

Silicon Processing

for the VLSI Era
Volume 1 - Process Technology
Second Edition

S. Wolf and R.N. Tauber

**SILICON PROCESSING
FOR
THE VLSI ERA**

**VOLUME 1:
PROCESS TECHNOLOGY
Second Edition**

**SILICON PROCESSING
FOR
THE VLSI ERA**

**VOLUME 1:
PROCESS TECHNOLOGY
Second Edition**

**STANLEY WOLF Ph.D.
RICHARD N. TAUBER Ph.D.**

**LATTICE PRESS
Sunset Beach, California**

DISCLAIMER

This publication is based on sources and information believed to be reliable, but the authors and Lattice Press disclaim any warranty or liability based on or relating to the contents of this publication.

Published by:

LATTICE PRESS

Post Office Box 340

Sunset Beach, California 90742, U.S.A.

Cover design by Roy Montibon, New Archetype Publishing, Los Angeles, CA.

Copyright © 2000 by Lattice Press.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system without written permission from the publisher, except for the inclusion of brief quotations in a review.

Library of Congress Cataloging in Publication Data

Wolf, Stanley and Tauber, Richard N.

Silicon Processing for the VLSI Era

Volume 1: Process Technology

Includes Index

1. Integrated circuits-Very large scale integration. 2. Silicon. I. Title

ISBN 0-9616721-6-1

9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

Chapter 15

MULTILEVEL INTERCONNECTS

For ULSI

Throughout the evolution of integrated circuits, the aim of device scaling has been twofold: (1) to increase circuit performance (mainly by increasing circuit speed), and (2) to increase the functional complexity of the circuits. At the outset, scaling down of active device sizes was a very effective means of achieving these goals. Eventually, the scaling of active devices became less profitable. The limitations of the circuit speed and maximum functional density came to depend more on the characteristics of the interconnects than on the scaled devices. In addition, the aspects of silicon utilization, chip cost, and ease and flexibility of IC design were also adversely affected by the interconnect-technology restrictions. The approaches to lifting these limitations have mainly involved the implementation of multilevel-interconnect schemes. In this chapter multilevel interconnect technology is covered.

15.1 THE NEED FOR MULTILEVEL-INTERCONNECT TECHNOLOGY

In this section how the speed and functional density of ICs can be limited by the interconnect technology is explained. Then why multilevel interconnects are effective in counteracting this problem is shown. Notation used throughout the chapter in descriptions of various features of a multilevel-interconnect structure will be introduced.

15.1.1 Interconnect Limitations of ULSI

15.1.1.1 Functional Density: In the course of integrated-circuit evolution, the maximum number of devices per chip has steadily increased, mainly as a result of the increase in functional density (although the growth in chip area has also played a role). *Functional density* is defined as the number of *interconnected* devices per chip area. While the *number* of devices per chip area is referred to as the *active-device density*. As the minimum feature size on an IC decreases, the active device density increases. The functional density, however, depends on how effectively these devices can be interconnected. Unless a larger number of the devices can be interconnected as the active-device density is increased, there will be no gain in functional density.

When integrated circuits contained only a relatively small number of devices per chip, the devices could be more easily interconnected. This situation persisted for some time, even as the active-device density increased. Nevertheless, the area occupied by the interconnection lines on the chip surface grew more quickly than the area needed to accommodate the active devices. Eventually, the condition was reached in which the minimum chip area became interconnect-limited. That is, the area needed to route the interconnect lines between the active devices exceeded the area occupied by these active devices. At this point, continued shrinking of active devices produced less circuit-performance benefits.

One way to overcome such a limitation is to implement a multilevel-interconnection system in which the area needed by the interconnect lines is shared among two or more levels. This allows the fractional area of the chip occupied by active devices to be increased, leading to an increase in the functional density.

15.1.1.2 Propagation Delay: As the minimum line width on an IC shrinks, the active-device density increases. For example, in 1-Mbit DRAMs 1.2- μm lines are used, while in 4-Mbit, 16-Mbit, and 64-Mbit DRAMs, 0.8- μm , 0.5- μm , and 0.35- μm lines are used, respectively. At these dimensions, the MOS transistor switching speed itself no longer limits the logic delay or access time of the IC. Instead, the time required for the transistor to charge capacitive loads becomes the limit to the performance of the IC. As devices shrink, the device contribution to the propagation delay of a digital signal also decreases. On the other hand, scaling of the interconnect line widths does not bring about a corresponding decrease in the propagation delay time through the interconnect lines. As the chip sizes increase, the interconnect-path lengths also increase. In fact most large ULSI circuits have become interconnect propagation-delay-time limited.

For example, as pointed out by Sah,¹ the intrinsic gate delay caused by electrons drifting through a 2- μm channel at their phonon-scattering-limited velocity of 10^7 cm/sec is 20 ps,² while the delay of a 256-kbit DRAM is ~ 100 ns. This is 5,000 times longer than the intrinsic gate delay. Thus, a chip with devices whose gate lengths are 2 μm or less cannot be limited by the intrinsic gate delay unless 5,000 devices are connected in series. This is certainly not the case in a memory chip, and it is highly unlikely that it would occur even in complex logic circuits.

The propagation delay of ICs becomes limited by the large RC delay of the interconnection lines when the minimum feature size decreases below about 1 μm (Fig. 15-1).³ The delay resulting from various interconnect materials per millimeter of length is given in Table 15-1, and is graphed for some of these materials (assuming a 1-cm-long interconnect line) in Fig. 15-1.* For example, if a polysilicon line having a 20 Ω/sq sheet resistance is used to connect one corner of a 0.7 x 0.7 cm chip to the diagonal corner, the data from Table 15-1 predict that the propagation delay will be 1.4 μs . This data implies that even if Al is used as the interconnect layer, the delay caused by 1-cm-long interconnection lines will limit the circuit speeds once device dimensions fall below $\sim 2 \mu\text{m}$.

The above indicates that for circuit performance to be increased as device dimensions shrink, two goals must be met. First, the materials used for transmitting signals over long distances on a chip must have the lowest possible resistance values. This has been the key motivation in the drive to replace polysilicon interconnects (20 Ω/sq) with polycide interconnects (1.7 Ω/sq for TiSi_2). It also points out one of the reasons why Cu is replacing Al as an interconnect material as devices shrink further.

Second, the length of interconnect lines on a chip must be made as short as possible. The RC delay is shown to be proportional to the square of the length of the interconnect line. That is:

$$R = (\rho l)/(w t_m) \quad (15.1)$$

* Note that in Fig. 15-1 a somewhat thicker field oxide is used, yielding a smaller value of C.

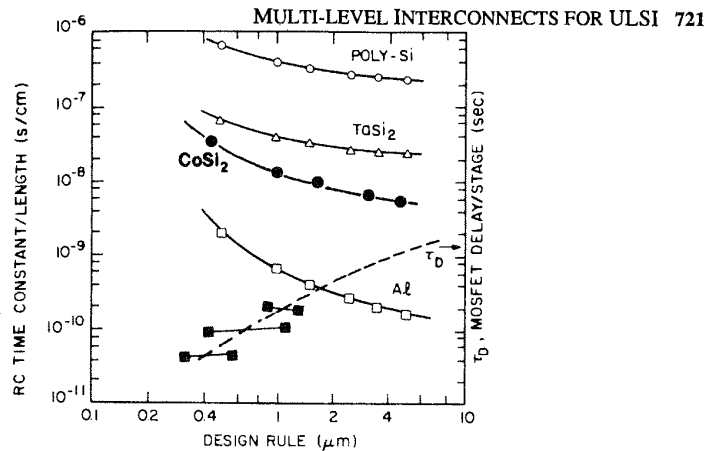


Fig. 15-1 RC time constant per unit length for several conductive materials as a function of feature size. Also shown is delay per stage of NMOS ring oscillators as a function of feature size. The RC time constant is calculated assuming a field oxide thickness of 1 μm .³ Copyright 1981 AIOP. Reprinted with permission.

and:

$$C = (\epsilon w l) / t_{\text{ox}}, \quad (15.2)$$

and therefore:

$$RC = \frac{\rho \epsilon l^2}{t_m t_{\text{ox}}} \quad (15.3)$$

where ρ is the resistivity, l is the interconnect line length, w is the line width, ϵ is the permittivity, t_m is the thickness of the metal, and t_{ox} is the thickness of the oxide. A multi level-interconnect structure is an effective way to allow the longest lines to be reduced in length.

There are several other propagation-delay-related considerations that make multilevel interconnects attractive. First, they make it more feasible to lay out all interconnect lines with as close to the average length as possible. This is important because if some lines are much longer than the rest, the overall delay time of the circuit will be increased. The length variation also

Table 15.1 Interconnection Delay in Silicon VLSI Chips⁴

Conductor Materials	Resistivity ($\mu\Omega\text{-cm}$)	Thickness (nm)	Sheet Resistance (Ω/square)	Delay (psec/mm ²)
Polysilicon	1,000	500	20	7,000
TaSi ₂	46	100	4.6	1,587
Ti	42.7	100	4.3	1,484
TiSi ₂	17	100	1.7	586
WSi ₂	12.5	100	1.3	448
W	5.3	100	0.53	183
Al	2.6	100	0.26	90

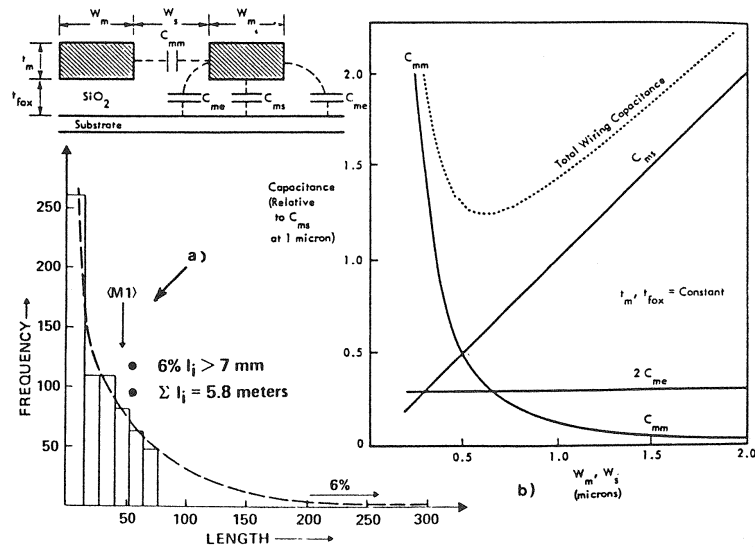


Fig. 15-2 (a) Example of lead length distribution of metal interconnects on an IC having a single level of metal.⁵ Reprinted with permission of Electronics. (b) Variation of components of interconnect capacitance with design rules. The various components are defined in the adjacent figure.⁶ (© 1982 IEEE).

introduces nonuniform on-chip RC response times, which can disturb the switch synchronization of the circuit.

Figure 15-2a shows a histogram of line lengths of a typical single-level-metal IC.⁵ Whereas most lines are short, a few are much longer (e.g., in this example ~6% of them are >7 mm long). Multilevel interconnects can help eliminate the incidence of the few long lines.

Second, as shown by McGreivy in Fig. 15-2b,⁶ if the line spaces become smaller than ~0.6 μm , the total parasitic capacitance of the interconnect lines will increase because the side-by-side capacitance increases with decreasing line pitch. The end result is an overall increase in the system cycle time. Again, the use of a multilevel-metal interconnect system, in which the wires run orthogonally in adjacent layers can reduce this problem.

Third, the layout should minimize cross-talk to reduce the level of inductance noise coming from the random switching of individual circuits. This goal can be accomplished through the use of a multilevel-interconnect structure in which all the wire tracks on one level run in one direction, while those on the adjacent level run in the perpendicular direction.

Finally, because the functional density of the chips can be increased, less chip-to-chip signal transmission is required. Since such off-chip driving is much slower than on-chip signal transmission, the overall system speed will be enhanced.

The above discussion implies that the lower levels of metal or polysilicon (which generally

have a finer pitch, and therefore higher resistance and capacitance) should be reserved for interconnecting neighboring devices with short interconnection lines. The upper interconnect levels, which are usually designed with a larger pitch and lower resistance, should be used to transmit signals across the entire chip.

15.1.1.3 Cost: If multilevel-interconnect processes are used to fabricate integrated circuits, the die size should decrease. Thus, more die per wafer can be manufactured. If the manufacturing cost per wafer remains the same and the yield is not impacted by the implementation of a multilevel-interconnect process, the cost per chip will decrease. In fact, smaller die sizes should imply higher yields, and enhancing the benefit of chip-size reduction. In addition, improved device performance may allow the circuit to command a higher market price.

However, the implementation of a multilevel-interconnect system requires that at least two additional masking steps be used for each additional level of interconnect. The extra process steps add to the manufacturing cost of each wafer. The number of defects/cm² is also generally proportional to the number of masking steps. In addition, the manufacturing yield and long-term reliability for a multilevel metal process are typically lower, since the process becomes more technically demanding. As a result, it must be determined whether the chip-size reduction and enhanced chip value will produce a margin of profit that is greater than the amount lost due to additional incurred process costs and yield and reliability loss.

15.1.2 Problems Associated with Multilevel-Interconnect Processes

As alluded to previously, adding a multilevel interconnect process to a fabrication sequence introduces a new set of difficulties. In addition to added process complexity and loss of topological planarity, there are several other concerns.

First, new materials must be used, which necessarily involves an extensive characterization of their properties to ensure they are compatible with all other aspects of the process technology. Second, new process-related manufacturing difficulties may be encountered that can adversely impact manufacturing yield (e.g., interlevel shorts due to pinholes; stringers due to incomplete etching over severe steps; failure to open vias due to difficulty in implementing reliable endpoint-detection techniques in the dry-etch process; film delamination due to poor adhesion or high stress; and difficulty in bonding to some metal alloys). Third, new failure modes may be encountered. For example, electromigration, corrosion, and hillock formation, which must also be characterized to determine whether they will significantly compromise circuit reliability.

The problems related to multilevel interconnects are listed at this point in order to show that benefits can be gained only by successfully pursuing a considerable technical-development effort. More specific details on these problems (and how they can be overcome) will be provided throughout the chapter.

15.1.3 Terminology of Multilevel-Interconnect Structures

Figure 15-3 shows the terminology associated with a double-level-metal structure for MOS technologies. The MOS structure has a dielectric layer between the polysilicon gate/interconnect level and Metal 1, which is referred to as the *pre-metal dielectric* (or *PMD*). The dielectric

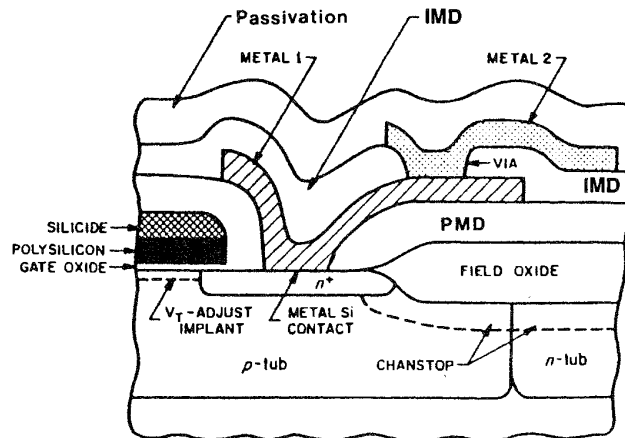


Fig. 15-3 Terminology of double-level metal interconnects.

layers between metal levels are called *intermetal dielectrics (IMD)*. The openings in PMD are referred to as *contact holes* (or *contacts*). They allow electrical connections to be established between Metal 1 and polysilicon, as well as between Metal 1 and the Si substrate. Openings in the intermetal dielectric layers are known as *vias*: these allow contact to be made between Metals 1 and 2, Metals 2 and 3, etc.

In bipolar technology, the dielectric layer between Metal 1 and the substrate is still referred to as PMD, despite the fact that it may not isolate Metal 1 from polysilicon. The openings in PMD are again called contact holes, although they are only used to allow contact to be established between Metal 1 and the substrate (i.e., not poly). The notation for the other metal and dielectric layers is otherwise identical to that used in MOS technologies.

A distinction exists between the use of the terms multi-level and multilayer. A multilayer-interconnect structure is a thin film consisting of more than one layer of material, but existing at just one level of the interconnect system. Hence, a multilayer film can serve as the conductor (or dielectric) at each level of a multilevel-interconnect system.

15.2 MATERIALS FOR MULTILEVEL INTERCONNECT TECHNOLOGIES

The two groups of materials employed in multilevel-interconnect technologies are thin-film conductors and thin-film insulators. In this section the properties of materials that have been adopted for use in ULSI applications will be described.

15.2.1 Conductor Materials for Multilevel Interconnects

There are a variety of conductor materials used in the interconnect structure of ICs. Aluminum and aluminum alloys (e.g., Al-Si, Al-Cu, and Al-Si-Cu) have been the most widely used metals. Their properties and deposition method (sputtering) are described in Chap. 11. Heavily-doped polysilicon has also been employed, primarily as a local interconnect. This material is discussed in Chap. 6. Polysilicon with an overlaid metal silicide layer (to reduce the resistance) is also used (with WSi_2 , $TiSi_2$, and $CoSi_2$ being the materials utilized in such *polycide* structures).

WSi_2 is described in Chap. 6, and TiSi_2 and CoSi_2 in Chap. 10. TiN is a conductor material that finds use as a diffusion barrier and adhesion layer. It is considered both in Chaps. 6 and 10. Tungsten (W) finds application primarily as a plug material. The properties of W and the deposition process for forming W-plugs is presented in Chap. 6. The etching of W is covered in Chap. 14. The CMP removal of W layers to form plugs is considered in this chapter. Copper (Cu) has recently been introduced into IC fabrication as an interconnect material, and technologies for its deposition and patterning by CMP are discussed in this chapter.

15.2.1.1 Requirements of VLSI Conductor Materials: The most important requirements of conductor materials for ULSI are listed in Table 15-2. The list is long and many requirements are quite stringent. Nevertheless, unless a conductor structure can satisfy virtually all, it is unlikely to find use in ULSI applications. Thus, the number of materials found suitable for ULSI interconnects is rather small (but growing). Table 15-3 summarizes the important properties of this group. The resistivities given in the table are typical for polycrystalline films with thicknesses of 10–1000 nm. The lower-resistivity values are exhibited by purer and thicker films, and with large-grains. Deviations from stoichiometry for alloy conductors usually lead to increased resistivity. As a result, the resistivity of a given type of film may vary from one deposition run to another, unless all deposition parameters are controlled to yield identical films.

Table 15.2 DESIRED PROPERTIES OF CONDUCTOR STRUCTURES FOR ULSI

- | | |
|--|---|
| <ul style="list-style-type: none"> a. Low resistivity (has to be $< 4 \mu\Omega\text{-cm}$) b. Surface smoothness (and hillock resistant) c. Resistance to electromigration failure effects d. Bondable (final level of a multilevel system). e. Adhesion to underlying substrate materials must be excellent f. Stable (mechanical and electrical properties remain main constant with time under the following conditions <ul style="list-style-type: none"> 1. throughout processing, including sintering, interlevel, and passivation dielectric deposition; 2. in an oxidizing ambient; 3. under normal operating conditions; 4. under ordinary storage (stress voids) g. Corrosion resistant h. Should not contaminate devices, wafers, or processing equipment i. Depositable to controlled thickness and uniformity, and free of deposited particulates j. Anisotropically etchable with high selectivity with respect to substrate and mask material k. Depositable over vertical walls with conformal coverage of steps ("good step coverage") l. Film reflectance should be controllable to allow effective photolithographic processing | <ul style="list-style-type: none"> m. If necessary, the metallization film should be depositable as a combination metal film (multilayer) n. Each required layer should be depositable in alloy form, with alloy composition tightly controllable o. Depositable pure - (i.e., without reaction with the gases present in the deposition chamber, and without incorporation of residual gases into the deposited films) p. Low film stresses q. Deposition and patterning process-should be economically viable, with <ul style="list-style-type: none"> 1. high throughput; 2. reasonable processing-equip-purchase, maintenance, and operating costs; 3. high reliability and up time of processing equipment; 4. low enough in complexity that highly skilled operators are not needed |
|--|---|

Table 15.3 PROPERTIES OF VARIOUS CONDUCTORS USED IN ULSI INTERCONNECTS

Metal or Alloy	ρ ($\mu\Omega\text{-cm}$)	Melting Point ($^{\circ}\text{C}$)	Reaction with Si ($^{\circ}\text{C}$)	Stable on Si up to ($^{\circ}\text{C}$)
Al	2.7–3.3	660	~250	~250
Mo	6–15	2620	400–700	~400
W	6–15	3410	600–700	~600
TiSi ₂	13–16	1540	–	≤ 900
WSi ₂	30–70	2165	–	≤ 1000
CoSi ₂	10–18	1326	–	≤ 950
Ti:W (10wt% Ti)	75–200		600–700	≤ 700
TiN	25–200	~2880		
Cu	1.7	1083		
Poly-Si (heavily doped)	1000	1412	–	

Another key property of ULSI thin-film conductors is their ability to adhere to Si and to SiO₂. In this respect, Al, Ti, and TiN films are the best. Al and Ti reduce SiO₂ to form interfacial metal-oxide bonds that promote adhesion and stability. WSi₂ and TiSi₂ also adhere well to SiO₂, but not as well as Al and Ti. On the other hand, W and Cu do not reduce SiO₂, and therefore exhibit poor adhesion to SiO₂ surfaces. As a result, W and Cu are not used as stand-alone conductor materials. Instead, a layer such as Ti, TiN or TaN is required under W or Cu films to promote adequate adhesion.

15.2.2 Dielectric Materials for Multilevel Interconnects

A variety of deposited dielectric films are used in IC interconnect applications. These include the following: a) atmospheric CVD SiO₂ from silane chemistries; b) low-pressure, plasma-enhanced, and ozone SiO₂ from TEOS chemistries; silicon nitride; c) spin-on-glass [SOG]; d) SiO₂ dielectrics deposited using high-density-plasma CVD [HDP-CVD SiO₂]; and e) new low-*k* dielectrics. All of these were discussed in Chap. 6, except SOG, HDP-CVD SiO₂, and the low-*k* dielectrics. These three are considered in this chapter.

15.2.2.1 Requirements of Dielectric Layers in Multilevel Interconnects: Dielectric layers must be used to electrically isolate one level of conductor from another in multilevel-interconnect systems. The list of properties that must be possessed by such dielectric layers is given in Table 15-4. This list of requirements is also long and stringent. When materials that have been developed for this role are described, Table 15-4 will serve as the point of reference.

It should also be mentioned that there can be a significant difference between the dielectric film referred to in Fig. 15-3 as PMD (used between polysilicon, or other local-interconnect level material, and Metal 1) and the dielectric films that are employed between the metal layers (intermetal dielectrics - e.g., IMD in Fig. 15-3). PMD films can be deposited (and densified if necessary) at a higher temperature than is possible for the intermetal dielectric layers.

Table 15.4 DESIRED PROPERTIES OF INTERLEVEL DIELECTRICS FOR VLSI ⁷

1. Low dielectric constant for frequencies up to ~20 MHz, in order to keep capacitance between metal lines low;
2. High breakdown field strength (> 5 MV/cm);
3. Low leakage, even under electric fields close to the breakdown field strength. Bulk resistivity should exceed 10^{15} Ω -cm;
4. Low surface conductance, surface resistivity should be $>10^{15}$ Ω -cm;
5. No moisture absorption or permeability to moisture should occur;
6. Films should exhibit low stress, and the preferred stress is compressive ($\sim 5 \times 10^8$ dynes/cm²), since dielectric films under tensile stress exhibit more of a tendency to crack;
7. Good adhesion to aluminum, and of aluminum to the dielectric. (Good adhesion is also needed to the other conductors used in ULSI, such as doped polysilicon and silicides.) In cases of poor adhesion (such as with Cu or W), a glue layer (such as Ti or TiN) may need to be applied between the conductor and the dielectric;
8. Good adhesion to dielectric layers above or below. Such dielectric layers could be thermal oxides, doped-CVD oxides, nitrides, oxynitrides, polyimides, or spin-on glasses;
9. Stable up to temperatures of 500°C;
10. Easily etched (by wet or dry processing).
11. Permeable to hydrogen. This is important for IC processing, in which an anneal in a hydrogen containing ambient must be used to reduce the concentration of interface states between Si and the gate oxides of MOS devices (see Chap. 8);
12. No incorporated electrical charge or dipoles, some polyimides in particular contain polar molecules that can orient themselves in an electric field and give rise to an electric field even when the externally applied field is removed;
13. Contains no metallic impurities;
14. Step coverage that does not produce reentrant angles;
15. Good thickness uniformity across the wafer, and from wafer to wafer.
16. In the case of doped oxides, good dopant uniformity across the wafer, and from wafer to wafer;
17. Low defect density (pinholes and particles);
18. Contains no residual constituents that outgas during later processing to the degree that they degrade the properties of other layers of the interconnect system (e.g., outgassing from some polyimide films, SOG films, or low-temperature TEOS films)

Pre-metal dielectric (PMD) films can be flowed and reflowed at temperatures in excess of 800°C. However, when Al is present on the wafer surface, the maximum temperature of the intermetal dielectric layers is limited to ~450°C.

15.3 PLANARIZATION OF INTERLEVEL DIELECTRIC LAYERS

As pointed out in the chapter introduction, the planarization of interlevel dielectrics is one of three critical issues that must be addressed when implementing a multilevel interconnect system for VLSI applications. This assertion is now justified and the various approaches developed for dealing with this issue are described.

15.3.1 Terminology of Planarization in Multilevel Interconnects

As additional levels are added to multilevel-interconnection schemes and circuit features are scaled to submicron dimensions, the required degree of planarization is increased. Such planarization can be implemented in either the conductor or the dielectric layers. In this section processes developed to planarize dielectric layers are described. In later sections, techniques for planarizing conductor layers and vias will be considered.

The term planarization is employed quite frequently (and loosely) both here and in other technical literature. Therefore it is useful to define this term more completely, especially as it applies to planarization of dielectric layers in multilevel-interconnect technology. The example case used for this discussion is that of a dielectric layer (IMD) deposited after Metal 1 is patterned. In the case where no planarization exists (Fig. 15-4a), the step heights on the DM1 surface closely approximate the step heights of the Metal 1 layer and the underlying topography. Furthermore, in this case the steps on the IMD surface also have steep slopes (i.e., vertical, or even reentrant).

15.3.1.1 Degree of Planarization: The steps on the surface of DM1 can be made less severe through various planarization processes. The degree to which this can be successfully accomplished differs according to the technique used. The degree of planarization is classified according to the following qualitative planarization criteria:

1. The first degree of planarization (smoothing) involves a lessening of the step slopes at the IMD surface (Fig. 15-4b). The step heights in this case, however, are not significantly reduced in magnitude.
2. In the second degree of planarization (partial or semi-planarization) the step heights are reduced (but not eliminated), and the slopes of the steps are also smoothed (Fig. 15-4c).
3. In the third degree of planarization (complete local planarization), the steps at the surface of IMD are completely eliminated wherever the spaces in the underlying topography are relatively close together (e.g., $<10 \mu\text{m}$ apart), but the steps at isolated, wide features still exhibit a step (Fig. 15-4d).
4. In the fourth degree of planarization (complete global planarization), the surface of the IMD is completely planarized over arbitrary topography (Fig. 15-4e).

