N_d/A) = k_M(c_P\theta_A)\theta(A/d_{\text{s}}^2)$$  \hspace{1cm} (25)

The overall polish rate is given by the expression

$$R = (R_D + R_M)/A = (k_D + k_Mc_P\theta_A)\theta(A/d_{\text{s}}^2)$$  \hspace{1cm} (26)

where $\theta$ contains the oxidizer concentration term and $\theta_A$ contains the abrasive particle concentration terms. Combining the various rate constants and concentrations into constants $\beta_i$, one can write the overall polish rate explicitly in terms of the concentrations of the abrasive or the oxidizer.

$$R = \beta_1[C]/([C] + \beta_2)$$  \hspace{1cm} (27)

$$R = \beta_3 + \beta_4[a]/([a] + \beta_5)$$  \hspace{1cm} (28)

The physical significance of the various $\beta$ terms as follows:¹⁶⁵ $\beta_1$ is the maximum polish rate attainable for the given set of experimental conditions and combines the dissolution and abrasion rates per MC* surface site area, $\beta_2$ is the ratio of the rate of film destruction (decomposition, dissolution, removal) to the rate of film formation, $\beta_3$ is the rate of metal dissolution and is proportional to $k_D$, and $\theta_A$ is the maximum rate for mechanical abrasion with a correction factor for material removal by dissolution and depends on $[C]$ through $\theta$ to account for the variation of mechanical removal rate at different concentrations of the oxidizer, and $\beta_4$ is the ratio of chemical and diffusion terms to chemical, diffusion, and mechanical terms.

The model provides a two-parameter data fit for rate vs $[C]$ plots and a three-parameter fit for rate vs $[a]$ plots. The model successfully fits the polish rate experimental data for tungsten for different oxidizers, abrasives, and polishing conditions.¹⁶⁶ For tungsten CMP data, $\beta$ values vary from 0.0056 ($\beta_2$) to 51 454 ($\beta_4$) for the various oxidizers and range of mechanical parameters investigated.¹⁶⁶ Figure 25 shows the model fit for the polish rate as a function of KIO₃ and alumina concentrations. At low oxidizer concentrations, the rate increases linearly with the concentration at constant film removal rate. At higher oxidizer concentrations, the rate reaches a limiting value set by the film removal rate. This value increases at higher pressures and rotation rates indicating the mechanical control of the limiting rates through film removal rate. Paul et al. have successfully extended their model to systems with inhibitors¹⁶⁸ and copper CMP¹⁶⁹ process demonstrating good data fit.

### 8. Alternative CMP Processes

Many of the problems associated with the CMP processes have been attributed to the mechanical stress and aggressive chemical action during polishing. The polishing pressure and the elastic deformation of the pad are viewed as the source of excessive dishing and erosion. Scratching and other mechanical damages have been blamed on the extremely high
localized mechanical stress created by the abrasives under high down force. The oxidizer and chelating agents required for high removal rates have been identified as the root causes of corrosion defects. In addition to the above, the fragile nature of low-k interconnects has prompted the development of alternative CMP processes such as abrasive-free polish (AFP). Early formulations were called “abrasive-less” simply because they contained low abrasive concentrations (less abrasive). Later formulations appear to be truly free of abrasives and contain only oxidizer, chelating agent, and inhibitor. Low down force (fraction of 1 psi) and low platen rotation rates (30 rpm) are used, and accordingly low polish rates (1000–1500 Å/min) are observed. Dishing and erosion are considerably better than that for conventional CMP processes; however the polish times are likely to be much longer. Matsuda et al. have investigated abrasive-free formulations containing heteropolyacids, surfactants such as poly(ethylene glycol), water-soluble polymers, and other additives. These are called micelle slurries, and polish rates up to 4000 Å have been achieved at down force as low as ~1 psi. It appears that abrasive-free CMP processes may have potential benefits in terms of defectivity; however significant process development efforts may be necessary before these new processes become viable options for a manufacturable planarization process for future copper interconnects.

9. Concluding Remarks

Despite early skepticism, the CMP process has succeeded far beyond the expectations and has become a critical part of semiconductor process technology. It is also considered as an enabling process in the fabrication of many other devices including thin film magnetic heads, MRAM, MEMS, and phase change memories (PCM). Many of the challenges are similar to those faced by the semiconductor technology: new materials, complex integration schemes, extremely thin layers, fragile structures, and increasingly stringent defectivity and planarity requirements. One of the more difficult challenges that researchers face is that many of the materials used in these devices are highly sensitive to the chemical environment and mechanical stress associated with CMP. The critical properties are significantly altered after CMP, and these materials need to be protected during post-CMP cleaning as well. The pattern dimensions and thickness are fast approaching molecular and atomic scales, and soon it will no longer be possible to know what is happening with today’s metrology techniques. We will have to planarize the layers we cannot “see” and eliminate defects and remove particles that we will not be able to “find”.

These challenges are providing unique opportunities for slurry research and planarization process development. It is necessary to think in a fundamentally different way, develop new strategies and novel concepts, invent new chemistries to enable molecular manufacturing, and build large devices with atomic precision, quickly and cheaply with virtually no defects.

10. Acknowledgments

The authors acknowledge S. K. Rangarajan, B. S. Haran, D. F. Canaperi, S. Papa Rao, S. M. Gates, J. M. Cotte, M. F. Lofaro, L. Hupka, J. Cummings, J. Choi, V. Anandan, and D. Penigalapati for their technical contributions, many useful discussions, critical comments, and valuable help during the preparation of the manuscript. The support and encouragement from IBM Research Silicon Science and Technology management is also gratefully acknowledged. The authors wish to thank the reviewers for their critical comments and suggestions.

11. References

Symposium F: Chemical Mechanical Planarization, San Francisco, CA, April, 2003, F7.5.


CR900170Z.