A quantitative measure of the step-height reduction, referred to as the planarization factor, β , is given by⁸

$$\beta = 1 - (t_{\text{step}}^f / t_{\text{step}}^i) \quad (15.4)$$

where t_{step}^i and t_{step}^f are the initial and final step heights, respectively. In complete-planarization cases, $\beta = 1$. If no planarization (or only smoothing) exists, then $\beta = 0$.

15.3.1.2 The Need for Dielectric Planarization: As the number of levels in an interconnect technology is increased, the stacking of additional layers on top of one another produces a more and more rugged topography. Consider, for example, a two-level metal, single-poly CMOS process. Assume the step height of the semi-recessed field oxide is $0.3 \mu\text{m}$, and the thicknesses of the poly and the first and second metals are 0.4 , 0.5 and $1.0 \mu\text{m}$, respectively. The maximum height of the steps on the wafer surface after each of these processes will be 0.3 , 0.7 , 1.2 , and

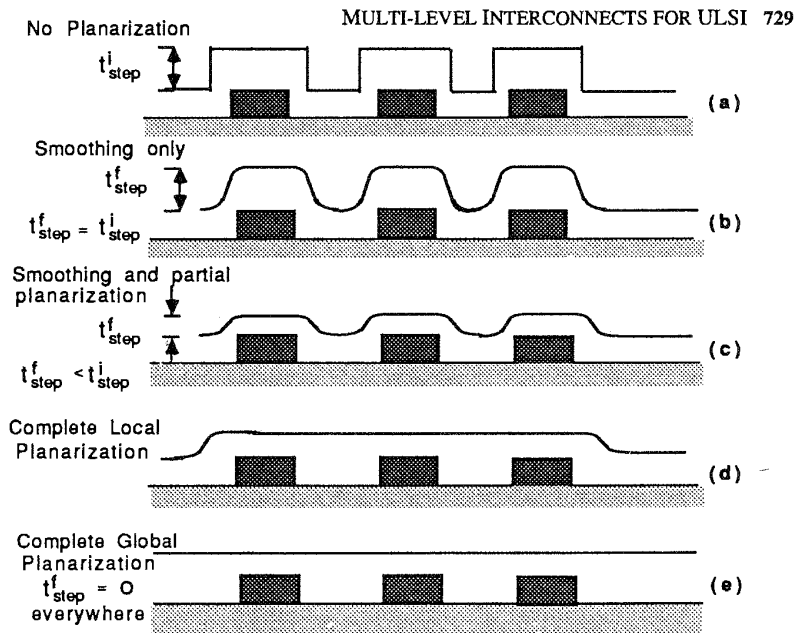


Fig. 15-4 Degrees of dielectric planarization.

2.2 μm , respectively (Fig. 15-5). It is apparent that the surface of the wafer must be planarized in some fashion to prevent the topography roughness from growing with each level. Without such planarization, the microscopic canyons that result on the wafer surface from stacking of device features can lead to topography conditions that would eventually reduce the yield of circuits to unacceptably small values.

A nonplanar wafer topography results in three conditions that prevent the fabrication of reliable electrical connections in multilevel interconnect systems. These conditions are as follows:

1. Poor step-coverage of the metal lines as they cross over the high and steep steps. A measure of how well a film maintains its nominal thickness is expressed by the ratio of minimum thickness of a film as it crosses a step, t_s , to the nominal thickness of the film on horizontal regions, t_n (Fig. 15-6a). This property is referred to as the step coverage of the film, and is expressed as the percentage of the nominal thickness that occurs at the step:

$$\text{Step coverage (\%)} = (t_s/t_n) \times 100\% \quad (15.5)$$

Step coverage of 100% is ideal for a conductor, but in general the height of the step and the aspect ratio of the features being covered impact the expected step coverage. (The aspect ratio is defined as the height-to-spacing ratio of two adjacent steps.) The greater the step height or the aspect ratio, the more difficult it is to attain coverage of the step without a corresponding thinning of the film that overlies it. Figure 15-6b shows an

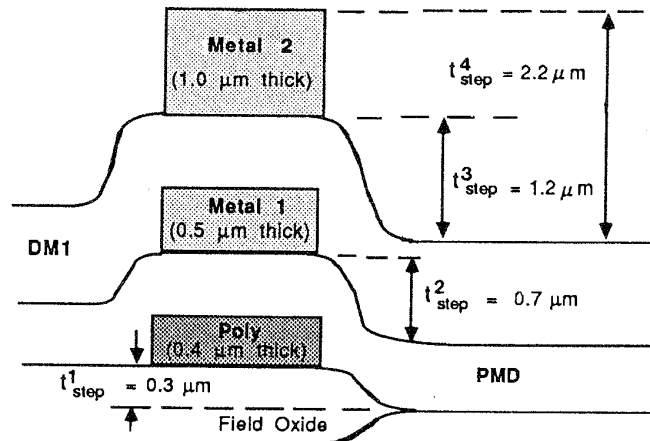


Fig. 15-5 As the number of levels in an interconnect technology is increased, the stacking of additional layers on top of one another produces steps of progressively greater heights.

example of poor step coverage. Hence, worse step coverage is expected under these conditions. In addition to the step height and aspect ratio, step coverage depends on two other topological factors: the contour and the slope of the step. In general, the smoother the steps contour and the less vertical its slope, the better will be the step coverage of any overlying films.

2. Metal stringers remain behind at the foot or sides of steep steps when anisotropic etching is used. This condition is shown in Fig. 15-6c. In addition, the resist is typically eroded by the same gases that etch the metal. The resist must therefore be thick enough in all locations that it will not be completely removed before the etch process is completed. If the dielectric surface has steps, in certain locations (i.e., over the upper corners of the steps in the dielectric, see Fig. 15-6d), there may not be sufficient resist thickness to survive during the time of overetch that must be used for removal of the stringers. The resist-erosion rate during the overetch time is also generally enhanced. Because most of the exposed metal has been etched away, the reactant gases that were being consumed during etching of the metal are now present in greater concentrations. In cases where the resist is completely eroded at thin spots, the metal beneath the resist will be exposed to the reactive species and may be etched away to an unacceptable degree.

Even if the above two limitations could be overcome, the following obstacle would eventually force planarization to be implemented:

3. The depth-of-field limitations of submicron optical-lithography tools will require surfaces to be planar within $\pm 0.5 \mu\text{m}$. As a result, if steps larger than $0.5 \mu\text{m}$ exist on the surface of IMD (or any other layer on the wafer), it will not be possible to pattern features in Metal-1 to the maximum resolution of the stepper. Thus, planarization will be mandatory if optical lithography is to be usable for fabricating ICs with such submicron feature sizes. (This topic is discussed in more detail in Chap. 13.)

Planarization of the intermetal dielectric layers is one of the chief approaches that have been taken to alleviate the problem of rough surface topography.

15.3.1.3 The Price That Must Be Paid as the Degree of Dielectric Planarization Is Increased: In the sections that follow, a progression of planarization techniques will be described, beginning with the smoothing of dielectric films, and going through complete global planarization. At least two significant penalties are encountered as the degree of planarization is increased, since the maximum variation in the thickness of the dielectric layer at different locations on the circuit will also increase. This is illustrated in Figs. 15-7a and b, in which the variations in the thickness of the dielectric in a non-planarized topography are compared to the variations that occur in a fully planarized interconnect structure.

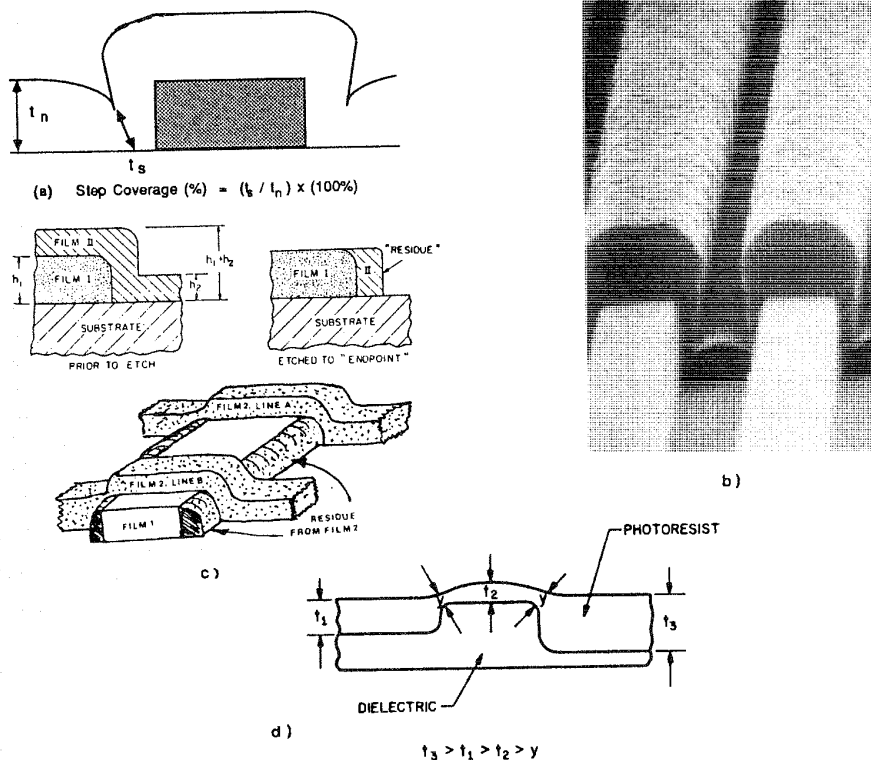


Fig. 15-6 (a) Definition of step coverage. (b) SEM micrograph showing poor step coverage. (c) Metal stringers remain behind at the foot or sides of a steep step when anisotropic etching is used. (d) Problem of resist thinning as film crosses underlying steps.

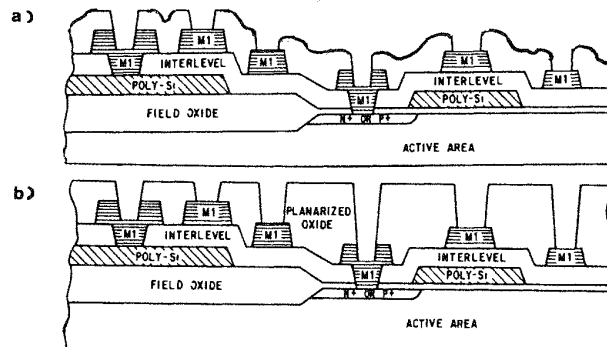


Fig. 15-7 (a) When no planarization is used, the via and contact holes between adjacent conductor levels all have approximately the same depth. (b) If planarization of the intermetal dielectric layer, IMD, is achieved, the via depths can vary very widely in depth.

The first penalty associated with this effect is the dielectric layer may end up being too thin over some of the underlying conductors after the dielectric has been planarized (Fig. 15-7b). If this occurs, it may be necessary to deposit an additional dielectric layer to increase the thickness of IMD before depositing Metal-2.

The second, and more serious, penalty involves the process of creating openings in the dielectric layers to allow selective contacting between conductors on different interconnect levels. As the degree of planarization is increased, such openings (referred to as *vias* when they are established in intermetal dielectric layers) will have different depths. For example, consider the two-level-metal structure shown in Fig. 15-9. In this structure, the LOCOS field-oxide step is 200 nm, the polysilicon thickness is 400 nm, and the PMD layer thickness is 400 nm. It can be seen that if a completely planarized IMD layer is used, via depths will vary from 0.3 to 1.3 μm . Such a large variation can lead to insurmountable via-filling problems.

If the technology for adequately covering vias with metal requires that the via sidewalls be sloped, an etch process for forming such sloped sidewalls will have to be used.* However, if the vias have significantly different depths, it may not be possible to implement a sloped-sidewall etch process. If a sloped-sidewall etch is used in the case shown in Fig. 15-8, the dimensions of the shallow vias will continue to increase during the time needed to completely etch the deep vias. If the lateral dimensions of the shallow vias exceed the width of the Metal-1 pattern below these vias, the etch process will begin to erode the PMD layer (Fig. 15-8b), even if a nested via is used (nested vias are discussed in Sect. 15.5).

In such cases, it may be necessary to avoid the use of a complete planarization process and to resort instead to a smoothing or semi-planarization procedure. With these methods it is possible to select a small enough maximum variation in via depths that a sloped-sidewall etch process can be successfully implemented.⁹ In general, however, the smoothing and semi-planarization techniques lose their effectiveness when three or more levels of metal are required and when

* Via sidewall sloping is usually needed when metal films are deposited by sputtering. In such cases, if the via sidewalls are too steep, metal step coverage becomes inadequate, and metal thinning (or even metal-film discontinuity) occurs. Therefore, it becomes necessary to slope the sidewalls as much as possible.

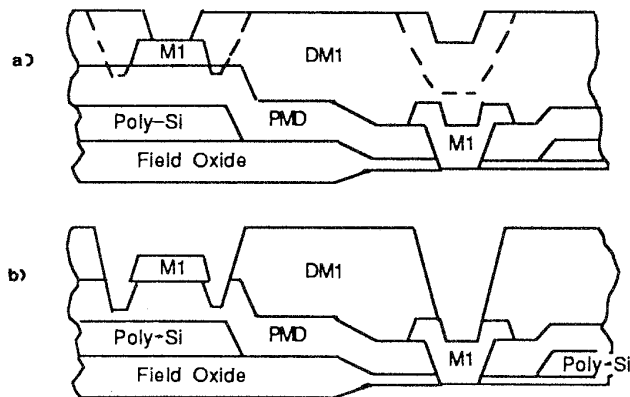


Fig. 15-8 (a) If a sloped-sidewall etch is used to etch vias, the dimensions of the shallow vias will continue to increase during the time needed to completely etch the deep vias. (b) If the lateral dimensions of the shallow vias exceed the width of the Metal-1 pattern below these vias, the etch process will begin to erode the PMD layer even if a nested via is used.

device dimensions are smaller than $1\ \mu\text{m}$. Here, the use of complete planarization processes becomes mandatory, and technologies must be employed that allow vertically sided contact holes and vias of varying depths to be completely filled. The techniques developed for such complete via filling will be discussed in subsequent sections.

15.3.1.4 Design Rules Related to Intermetal Dielectric-Formation and Planarization Processes:

The type of intermetal dielectric material, its thickness, and the method of its deposition and planarization can all have an impact on the design rules that must be applied when laying out an integrated circuit with a particular multilevel-metal system. Here a two-level-metal system with an underlying polysilicon level. Assume that the topography beneath the polysilicon is completely planar. In this system, design rules dealing with the following conditions must be specified: (1) minimum polysilicon-to-Metal-1 spacing, assuming no metal overlap of the poly, (2) minimum distance between coincident edges when Metal-1 overlaps polysilicon, and (3) minimum spacing of Metal-1 to Metal-1.

If polysilicon and Metal-1 come too close together, a gap will exist between them that will lead either to void formation in the CVD oxide PMD layer following deposition (Fig. 15-9a) or to the formation of a deep narrow crevice that cannot be covered by the next level of metal, even after some planarization procedures have been implemented (Fig. 15-9b).

In the second case, if polysilicon and metal edges are coincident, the total step height may be too high for adequate coverage by Metal-2 to be achieved (Fig. 15-9c). Figure 15-9d illustrates the case in which the problems of coincident edges and a narrow crevice are exacerbated by their simultaneous occurrence.

In the third case, the spacing between adjacent Metal-1 lines must be large enough to prevent void or crevice formation in the deposited dielectric film. The step that exists in the LOCOS field oxide must also be considered in real situations when such design rules are being

formulated. One possible way to prevent the formation of voids, crevices, and excessively high steps, is to allow the spacings between Metal-1 and underlying features to be sufficiently large that any possible combination of misalignment cannot lead to the problems listed. However, this alternative is impractical, because it increases chip size and causes excessive complexity in layout design.

The trend in advanced IC designs today is toward the elimination of all such restrictive dielectric-layer-related design rules. The development focus is on dielectric planarization processes

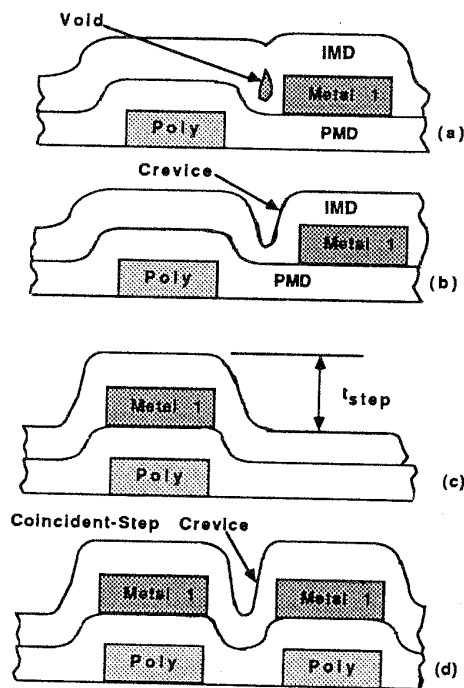


Fig. 15-9 Conditions associated with the relative spacing of polysilicon and Metal-1 that lead to layout design rules of these layers

that allow all spaces between any adjacent or overlapping features (and all possible step heights), to be covered with a dielectric layer whose surface does not exhibit any voids, crevices, or excessively high steps. Even though this is an actively pursued goal, it has not yet been attained in all processes and technologies. Various planarization approaches that have been developed along the path toward the condition in which no restrictive design rules need be invoked will be described.

15.3.2 Step-Height Reduction of Underlying Topography as a Technique to Alleviate the Need for Planarization

In two-level-metal processes, it may be possible to avoid performing any planarization procedures with respect to the IMD if the topography underlying IMD is planarized by the process steps that precede IMD deposition. Although this approach may appear to be obsolete for ULSI and ULSI circuits, several of the step-height reduction techniques turn out to be useful for providing a planar surface under Metal 1. Such a condition will ultimately be required for three and four-level-metal processes. The survey of planarization techniques of dielectric layers begins by considering this approach. In this section two older step-height reduction techniques, are described, namely: 1) CVD SiO_2 and bias-sputter etchback; and 2) planarization through sacrificial-layer etchback.

15.3.2.1 Provide a Semi-Planar Surface over Local-Interconnect Levels: Polysilicon or polycide interconnect structures are often used as a local-interconnect layer. The oxide spacers formed at the edges of the polysilicon during the fabrication of lightly doped drain (LDD) structures and salicides (see Chap. 16 for a discussion of both of these structures) result in some smoothing of the steps in the dielectric layer between the poly and Metal 1.

Since this local-interconnect level is made of materials that can normally tolerate relatively high temperatures (e.g., $> 800^\circ\text{C}$), the CVD glass that covers it can be flowed to smooth the steps at its surface. As shown in Fig. 15-10, a 30-minute 900°C anneal in N_2 of a BPSG film (4.8 wt% B and 4.6 wt% P) over patterned polysilicon can result in an almost completely planar surface.¹⁰

15.3.2.2 CVD SiO_2 and Bias-Sputter Etchback: An early method for forming partially planarized dielectrics is the CVD/bias-sputter etchback technique (CVD/Etch), is presented here because it was a predecessor of the high-density plasma (HDP) CVD dielectric deposition process more

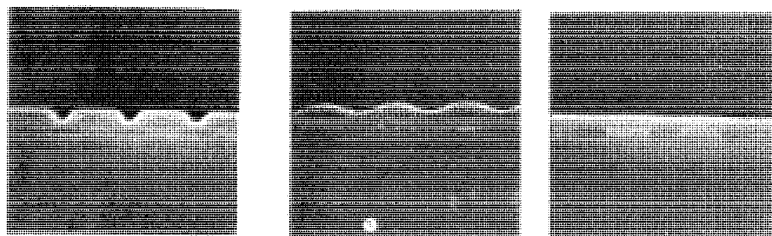


Fig. 15-10 Planarization of the PMD layer (BPSG) by thermal flow.¹⁰ Reprinted with permission of Semiconductor International.

recently developed for filling high-aspect-ratio features (discussed in Sect. 15-11). In the CVD/Etch process, an SiO_2 layer is formed by the use of plasma-enhanced CVD (PECVD [Figs. 15-11a and b]).

Initial attempts to implement this technique involved deposition of the film in one machine, followed by sputter etching of the deposited layer in another. An enhancement of this technique involved reactive facet tapering of the oxide by the use of a reactive gas (CHF_3) in conjunction with Ar.^{11,12} This produced steps on the surface of the oxide with slopes that were not as gentle as they would have been if they had been created by purely physical sputtering. However, the erosion rate was much faster, which significantly increased the throughput. Subsequent work integrated the process into a single reactor, first with a parallel-plate reactor,^{13,14} and later with a multi-chamber tool that could perform deposition and etching without exposing the wafers to atmosphere between process steps.^{15,16}

One disadvantage of the CVD/Etch method is that at low temperatures the CVD film does not fill high aspect ratio openings (e.g., a $0.5\text{-}\mu\text{m}$ space between $0.7\text{-}\mu\text{m}$ -thick metal lines) with just a single sputter-etch step. Instead, for high-aspect-ratio applications, the deposition and etchback procedure must be performed in several sequential deposition/etch sequences to prevent the formation of voids in such spaces.¹⁶ First, an oxide layer is deposited that is not thick enough to cause seam closure in the opening. Sputter etching is then used to form sloped surfaces in the small openings. This keeps the small trenches open during deposition, allowing them to eventually be completely filled without the formation of keyholes (voids). This approach also tends to stop the propagation of pinholes.

At the end of this multistep process, the surface topography will have pyramid structures over narrow lines, as shown in Fig. 15-12. To eradicate these pyramids, an erosive-mode step such as a resist-etchback technique can be used.¹⁷ A report has been made of a two-level-metal

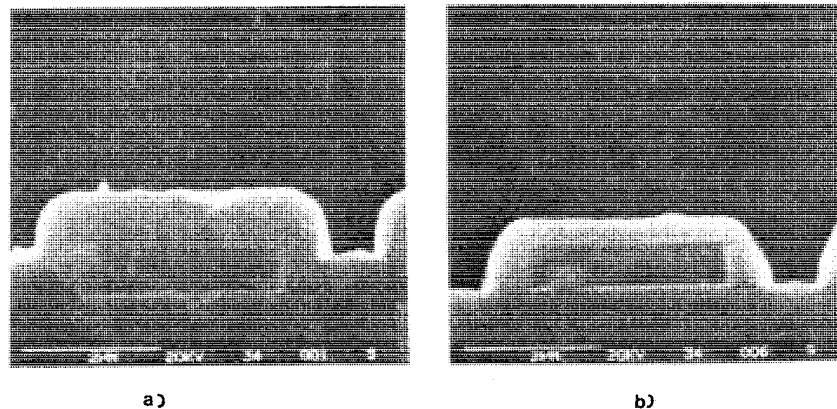


Fig. 15-11 CVD SiO_2 and *in situ* etchback. (a) Photograph of an as-deposited PECVD oxide over vertical sidewalled Metal-1 lines. (b) Shape of PECVD oxide after reactive ion etching using energetic ion bombardment.¹¹ (© 1987 IEEE).



Fig. 15-12 At the conclusion of a CVD/Etch process, the surface topography over narrow lines will have a pyramid structure.¹⁶ (© 1987 IEEE).

wiring technology with metal pitches of $1.9\ \mu\text{m}$ for Metal-1 and $2.4\ \mu\text{m}$ for Metal-2 that uses this method to form the interlevel dielectric.¹⁸

There are three main drawbacks to this multiple CVD/Etch process: 1) the throughput is very low, even when batch CVD and etch reactors are used; 2) an erosion step must be used to planarize the pyramid structures over narrow lines (further reducing throughput); and 3) global planarization is not achieved (i.e., complete planarization is not achieved over wide steps).

15.3.3 Planarization through Sacrificial-Layer Etchback

Planarization of CVD interlevel-dielectric films can also be achieved using the *sacrificial-layer etchback technique* (Fig. 15-13). This method gained the most widespread acceptance in two-level-metal processes down to $\sim 1\ \mu\text{m}$ device technologies (and even in some reported three-level-metal bipolar processes). Using this approach, it is possible to achieve a high degree of planarization between steps that are $\sim 2\text{--}10\ \mu\text{m}$ apart. With more closely spaced features, the technique runs into problems. For more widely spaced steps, planarization is less complete. The process is carried out by first depositing the CVD film that will serve as the PMD or IMD. This layer is then coated with a film that will later be etched off (sacrificed). In most cases, such sacrificial layers are photoresists, spun on in liquid form. Upon being baked, these liquids become solid thin films that are thick enough so they provide a film surface that is close to being flat, as shown in Fig. 15-14. As long as the feature size is smaller than $10\ \mu\text{m}$, the degree of planarization can exceed 70% with most photoresists.

In the next step, the sacrificial layer is first rapidly etched back in a plasma (typically, O_2 or O_2 mixed with CF_4) until the topmost regions of the dielectric layer are just exposed (Fig. 15-13b). The etch chemistry is then modified so the sacrificial-layer material and the dielectric are etched at approximately the same rate. The etch is continued under these conditions until all of the sacrificial layer has been etched away. At this point, the surface of the dielectric film is highly planarized since the profile of the sacrificial layer is transferred to the dielectric layer by the etchback procedure. The thickness of the dielectric film over underlying features, such as metal lines, may be thinner than desired after the etchback is completed. In some processes, in fact, the etchback step is allowed to proceed until the Metal-1 lines are exposed. In any case, an

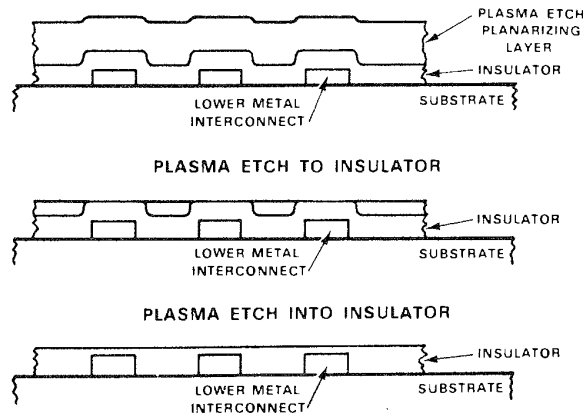


Fig. 15-13 Sequence of steps used in a sacrificial-layer etchback process for planarization.

additional layer of CVD dielectric is generally deposited in order to establish a minimum adequate thickness everywhere on the wafer surface.

15.3.3.1 Sacrificial-Etchback Process Problems: To achieve the desired degree of planarity and final dielectric film thickness, very tight process control is necessary for such parameters as the magnitude of the etch rate, the etch-rate uniformity across a wafer, end-point detection, and oxide-thickness uniformity under the resist layer. The resist etch-rate is sensitive to the cure cycle (especially if a high-temperature bake step [$>150^{\circ}\text{C}$] is used to improve surface planarity), and to chamber preconditioning. In batch processes, the etch rate is also sensitive to the number

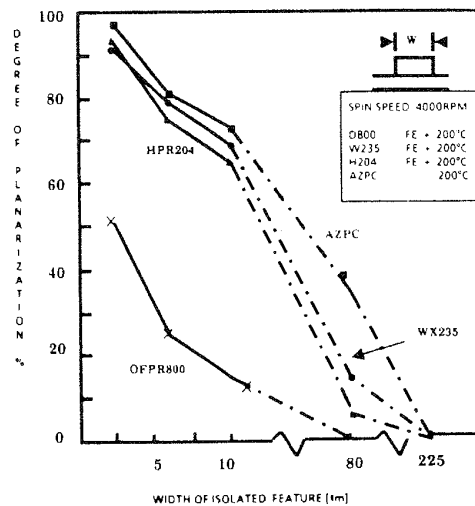


Fig. 15-14 Degree of planarization vs. various feature sizes for different polymers.¹⁴² (© 1987 IEEE).

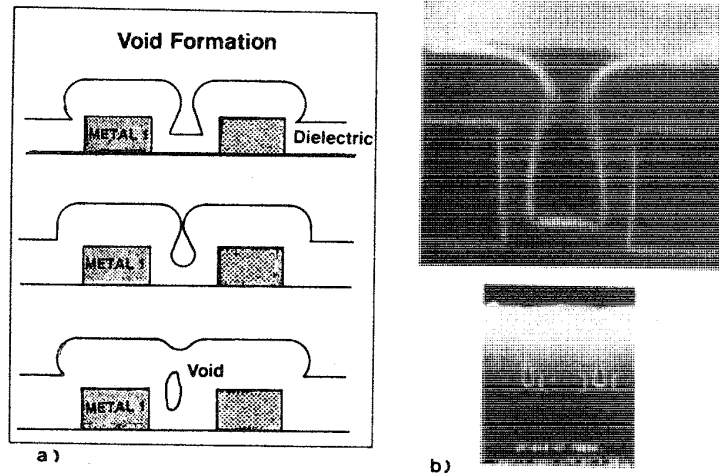


Fig. 15-15 (a) Diagram of void formation due to cusping of the dielectric. (b) SEM of voids left between $1.5\text{-}\mu\text{m}$ -spaced Metal-1 lines in a $3.0\text{-}\mu\text{m}$ -thick PMD PECVD oxide.

of wafers present in the chamber. Such tight process control of a complex process is hard to maintain in production line.

The process also requires a multitude of steps, which implies a lengthy process, and thus a low throughput. Etchback is a difficult process to successfully implement in a batch-etch mode, resulting in an even lower throughput. Finally, if the aspect ratio between adjacent Metal-1 lines with vertical sidewalls exceeds 0.4, voids will form when most CVD- SiO_2 layers are deposited (Fig. 15-15). These voids can trap resist and moisture, which can subsequently be released if the voids are opened during the etchback process. Opened voids will also trap metal during the Metal-2 deposition step. It will remain in the form of metal stringers following Metal-2 patterning. Thus, this process is not feasible for Metal-1 spacings of less than $\sim 1.25\text{ }\mu\text{m}$ (unless a process is available for deposition of SiO_2 layers at low temperatures without formation of voids in high-aspect ratio spaces [e.g., PECVD TEOS films, O_3 :TEOS films, or HDP CVD films]).

15.4 DOUBLE-LEVEL-METAL (DLM) INTERCONNECTS FOR $1\text{-}\mu\text{m}$ CMOS

Double-level-metal (DLM) interconnect systems were used in CMOS technologies ranging from $2.0\text{ }\mu\text{m}$ down to about $0.8\text{ }\mu\text{m}$. In this section an example of the kinds of materials and sequence of process technologies employed to fabricate such DLM systems is described. In later sections (after some additional interconnect processes are covered), examples of a *triple-level metal (TLM)* system (used in the $0.65\text{--}0.35\text{ }\mu\text{m}$ CMOS generations), and a 5-level (or more) interconnect technology (used for $0.25\text{--}0.18\text{ }\mu\text{m}$ CMOS generations) are presented. These examples are merely *academic constructs* of a multi-level interconnect system. That is, actual systems used in ICs are likely to differ in various details from the examples given here. But the process flows and materials used for descriptive purposes are reasonably representative of the way such interconnects are actually implemented (see Figs. 16-17 to 16-19).

The fabrication of DLM interconnect structures begins after the contacts in a PMD BPSG layer are opened and a reflow step has been carried out. At that point Metal-1 is deposited. In this DLM technology an alloy of Al, such as Al-Si (e.g., 1wt% Si) or Al-Cu (e.g., 1–2wt% Cu) is the material of choice, deposited to a thickness of about 500 nm by sputtering. The Al-Si alloys were used in earlier NMOS and CMOS technologies (i.e., 2- μ m CMOS), with the silicon being added to the aluminum to prevent spiking of the contacts during subsequent annealing steps (see Chap. 11). In 1.2- μ m (and smaller) CMOS, such Al:Si alloys were replaced with a metallization system that included a diffusion barrier (typically Ti/TiN) and an Al:Cu alloy film for the main interconnect layer. The Cu was added to the Al increase the electromigration resistance of the films. Dry etching was used to pattern the Metal-1 films.

An IMD layer is deposited next, to a thickness of about 1.5 μ m. This layer electrically isolates Metal-1 from Metal-2. A variety of CVD processes for depositing such dielectric films have been developed (see Chap. 6). Deposition of the IMD layer generally makes the wafer topography too severe to allow good step coverage by the Metal-2 film. In the DLM era, partial planarization by sacrificial resist etchback was the most widely used technique for alleviating this topography problem. The resist-etchback technique can provide wafer topography in which the steps in the IMD layer are more gently sloped and reduced in height (compared to the topography conditions that exist prior to resist etchback). After resist etchback, another thin layer of dielectric is deposited (~400-nm thick) because the resist etchback step thins the original dielectric film to a degree which might make it too thin over some areas of the circuit. The next step is to open vias in the intermetal dielectric layer (to allow electrical connection between Metal-2 and Metal-1 at desired locations). To help ensure adequate step coverage of Metal-2 into these vias, a two-step “wine glass” etching process was widely practiced. (Chapter 14 provides more details on the “wine-glass” SiO₂ etch process.) The shape of the via sidewalls after such a two-step etching procedure is shown in Fig. 16-17.

Metal-2 is then deposited (again by sputtering a layer of Al:Cu). This layer is generally thicker than Metal-1 (e.g., Metal-2 is ~1000 nm [1 μ m] thick). The sputtering process usually involves heating the wafer during sputtering (to 300–400°C) to further improve the step coverage of Metal-2 over IMD topography (and into the vias). Metal-2 was then patterned by dry etching. Figure 15-16 shows an SEM of such a DLM interconnect system.

A *passivation* (or *overcoat*) layer, such as CVD PSG or plasma-enhanced CVD silicon nitride (or both), is then put down onto the wafer surface. This layer seals the device structures on the wafer, protecting them from contaminants and moisture. It also serves as a scratch protection layer. Openings are etched into this layer so that a set of special metallization patterns under the passivation layer is exposed. These metal patterns are normally located in the periphery of the circuit and are called *bonding pads* (see Fig. 16-20 again). Bonding pads are typically about 100 x 100 μ m in size and are separated by a space of 50 to 100 μ m. Wires are connected (bonded) to the metal of the bonding pads and then bonded to the chip package. In this way connections are established from the chip to the package leads (see Chap. 17 for details on chip bonding).

A sintering step is carried out after the passivation layer is deposited. This step brings the metal and the n^+ and p^+ regions in the silicon into intimate contact. Such intimate contact between the metal and the heavily-doped Si regions establishes a low-resistance ohmic contact.

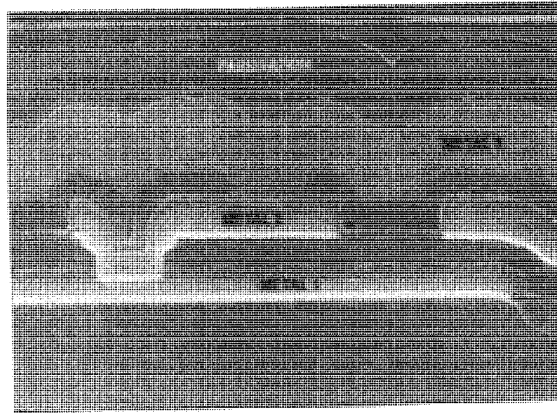


Fig. 15-16 SEM of the cross-section of a double-level metal interconnect system.

The annealing process exposes the wafer to a 375–500°C temperature in an ambient of H_2 , or $N_2 + H_2$ (5%) for about 30 minutes. This step is also used as the annealing process for reducing the interface trap density in the gate oxide/silicon interface (see Chap. 8). Note that in such DLM processes, vias were typically not allowed to be stacked directly above contact holes. The incomplete filling of the contact hole by Metal-1 would result in poor contact between Metal-2 and Metal-1 through the via if this occurs (as shown in Fig. 15-17a). The inability to stack vias on top of contacts represents a significant reduction in routing flexibility and decreases the effectiveness of using the available silicon real estate for interconnection.

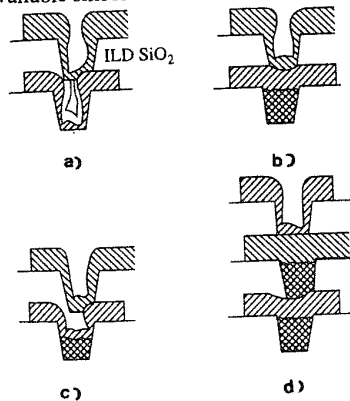


Fig. 15-17 Stacked contact/vias. (a) If no metal plugs are used in the lower contact hole, poor contact occurs between Metal-1 and Metal-2. (b) If the contact hole is filled with a metal plug, good contact can be made between Metal-1 and Metal-2. (c) If the metal plug has a recess (or "dimple"), poor contact between Metal-1 and Metal-2 can still result. (d) Good metal plugs can allow multiple stacking of vias.

15.5 CHEMICAL MECHANICAL POLISHING (CMP)

As the number of interconnect layers in ICs has increased, the planarization of dielectric and metal layers has become more critical. In general, planarizing techniques such as thermal flow, sacrificial-resist etchback, and spin-on glass are inadequate for interconnect systems with more than three layers of metal. All of these techniques ultimately provide only a limited degree of smoothing and local planarization (i.e., over distances smaller than $10\ \mu\text{m}$). They are not able to provide global planarization.¹⁹ Global planarization of both dielectric and metal layers, however, is possible with chemical-mechanical polishing (CMP), a process that was pioneered at IBM in the 1980s. In this technique high elevation features are selectively removed (i.e., material from high elevation features is removed more rapidly than material at lower elevations), resulting in topography with improved planarity (see Figs. 15-18 and 15-47). Mechanical polishing, assisted by chemical action, produces such selective material removal. The process is performed by mounting the wafer face down on a carrier. The carrier is then pressed against a rotating platen containing a polishing pad. The carrier itself is also rotated. An abrasive-containing aqueous slurry is dripped onto the table, saturating the pad (Fig. 15-19). The degree of planarization is expressed by Eq. 15-4. A step-height reduction of 90–95% is achievable using CMP, yielding values of β in the 0.90–0.95 range. CMP is also capable of planarization of distances up to several millimeters, far in excess of that possible with any other technique thus far developed.²⁰

CMP offers three key advantages for fabricating sub-micron ICs. First, for logic devices (such as microprocessors and ASICs), CMP makes it possible to achieve high manufacturing yields and high device speeds with multilevel interconnects having up to eight levels of metal. By planarizing IC topography, CMP avoids the problem of metal thinning over steep topographies in multi-level interconnect structures. Second, memory devices (such as DRAMs) benefit from CMP even though they use interconnect systems with only two levels of metal. For memories, the flat wafer surface after CMP increases the depth of focus budget available for lithography (see Chap. 13). This allows memory circuit designers to employ smaller critical dimensions, thus reducing chip sizes without decreasing yield. Smaller chip sizes provide a

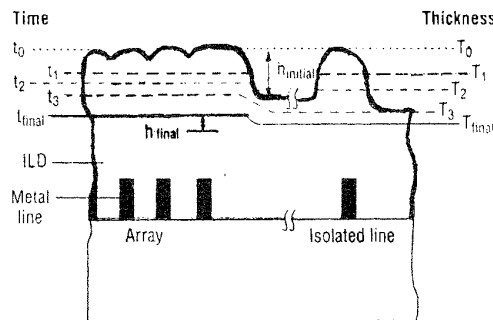


Fig. 15-18 The basic CMP process is one of reducing film thickness selectively over the wafer, ending with improved planarity.⁴⁶ Reprinted with permission of Solid State Technology. Published by PennWell.

Slurry
disper

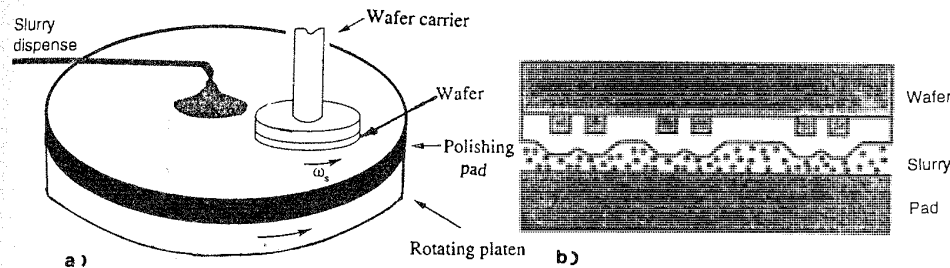


Fig. 15-19 Schematic drawing of (a) a basic CMP polishing tool; (b) the basic CMP process.

significant cost advantage in the competitive realm of DRAM manufacture. Third, CMP is capable of reducing (rather than increasing) defect density.

In this section the following topics will be discussed: a) the history of CMP; b) the mechanisms involved in CMP of metals and oxides; c) CMP equipment and consumables; and d) challenges facing CMP processing. The latter include: inadequate understanding of the polishing mechanisms, low throughput, inadequate reliability of the polishing tools, degradation of removal rate with usage time of a polishing pad, non-uniformity of film thickness removal across the wafer, high cost of ownership, and high defect density.

Before beginning this discussion the goals desired of a CMP process are listed. They are:

1. Good planarity across a die (highest spot to lowest spot should be less than 150 nm);
2. Good film thickness uniformity within the wafer ($\pm 5\%$, 1σ);
3. High removal rates (100–140 nm/min), for high throughput and low cost-of-ownership;
4. Good selectivity and/or good endpoint technology to avoid under or over-polishing;
5. The CMP process should not add excessive defects (e.g., for 0.18 μm CMOS, fewer than 0.125 particles/cm² with a size larger than 0.06 μm are allowed at each mask level).

15.5.1 The History of CMP

The early history of CMP was chronicled by Fury.²¹ He recounts that the process of chemical-mechanical polishing originated at IBM in 1983. Research and development was exclusively performed within IBM until 1988. During those five years IBM closely guarded all information about the process. By the late 1980s several factors changed their secretive attitude. First, Intel microprocessors were selected for IBM's PC product line, and IBM decided to share some CMP data to help improve Intel's interconnect technology. Second, at that time the DRAM market was dominated by Japanese manufacturing companies. IBM was wary of having to rely entirely on foreign suppliers for such a critical component of its computer systems. Hence, they elected to help strengthen a domestic DRAM supplier (Micron Technology, Inc.) by teaching them about the CMP process. Information about CMP was included in the process for the 4 MB DRAM that IBM sold to Micron Technology in the mid-1980s. In 1988 IBM agreed to disclose to SEMATECH its process scheme for multi-level interconnects, although they did not give all

the details on CMP. They did, however, provide one of their CMP polishing tools (a Si-wafer polishing tool sold by Westech). Motorola was introduced to CMP through the Power-PC microprocessor project, a joint venture of IBM, Motorola, and Apple Computer. In 1989 IBM also published the first paper which publicly discussed the CMP of oxide films.²² These events initiated the rapid proliferation of CMP throughout the IC industry.

The spread of CMP was hastened by some other events, chiefly the downsizing of IBM from 1990–94. Many engineers with expertise in CMP left IBM and went to work for other IC companies who were seeking to implement the technology. The effort at SEMATECH involved improving the CMP tool performance and reliability. Their work with Westech tools led to the development of a popular CMP tool, the IPEC 472. (Westech was acquired by IPEC.)

Although early development of CMP was centered in the US, worldwide efforts soon got underway. Polishing tools were initially offered by companies that had been building machines for polishing silicon wafers (Speedfam and Strasbaugh, as well as IPEC [Westech]). Overseas suppliers soon joined them (e.g., Ebara and Cybeq from Japan, and Wolters from Europe). Eventually the larger semiconductor equipment vendors entered the market. Applied Materials introduced its tool in 1996, and Lam Research acquired On-Trak Systems, another commercial CMP vendor. The large number of vendors gave rise to some consolidation as well (an example of which was the merger of Speedfam and IPEC in 1999).

CMP was the fastest growing IC process technology of the 1990s, and it was also the most rapidly growing segment of the semiconductor process equipment market in that time frame. From 1995 to 1997 sales tripled, reaching \$520 million in 1997. Sales were expected to exceed \$1 billion by 2000.²³ The number of CMP polishing steps per year was also estimated to reach 125 million by that date. The first application of CMP was the planarization of dielectric films (mainly CVD oxide), but the number of applications has increased. Currently it is being used to planarize metal films (including tungsten, aluminum, and copper), oxides and low-*k* dielectric films. By the end of the century, CMP had evolved from being a cleanroom engineers nightmare, to a critical process technology without which production of ULSI chips with multilevel interconnects might not be possible.

15.5.2 Modeling the Mechanisms of CMP

The mechanisms of CMP are quite complex, and differ for each type of film being polished. Many variables also impact the removal rate, including: film type, pressure (force applied to the polishing pad), velocity, temperature, slurry feed rate, the type of polishing motion, slurry particle size, pad hardness, and pad conditioning. The pressing need for production-worthy CMP capabilities forced early researchers to ignore the complexity of the process and go straight to an empirical trial-and-error approach. But, to logically optimize a CMP process, a more fundamental understanding of the mechanical and chemical forces at work is needed. While a complete understanding of CMP mechanisms is not yet available, some insight has been gained. A brief introduction as to what is known about CMP mechanisms is presented here, divided into two categories: metal CMP and oxide CMP.

15.5.2.1 Metal CMP Mechanisms: The mechanisms involved in CMP of many metals are relatively simple. In general, an oxide of the metal is chemically formed on the metal film surface. Then it is removed mechanically with abrasives in the CMP slurry. Thus, for effective

polishing, it must be possible to both oxidize the metal and closely match the hardness of the slurry abrasive to the hardness of the material being removed. This prevents scratching and deforming the softer metal beneath the oxide being formed. The goal is to find slurries with just the right combination of chemical and abrasive aggressiveness to produce high removal rates, good planarity, local uniformity, and high selectivity. Successfully carrying out the above process for specific metals can be challenging. The slurry's chemical composition tends to play a more important role in metal CMP than in dielectric CMP. As a result, it is possible to achieve fairly high metal-to-oxide selectivity (e.g., 120:1 for tungsten to silicon dioxide). However, the slurry solution may also etch the metal oxide that is formed. A careful balance between etching and oxidation must therefore be maintained. If etching dominates, then the metal will etch isotropically and planarization will not be obtained. The metal deposition step may have to be modified to prevent its deposition near the wafer edge. Otherwise, after CMP, metal stringers may remain on the wafer's beveled edges, causing yield and reliability problems.

CMP of aluminum is being examined for use in damascene Al interconnects (see Sect. 15-12). One of the challenges facing Al-CMP is that the metal is soft and susceptible to scratching, but its oxide, Al_2O_3 , is hard and adheres tightly to the metal. This makes Al_2O_3 difficult to remove. If a strong oxidizer is used, the entire Al film (not only its elevated regions) tends to disappear. It has been shown that if a correct set of process conditions is established (e.g., an appropriate combination of pad, slurry type, and applied pressure), Al films can be planarized with CMP. While removal rates of about 200 nm/min have been demonstrated, the resolution of the issues concerning the Al-CMP process had not been achieved by 1997.²⁴

Tungsten (W) CMP has been developed to replace the W-etchback step of the W-plug formation sequence (Chap. 6). The W-CMP process, however, requires a surface that is planar prior to via formation and metal deposition (a condition not required in processes that form W-plugs by W-etchback). W-CMP has been described as a balance between the electrochemical oxidation of W to form WO_3 , and the mechanical removal of that reaction product by abrasive slurry particles.

The polishing of tungsten films raises two key issues. The first involves the selection of an appropriate slurry. Tungsten polishing at IBM was originally done with a slurry mixture of potassium ferricyanide and potassium phosphate (having a pH adjusted to about 5.0 to 6.5), together with colloidal silica or dispersed alumina silica, Al_2O_3 .²⁵ However, potassium ferricyanide is highly toxic, and its disposal causes severe environmental impact. As a result, this chemical has been replaced, with the most widely used substitutes being ferric nitrate, potassium iodate, and hydrogen peroxide.²⁶ An example of commercially available slurries for W-CMP are Cabot's SEMI-SPERSE® WA-355 and SEMI-SPERSE® FE-10.

The second key issue of W-CMP arises because blanket-W films are deposited over a liner film of TiN/Ti (Chap. 6). When removing the W film, this liner film must also be polished away. A high removal rate of the TiN/Ti layer is essential to prevent erosion of the oxide underlying the metal layers, and to also prevent plug recessing (Fig. 15-20a). However, the polish rates of TiN/Ti are typically three to eight times slower than that of W (which is removed at rates exceeding 350 nm/min). It is also necessary to have a high selectivity to the underlying oxide to obtain local planarity of the plug surface and to prevent excessive loss of the surrounding oxide by dishing. To remove the liner film, *over-polishing* is typically used (named

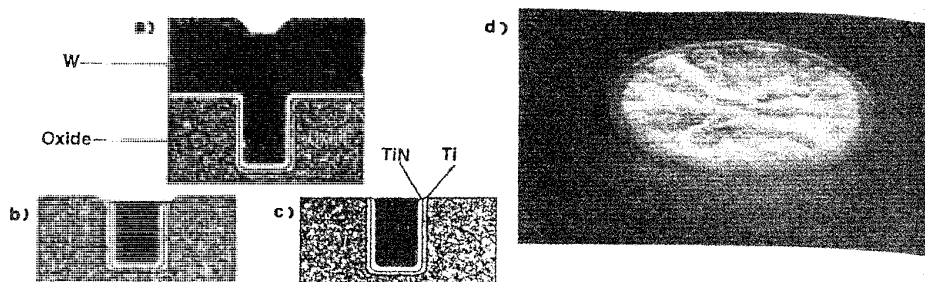


Fig. 15-20 W plug formation using CMP a) W deposition over TiN/Ti barrier layers b) CMP of W and TiN/Ti leaves recessed plugs. c) Oxide buff eliminates recess.²⁶ Reprinted with permission of Solid State Technology. Published by PennWell. (d) SEM of W plug after W and TiN/Ti CMP steps and oxide buff step.²⁷ Reprinted with permission of Semiconductor International.

because it causes some recessing of the W-plug as shown in Fig. 15-20b). During this step the field oxide is also damaged to some degree.

To solve the above problems, multi-step polishing sequences are often used in W-CMP processes.^{26,27} For example, in the first step the W layer is removed using one platen of a polishing tool. In the second step, the wafer is moved to a second platen where the liner film (e.g., TiN/Ti) is polished away. Finally, a third oxide-buff step (on yet a third platen) removes the oxide damage by polishing away a small amount of the now-exposed underlying oxide (using a silica based slurry that minimizes surface roughness and facilitates the removal of alumina slurries). This also restores the level of the W-plug to that of the oxide surface by eliminating the recess (Fig. 15-20c). Such CMP W-plug processes reportedly produce higher yields than processes which employ a W-etchback process.^{28,29} The higher yields with W-CMP are attributed to improved defectivity and superior final plug morphology.

The CMP of copper (Cu) interconnect films involves a different set of challenges.^{30,31} The removal rate of Cu is about 350 nm/min with a selectivity of 150:1 to the underlying oxide. However, Ta is commonly used as the liner layer for Cu, and removing it by CMP represents some difficulties. Cu is relatively soft and oxidizes easily with standard chemistry (such as hydrogen peroxide or hydrogen persulfate). On the other hand, Ta is very hard and is also much more chemically inert. This typically leads to large differences in selectivity when simultaneously polishing Cu, Ta, and the underlying oxide (i.e., the Ta removal rate is only about 30 nm/min). This leads to significant *dishing* of wide Cu and oxide features (defined in Sect. 15.5.4.1). Damascene CMP applications require low selectivity between the primary conductor material and barrier films (1:1) and a high selectivity between the primary conductor and insulator (> 100:1) to essentially stop the process. Since Ta is the harder and more inert metal, it is thought to be removed primarily by abrasion. An aggressive oxidizer could be used, but this would create the possibility of polishing through the Cu and removing device features. Work has been proceeding to find an oxidizer that selectively attacks the Ta without removing Cu.³²

Another related problem in Cu/Ta-CMP is the electrochemical interaction between Cu and barrier materials during CMP.³³ Two dissimilar metals in simultaneous electrical contact and contact with an electrolyte solution constitute a galvanic cell. This chemistry frequently leads to anodic corrosion. Such a situation arises near the end of a polishing cycle of Cu and Ta (when the barrier metal becomes exposed in the field regions, and Cu is exposed in the trenches). Cu in the trench can be removed (sometimes entirely), even if the trench was initially well-filled with it. Obviously, this effect must be suppressed to enable successful CMP of Cu/Ta films.

Use of multi-step polishing processes has also been pursued for CMP of Cu interconnects. As in W-plug CMP, a three-step damascene-Cu-CMP process uses three platens, each with its own slurry mixture. Copper is removed with the first platen, Ta layers with the second, and an oxide buffing step is carried out on the third. One must ensure that all the Cu in the field regions is removed during the first step, since the removal rate of Cu in the second step is deliberately kept very small. In another approach used by IPEC-Speedfam, multiple slurries are simultaneously delivered through-the-pad to each of four autonomous polishing modules. This permits a multi-step Cu CMP process to be carried out without transferring wafers from one platen to another. Increases in throughput and yield are thereby obtained.

For multi-step Cu-CMP processes one of 3 or 4 types of slurry might be best for the second polishing step. The selection of a particular slurry might depend on the liner layer being used and the type of intermetal dielectric material beneath the metal film. A large number of companies are developing slurries for Cu-CMP, hoping to capture a significant market share of this important emerging process. They are listed in Sect. 15.5.3.4.

15.5.2.2 Silicon Dioxide CMP Mechanisms: The mechanisms that underlie CMP of silicon dioxide are more complex than the mechanisms of metal CMP. Since SiO_2 is itself an oxide, the mechanism of oxidation used in metal CMP is absent from the CMP of SiO_2 . In fact, in 1997 Tomozawa stated that, like optical-glass polishing, much of the oxide CMP process was still an art rather than a technology. To qualify as a technology, the basic mechanisms of a process must be understood, allowing it to be logically optimized. By 1997 the basic mechanisms of oxide CMP were not yet well understood.³³ The description provided here briefly summarizes what is known about the phenomena associated with oxide CMP. Since oxide CMP is similar to optical-glass polishing, some things about it can be learned from the long history of glass polishing. Optical-glass and SiO_2 films on Si wafers are polished with a slurry consisting of abrasive silica particles (whose hardness is about equal to that of the oxide) suspended in an aqueous alkaline solution (typically KOH or NH_4OH , diluted in DI water).

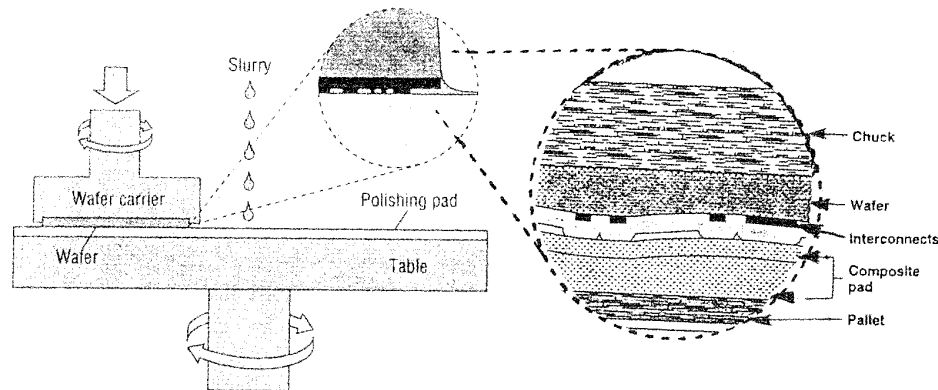
A relationship that models the optical-glass polishing rate, R (defined as the thickness of glass removed per unit time), was developed by Preston, and is known as Preston's law.³⁴ This law states that R is proportional to the applied pressure, P , and linear velocity, v , at which a polishing pad moves relative to the optical-glass specimen, or:

$$R = K_p P v \quad (15.6)$$

where K_p (Preston's coefficient) is a constant under a set of given experimental conditions. K_p is a function of glass hardness, Young's modulus, the polishing slurry, and the composition and hardness of the polishing pads. The same law appears to apply to the rate of removal of oxide by CMP. (Typical values of P and v in CMP processes are 7 psi, and 30-40 ft/min. respectively.)

Various models have been proposed to explain why oxide removal proceeds according to Preston's law, as well as how this mechanism allows CMP of oxide to selectively remove the "up regions" of an oxide film on wafers faster than the "down regions" (thereby allowing planarization of the film to occur). Izumitani proposed that optical-glass polishing "proceeds by the formation of a hydrated layer by means of a chemical reaction between the glass surface and water and then the removal of this hydrated layer by the abrasive particles."³⁵ This proposed mechanism is consistent with the observation that the polishing rate is extremely low without the presence of water to cause the hydration. (Polishing with a slurry of silica particles in a non-aqueous solution exhibits substantially lower polishing rates.) The hydrated layer that thus forms on the oxide surface generally has a much lower hardness and mechanical strength than the SiO_2 beneath it. Therefore, plastic deformation results in mechanical removal of the hydrated layer (by a plowing effect) when it is rubbed by the abrasive particles.

The same type of mechanism is expected for oxide CMP. Removal of the oxide takes place by a combination of mechanical and chemical action, using an abrasive slurry dispersed in alkaline solution (Fig. 15-21). Since silica particles are used in the slurry of oxide CMP, they have about the same hardness as the SiO_2 . Thus, the unreacted SiO_2 is not appreciably removed (nor damaged) by the plowing effect when it is rubbed by the silica particles in the slurry. However, the silica slurry particles still cause the oxide to plastically deform. This deformation is likely assisted by frictional heating. It has been determined that oxide readily undergoes hydration (water entry) if it is plastically deformed by a diamond indenter. When entered by water, the hardness and mechanical strength of the oxide is generally decreased. Thus, during CMP, it seems that hydration of the oxide film surface is caused by the plastic deformation of the oxide by the abrasive particles. The degree of deformation is increased as the pressure of these particles on the surface is increased, and by frictional heating. The softer hydrated layer is then removed by the plowing action of the abrasive particles, which increases as the linear velocity of the polishing pad is increased. These conclusions are substantiated by Preston's law. The properties of the pad also appear to play an important role in CMP. It is probable that the



15-21 Detailed drawing of a wafer being polished by CMP.

polishing pad has to be soft enough to grip the abrasive particles, but hard enough to transmit the pressure to the oxide. It is generally agreed that some sort of chemical mechanism (such as SiO_2 dissolution) is a cofactor in causing oxide CMP. Various chemical processes have been proposed.³⁶ The larger polishing rate in higher pH slurry is often cited as evidence for silica dissolution.

While the above model gives some insight into how oxide is polished with a silica-based slurry, it does not address the issue of how planarization of the oxide by such polishing is achieved. To obtain planarization of a film with topography, the surfaces of elevated features ("up areas") must be eroded faster than those at lower elevations ("down areas"). The selective removal of material in "up areas" is empirically observed to depend on many factors, including the feature dimensions, the structure (and surface condition) of the polishing pad, and the height of the step being removed.

Selective removal of material that protrudes above the surrounding topography is thought to arise from the compliant nature of the polishing pads.³⁷ Figure 15-21 depicts the localized compression occurring within a pad as features on the wafer surface are driven across it. The pad applies pressure to the wafer surface through the slurry, with the largest pressures being exerted where the pad is most compressed. Higher pressure is applied to elevated regions and the leading edges of the features, than to lower-lying regions and trailing edges.³⁸ The difference in local applied pressure causes differences in material removal rates. Smaller features are rounded off and polished faster than wider features.

As the steps erode, the "down areas" are exposed to increasing pressure, and the removal rate there increases (approaching that of the "up areas"). As this happens, the rate of step height reduction decreases toward zero. In the latter stages of the polishing process the selectivity is so low that small residual steps remain, and in practice cannot be fully removed by continued polishing. Material is also removed faster from wide down areas than from narrower ones. (See Sect. 15.5.5.1 that deals with such "dishing" effects.) This results in slower planarization of steps adjacent to wide field areas and faster planarization of steps adjacent to narrow field areas.

As first discussed by Burke,³⁹ the time to planarize an up-area structure (or to reduce its

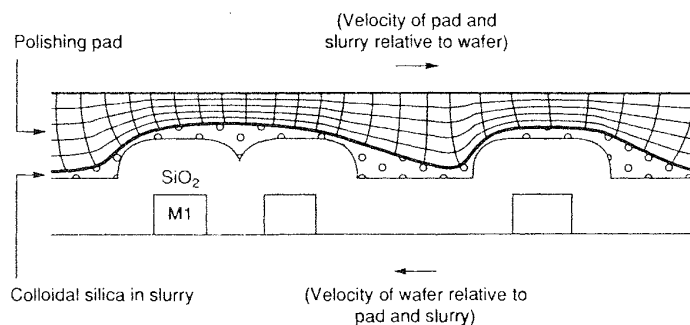


Fig. 15-22 Depiction of polishing pad compression and localized pressures applied to topographic features. Pressures are greatest at the top leading edges of steps, and are lowest in the wakes of the steps.

height to some fraction of its initial value) is dependent on the down area adjacent to it. Specifically, the time to planarize depends on the width of the area as well as the ability of the polishing pad to rigidly span that distance. The stiffer the pad, the more rigid the span and the slower the rate of material removal from down areas. This gives good planarization. The softer the pad, the less able it is to bridge larger distances. Thus, more material is removed by softer pads in down areas between adjacent up areas. This gives poorer planarization. When features to be polished are separated by distances greater than the leveling distance of the pad, there is no up-area to down area selectivity, and planarization does not occur. Ideally, the leveling length of a pad is large enough to planarize the widest down areas on a chip. On the other hand, the greater the leveling length, the more the pad will respond to wafer-scale non-flatness parameters, such as TLV and bow. This results in non-uniformity of the polished film thickness across the wafer (discussed in more detail in Sect. 15.5.5.2).

In summary, planarization will occur most slowly for step heights in an oxide over the largest or densest interconnect structures adjacent to wide unpatterned field areas, scribe lines, and wafer edges. It will occur most rapidly for step heights in the oxide over smaller arrays, smaller gaps, and narrow isolated features. It is also necessary to specify an acceptable degree of planarization, because in the latter stages of the CMP process, material is removed nearly as fast from down areas as from up areas. Step heights no longer decrease at appreciable rates.

Like metal-CMP processes, oxide-CMP in a production environment is often a multi-step process. When in-line endpoint detection is employed (as will be discussed in Sect. 15.5.5.4), two or more polishing steps may be needed to obtain the correct oxide thickness. In addition, a separate final polish (or buff) step is frequently used to provide light polishing and to remove slurry particles and small wafer defects. A small amount of oxide can also be buffed away with the step (typically 30–40 nm), which aids in removing defects (especially residual slurry particles and the small scratches created by the primary polish). Use of a buff step is one of the reasons why CMP can actually reduce defect densities (including eliminating some types of defects created in other processes), rather than increase them.⁴⁰

15.5.2.3 CMP of Low- k Dielectrics: Other materials besides CVD-SiO₂ are being investigated for use as intermetal dielectric films (especially those with a lower dielectric constant - low- k dielectrics). To be used in damascene interconnect structures, such materials must be compatible with CMP. The CMP of fluorinated oxide (F-SiO₂) has been studied.⁴¹ It was found that both mechanical abrasion and chemical erosion are intensified by fluorine doping (relative to undoped TEOS oxides). This happens because of the reduced hardness and chemical reactivity of fluorinated oxides. One way to avoid these problems is to cap the F-SiO₂ layers with a CVD SiO₂ layer. This cap layer must be thick enough so that the dielectric stack is planarized by CMP and the F-SiO₂ layer is not reached. However, the cap layer increases the overall dielectric constant of the IMD stack. Furthermore, depositing the stacked film is a multi-step deposition process. Another approach is to pursue the development of more chemically stable low- k materials, and such efforts are continuing.

15.5.3 CMP Equipment

There are number of different tools and consumables used in a complete CMP process, including: 1) polishing tools; 2) pads; 3) slurries; 4) slurry distribution systems; 5) endpoint detection

systems; 6) cleaning tools; and 7) metrology tools. These items are described in this section.

15.5.3.1 CMP Polishing Tools: The basic tool used for CMP evolved from the machines that had been used to polish bare silicon substrate surfaces. When the idea for CMP was first conceived at IBM, they purchased silicon polishing tools (used by wafer manufacturers) from Strasbaugh and Westech (now SpeedFam-IPEC). The basic configuration of the silicon polisher was retained, but these tools were modified to make them suitable for developing CMP. As mentioned earlier, in such polishers, the wafer is mounted upside down on a wafer carrier, and the carrier is pressed downward against a polishing pad (Figs. 15-2 and 15-5). Both the wafer carrier and the polishing pad are rotated. Slurry is dripped onto the polishing pad, which becomes saturated. The pad thus drags the slurry beneath the rotating wafer carrier, where it can polish the wafer surface (as described in the section on mechanisms of CMP). The first polishers adapted for CMP had a single wafer carrier.

The CMP polisher differs from the Si polishing tool in a number of ways. In Si polishing several tens of microns of material are removed, while the CMP process removes only about $0.5\text{--}1.0\text{ }\mu\text{m}$ of material. The uniformity of material removal across a wafer in CMP is also much more stringent than for Si polishing ($\pm 0.5\text{--}0.1\text{ nm}$ for CMP). Furthermore, CMP tools require a higher level of automation, throughput, reliability, and critical process-parameter control. Thus, despite more than 30 years of experience in Si wafer polishing, new processes and equipment had to be developed by the IC industry for carrying out CMP.

The Westech and Strasbaugh polishers first used (and modified for CMP) by IBM were also the tools that found widest use in the early CMP processes. The Westech Si polishing tool provided to SEMATECH was improved by that organization for CMP in a program that also focused on increasing CMP tool reliability. The results of this program were used in designing the IPEC model 472[®] CMP tool, which achieved wide adoption during the mid-1990's.

In Fig. 15-23 the wafer is shown mounted on the carrier. Instead of the wafer being held directly on the chuck, a backing film is sandwiched between the wafer and the steel chuck. This film provides elasticity between the wafer and chuck. Without a flexible backing film any defect or particle on the chuck (or on the back of the wafer) will cause a thin spot in the film being polished. In addition, the possibility of wafer breakage is increased. The depth of the pocket in

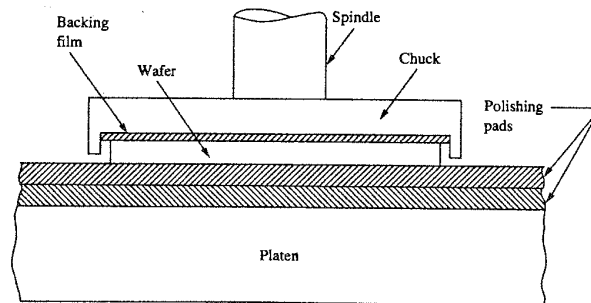


Fig. 15-23 Schematic of the components of a CMP wafer carrier.

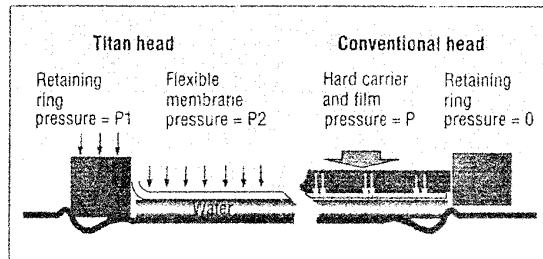


Fig. 15-24 Comparison of a conventional carrier and one in which pressure can be applied to the wafer retaining ring.²⁶ Reprinted with permission of Solid State Technology. Published by PennWell.

which the wafer resides is chosen to allow the top surface of the wafer to extend beyond the pocket, allowing its surface to be polished. In conventional wafer carriers the retaining ring does not make contact with the pad. As a result, there is no pressure applied to the pad by this ring. However, the pad pressure at the edge of the wafer is not well controlled, and the polishing rate of the film is decreased at the wafer's edge. Hence, an edge-exclusion ring of 5-7 mm must be enforced for wafers with such carriers. Carriers used in some polishing tools (e.g., Applied Materials' Mirra® and Strasbaugh's Symphony®) are designed with a retaining ring that can apply variable pressure to the pad (Fig. 15-24). This allows the pad profile at the wafers edge to be controlled. A smaller edge-exclusion zone on the wafer can therefore be used (2-3 mm).

While early CMP tools used a single wafer carrier, more recent designs contain multiple carriers (and multiple platens). This increases the throughput of the polisher and allows multi-step CMP processes to be carried out on a single tool. The following tools are examples of rotary motion CMP tools with multiple wafer carriers: Strasbaugh Symphony® which has four carriers and three platens; SpeedFam-IPEC Auriga® which has five wafer carriers and two platens (for primary and secondary polishing, Fig. 15-25a); SpeedFam-IPEC AvantGard 776® which has four wafer carriers and four platens; Cybeq Nano Tech Isoplanar 8000® which has six wafer carriers; and Applied Materials Mirra® which has four wafer carriers and three platens (each with an in-situ pad conditioning arm [Fig. 15-25b]). In the Mirra®, a central spindle rotates the wafer carriers from one platen to the next. A three-platen configuration (such as is available on the Symphony® and the Mirra®) permits a three-step CMP process to be performed on a single tool.

CMP tools that use a pad that moves past the rotating wafer carrier in a linear (rather than rotary) motion have also been developed (*linear CMP tools*). These tools offer the potential benefits of higher linear speeds 400–500 ft/min (versus 34 ft/min in rotary CMP tools), harder pad materials, and reduced polishing pressure (i.e., because the polish rate can be maintained at a high rate with reduced applied pressure as a result of increased pad velocity). This set of conditions may be able to produce better planarization of oxides in a shorter time and reduce the problem of dishing in CMP (see Sect. 15.5.5.1). Linear CMP tools are offered by Lam Research (Teres® Fig. 15-26), Apex, and Doosan. The Teres® has two linear polishing belts, and one station with a rotating platen (for carrying out buffing/ cleaning steps). The wafer carriers move independently from one station to another. The Teres® also employs an *air bearing* beneath the

MULTI-LEVEL INTERCONNECTS FOR ULSI 753

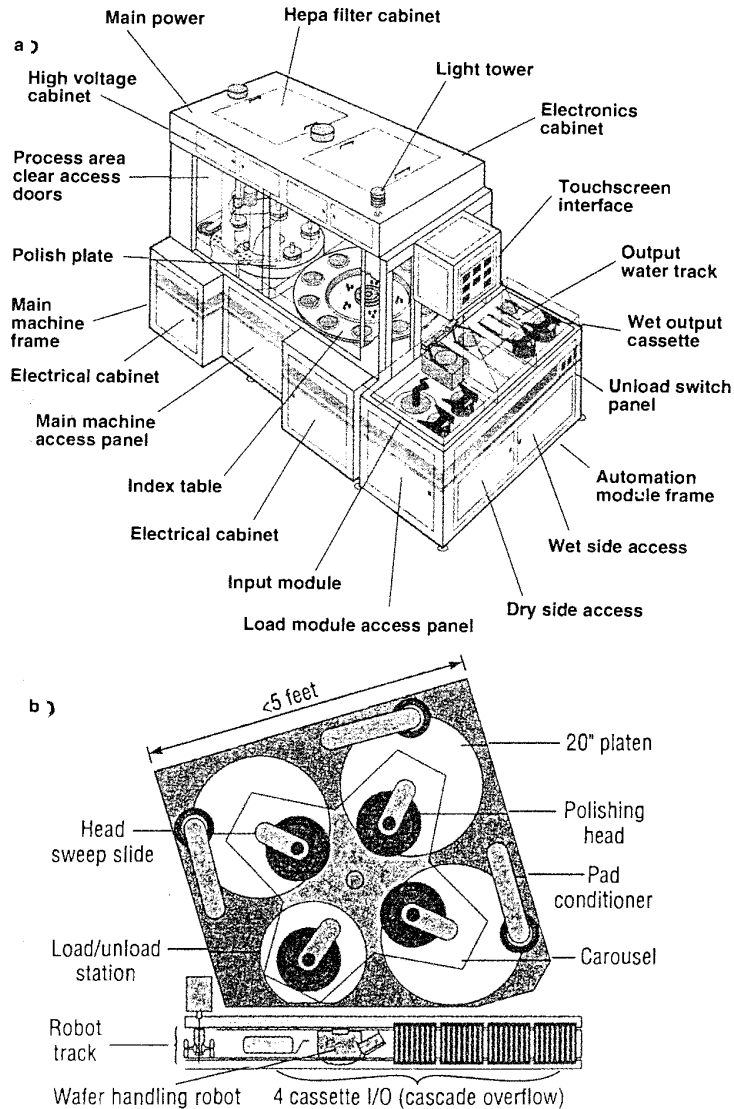


Fig. 15-25 (a) Detailed view of the SpeedFam-IPEC CMP-V® five-carrier polishing tool. Courtesy of SpeedFam-IPEC, Inc. (b) Architecture of the Mirra® three-platen CMP tool. Courtesy of Applied Materials, Inc.

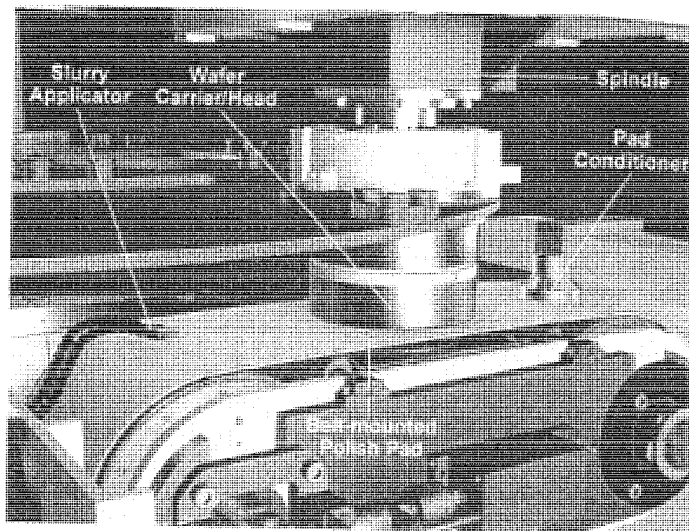


Fig. 15-26 Schematic drawing of the Lam Teres® linear CMP tool. Courtesy of Lam Research, Inc.

polishing belt. This allows the pressure distribution across the wafers to be profiled. At a given pressure and speed, adjustments to the pressure distribution profile can be made to significantly reduce film thickness non-uniformities (Sect. 15.5.4.1 and 15.5.4.2).

A tool with a stationary platen and a rotating wafer carrier was introduced by Obsidian Corp. in 1999. It employs a slurry-less pad supplied by the 3M Corp. (see Sect. 15.5.3.3). The pad material is mounted on a roller (akin to a roll of paper towels). After each wafer is polished, the roll is advanced to remove the used pad surface, making available a new segment of pad material to polish the next one.

15.5.3.2 CMP Consumables (Polishing Pads): Pads and slurries are the consumables of a CMP process. Historically, the polishing tools were considered to be the technology's major contributors. But consumables are now being recognized as being equally important. With a new group of consumables being made available, some customers select the slurries first, then review their compatibility with existing tools, and based on the results, select a polishing tool. The importance of CMP consumables is expected to increase with the advent of Cu-CMP and low-*k*-CMP. The general characteristics of polishing pads are described here, and those of slurries are considered in the next section.

Polishing pads are made of cast and sliced polyurethane (with filler), or polyurethane-impregnated felt (Fig. 15-27). While there is no consensus, it is generally agreed that the hardness and porosity of the pad are important parameters and that the pad plays a crucial (but still poorly understood) role in determining the final polish results. Polishing pads are supplied by a number of vendors, with Rodel Corp. having the largest market share. Other suppliers of

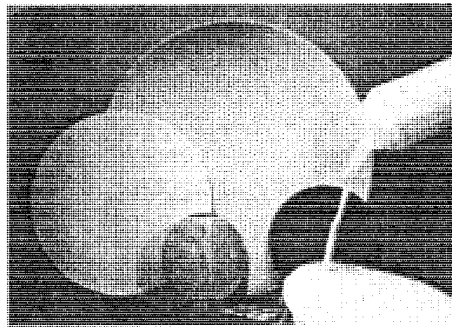


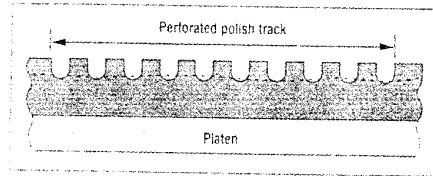
Fig. 15-27 Photograph of polishing pads. Courtesy of Rodel, Inc.

pads include Cabot, Fujimi, and Teijin. The Rodel Suba IV and Rodel IC-1400 are families of polyurethane pads with different hardness and porosity characteristics. The Rodel Politex pad made from a flexible felt-like material.

In many CMP applications two pads are simultaneously used, because a hard pad gives better local (within die) planarity, but a softer pad give better uniformity of material removal across the entire wafer. Using a combination of two pads provides a compromise between these extremes. For example a hard IC-1400 pad is mounted onto a softer SUBA IV pad to form a stacked pad. As shown in Figs. 15-27 and 15-28, the surface of the pad is either cut with concentric grooves or contains perforations 1-mm wide and 250- μ m deep punched into the pad. These grooves (or perforations) form channels to help transport the slurry between the pad and wafer. Pores present on the pad surface also aid in slurry transport. Pad life has been extended as CMP process understanding has improved. Early pads had to be changed about every 100 wafers, but now from 400 to 800 wafers can be polished before the pad is replaced.⁴²

One crucial aspect of CMP is that unless a pad is continuously conditioned, removal rates decrease rapidly with pad usage. For example, starting with a new pad, the removal rate can drop from 210 nm/min to 75 nm/min after only 50 wafers have been polished. This occurs because the surface of the pad rapidly glazes during the polishing process. The pores of the pad become closed, reducing slurry delivery to the wafer surface. A consistent (and rapid) removal rate is necessary for a production-worthy CMP process. This can be achieved by pad conditioning, which opens the pores of the pad by forming microscratches on the pad surface. (Figure 15-29 shows SEM micrographs of the pad surface with and without conditioning. The pores of the non-conditioned pad are completely clogged.⁴³) Pad conditioning is carried out by sweeping the pad surface with a wheel containing a diamond abrasive surface. Alternative conditioning abrasives have been pursued, since diamond particles may break off the conditioning wheel and become embedded in the pad. This can lead to severe scratching of the surfaces being polished. Recently, however, superior types of bonding mechanisms have been developed, which chemically bond the diamonds to the base of the conditioning pad. Diamonds

a)



b)

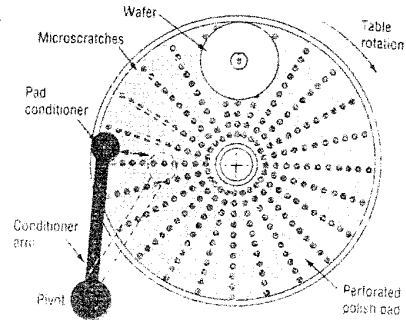
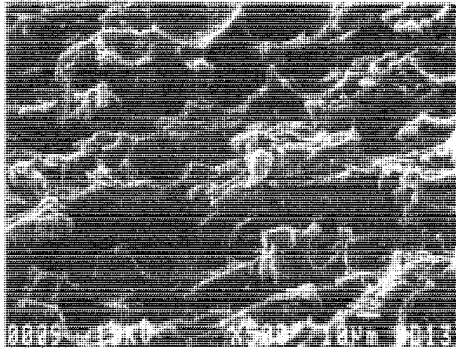


Fig. 15-28 (a) Polishing pad cross section. (b) Effect of pad conditioning.⁴³ Reprinted with permission of Solid State Technology. Published by PennWell.

adhere better to such pads than to those which employ a plating technique for bonding. Thus, diamond loss (and resultant microscratching), is now much less of a problem. Many polishers incorporate a continuous *in-situ* pad conditioning capability (i.e., a unit that moves across the pad during polishing, see Fig. 15-25b). This allows each wafer on the platen to be exposed to a pad with the same amount of conditioning, resulting in a more uniform removal rate from wafer to wafer.

New pads also require a break-in conditioning cycle to establish a constant removal rate. In this procedure a set number of dummy wafers are polished for the purpose of stabilizing the pad condition and performance. Break-in cycles can consume a significant fraction of the pad's useful life, and the polishing tool is not available for productive use while this is in progress.

a)



b)



Fig. 15-29 SEM of: (a) new polishing pad. (b) Polishing pad with a "glazed" surface.⁴³ Reprinted with permission of Solid State Technology. Published by PennWell.

15.5.3.3 CMP Consumables (Slurries): The goal for each CMP process is to find a slurry that produces high removal rates, good planarity, local film uniformity, and high selectivity. Also important are the ease with which the slurry can be cleaned from the wafer, the aggressiveness with which it attacks the polishing equipment, and the cost of disposing of slurry waste. Some of these issues have been discussed in Sect. 15.5.2, and others will be addressed Sect. 15.5.3.4. The remaining slurry topics are considered here.

Slurries consist of small, abrasive particles of specific size (typically 10–100 nm) and shape suspended in an aqueous solution. The abrasive particles have roughly the same hardness as the film being polished. Acids or bases are added to the solution, depending on the material to be polished. The slurry parameters that impact the polishing rate include the chemical composition of the slurry solution (and its pH) and the concentration (in wt %) of the solid particles in the slurry. Also important are the shape, size, and size distribution of the slurry particles. An automatic slurry feeding system is used to ensure uniform wetting of the polishing pad and the proper delivery of the slurry. Slurries are marketed by a number of suppliers, the largest of which is Cabot Corp. More slurry vendors are entering the market, including Rodel Inc., Fujimi Inc., Olin Microelectronic Materials, Praxair Surface Technologies, and EKC Technology, Inc.

Slurries for oxide CMP traditionally consisted of colloidal silica or fumed silica suspended in an alkaline solution (typically potassium hydroxide [KOH] or ammonium hydroxide [NH_4OH], diluted in DI water). The solids content of oxide CMP slurries is 10–15 wt%, and the pH is between 10.0–11.0. By controlling pH within this range, agglomeration is minimized. The polishing rate of oxide increases with increasing pH, with increasing slurry particle concentration, and with increasing particle size.

Slurries for metal CMP applications typically use alumina particles suspended in various liquids (depending on the specific metal being polished). As noted, by controlling the pH of the slurry, colloidal silica particles in oxide slurries can be kept well dispersed for relatively long periods of time (giving them longer shelf-lives). However, alumina-based slurries have traditionally exhibited the drawback of poor colloidal stability (Sect. 15.5.3.3). This causes slurry particle agglomerates to form. They produce unacceptable microscratches in the films being polished. Recently, improvements have been made in the colloidal stability of alumina based slurries.⁴² Nevertheless, development of slurries based on other particle types (e.g., cerium oxide and zirconium oxide particles) for CMP has been pursued.

Slurries tend to dry rapidly on a wafer once it leaves the wet environment of the polishing table. Once the slurry dries, it is very difficult (if not impossible) to remove all of its particles from the wafer surface. To prevent this from occurring, wafers are quickly placed into megasonic wet-cleaning tanks to remove the slurry before it dries. In fabs that do not use polishing tools with integrated cleaning systems, logistical difficulties may delay the transport of wafers from the polisher to the cleaning system.

Slurry waste disposal also must be addressed. Slurry always contains aggressive chemicals. In some cases these chemicals are also highly toxic. In addition, the material removed by a CMP process ends up in the expended slurry and rinse water. For example, copper is a known toxin in water. Therefore, fabs require waste water treatment systems to recycle the drain water of CMP processes. Similarly, as slurry volumes go up, large amounts of solid waste are generated (estimated to be 100,00–400,000 pounds per year per typical fab). A decrease in slurry

consumption by the CMP process will be needed to reduce this waste disposal burden. The task of slurry disposal is predicted to become the responsibility of the slurry manufacturer.⁴²

Slurry-less CMP processes have also been developed. A pad supplied by 3M embeds the abrasive material in the pad itself, and by 1998 this pad could be used for CMP of oxides in ILD and STI applications. Metal CMP pads were in development. These pads are composed of four layers: a microreplicated abrasive, a rigid layer, a resilient layer, and a self-adhesive backing. The use of this slurry-less pad in a CMP system was pioneered by Obsidian Corp. In Obsidian's tool, a roll of abrasive pad material is spooled across a platen and incremented after each wafer. No slurry or pad conditioning is required. However, one concern about the slurry-less approach is the uniformity of polishing performance from roll-to-roll.

15.5.3.4 Slurry Distribution Systems: Slurry management (i.e., mixing, storage, distribution, and dispensing) is another key issue of a CMP process. Reliable slurry distribution systems are an increasingly important component of the overall CMP equipment set. Slurry management is challenging for a number of reasons.⁴⁴ Precise mixing and consistent batch blends are necessary to achieve the repeatability demanded by process engineers. The slurry consists of particles in a liquid, and they must be kept suspended in it in a homogeneous state. However, these particles will agglomerate into clumps unless steps are taken to suppress this tendency. The slurry pH plays a role in the dispersion of silica particles. At a pH above 7.5 the silica particles attain sufficient surface charge to generate electrostatic repulsion and effectively disperse the slurry.⁴⁵ (Above a pH of 10.7, however, the particles dissolve and form silicate.) In addition, from the time the raw slurry is introduced into the distribution system it is also critical that a blanket of humidified air be provided over the slurry to prevent agglomeration in the dispense system. Finally, slurries for Cu and W-CMP require constant agitation to maintain particle suspension. Slurries have a finite shelf life (after which the agglomerate level becomes excessive). This shelf life ranges from less than one week to about one month. Filtering is therefore necessary to prevent any agglomerates (which might have formed during that time), from reaching the wafer being polished.

The distribution system must continuously provide slurry to a number of polishers in a sustained manufacturing production environment (i.e., 24 hours a day, seven days a week, and in high volumes). The pH and solids content of the slurry must be monitored in-line to ensure consistency. Furthermore, since slurries are abrasive particles contained in chemically aggressive liquids, they present a harsh environment to the components of the distribution system. That is, pumps, piping, valves, and measurement apparatus in contact with slurry can be stressed to the point of failure in a short period of time. Nevertheless, bulk slurry distribution systems have been designed to handle these conditions. They are available from such vendors as Mega Systems & Chemicals and FSI International.

15.5.3.5 Endpoint Detection: Since CMP is a process for reducing thickness at selected locations on a wafer, it is necessary to be able to determine when the process has reached its *endpoint*. This can be achieved by making measurements on the film to be polished before the CMP process, and again after the wafer has been polished for some pre-determined time (based on the pre-polish film measurements). This is termed an *in-line endpoint detection* process. Another

approach is to measure the film thickness as the polishing process is proceeding. This is called a *real-time, in-situ endpoint detection* process. Both techniques will be described here.

In *in-line endpoint detection* used for dielectric CMP, a pre-CMP measurement of the dielectric film to be polished is performed.⁴⁶ This measurement is made over an area containing dense device features to determine the thickness, T_0 , shown in Fig. 15-18. Based on the measured value of T_0 , the polishing time (t_{final}) needed to reduce the film to a thickness, T_{final} (see Fig. 15-18) is calculated. The value of t_{final} is chosen from CMP calibration data based on applied pressure, rotation speed, and average pad-life-degradation. This produces a value of T_{final} centered within the allowed process window. Various factors may alter the actual value of T_{final} realized at t_{final} (e.g., by as much as 20%, 1- σ , during process development). If the material removed by t_{final} exceeds the allowed limit, the wafer must either be scrapped or more dielectric material must be deposited. If the remaining thickness is excessive, the wafer can be returned to the polisher for more removal. In fact, a typical manufacturing approach to oxide-CMP in which in-line endpoint detection is employed, is to deliberately *underpolish* the SiO_2 layer. The film is then measured and polished a second time, using an adjusted polishing time. The cycle is repeated until the measured thickness is within specification limits.

This loop is the basic approach used in dielectric CMP (Fig. 15-30), but it has some drawbacks. The most significant are that the technique is time consuming, measurement intensive, and it reduces the overall time the polisher can be kept in use. In addition, there is a risk of wafer breakage when being removed for measurement. Some thickness metrology tools are designed to be placed inside the polishing tool housing as another subsystem (e.g., the ITM® tool of Nova Measuring Instruments). A central handling mechanism transports the wafer to the measurement position and then to a scrubber, back to polish or to the output cassette. Multiple points on the wafer can be measured with a small optical head. This head (rather than the wafer stage), is moved, providing about 5 times faster movement.

The use of a *real-time, in situ endpoint detection* method allows the problems of in-line endpoint to be circumvented. *In-situ* endpoint detection would permit the user to stop at the desired remaining thickness without having to repolish. This would also improve throughput and increase the process window. Several methods have been developed to perform *in-situ*

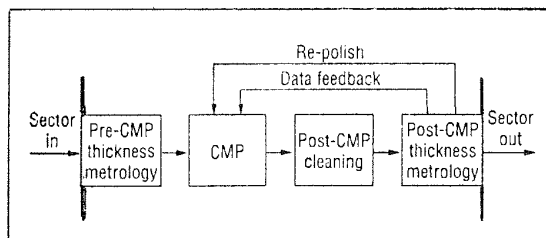


Fig. 15-30 Polishing, followed by measurement to confirm the thickness (and repolishing if necessary) is the cycle used in *in-line* endpoint detection.⁴⁶ Reprinted with permission of Solid State Technology.

endpoint detection in CMP, including: 1) front side laser interferometry or spectrometry, 2) optical sensing of multiple points on the backside of the wafer, and 3) monitoring the current supplied to the motor of the wafer carrier.

The *wafer frontside laser interferometry technique* is used by Applied Materials in their Mirra® tool, and the *spectrometer reflectivity technique* on the Lam Research Teres® tool and the Speedfam Auriga® tool. The name of the Applied Materials endpoint system is the *in-situ removal monitor (ISRM)*.^{47,48} Each polishing platen has a laser and a detector mounted on the polishing table. A window in the polishing pad allows the laser to be aimed at the top surface of the wafer, and the detector to sense the reflected light. As the laser/detector passes under the wafer, the laser light passes through the clear window in the polishing pad, and reflections are sampled. The Mirra® uses sophisticated signal processing/detection algorithms to reduce noise from the slurry and pad surface condition. In development tests, endpoint accuracies to within 20 nm were reportedly achieved. In W-CMP polishing, the ISRM system measures the reflectivity of the metal film and signals the transition from metal film to oxide. An over-polish time is specified to ensure complete removal of the liner film.

In the Lam Research Teres® tool, a *multiple-wavelength, spectrometry-based endpoint system* is employed. It can continuously monitor the thickness of transparent films as they are being polished. It can also measure the thickness of very thin metal films. Thus, during a Cu CMP process, the endpoint of the Cu polishing step can be sighted before all the Cu is removed. This allows a set of process conditions that rapidly removes Cu to be initially used. When the point of complete Cu removal is detected, the polish conditions can be switched to those that result in a slower Cu removal rate. This allows the Cu to be completely removed over the underlying Ta liner film without significantly eroding the exposed Cu in the trenches. In the Speedfam Auriga® multi-head polisher, a halogen lamp is used as the light source and diode arrays are employed to detect the light reflected from the surface of the wafer being polished. Endpoint in W-CMP is determined from an intensity change in the reflected light as the exposed surface changes from tungsten to oxide.

The *optical monitoring of the wafer backside* as an endpoint detection technique for monitoring the oxide-CMP process has been developed by Luxtron.⁴⁹ In this approach, the wafer's backside is illuminated with infrared radiation to which the silicon is substantially transparent. The radiation is therefore able to sense the film thickness on the wafer's front side as it is being polished.⁵⁰ By mounting several sensors onto the wafer carrier, multiple points on the wafer can be monitored. The film thickness is measured using the basic principle of *optical interferometry* (discussed in Chap. 8, Sect. 8.13.1) for measuring the thickness of dielectric films. For wafers with no metal on the front surface this technique is quite straightforward. If there is some metal on the front surface (e.g., in a multilevel-metal process), the situation is more complicated. The metal will block some of the incident light. Fortunately, there is generally sufficient space between device features to allow light to pass. Figure 15-31 shows a schematic of a single sensor probe. Two optical fibers connected to an infrared emitter and detector are mounted in a modified wafer carrier of a commercial CMP tool. Electrical communication from the sensors is accomplished via a wireless link. As many as four sensors can be mounted in a single carrier.

Fig
light

wc
fil
ox
cu
ro
m
ch
su
w
Si
cu
ve

11
re
co
cl
R
fo
th
a
n

s
b
a
c
s

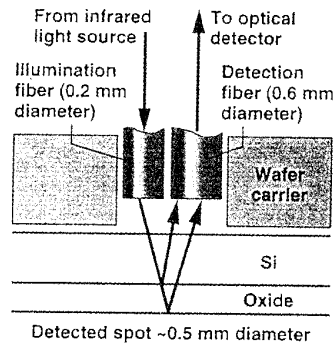


Fig. 15-31 Schematic of an optical sensor embedded in a wafer carrier.⁴⁹ Wafer is backlit with IR light to measure films on the front surface. Reprinted with permission of Semiconductor International.

The *motor-current-monitoring technique*⁴⁹ is a production-proven endpoint detection that works well when polishing a material down to a stop layer (such as the polishing a CVD-SiO₂ film on a silicon nitride stop layer in shallow-trench isolation,⁵¹ and the polishing of W over oxide in W-plug CMP processes⁴⁹). It is also relatively easy to implement. Circuitry such as a *current shunt* or a *Hall effect probe* is used to monitor the current supplied to the motor that rotates the wafer carrier. Since the wafer carrier is driven at a constant rotational speed to maintain a constant polishing rate, the drive current is varied to compensate for any load changes on the motor. This makes the motor current sensitive to frictional changes at the wafer surface. The largest changes occur when one material has been polished away, exposing a layer which has different polishing characteristics (e.g., that could function as a CMP-stop layer). Since there is no layer breakthrough in oxide-CMP (as can be seen in Fig. 15-18), the motor-current-monitoring technique is not appropriate for such applications. Luxtron and Apex are vendors of such motor-current endpoint controllers (e.g., the Luxtron Xinix Model 2450®).

15.5.3.6 Cleaning Issues in CMP: CMP is an inherently dirty process that can introduce yield-reducing defects to the wafer. Some early resistance to the adoption of CMP stemmed from contamination concerns. As a result, significant effort has been expended to develop post-polish cleaning techniques so wafers returned to the cleanroom after CMP are not contaminated.⁴⁰ Residual slurry introduces foreign particles, metal contaminants, and chemicals, all undesirable for the remaining steps of chip fabrication. Wafers are subjected to post-CMP cleaning steps that return the wafer surface to an acceptable cleanliness level. The main objective is to provide a wafer surface that is free of slurry particles, but it must also remove organic residues and trace metal ions, without introducing further defects.

The current technology of choice for both oxide and metal CMP uses double-sided brush scrubbing (Fig. 5-13 in Chap. 5), although megasonics immersion (Chap. 5, Sect. 5.6.1) has also been used.³⁴ Such scrubbing must function reliably with bases (e.g., KOH and NH₄OH) and acids (e.g., citric and HF) in order to optimally remove both metals and mobile ions. Adhesion of particles to the wafer depends on the surfactant properties of the polish. For oxide polishing, silica-based slurries such as NH₄OH and KOH decrease silica particle adhesion (as long as the

wafer is kept wet). Alumina-bearing slurries used for tungsten polishing more readily adhere to the wafer surface and are generally more difficult to remove. To prevent additional wafer and slurry contamination, the number of airborne particles can be minimized by filtering the air fed to the CMP tool with the same HEPA filtering techniques used for the cleanroom environment. The cleanliness of the DI water used to rinse or submerge the wafers is also an issue. Tubs or baths in which wafers are kept wet must be continuously refreshed with new DI water. The CMP tool should automatically rinse all surfaces that come in contact with slurry, including the polishing table, drain tubs, carriers, and pad conditioner. These areas should also be cleaned periodically to prevent dried slurry buildup. As mentioned previously, slurry is composed of very small abrasive particles, and filtering by conventional methods is not easy. Nevertheless, such filtering is needed because large agglomerates can develop within the slurry and have serious device defect consequences.

Wafer rinsing should be done while the wafers are still on the pad. Spray rinsing should be done as soon as the wafers are lifted from the pad. If wafers are exposed to air for even one or two minutes before such spraying (e.g., while waiting to be unloaded), subsequent cleaning becomes difficult (and perhaps impossible). Immediate removal of metal-polish slurries is also critical, because these can continue to react and remove material even after polishing has ceased.

Double-sided scrubbing follows this rinsing step (although in many sequences a final buffing step is used to remove residual slurry particles and such defects as small scratches). In some polishing tools the cleaners and scrubbers are integrated into the tool design. In others they are stand-alone units. Wafers are loaded into wet cassettes at the polishing tool's output module. In non-integrated systems they are then manually transferred to one or more stations where they are cleaned, dried, and then sent back to the clean room for subsequent processing. On-Trak Systems (a division of Lam Research), supplies stand-alone double-sided scrubbing units for the cleaning step, as well as does Dainippon Screen. Other LAM post-CMP cleaning tools (e.g., the Synergy Integra® model, Fig. 15-32), are designed to be integrated with the polishing tools of other CMP vendors (including Applied Materials, SpeedFam, and Strasbaugh). Lams's share of the post-CMP cleaning tool market was about 70% in 1999.

However, the Mirra® can also be integrated with an Applied Materials cleaning tool (the Mesa,® Fig. 15-33).⁵³ It consists of four separate process modules: a megasonics bath, two-double-sided brush-scrub modules, and a spin-rinse-dry module. Each module cleans both sides of the wafer, and all handling is done by edge grip.

The scrubbing brushes are made of a soft polyvinyl alcohol (PVA) material that is brought into direct contact with the wafer surface. The PVA material has an open structure consisting of interconnecting cells that allow the brush to be constantly flushed with DI water during wafer cleaning. The soft brushes clean efficiently when compressed against the wafer. PVA brushes are compatible with the basic liquids used in oxide slurries (NH_4OH and KOH , with pH of ~ 11). They are widely used for post oxide-CMP cleaning. When using PVA brush scrubbing to clean wafers after oxide CMP, it has been reported that adding NH_4OH to the DI water used while scrubbing reduces residual particles on the wafer. The oxide surface, the PVA material, and the silica slurry particles all have negative zeta potential at the neutral pH of DI water.

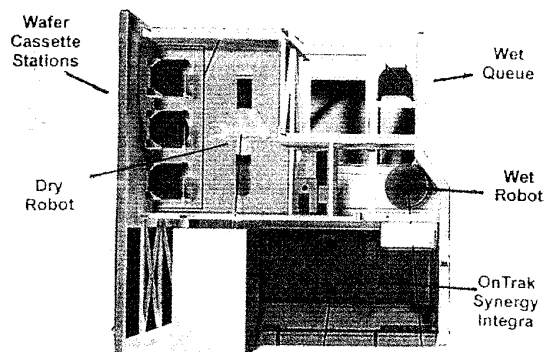


Fig. 15-32 Schematic drawing of the Lam Research Synergy® post-CMP cleaning tool. Courtesy of Lam Research, Inc.

Mechanical energy supplied by the brush is high enough to remove all the slurry from the wafer surface.

In post W-CMP cleaning, brush scrubbing is also effective. In this application, the wafers come from an acidic environment (2–4 pH) to a water rinse and a PVA brush clean. Some PVA materials cannot endure this chemistry well and must be frequently replaced. Other PVA materials (believed to be free of unreacted residues that hydrolyze in the acidic environment) appear to render them stable over an operating pH range of 2–12 pH. These materials are reportedly compatible with post W-CMP scrubbing.⁵⁴ It is reported that rinsing the wafers with dilute NH_4OH during post W-CMP scrubbing is most effective in removing alumina slurry particles.⁵⁵ An additional benefit of using NH_4OH is that it prevents loading of the PVA brushes by alumina particles or tungsten oxides.

Cleaning wafers after copper CMP presents a unique set of challenges. Not only must particles and chemicals from the slurry be removed, but also the Cu being polished must not be

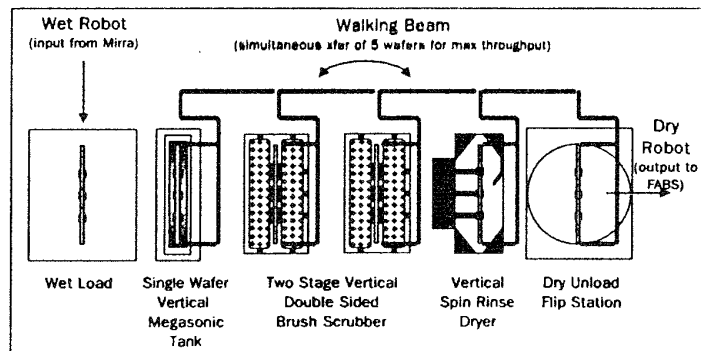


Fig. 15-33 Mirra Mesa® cleaning module layout. Courtesy of Applied Materials, Inc.

allowed to enter the fab cleanroom. It must be removed from the front, back, and beveled edge of the wafer. The copper oxides formed and removed as part of the CMP process do not dissolve in DI water and may be easily transferred to the PVA brushes. If the brushes become "loaded" with these copper oxides, they may transfer them to subsequently processed wafers, thus producing severe Cu contamination in the fab. Proprietary chemicals have been developed to dissolve the copper oxides and prevent such brush loading.⁵⁶ In the On-Trak Systems brush scrubbing tool, such a chemical is delivered to the wafer surface through the porous core of the brush itself (Fig. 15-34).

CMP processes are generally located adjacent to the cleanroom, connected through a bulkhead load, similar to etchers or CVD systems. In some cases, the CMP tool configuration consists of a polisher and a stand-alone cleaning system with a robot that transfers wafers between them. A problem with this configuration being adjacent to the cleanroom is the potential for water and slurry spillage (and then leakage into the cleanroom) as the transfers are made. A better solution is a *dry-in/dry-out CMP system*, in which the polisher and cleaning system are completely integrated. Wafers are planarized, cleaned, dried, measured, and returned back to the cassette located in the cleanroom.

15.5.3.7 CMP Metrology: There are a number of metrology requirements in CMP. First, it must be determined when the CMP process is completed (*endpoint detection*, Sect. 15.5.3.4). Second, the thickness of the films after CMP must be measured, including thickness uniformity across a wafer. Third, the degree of film planarity must be measured after CMP. Finally, films subjected to the CMP process must be inspected for defects and damage. The latter two CMP metrology topics are discussed here.

Film thickness after CMP (and before CMP in in-line endpoint detection), is performed with a number of different techniques, including: spectrophotometry, ellipsometry, and beam-profile reflectometry. One problem associated with these techniques is that of *cycle skipping* (or *order skipping*). Since the signal obtained from the thickness measurement is periodic, the same

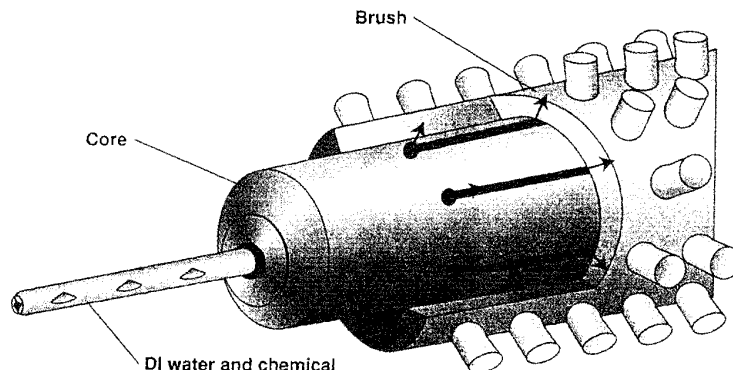


Fig. 15-34 Through-the-brush chemical delivery. Courtesy of Lam Research, Inc.

measured data can fit more than one thickness value. For oxide, the resulting errors are multiples of about 240 nm, enough to lead to incorrect conclusions about the post-CMP oxide thickness (which might cause properly processed wafers to be scrapped). Different strategies by each measurement technique are employed to avoid such problems. Ellipsometers use additional wavelengths, spectrophotometers rely on better signal-to-noise data collection, and beam-profile reflectometry is combined with spectrophotometry to avoid cycle skipping errors.

One factor that impacts *thickness uniformity after CMP* is the uniformity of the wafer itself. It is observed that the *within wafer nonuniformity (WIWNU)* of the film removed by a CMP process increases linearly with an increase in the total thickness variation, TTV (Chap. 1, Sect. 1.6.2) of the wafer.⁵⁷ In addition, wafers may not retain their initial flatness as they undergo the various steps of the process sequence. In particular, stress arising from thin film deposition processes can deform a wafer into a concave shape (negative bow) or a convex shape (positive bow, Fig. 4-7, Chap. 4). Although bow is reduced by the pressure applied to the wafer during CMP, some remains (as the degree of bow can vary from wafer to wafer). In addition some residual local warpage persists after the wafer has been pressed against the wafer platen. Consequently, the CMP process may polish the edges faster than the center in a concave wafer. Wafers with lower pre-polish bow reportedly had a smaller WIWNU than wafers with higher pre-polish bow. As a result of the above issues, measurement of the wafer TTV as received from the wafer suppliers has been recommended.⁵⁷ Some CMP users have set an empirical TTV cut-off value of 5 μm for 200-mm wafer pre-polish sorting. Capacitive gauging tools are used to perform such TTV and bow measurements. More than 16,000 data points can be collected on a 200-mm wafer in about one minute, and the results plotted to give a high-resolution map of the wafer geometry. Such gauging tools are available from ADE Corp.

The *degree of planarity* achieved with a CMP process is typically measured on a single die or a few dies on a wafer using either profilometry or atomic force microscopy. Such planarity can be location and pattern dependent. Within-die planarity is influenced by the tendency of CMP to polish smaller, individual features faster, and larger and densely packed features more slowly. Thus, CMP thickness measurements made with the optical thickness tools commonly include several features at each selected site (from the smallest that the tool can conveniently measure, to a large [field] area where the polishing rate reaches its lowest limit).

Detection of damage, contamination, and particles on post-CMP wafers is another aspect of CMP metrology.⁵⁸ The acceptable density of defects is becoming smaller for each device generation. In 0.18- μm CMOS technology the SIA roadmap calls for a defect density of 0.14 defects/ cm^2 , and the minimum detectable particle size is decreased to 0.06 μm from 0.08 μm for the 0.25 μm generation. In oxide CMP, defects can be separated into two categories: 1) *foreign material* such as residual slurry, surface particles, embedded particles, and residual metal; and 2) *voids in the surface material* (such as microscratches and dishing). *Microscratches* can occur during CMP when a particle is dragged across the wafer surface, causing a shallow void in the oxide layer. In metal CMP, defect types include puddles (which can cause shorts if they remain between metal lines), surface voids, residual slurry, surface particles, metal-filled microscratches, and plug holes from which metal has been removed.

The detection signals caused by such defects are obscured to some extent by noise sources associated with the CMP process. The three prevalent noise sources that impact defect

inspection at post-CMP processing are: 1) layer thickness variation, 2) grain noise, and 3) pattern noise. *SiO₂ layer thickness variation* is caused by nonuniform polishing across the wafer. *Grain noise* arises from the surface roughness caused by the polycrystalline structure of materials beneath the oxide being polished (e.g., tungsten, aluminum, and polycrystalline Si). The noise occurs from randomly distributed pyramidal peaks at the surface of the polycrystalline films. These scatter the detection signal in a non-consistent manner so it is not completely subtracted when two adjacent die are compared. *Pattern noise* arises from metal features beneath the oxide which distort the incoming signal by causing background scattering. Defect-detection tools employ image-processing algorithms to limit false alarms caused by such noise (i.e., that would indicate the presence of a defect where none actually exists).

Two illumination modes (*bright-field* and *dark-field*, Chap. 13, Sect. 13.6.1) are commonly used in the laser scattering defect counters employed for in-line automated defect inspection (such as the KLA-Tencor Surfscan AIT® model). Since reflected light is used in *bright-field inspection*, one of the main challenges for CMP metrology is to reduce the interference between reflection at the oxide surface and reflections that arise from underlying features. Some inspection tools use an ultra-broadband illumination source (in place of a monochromatic source) to suppress such interference effects. In *dark-field systems* only the scattered light is collected, while the spectrally reflected light is rejected. Dark-field systems with a low angle of illumination appear to work best for detecting slurry particles on an oxide layer over a grainy sub-layer. Dark field systems with optical repetitive-pattern-filtering are reportedly more effective for detecting defects over dense repetitive arrays.⁵⁸

Some 0.18 μm device inspections for detecting CMP defects may require use of SEM technology (such as the KLA-Tencor 2710). Because of their high resolution and large depth of focus they can locate ultra-small defects hidden between dense geometries or in high-aspect ratio areas. Small CuO particles $\sim 0.1 \mu\text{m}$ from a dual damascene process are easily detected using a SEM.

15.5.3.8 CMP Polisher Tool Reliability: The CMP process creates a harsh set of working conditions for the process tools. Some have compared the operation of a CMP polisher to "flying a helicopter in a sandstorm."⁴⁸ In the early days of CMP, production polishing tools could have a mean-time-between-failure (MTBF) as small as a few days. Some of the problems of early CMP tools included the following:⁴⁵ insufficient vacuum for wafer handling, wafer sensor failure, misalignment of load cup and sender and receiver cassettes, wafer release on polish pads (resulting in wafer breakage), equipment hangup (causing slurry to dry on the wafer surface), and leakage of slurry and de-ionized water because of lack of quality control on valves and rotating unions. To make the systems more reliable, tools were designed to protect moving mechanisms from the abrasive and corrosive slurries (including sealing of parts and the use of air and water labyrinths). Exposed materials were made with corrosion-resistant materials. Improved robotics for wafer loading, unloading, and handling were implemented. New system components were subjected to accelerated testing to simulate one-year of production in three months. As a result, MTBF rates in CMP tools have been continuously improving.

15.5.3.9 CMP Systems and Process Integration: The CMP process requires a host of tools and consumables (Fig. 15-35). At the outset of the proliferation of CMP throughout the IC industry,

each
per
for
for
off
pro
col
toc
vel
La
de:
cle
Ap

Sp
ac
Ap
CM
1E
15
ex
thi

FI
C

each of the components of the CMP process was supplied by individual companies. In that time period the burden of integrating these various components fell to the user. As the overall market for CMP tools and consumables grew, the equipment and materials suppliers realized the need for a better CMP infrastructure. They understood that to gain market share, they would need to offer CMP users a kind of off-the-shelf, "one-stop shopping" package, perhaps with guaranteed process results. One way to accomplish this was to form strategic partnerships with other (non-competitive) suppliers. Most polisher suppliers included slurry distribution systems, cleaning tools, pads, slurries, endpoint detection, and film thickness measurement tools from other vendors as available options compatible with their polishing tools. Other companies (including Lam Research, Applied Materials, SpeedFam-IPEC, Ebara, Strasbaugh, and Peter Wolters) designed CMP systems as cluster tools. Such systems integrate polishing, endpoint detection, cleaning (and perhaps inspection), into a single system. The Lam Research Teres® and the Applied Materials Mirra Mesa® are examples of such cluster CMP tools.

These integration trends have been accompanied by business consolidations as well. Speedfam and IPEC (two large polishing tool suppliers), merged in 1999. Lam Research acquired On-Track Systems (a maker of polishing tools and post-CMP cleaning systems). Applied Materials announced it would acquire Obsidian, Inc., a supplier of slurry-less pads for CMP. As CMP technology matures it is predicted that such consolidations will continue.⁵⁹

15.5.4 Miscellaneous Problems of CMP

15.5.4.1 Dishing: *Dishing*, another CMP problem, involves the thinning of regions that remain exposed when a film is being polished down to a CMP-stop layer (Fig. 15-35). One case where this occurs is in the CMP of a CVD-oxide layer, deposited to fill trenches in the course of

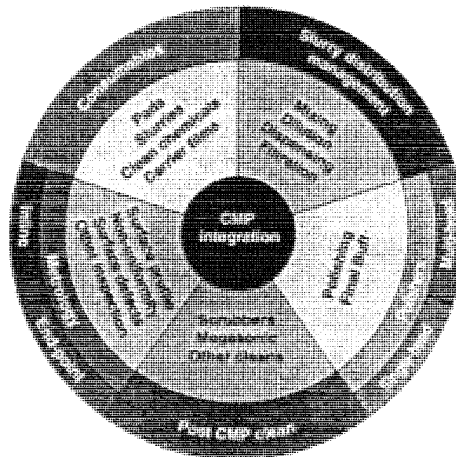


Fig. 15-35 Many components of CMP must be successfully integrated to develop a production-worthy CMP process.⁶⁴ Reprinted with permission of Semiconductor International.

fabricating shallow trench isolation (STI) structures. A silicon nitride layer is typically used in the trench-filling and polishing sequence to act as a CMP-stop layer. (Section 16.3.2, Chap. 16 of this book, and Sect. 6.5 of Vol. 3, give details on the process sequence used to form STI structures.) The thinning of these exposed areas (dishing) is known to be highly dependent on feature size, with the dishing effect become larger as the width of the field oxide is increased (Fig. 15-36). For example, in the CMP process of Fig. 15-37, dishing causes the field oxide thickness to decrease by more than 50 nm when the field oxide width exceeds 100 μm .⁶⁰ Dishing occurs because the polishing pad is not perfectly rigid. If a pad exhibited perfect rigidity, CMP of the exposed oxide regions would produce a perfectly planar surface (because the exposed recessed regions would not be subjected to any polishing pressure). Since actual pads exhibit some flexibility, above-zero polishing pressures are produced in recessed areas, especially in the middle of wide exposed regions. Thus, oxide is removed from these regions, and a "dished" region exists after planarization.

Dishing must be minimized to maintain planarity for subsequent wafer processing. It has remained a serious impediment to the development of Damascene processes (see Chap. 16). A number of approaches have been proposed to decrease dishing, including: 1) a combination CMP/RIE process; 2) dummy active regions; 3) using a linear polishing tool; 4) using a rotating polishing tool with a harder polishing pad and an efficient endpoint detection system.

In the *CMP/RIE approach*, the oxide is polished by CMP to a point where the nitride is not yet exposed, and the surface is thus still completely planar (no dishing has yet occurred). Then RIE is used to etch the oxide layer, with an endpoint algorithm used to stop the etching just as the nitride is exposed to plasma.⁶¹ The drawback of this method is that increases the complexity of the process used to planarize the trench oxide.

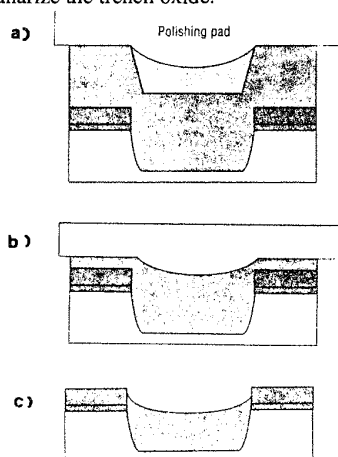


Fig. 15-36 Origin of the dishing effect in CMP. (a) polishing pad flexes into the trench opening; (b) formation of a concave surface; (c) trench area remains "dished" after planarization. Reprinted with permission of Solid State Technology. Published by PennWell.

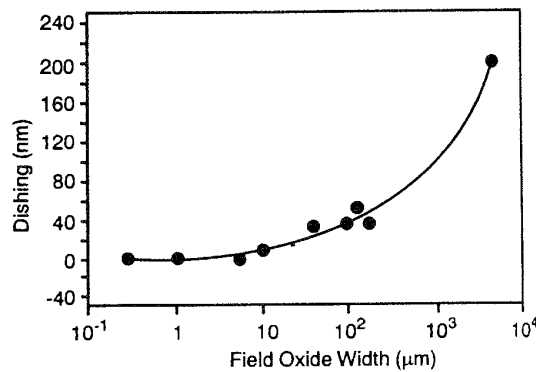


Fig. 15-37 Dishing versus field-oxide-region width for a specific oxide CMP process⁶⁰ (©1993 IEEE).

The *dummy-active region approach*, some areas of the trench are not etched. Instead, as shown in Fig. 15-38, regions of Si remain in the center of trench. This essentially reduces trench width so that dishing effects are decreased to an acceptable degree. The CMP-removal rate over regions that are densely packed with active areas is also slower than over regions containing small isolated areas. Thus, to expose the nitride CMP-stop layer over such densely packed regions, the nitride over small isolated active area gets stripped, causing damage to Si in areas beneath the stripped nitride. Thus, dummy active regions must also be added to areas around isolated active regions. Finally, since dishing also occurs when wide areas of metal are polished (e.g., bond pads, power busses, and wide interconnect lines), such dummy structures can also be added to these features. The technique, however, poses electrical, design, and space constraints, with the main issue being that several times as many structures must be added within a die over

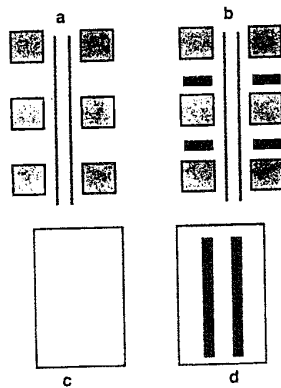


Fig. 15-38 Dummy (b) and dishing support features of small (a) and large wafer (c) features can prevent dishing.³¹ Reprinted with permission of Semiconductor International.

previous designs. This complicates the design and mask making process, and significantly raises defect levels. Thus, some see this approach as merely a stop-gap measure.

The use of *linear polishing tools* to suppress dishing offers the benefit of avoiding the drawbacks of the first two methods. Linear polishers employ higher pad velocities (300–500 ft/min versus 34 rpm for rotary pad speeds) and lower polishing pressures (3–5 psi versus 7 psi for rotary polishers). A highly selective slurry (with cerium oxide particles dispersed in an aqueous suspension of material that inhibits silicon nitride removal) is also employed. This set of process conditions significantly reduces dishing effects (maximum oxide removal across a die was 20 nm) and decreases the nitride removal over isolated active regions.⁶² It is reported that dishing can also be decreased in *rotary polishing tools* by selecting the correct set of process conditions and CMP consumables. Dishing effects are reduced by using a harder polishing pad, a relatively non-selective slurry, and efficient endpoint detection method, and an optimized CMP process.⁶³

15.5.4.2 Thickness Non-Uniformity Within a Wafer After CMP: Rigid polishing pads provide good planarity within a die (chip), but poor thickness uniformity across a wafer. The reason is explained here. If wafers were perfectly flat, this problem would not occur. A film on the surface of a flat wafer would be polished to a uniform thickness across the wafer (Fig. 15-39a). But not all wafers are equally flat. For example, if a wafer is tapered, and the pad is rigid, the film on the thicker end of the wafer will be polished more than on the thinner end (Fig. 15-39b). If the wafer thickness undulates (with some regions thicker than others), a film covering the thicker wafer regions will be polished more (and will end up thinner) than the film on the thin parts of the wafer (Fig. 15-39c). Softer (i.e., more compliant) pads produce improved across the wafer thickness uniformity, but at the price of degraded planarity. As discussed earlier, one solution is to use a composite pad (a rigid pad backed by a softer one). Another approach is to use a linear polishing tool with a flexible (e.g., air) bearing beneath the polishing belt. This allows the belt to conform better to the wafer surface, and reportedly produces improved across the wafer thickness uniformity (and significantly less dishing).

15.5.4.3 Economic Considerations: Throughput and Cost of Ownership: While CMP is an enabling process for deep submicron IC fabrication, it is also a relatively expensive one. The cost of a single CMP pass in 1999 ranged from \$6 to \$15, depending on the material polished. The main contributors to this cost include the polishing (and cleaning) tools, the slurry, and the CMP fabrication facilities. The price of a fully-integrated polishing tool in 1999 ranged from \$2–2.5 million, with the cleaning module costing about \$500,000. These tools are capable of polishing 50–60 200-mm wafers per hour. Slurries are also expensive. Oxide slurries cost about \$10/gallon, tungsten slurries about \$30/gallon, and Cu slurries from \$40–50/gallon. One gallon is consumed for about every ten wafers. The CMP facilities, labor costs, utilities, DI water, pads, and waste disposal are additional expenses.

15.6 METAL DEPOSITION AND VIA FILLING

In this section, methods of filling contact holes and vias are discussed. First conventional approaches are described. Then more advanced techniques are considered. These include tungsten plug formation by CVD and novel sputtering processes.

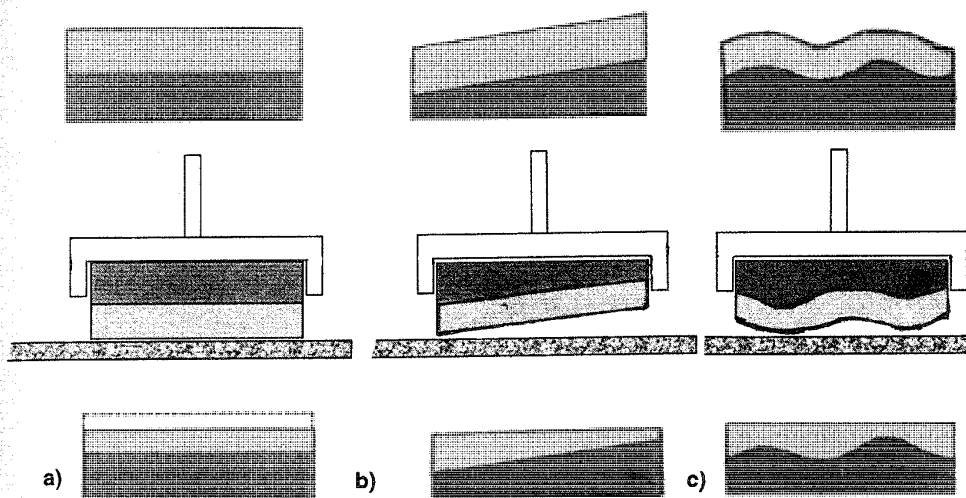


Fig. 15-39 If a rigid pad is used for CMP: (a) Film on flat wafer is uniformly polished; (b) Film on a tapered wafer is polished thinner over the thicker wafer end; (c) For wafers with variations in wafer thickness across the wafer, the film is polished thinner over the thick wafer regions.

15.6.1 Conventional Approach to Contact and Via Fabrication and Formation of Metal-to-Metal Contacts through Vias

One aspect of minimum feature size that applies to via fabrication is the issue of the contact hole and via sidewalls: for the same contact-opening sizes at the bottom of contacts and vias, straight sidewalled holes would require less area than those with sloped sidewalls. Furthermore, when dry etching is used to open the contacts and vias, it is often easier to produce straight rather than sloped sidewalls.

Unfortunately, when a conventional sputtering process is used to deposit metal into a hole, straight sidewalls result in worse step coverage by the metal sloped sidewalls. This has been shown both theoretically⁶⁵ and experimentally (Fig. 15-40a).⁶⁶ While adequate step coverage of the metal is critical, there is no single value of minimum step coverage that is considered adequate for all conditions. This is because the minimum value depends on many factors (including the current density passing through the metal, its absolute film thickness at the thinnest spot, the type of metal, etc.). In general, a range of 20–50% minimum step coverage is acceptable, depending on the application.

The step coverage of a *sputtered metal film* can be improved in a number of ways. The first is to optimize the deposition conditions by increasing the surface-migration ability of the atoms that have already arrived on the surface. These atoms will then rearrange themselves until the film has a uniform thickness over all topographical features. Surface mobility can be increased by raising the substrate temperature. Sputtering processes during which the wafers are heated

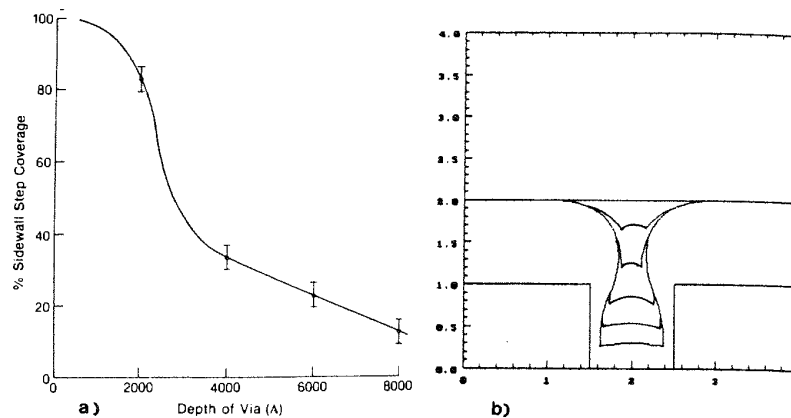


Fig. 15-40 (a) Percent sidewall step coverage of a standard dc sputtered aluminum film (0.8 μm thick) into vertically-sided 1.2 μm diameter vias of different depths.⁶⁶ (© 1986). Reprinted with permission of the AIOP. (b) Simulation of a sputtered Al film into a via with an aspect ratio of 1.0, with the via being partially filled and unfilled. The profiles from the top represent the metal film deposited into a 100% filled, 80% filled, 60% filled, 40% filled, 20% filled and unfilled via.⁶⁷ This paper was originally presented at the Fall 1988 Meeting of The Electrochemical Society, Inc. held in Chicago, IL.

have been shown to produce improved step coverage. Assuming that a sputtering process is developed that optimizes step coverage, this coverage will still depend on the sidewall slope and aspect ratio of the contact hole or via.

The aspect-ratio dependence of the step coverage into contact holes and vias is critical as feature sizes are scaled into the submicron regime. In order for the interlevel capacitance of the interconnections to the underlying and overlying metallization lines to be maintained (or even reduced), the thickness of the dielectric layers in which the contact holes and vias are formed may not be decreased at all. Otherwise, the aspect ratio will grow larger. For example, when an Al film is sputter deposited at < 300°C, the step coverage into a vertically-sided via whose aspect ratio is 0.5 is typically ~20%. As the aspect ratio increases to 1.0 (so that the via depth equals the width), the step coverage decreases to less than 5%.

An example of a simulation of a sputtered Al film into a via with an aspect ratio of 1.0 is shown in Fig. 15-40b. As observed in Fig. 15-41, sputtered-Al depositions used in production in 1988 provided more than 50% step coverage into vertically sided vias only if the aspect ratio was smaller than 0.25 (e.g., for 1.2-μm-wide vias, the depth can be no more than 0.3 μm). Since the thicknesses of the PMD and IMD layers generally exceed 0.5 μm each, adequate step coverage using PVD may not be possible for small vertically-sided vias. The conclusion to be drawn is that in conventional via processing, sloped vias are necessary to ensure adequate step coverage in most applications when contact holes and vias are less than 1.5 μm wide.

Two other important aspects of conventional via processing involve: 1) the planarization of the intermetal-dielectric films, 2) the decision of whether the design rules allow *stacked vias*

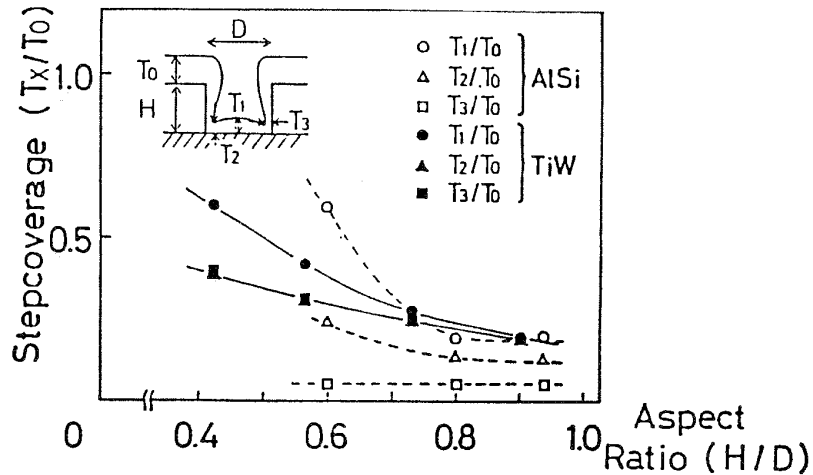


Fig. 15-41 Comparison of the step coverage of Ti:W ($0.6 \mu\text{m}$ thick) and Al:Si ($0.9 \mu\text{m}$ thick) deposited into a $1.2 \mu\text{m}$ -wide vertically sided, $0.9 \mu\text{m}$ -deep contact hole.¹³⁸

(i.e., vias located on top of silicon and polysilicon contacts, Fig. 15-42b), or only *staggered vias* (Fig. 15-42a). If complete planarization of the IMD is achieved, and if stacked vias are also permissible, via depths can vary by as much as $1 \mu\text{m}$ (Fig. 15-7). Such large variations have two significant impacts on the via fabrication process. First, the deepest vias will have poor (and generally inadequate) step coverage unless sloped sidewalls are formed. Second, since a sloped-

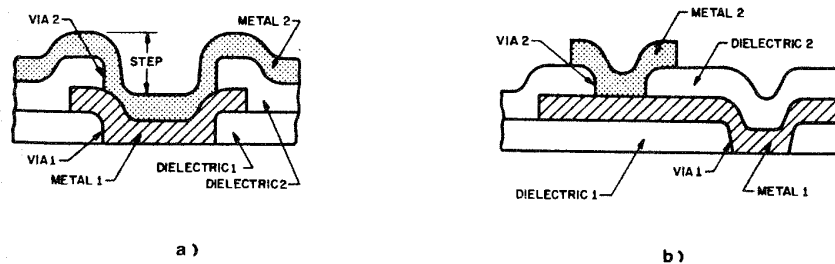


Fig. 15-42 Contact schemes: (a) stacked vias; and (b) staggered vias. From A. G. Sabnis, in G. S. Gilenblat and S. S. Cohen, Eds., *VLSI Electronics: Microstructure Science*, Vol. 15, Metallization, Chap. 7. Copyright 1987, Academic Press. Reprinted with permission.

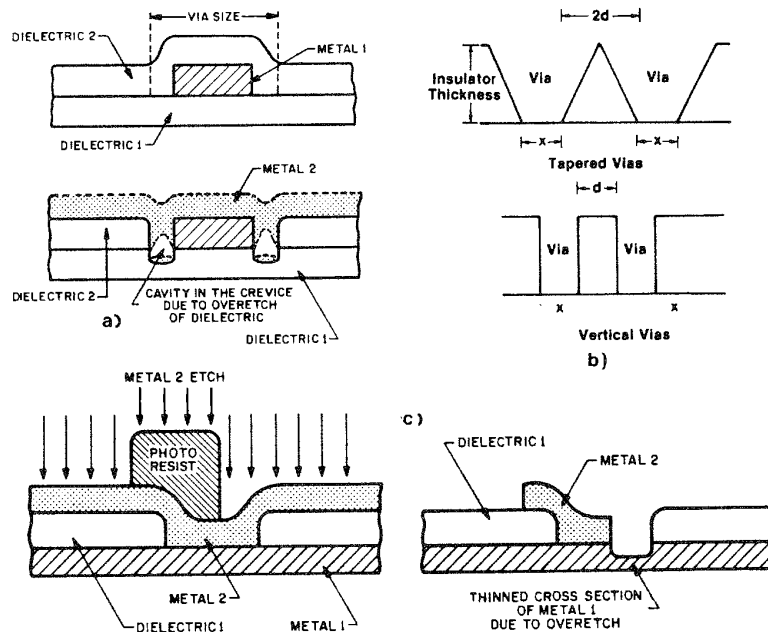


Fig. 15-43 (a) If the via dimension in the direction of the Metal-2 current flow is larger than the Metal-1 width, then crevices may be formed into the dielectric below Metal 1, and cavities can result in Metal 2. (b) If sloped vias are used, the minimum spacing between vias must be larger than if vertical vias are used. (c) If the via dimension in the direction of Metal-1 current is larger than the Metal-2 width, then Metal-1 can get overetched, thus the overlying metal must be larger than the top of the via opening. From A.G. Sabnis, in G.S. Gilenblat Ed., *VLSI Electronics: Microstructure Science*, Vol. 15, Metallization, Chap. 7. © 1987, Academic Press. Reprinted with permission.

via process must be used, the Metal-1 patterns under the shallowest vias must be wide enough to prevent the etching of grooves along the pattern sidewalls when the deep vias are being opened (Fig. 15-43a). This increase in pattern size results in an area penalty that can significantly decrease packing density.

As noted in Sect. 15-4, such conventional via processing was employed in double-level-metal (DLM) processes for CMOS. The rules that had to be followed because of the difficulty in filling vertical vias included:

1. No step-height reduction planarization process could be used with respect to PMD and IMD layers (i.e., only the first degree of planarization - smoothing - was employed), so that all vias could be maintained at the same depth.

2. Stacking of vias on top of contacts (or other vias) was not allowed.
3. A connection to Metal-2 was allowed only through Metal-1.
4. The vias had to be sloped to allow adequate step coverage by sputter-deposited metal.

However, vias with sloped sidewalls give rise to a new set of design rules that must be followed when laying out patterns of the metal and via structures in a multilevel metal system.

In the ideal case, metal pitch is determined by the minimum line and space dimensions that can be patterned using the most recent advances in lithographic techniques. In practice, metal pitch is also limited by the via size and the underlying metal pad size. The following design rules must thus be obeyed:

1. The underlying metal pad must be larger than the via opening. This condition is referred to as a *framed* or *nested* via. If the via were larger than the metal pad, grooves would be etched alongside the pad during the via-etch step (Fig. 15-43a) and could cause microcracking or thinning of Metal-2 when it is deposited over the via to contact Metal-1. The minimum dimension by which the metal pad must frame the via is dependent on the misalignment tolerances of the lithography step. If vias with varying depths must be etched, the enlargement that will occur when shallow-sloped vias continue to be etched during etching of the deep vias must also be taken into consideration.
2. The slope of the vias must be taken into account when deciding on the minimum spacing between vias (Fig. 15-43b). A minimum space is needed between the tops of the etched via openings so that metal space on this horizontal surface can be reproducibly patterned.
3. The overlying metal pad must be larger than the top of the via opening to ensure full coverage of the via (Fig. 15-43c).

15.6.2 Advanced Via Processing (Vertical Vias and Complete Filling of Vias by Metal)

As long as an adequate sidewall taper of contact windows and vias can be implemented (and the aspect ratios do not become too great), sputter deposition can produce metal films with adequate step coverage into vias. However, design rules have evolved to where sputtered metals can no longer acceptably fill contact and via holes. That is, it became necessary to be able to completely fill contact holes and vias with metal. In submicron technologies the extra area needed to accommodate sloped vias severely limits the maximum packing density (particularly since it is also difficult to tightly control the slope of the sidewalls in a production environment). Finally, if sputter etching is used to clean the native oxide from the Al at the bottom of the vias prior to deposition of the next Al film, SiO_2 from the via sidewalls will be redeposited in the bottom, causing contact resistance problems. The problem is minimized through the use of vertically sided vias (or vias with a slight sidewall slope, e.g., 85°). Once the limits that mandate the use of such vertically-sided vias were reached, via processing techniques which result in completely filling vertical vias with metal were required.

Several space-saving benefits are achieved through the use of this technology. First, the minimum distance between adjacent metal lines at the via is reduced (since the via has no slope,

as shown in Fig. 15-44b). Second, the pad area of Metal-2 can be decreased, since the plug in the completely filled via provides ample overetch protection to underlying metal structures without mask coverage (Fig. 15-44a). Third, the minimum size of the Metal 1 pad areas can be reduced, since any small grooves that might be etched alongside the pad during the via etch will be easily filled during the plug formation process. As a result, the pitches of both Metal 1 and Metal 2 can be reduced, allowing significantly increased packing density.

15.6.3 Processing Techniques that Allow for Vertical Vias

Several fabrication approaches have been investigated for completely filling vertical vias. The most important of these are discussed below.

15.6.3.1 CVD-W Plugs: The most widely used technology for filling contact holes and vias for CMOS technologies down to about $0.35\ \mu\text{m}$ involves the fabrication of *tungsten plugs*. In this approach, a W film is blanket deposited by CVD (described in Chap. 6). The tungsten film thus fills the contact or via hole, and also covers the flat areas on the remainder of the wafer surface. The W film on the flat areas must be removed, to leave only the tungsten in the holes, flush with the top surface of the surface dielectric film (i.e., tungsten plugs have been formed). The removal step may either be an etchback process (carried out in a plasma etcher, see Chap. 14), or a CMP removal step (described in Sect. 15-5.2). The simplest dry-etchback process involves etching back the W directly (i.e., without using a sacrificial resist layer). However, when vias of varying widths must be filled the "dimples" that occur at the center of the W-filled vias will be deeper in those that are wider and will also be replicated into the final film. In some cases, faster etching of W will also occur along the seam. When a metal layer is deposited over these via plugs, poor step coverage may result.

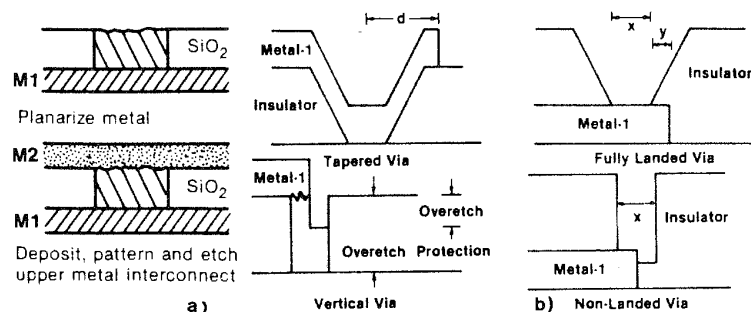


Fig. 15-44 Metal-1 and Metal-2 pad areas can be smaller when vertical vias and fully filled vias are used because (a) such filled openings provide protection during any overetch that may occur during the etching of Metal 1 or Metal 2; (b) any grooves that may be created during the overetching that occurs when the via openings are etched are easily filled by the via-filling process.⁶⁸ (© 1987 IEEE).

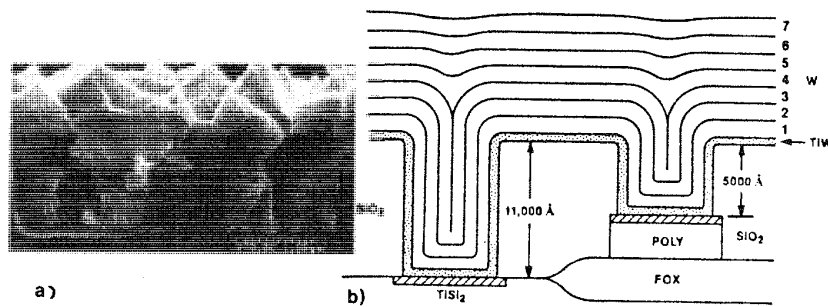


Fig. 15-45 (a) Surface roughness of conventional CVD W films formed using the hydrogen reduction process. (b) Smoother CVD W films can be formed using a multi-layer deposition technique in which thin silicon layer is deposited between the W layers.

In both the resist and non-resist etchback processes the dielectric surface must be highly planarized; otherwise, the overetch required to remove W stringers at steps will lower the height of the W plugs in the vias. The surface roughness of the CVD-W film (Fig. 15-45a) worsens this problem, since a longer overetch time is needed to clear all the W residue from the field. Another issue that must be addressed is that the glue layer deposited under the blanket-W film must also be removed, whether a dry-etch or CMP step is used to etchback the film. If a CMP step is used high TiN and Ti removal rates are essential to prevent oxide erosion and plug recessing. It is also necessary to have high selectivity to the underlying oxide to obtain local planarity over the plug surface, and to prevent excessive loss of the surrounding field oxide.

15.6.3.2 Filling Vias with Al Deposited by Sputtering: Various techniques have been attempted to develop a method of completely filling the vias with aluminum using sputter deposition. These approaches generally involve heating the wafer during the Al sputter deposition step (and possibly afterwards). In one technique, a two-step deposition process is employed, the first involves CVD and the second PVD. That is, a layer of pure Al is first deposited by CVD, and then a second layer of Al:Cu is sputtered onto the first Al layer. The first layer conformally covers the sidewalls and bottom of the vias, and serves as a wetting layer for the second film. The second PVD film contains the Cu, and by sputter depositing at an elevated temperature (300°C), complete via filling is reportedly achievable.

In another approach, dubbed *force-fill Al*, the Al is deposited into the via under sputtering conditions that deliberately forms a void in the Al film that fills the via (Fig. 15-46).⁷⁰ This deposition step is followed with a process in which the wafer is put into a chamber where it is simultaneously subjected to high pressure (an Ar atmosphere at 600 MPa) and temperature (350–400°C). This set of conditions makes the Al soft and forces the Al film into the via, whereupon the void is collapsed and the via is completely filled.

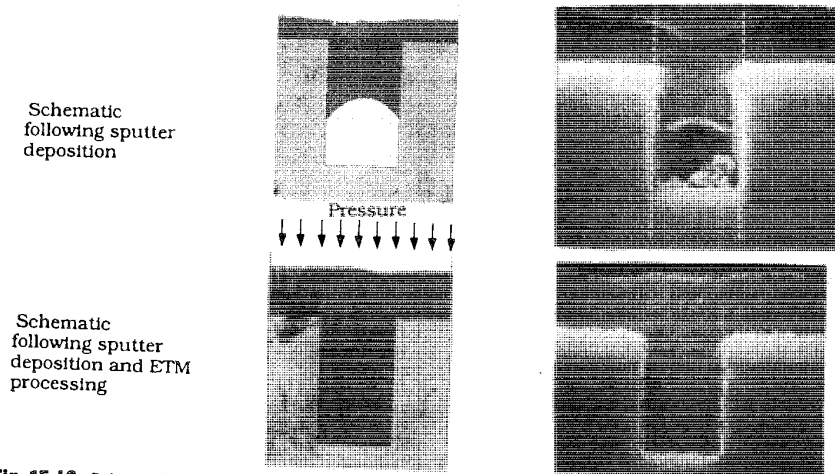


Fig. 15-46 Schematic drawing of the method used to fill contact holes and vias with the sputter-deposition and force-fill method.⁷⁰ (© 1994 IEEE).

Three other approaches using thermal procedures to improve the via filling of an Al-sputter process are the following. In the first, a low-temperature Al sputter deposition step (50–100°C) is used to put down a thin Al layer. This covers all the via surfaces with an Al film that has very small grains. To reduce overhang, a collimator is used in this step (see Chap. 11). This is followed with a high temperature sputter step (500–550°C) during which the Al can migrate and accumulate on each of the small grains to form a conformal Al layer that then fills the via hole to a high degree.^{71,72} For this process to be effective, the pressure in the sputter chamber between sputtering steps must be in the ultra-high vacuum range of 1×10^{-8} torr, with low residual oxygen and water vapor. If done in two sputter chambers, the wafers should not be exposed to atmosphere between step one and two. In the second approach, the Al is deposited by sputtering, but is then subjected to an RTP step to flow it into the via. The deposition may be carried out using two depositions and RTP steps.¹³⁵ At the time of publication, the complete filling of vias with Al was still not a widely used production process. In the third, a thin seed layer of Al is deposited by CVD, to a thickness of ≤ 100 nm. This layer provides excellent conformality and good wettability. A PVD Al:Cu layer is then deposited at 430°C. This reportedly can fill high aspect ratio vias with Al without forming voids.¹²⁸

15.6.3.3 Filling Vias by Electroplating of Cu: In the damascene process of fabricating interconnect structures, the metal film must be deposited into a recess in the dielectric layer so the recess is completely filled (i.e., no voids or seams should be present). The metal deposited outside the recess must then be polished back to leave only the metal in the via or groove. Deposition of Cu into such grooves can be done by CVD or electroplating. As we be discussed in Sect. 15-8, electroplating appears to be the most promising process for Cu line deposition in the damascene process. This is due to its excellent gap-filling, high deposition rate, low temperature, and system simplicity.

15.7 TRIPLE-LEVEL-METAL (TLM) INTERCONNECTS FOR 0.5- μm CMOS

In CMOS technologies from about 0.65 μm down to 0.35 μm , three or four levels of metal were used on the most advanced ICs. Here we describe the sequence of process technologies used to fabricate interconnect structures in these generations of CMOS. The interconnect fabrication sequence begins with a CMP planarization of the PMD layer. Contact holes are then etched into this planarized PMD dielectric. An anisotropic SiO_2 etch process is used so that vertical contact holes are formed. A blanket CVD-W process is used to fill the contact holes (which begins with a sputter deposition of a Ti/TiN liner film under the CVD-W). The W and liner layer is then etched back (or removed by CMP) to produce W plugs in the PMD layer. Next, a multilayer film consisting of Ti/TiN/Al:Cu is sputter deposited (to form interconnect lines of Metal-1 on the flat PMD layer). A capping layer of TiN atop the Al:Cu film is also sputter deposited. This TiN will serve as an ARC for the lithography step that is used to define Metal-1. Metal-1 is next dry etched and an IMD layer between Metal-1 and Metal-2 is deposited (either with a PECVD TEOS, an O_3 :TEOS, or a PECVD TEOS:SOG:PECVD TEOS sandwich layer) to achieve adequate gap fill. This layer is next polished flat by CMP, and vertical vias are etched into the planarized IMD film. The process sequence to form W plugs in the vias of the IMD is carried out again, with the Metal-2 deposition and etch processes mirroring those used for Metal-1. The sequence continues in the same manner to form Metal-3. A passivation layer is deposited on top of Metal-3 to complete the interconnect structure. Note that a fourth layer of metal can be added by continuing the process sequence after Metal-3 is patterned. Figure 15-47 shows a cross-section of a multi-level metal interconnect structure using W plugs and CMP. In Fig. 15-48 an SEM of the same type of multilayer interconnect system is from a perspective view. Note that this interconnect structure is implemented with stacked vias.

15.8 COPPER FOR ULSI INTERCONNECTS

As of 1999 copper was poised to take over as the main on-chip conductor for all types of ICs.^{73,136} Copper is an attractive substitute for standard aluminum interconnects for two reasons

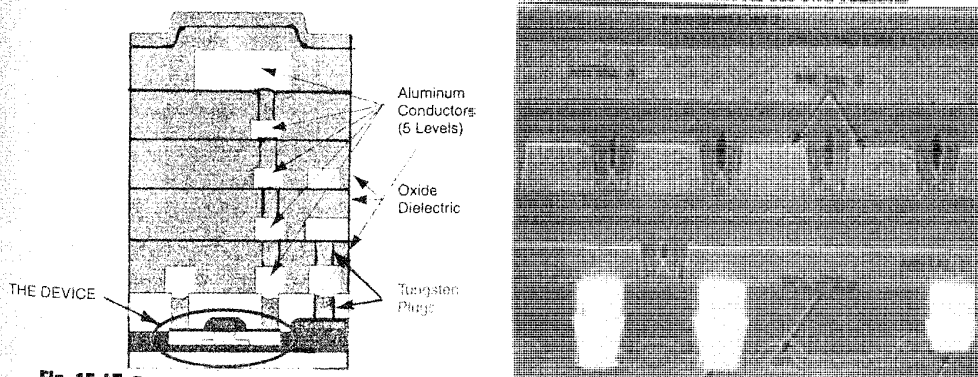


Fig. 15-47 Cross-section of a multi-level-metal interconnect with tungsten plugs and CMP of the dielectric layers: (a) Schematic drawing; (b) SEM.⁴⁸ Reprinted with permission of Semiconductor International.

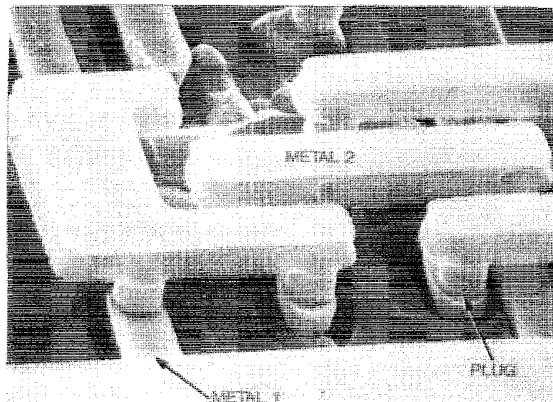


Fig. 15-48 SEM of a perspective view of a multi-level interconnect structure that uses tungsten plugs and CMP if the dielectric layers.¹⁴⁵ Reprinted with permission of Semiconductor International.

First, Cu films exhibit much better electromigration (EM) and stress-void resistance than do Al films. The upper limit of current density to prevent electromigration for Cu is 5×10^6 A/cm², whereas for Al it is 2×10^5 A/cm². About an order of magnitude longer EM lifetimes are obtained for Cu lines of the same cross-section as Al lines, primarily because Cu atoms are more massive than Al atoms.⁷⁴ These two benefits can be exploited in one of two ways to reduce the RC time constant of a Cu interconnect system. Either the total resistance, R , of an interconnect line can be reduced (by retaining the same metal thickness), or the side-by-side capacitance C_{intra} of interconnect lines can be reduced by decreasing the metal thickness (i.e., by keeping R constant, but reducing C , while carrying the same current in a smaller feature without fearing EM failure). The primary clock-speed benefit derived from using Cu is gained by using Cu at the upper interconnect levels, where the conductor lengths can be of the same order as the chip size. For IC generations at $0.18 \mu\text{m}$, and smaller, current densities in the lower interconnect levels can lead to electromigration failure of Al:Cu conductor structures. The increased electromigration resistance of Cu helps to overcome this limitation.

As shown in Fig. 15-49, by using Cu together with a low- k dielectric to replace Al and SiO_2 , the overall delay (and power consumption as well) in technologies below $0.25 \mu\text{m}$ is dramatically reduced.¹⁴⁰ Finally, by using Cu fewer levels of metal are required in an IC, potentially by as much as one half (Fig. 15-50).

If Cu is substituted for Al, however, a new set of technical challenges arise, including:

1. Cu does not form volatile copper fluorides and chlorides, making plasma etching of Cu impractical at temperatures below 200°C . Thus, instead of patterning and etching Cu films, it is more advantageous to blanket-deposit Cu films into recesses in the dielectric layers and then polish them away by CMP in the flat areas above. This is the basis of *damascene* interconnect structure formation (discussed in Sect. 15-12).
2. Unlike Al, Cu does not form a dense, self-passivating oxide on its surface. Thus, chemical reactions can continue to attack exposed Cu lines, making them more vulnerable to corrosion than Al lines.

Delay as a Function of Feature Size

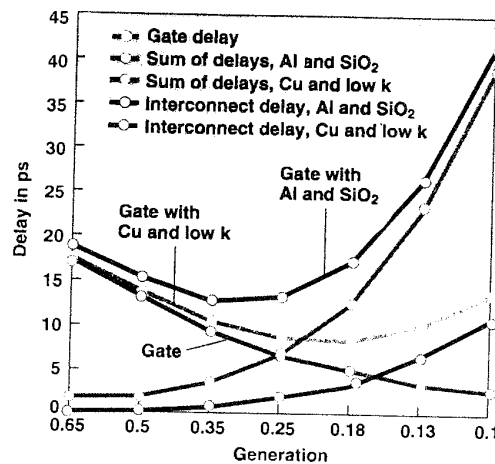


Fig. 15-49 Delays caused by interconnects are charted as a function of feature size for both Al/SiO₂ and Cu/low-*k* approaches.¹⁴⁰ (© 1997 IEEE).

3. Copper readily diffuses through SiO₂.

4. If the Cu manages to diffuse through the SiO₂ and then enters the silicon lattice, it will create deep impurity levels that degrade transistor performance and cause junction leakage. As a result, special Cu processing areas are set up in wafer fabs to keep trace amounts of Cu from other process steps.

5. Unlike Al, Cu does not adhere well to SiO₂.

To counteract the last three problems, a film that is both an adhesion layer and a diffusion barrier must be deposited between the Cu and the SiO₂. Such adhesion/diffusion layers for inlaid Cu processes must meet several criteria:

Reduced Complexity

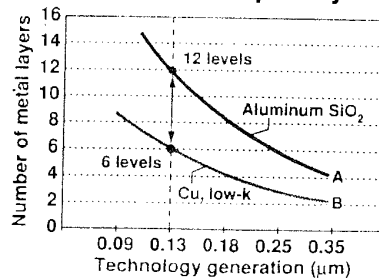


Fig. 15-50 One of the main benefits of copper, beyond the ability to increase chip speed and power consumption, is that the number of metal levels can potentially be reduced by as much as half.⁷³ Reprinted with permission of Semiconductor International.

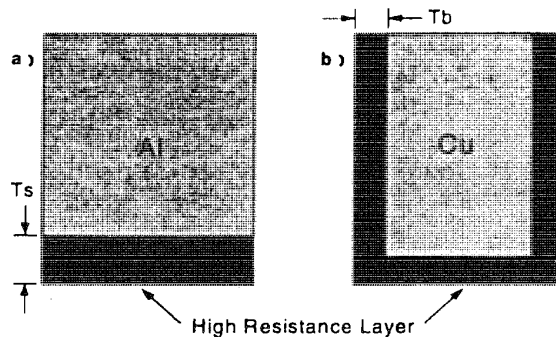


Fig. 15-51 Interconnect cross-sections showing the impact of higher resistance shunt or barrier layers: (a) Al interconnect; (b) Cu interconnect.¹⁴¹ (© 1995 IEEE).

1. They must be a good adhesion layer for Cu onto SiO_2 ;
2. They must be effective barriers against Cu diffusing through them at all post-copper-deposition temperatures;
3. They should exhibit low resistivity. That is, diffusion barrier materials will be expected to exhibit resistivities of less than $1000 \mu\Omega\text{-cm}$ to meet the needs of advanced interconnect processes. This will keep the total resistivity of the composite Cu/barrier interconnect structure below the resistivity of a comparably-thick Al-alloy/barrier interconnect line.
4. They should retain their diffusion-barrier properties when they are very thin, because they will not have the low resistivity of Cu. That is, thinner diffusion barriers for Cu will be needed than for Al because inlaid Cu interconnects require the barrier layer to be formed on at least three sides of the Cu conductor (Fig. 15-51). To keep the lower-conductivity barrier layers from impacting the resistance of the interconnect line, they must thus be relatively thinner for such Cu lines.

15.8.1 Process Integration Issues of Cu

As of 1998, it was stated that considerable work remained ahead to achieve thin, conformal liners to prevent Cu from diffusing (Fig. 15-52).⁷⁵ Among the materials that are candidates for the liners under Cu lines are TiN and Ta/TaN⁷⁶ deposited by sputtering or CVD. (Ta is an attractive barrier material because of its high melting point and immiscibility with Cu. It is also a highly reactive metal that forms strong metal-metal bonds, much like titanium. Thus, it provides a low-resistance ohmic contact and excellent adhesion to the copper above [and below] it. Doping the Ta with a few percent nitrogen blocks grain boundary diffusion paths. More heavily nitrided Ta, produced by reactive sputtering of Ta in the presence of nitrogen, can also be used as a barrier layer.) A *thin Cu seed layer* (typically deposited by sputtering) is also needed to be deposited on top of the liner.⁷⁷ This layer is necessary for the subsequent electroplating operation, in which it provides a low-resistance conductor for the plating current that drives the process, and also facilitates film nucleation. The seed layer carries current from

Fig. 1
seed l

the ex
near
vias
will p
into h

In
interc
comp
excee

In
barrie
proce
will t
encap
Manu
exam
prohit

Cop
electro
is des
underl

15.8.

Organ
Copper
ateL c
has be

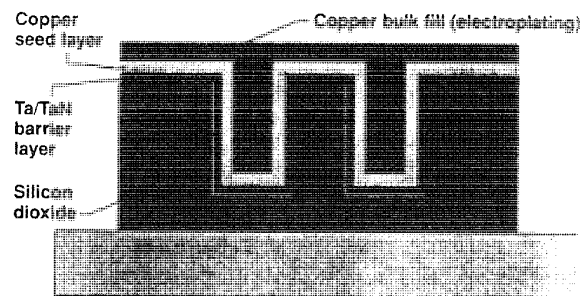


Fig. 15-52 Copper interconnect structure showing a PVD-deposited Ta-based diffusion barrier and copper seed layer, and electroplated copper fill.⁷³ Reprinted with permission of Semiconductor International.

the edge of the wafer to the center, allowing the plating current source to contact the wafer only near the edge. In addition, the seed layer carries current from the top surface into the bottom of vias and trenches. If there is insufficient seed-layer thickness at the bottom, the trench or via will prematurely close during deposition, leaving a center void. To get adequate seed thickness into high aspect-ratio vias, either an ionized-PVD-Cu or a CVD-Cu process may be needed.

In the late 1990's several companies announced the development of Cu multilevel interconnect systems that will be used on commercially available ICs.^{78,79} The resistivity of composite copper interconnects in these reports (with the diffusion barrier in place) did not exceed $2.2 \mu\Omega\text{-cm}$, a 40% improvement over that of comparable Ti/Al:Cu lines.

In the announced Cu processes tungsten plugs are used in the contact holes as a secondary barrier that keeps the first layer of Cu physically separated from the Si substrate. Damascene processes are used to pattern the Cu. In this approach, the Cu lines residing in the SiO_2 recesses will be capped with a thin layer of low-stress silicon nitride, so that the Cu will be totally encapsulated (with, for example, a TaN diffusion barrier on the sides and bottom). Manufacturing protocols must also be redesigned to tightly control cross-contamination. For example, sharing of process and metrology tools between front end and back end must be prohibited, as opposed to Al processing where some tolerance of sharing such tools exists).

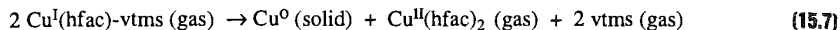
Copper films can be deposited in a number of ways, including PVD, CVD, electroplating and electroless plating. In Sect. 15.8.2 CVD-Cu technology is discussed, and the plating of Cu films is described in Sect 15.8.3. PVD of Cu is covered in Chap. 11, and CMP of Cu (and the underlying barrier layer) in Sect. 15-5.

15.8.2 CVD of Copper (Cu)

Organo-metallic sources are the most widely used precursors of CVD copper. These include Copper(II) hexa-fluoro-acetyl-acetonate $[\text{Cu}^{\text{II}}(\text{hfac})_2]$ and Copper(I) hexa-fluoro-acetyl-acetonate L or $\text{Cu}^{\text{I}}(\text{hfac})\text{L}$, where L is a neutral ligand weakly bonded to the Cu. One such ligand that has been used is vinyl-tri-methyl-silane (or *vtms*). $\text{Cu}^{\text{II}}(\text{hfac})_2$ is a solid at room temperatures

and must be dissolved in isopropyl alcohol in order to form a liquid precursor that can be delivered as a vapor (with hydrogen used as the carrier gas). High quality Cu films with low resistivity can be formed by the hydrogen reduction of $\text{Cu}^{\text{II}}(\text{hfac})_2$. However, to get the lowest resistivity Cu films the deposition temperature must be between 350–450°C. Since Cu is likely to be used with thermally-fragile low- k dielectrics, such high deposition temperatures are undesirable. A low-pressure plasma process has reportedly been developed which allows the deposition temperature to be reduced to about 170°C and which produces films with resistivities of $2.2 \mu\Omega\text{-cm}$.

The $\text{Cu}^{\text{I}}(\text{hfac})\text{-vtms}$ precursor is commercially available as a product called Cupra Select,TM developed by Schumacher.⁸⁰ It is a liquid at room temperature and is fed into the process chamber by direct liquid injection (i.e., the liquid is pumped to a vaporizer and then injected, see Chap. 6). This delivery method is needed for such precursors as $\text{Cu}^{\text{I}}(\text{hfac})\text{-vtms}$, which decompose after long periods of heating in a bubbler. Low resistivity Cu films can be formed at deposition temperatures of 150°C. The Cu is then deposited through a disproportionation mechanism:



Both the $\text{Cu}^{\text{II}}(\text{hfac})_2$ and the vtms ligand are stable compounds and readily desorb from the surface at the deposition temperature, leaving behind a Cu film relatively free of residues. Issues that must also be addressed with CVD-Cu are wafer-backside Cu deposition and void-free filling of contact holes and vias with the CVD process.

15.8.3 Electroplating and Electroless-Plating of Copper

Electroplating and electroless-plating of metals are processes widely used in the printed-circuit-board and multichip module technologies. They have only recently begun to find use in IC chip manufacture. That is, in future ULSI processes, low-temperature, low-cost, metal-deposition processes compatible with thermally fragile low- k dielectrics will be needed. Plating of metals (especially Cu) becomes a very attractive process for such applications for the above two reasons. Deposition of Cu by metal plating offers high throughput (50 wafers/hr) and thereby potentially much lower cost of ownership than PVD or CVD. In addition, such processes can be carried out at temperatures $\leq 40^\circ\text{C}$, and a well-tuned electroplating process can fill high-aspect ratio structures (i.e., trenches with aspect ratios of 5.0 were completely filled).

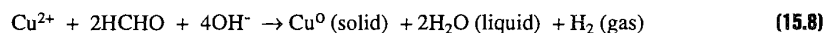
Electroplating processes utilize solutions containing ions of the metal to be deposited.^{81,144} In the case of copper, the solution contains copper sulfate (CuSO_4), sulfuric acid (H_2SO_4), and water. The CuSO_4 in this acidified solution breaks up into Cu^{2+} (cupric) and SO_4^{2-} ions. The wafers (with a conductive top surface, which in the case of Cu plating is the PVD-Cu seed layer mentioned earlier) are immersed in the solution. The wafer surface is also electrically connected to the negative side of an external dc power supply. Thus it becomes the cathode. As the positive cupric ions arrive at the negative-biased cathode they acquire two electrons and are reduced to Cu metal, which plates out on the wafer surface (with the reaction for Cu being $\text{Cu}^{2+} + 2\text{e}^- \rightarrow \text{Cu}[\text{O}]$). All cupric ions removed from solution by such plating are replaced by a reaction at a solid copper anode that is also immersed in the solution. This anode is attached to

the positive side of the power supply, and the copper atoms of the anode are oxidized to form the Cu^{2+} ions. These ions are released from the anode and dissolve into the solution. The metal ion formation reaction is thus sustained by passing an electrical current through the bath. The electroplating rate is a direct function of the current density. To obtain uniform deposition across a wafer the plating cell must be designed to give uniform potentials across the wafer surface. Because of concerns about Cu contamination, it is also critical that no copper be deposited on the backside of the wafer.

Generally, electroplating processes have high deposition rates (e.g., 400 nm/min) and form desirable metallurgy. However, electroplating requires a thin seed layer of Cu (about 50-nm thick) which must first be deposited by some other method (e.g., CVD or PVD) onto the adhesion layer. The correct choice of the seed layer is critical to complete recess-filling by electroplating. It would be ideal if a single layer could serve as both a barrier and seed layer, but currently two materials are used. An adhesion/barrier layer like TiN or Ta/TaN or TaSi_xN_y is first deposited, with a Cu seed layer applied in a second step. To ensure that low-resistance vias are fabricated with high yield, the via pre-clean (carried out by Ar sputter etching), barrier-layer deposition, and seed-layer deposition steps are likely to be performed on a single vacuum-integrated cluster tool. A schematic drawing of a Cu-electroplating tool is shown in Fig. 15-53.

An electroplating process for depositing Cu in a damascene process was described in 1995. It utilized a Ta diffusion barrier layer, a seed Cu layer deposited by collimated PVD Cu, and a Cu-plated layer in which a sulfuric acid/copper sulfate solution was used as the plating bath. The plated Cu films were formed to a thickness of about 1.5–2.5 μm . Then CMP was used to remove the Cu in the field regions, leaving Cu behind only in trenches and vias.⁴⁷ Good trench filling was observed for aspect ratios of up to 3:1 and trench widths of 0.4 μm .

An *electroless* process involves the formation of a thin film from an electrolytic solution without externally applied voltage. The deposition is due to an electrochemical reaction among the metal ions, reducing agent, complexing, and pH adjusters on catalytic surfaces. The chemical process involves CuO_4 and formaldehyde reduction.⁸³



An electroless Cu plating process for ICs was described in 1993 to give films with resistivities of 1.9 $\mu\Omega\text{-cm}$, close to that of bulk-Cu (1.7 $\mu\Omega\text{-cm}$).⁸⁴ Later work demonstrated filling of vias with aspect ratios of greater than 4 onto diffusion-barrier films of TiN.⁸⁵ Potential concerns of using both electroplating and electroless plating processes include void formation (which would trap pockets of the electroplating solution), and the environmental impact and disposal of the waste baths.

15.9 SPIN-ON GLASS (SOG)

Spin-on glass (SOG) is an interlevel-dielectric material that is applied in liquid form. It thus exhibits planarization and gap-fill capabilities. SOG films fill narrower spaces without causing voids than conventional CVD intermetal dielectric films (other than perhaps ozone-TEOS CVD films). Even such crevices as those caused by closely spaced Metal-1 and polysilicon edges (Fig. 15-9b) can be planarized by SOG to a degree that allows adequate step coverage. Other advantages of SOG films include: (a) simpler processing; (b) lower defect density; (c) higher

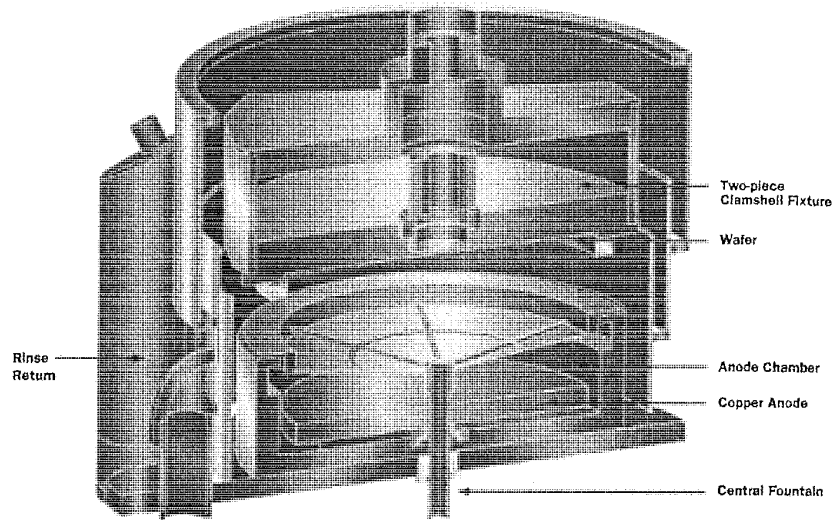


Fig. 15-53 Schematic drawing of the Novellus Systems SABRE® Cu-electroplating tool. Courtesy of Novellus Systems, Inc.

throughput; (d) relatively low cost; and (e) no use of hazardous gases. SOGs in the late 1990s were used primarily in memory applications. This is because memories are more cost sensitive than logic chips, but also because memory chips have a regular, densely packed array structure that lends itself to SOG processing.²⁵ SOGs are available from such vendor as Allied Signal, Tokyo Ohka, Hitachi and Dow Corning.

SOG materials are *siloxanes*, *silicates*, or *hydrogen silsesquioxanes* (HSQ) mixed in alcohol-based solvents.⁸⁶ In siloxane SOGs a small percentage of Si-C bonds remains in the film following the final cure cycle. Upon baking, the solvents are driven off and the remaining solid film exhibits properties similar to those of SiO_2 . Silicate SOGs can also be doped with such compounds as P_2O_5 to improve the dielectric film properties. However, adding P increases the water absorption and the modified SOG is mechanically and electrically less stable than the undoped silicate SOG. Case studies on the use of SOG materials can be found in references 87 and 88 (silicate based) and in references 89 and 90 (siloxane based). Until the late 1980s silicate SOGs were more widely used. As the industry transitioned to geometries smaller than $1\ \mu\text{m}$ the trend was to use siloxane SOGs.

HSQ SOGs⁹¹ do not contain Si-C bonds, and hence are carbon free. Instead they consist of a network of H-Si-O atoms in a random structure. These HSQ polymers are dissolved in a hydrocarbon solvent to allow them to be spun onto Si wafers. Upon hot-plate curing, they evolve the solvent and the Si-H bonds are dissociated. Subsequent molecular rearrangement during the curing forms SiO_2 bonds. Thus, the film after curing can be considered to consist of

two components SiO_2 and residual $\text{H}_x\text{SiO}_{2-x}$. Furnace temperature and oxygen concentration are two key variables that influence these reactions. HSQ films can be cured up to 750°C without exhibiting cracking. Since HSQ films can be used in a non-etchback process (see Sect. 15.9.3) they are being investigated as replacements for siloxane-based SOGs in PMD and IMD films.

Desirable properties of an SOG material are good consistency and shelf life, a simple cure cycle, excellent thermal stability, low stress and good crack resistance, spin-on uniformity, good adhesion, and an appropriate film-thickness range. Because early SOG products were proprietary mixes of organic monomers and attached organic groups mixed in solvents, it was difficult to predict their properties based on fundamental ideas of chemistry and materials science. The material properties of commercially available SOGs have subsequently been described in more detail in references 92, 93, 94, 95 and 97 (siloxane based), and in references 96 and 97 (silicate based).

SOG films provide somewhat lower degrees of planarization than do photoresists. However, their smoothing effect on sharp, vertical edges is satisfactory; enabling good step coverage of Metal 2 lines without line width variation. Figure 15-54 shows the planarization capabilities of SOG films as a function of metal pitch. The planarization can be as high as 80% for very closely spaced features, while for larger spacings it drops to less than 50%.⁹⁸

15.9.1 SOG Process Integration:

SOG processing involves some special considerations. Since the solvents evaporate from the SOG very quickly after being dispensed, an optimum dispense/spin cycle must be developed so adequate thickness uniformity can be obtained. The exhaust system is also very important for controlling film uniformity. In addition, the dispense nozzle must be cleaned frequently to prevent clogging, and the exhaust bowl must be rinsed so any particles spun off the wafer will be washed away. Specially designed SOG track equipment with spinners and integrated hot plates is commercially available to meet these requirements. A static dispense is recommended for coatings with thicknesses of 200–400 nm.

After being spun on, the SOG is baked first at a low temperature (e.g., $150\text{--}250^\circ\text{C}$ for 1–15 min in air), and then at a higher temperature (e.g., $400\text{--}425^\circ\text{C}$ for 30–60 min in air). The solvent is first driven off, and water is evolved from the film (due to polymerization of the silanol $[\text{SiOH}]$ groups). The loss of considerable mass together with material shrinkage creates a tensile stress in the film (~ 400 Mpa). If too thick a layer is applied, the stress can lead to film cracking. This problem is avoided by applying SOG layers no thicker than about 300 nm. The ability to fill narrow spaces is limited since low-viscosity SOGs (which have better gap filling properties) usually have higher shrinkage. Organic groups, such as methyl (CH_3) or phenyl (C_6H_5) have been added to siloxane films (from 1 to 12 wt% carbon) to improve cracking resistance. (While a higher carbon content would be more effective for this purpose, it would also tend to degrade adhesion and steam resistance. Hence, a tradeoff is usually made in selecting an optimum carbon content.)⁹⁹

The simplest technique for integrating an SOG film into a multilevel interconnect process would be to use it as a stand-alone film. Unfortunately, when an SOG film is formed to the thickness required of an interlevel dielectric, it exhibits an intolerable degree of cracking. SOG can also be spun on over a CVD-oxide layer, or CVD SiO_2 can be applied over a thin spun-on

layer of SOG. Such approaches, however, have proved to be unworkable, since frequent adhesion failures are observed when SOG is in contact with metal or resist layers. In addition, resist stripping procedures may attack the SOG film and degrade its stability.¹⁰⁰

The most widely reported implementations of SOG material involve sandwich-type structures, in which SOG is deposited over an initial CVD dielectric layer (e.g., pre-metal dielectric layer-1 or PMD-1), and a second CVD dielectric layer (PMD-2) is then applied. There are two variations of this sandwich: etchback SOG, and non-etchback SOG. In the first technique, the SOG is partially etched back after being cured so it remains only in the troughs between metal lines (Fig. 15-54a).^{99,101,102,103} No SOG remains over locations at which vias will be etched. After the etchback step, PMD-2 is deposited. In the non-etchback SOG structures, PMD-2 is deposited over the SOG following curing. The SOG remains permanently as a continuous thin film between the two layers of CVD SiO₂ (Fig. 15-55).^{104,105,106,107}

15.9.2 The Etchback SOG Process

Despite its greater complexity, the etchback approach has been more widely adopted, primarily because it avoids the so-called poisoned-via problem (that can be encountered with the non-etchback method). The etchback method is normally implemented with the less crack-susceptible siloxane-based SOG films, and the process sequence is illustrated in Fig. 15-54a. Since sufficient planarization is normally not obtained with a single coating of SOG, two or more layers are spun on before etchback is performed. A detailed description of an etchback SOG process is given in reference 108.

One drawback to the etchback process is that a thicker PMD-1 film must be deposited than will ultimately remain, since some is removed during the etchback step. For a semi-recessed field-oxide thickness of 0.6 μm , a polysilicon layer 0.4 μm thick, and a 0.2- μm -thick SOG film, a 1.0- μm -thick PMD-1 layer is needed. A minimum spacing of 1.2 μm is required between Metal-1 steps to prevent the formation of voids with a conventional-PECVD SiO₂ PMD-1 layer. For smaller Metal-1 spacings, the etchback process becomes impractically complex with conventional PECVD PMD-1 layers.

A plasma-TEOS/SOG process has been reported to allow spaces as small as 0.8 μm to be planarized by means of this etchback-SOG method,¹⁰⁹ with a 300-nm PECVD layer and a plasma-TEOS film used as a bilayer PMD-1 film (Fig. 15-54b). The plasma TEOS allows the small spacings to be filled without the formation of voids. The SOG is next applied to fill the "intermediate-width" gaps (i.e., near 2.5 μm) that are not sufficiently filled by the bilayer PMD-1 film. Following an SOG etchback step, a thin PMD-2 layer of PECVD SiO₂ is used to complete the sandwich structure.

Other problems of the etchback technique include:

1. SOG-to-SOG adhesion problems are sometimes observed when more than one layer of SOG is used to improve the degree of planarization. The cause of such SOG-to-SOG delamination is thought to be the slow out-gassing of the first SOG layer during the cure of the second layer, which pushes up the top layer.¹¹⁰
2. Lifting of the SOG from small gaps, occurs due to stress caused by volume shrinkage of the SOG and poor adhesion to the PMD-1 film.¹¹⁰

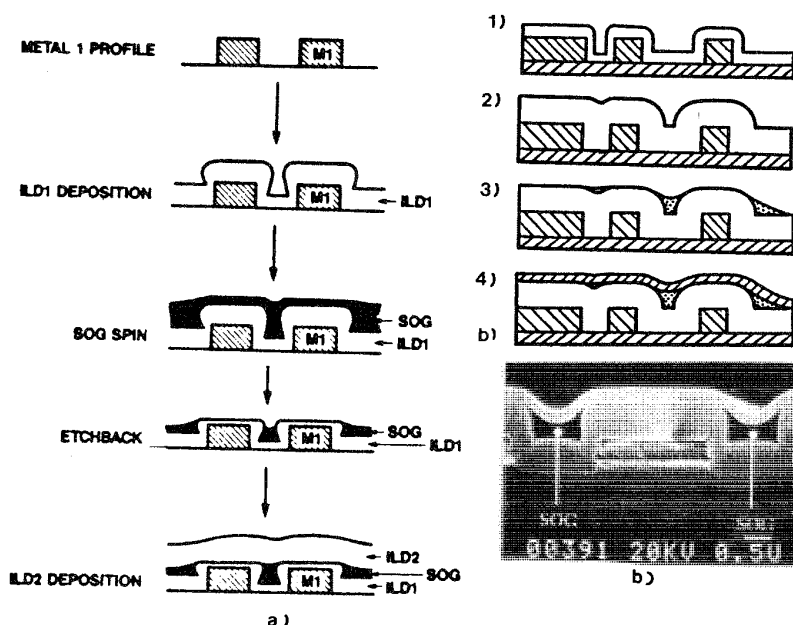


Fig. 15-54 (a) Sequence of steps used in the etchback SOG process (b) Etchback SOG process which uses plasma TEOS as part of the CVD film that is first deposited. This allows smaller spaces between adjacent steps to be filled without void formation.¹⁰⁹ (© 1988 IEEE).

3. Etch-rate differences exist between the SOG and the PMD-1 layer (this problem is compounded by the added complication that the SOG etch rate is apparently also stress dependent).

4. Microloading effects occur during the etchback step (similar to those described in the sacrificial etchback section).¹¹¹

5. In submicron technology, etchback processes have very small process margins. The CVD layer under the SOG must be thick enough to prevent Metal-1 from being exposed during the etchback step. The SOG thickness varies with topography and the amount of SOG to be removed has to include the thickest area where vias are allowed to be opened. A thick CVD layer will reduce the spacing between Metal-1 lines such that the SOG cannot flow in or the CVD layer pinches off, leaving voids between the metal lines.

15.9.3 The Non-Etchback SOG Process

Although there are fewer steps with this process, and the difficult etchback step is eliminated, the non-etchback process suffers from the poisoned-via problem if siloxane-type SOG materials are used. During Metal-2 deposition, this type of SOG outgasses, causing higher contact

resistance (e.g., 300–400 m Ω /via versus 60 m Ω /via for non-SOG controls) between Metal-2 and Metal-1, or even an electrical "open" (Fig. 15-55c). The source of the gases has been found to be primarily moisture.¹¹² Moisture remains within the SOG due to insufficient curing, or is reabsorbed during the wet process steps. This step is performed following the stripping of resist after the via etch process or during pre-metal-deposition cleaning steps.

Although it has been suggested that the problem could be overcome through the use of an all-dry process sequence following SOG spin-on coupled with complete curing of the SOG film, most non-etchback approaches instead use a silicate-type SOG. Special crack-resistant silicate SOG materials (phosphorus-doped to increase crack resistance) have been developed, since silicate SOG is inorganic. When such phosphosilicate glasses (containing ~9 wt% P₂O₅) are properly cured, the poisoned via effect is apparently eliminated.^{105,106,107} Another solution to the poisoned via problem is to treat the exposed SOG on the via sidewalls with an e-beam or ion beam. This apparently vitrifies the surface of the exposed SOG, forming a thin layer or "skin" of silicate like material. One company, Electron Vision, is developing an e-beam curing technique for spin-on dielectrics that will enable non-etchback processing.⁸⁶ Such non-etchback processes (with multiple coatings of SOG applied for adequate planarization) have been reported for Metal-1 spaces as small as 1 μ m.¹¹³ Two detailed reports of how a non-etchback phosphosilicate SOG process was integrated into a 1- μ m CMOS processes for both polysilicon and first metal interconnect planarization is given in references 114 and 115.

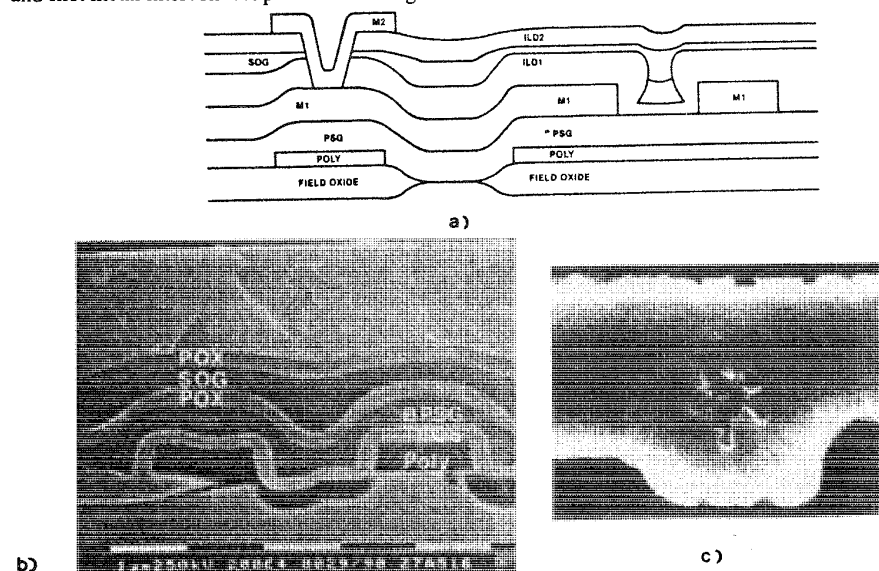


Fig. 15-55 (a) Drawing of a cross-section of structure formed using the non-etchback SOG process. (b) SEM micrograph of a structure formed with the non-etchback SOG process. Courtesy of Allied-Signal, Inc. (c) SEM micrograph of a cross section of a "poisoned via."¹⁰⁷ (© 1987 IEEE).

15.10 LOW- k DIELECTRICS

Interconnect delay can be reduced not only by decreasing R of the conductor structures in ICs, but also by decreasing C of the dielectric layers. The value of C , in turn, can be reduced by using dielectric materials with smaller permittivity values (i.e., *low- k dielectric materials*).¹¹⁶ The generic dielectric film material for interconnects has been SiO_2 , which has a k -value of ~ 4.0 . Air has the lowest k -value (i.e., its value is $k = 1.0$). There has thus been a push to integrate new dielectric materials into IC fabrication with k -values smaller than that of SiO_2 . (Note: the k -values of LPCVD and PECVD nitrides range from 6.0–7.0 and 6.0–9.0, respectively.) The National Technology Roadmap for Semiconductors urges IC manufacturers to identify, characterize, and integrate interlevel dielectrics with values of $k = 2.5$ –3.0 for 0.18 μm devices, and $k = 2.0$ –2.5 for 0.15 μm devices. However, the industry is being more conservative. These technology generations will probably be implemented with dielectrics having ranges of $k = 3.0$ –4.0, and $k = 2.5$ –3.0, respectively. There are many process integration issues that must be satisfied before new materials are successfully incorporated into an IC fabrication sequence.¹¹⁸ The new low- k materials must meet these requirements to be useful in IC fabrication. Here such integration issues and the candidate low- k materials being analyzed for the suitability for this task are described. The classes of low- k dielectric materials are divided into three k -ranges: 1) $k = 2.8$ –3.5; 2) $k = 2.5$ –2.8; and 3) $k < 2.0$.

15.10.1 Process Integration Issues of Low- k Dielectrics

If a dielectric material is to be successfully used in an IC, it must not only exhibit the desired k value, but it must also 1) exhibit adequate material characteristics (i.e., thermal, electrical, and mechanical); 2) be able to work with the other materials of the interconnect structure; 3) be compatible with the IC processes of cleaning, etching, CMP, and thermal treatments; 4) be available in high purity form, and at low cost; and 5) be able to operate reliably over the life of the product under the specified device operating conditions.

There are several important *thermal characteristics* of a dielectric material that must be considered. First, it should be stable up to temperatures of 425°C for at least short periods of time (to make it compatible with the thermal excursions of metal deposition, assembly, and packaging). Second, it is desirable that these dielectrics have a thermal conductivity close to that of SiO_2 (12.0 mW/cm°C) for purposes of heat dissipation during device operation. Third, they should exhibit low thermal expansion coefficients (to prevent cracking during thermal excursions, e.g., < 10 ppm/°C). As far as *electrical characteristics*, such films should have leakage current and breakdown field behavior (at least 2–3 MeV/cm) similar to SiO_2 and low charge-trapping characteristics. As far as *mechanical properties* are concerned, the films should exhibit low film stress and have the capability of being deposited to a thickness greater than 2 μm without cracking.

The *material compatibility* characteristics of such low- k dielectrics involve adhesion and chemical reactivity issues. The films must adhere well to the materials that are deposited on them. They also must adhere tightly to the materials on which they themselves are deposited. They should also not absorb moisture nor exhibit excessive permeability to moisture. Moisture can adversely impact the other materials of the interconnect structure, and can be outgassed during other processes, which degrades the process ambient.

There are several processes that need to be performed on the dielectric films after their deposition. The films must be able to be properly integrated into such processing. First, the films will need to undergo CMP, and such materials will thus need to be polishing-process-compatible. Next, vias (and perhaps grooves in a damascene technology) will have to be etched in the films. Issues of etch chemistry, etch selectivity to oxides, nitrides, and oxynitrides will need to be addressed. After etching, the resist mask must be removed. It must be possible to perform any resist-removal processing (and any resist-residue-removal steps) without damaging the dielectric layers.

Other *miscellaneous process integration* issues of low- k dielectrics include the following. First, although the gap-fill characteristics of the low- k films are not a consideration for damascene structures (as the dielectric films are deposited on surfaces that have been planarized by CMP), they will need to be deposited in a single pass at a thickness of about twice that used in dielectric films in non-damascene structures. Thus, they will need to exhibit crack-resistant characteristics for such film thicknesses. In addition, for some dielectric materials it might be necessary to use a capping layer of conventional CVD oxide or oxynitride to make the overall layer able to withstand the Cu CMP process. For non-damascene structures (i.e., in which the metal [Al] patterns are etched and then covered with a dielectric layer), the gap-fill characteristics of the low- k material will be important. It is asserted that both damascene and non-damascene interconnect structures will continue to find use as devices are scaled, depending on the circuit application. Finally, use of spin-on dielectric films that evolve toxic vapors will be avoided.

15.10.2 First Generation Low- k Dielectrics ($2.8 < k < 3.5$)

Dielectric materials with k values that are moderately lower than SiO_2 were the first materials to be adopted for this function. Three types of such moderately low- k films, *HSQ films* (introduced in Sect. 15-9), *fluorinated oxide films** and *low- k SOGs* found their way into IC production by 1999. Their main use was reported to be in the $0.5\ \mu\text{m}$ to $0.25\ \mu\text{m}$ CMOS generations.

The *spin-on HSQ films* were initially investigated as replacement films for siloxane SOG films in non-etchback SOG planarization processes. But they also exhibit a lower k -value than SiO_2 films (HSQ films have a k value of < 2.9). The HSQ material is available from Dow Corning under the tradename of FOx (Flowable Oxide).¹³⁷ HSQ was reported to have been integrated into products by Texas Instruments in the $0.5\ \mu\text{m}$ generation, and in 1997 an article describing its use in a five-layer Al-interconnect structure was published.¹¹⁸ An oxide cap was used on the HSQ film to improve its low- k value stability during W-CVD and Al fill.

FSG films are silicon oxyfluorides (F_xSiO_y) deposited by CVD. They exhibit a k value of ~ 3.5 . FSG films are formed in PECVD or HDP-CVD tools by adding SiF_4 to the process gas ambient used to deposit SiO_2 by CVD (silane and oxygen). By increasing the SiF_4 flowrate, more F is incorporated into the film. Higher concentrations of F cause the value of k to decrease. However, a maximum of about 6% F is permitted to be incorporated into the FSG films because higher concentrations would cause F to be evolved during RIE etch of these films. This would corrode the metal in the interconnect structure. FSG films were reported to be successfully integrated into 0.35 and $0.25\ \mu\text{m}$ circuits by IBM.⁷⁵ The evolution of F from the FSG oxide has

* These are also referred to as *fluorosilicated glass (FSG)*, or *silicon oxyfluoride (SiOF)*.

raised concerns about using this material in Cu interconnect systems. The F readily attacks the Ta-based liners of Cu systems, leading to volatile TaF_5 formation, and the consequent loss of adhesion between the low- k film and the Ta. A low- k SOG film is available from Allied-Signal with a k value of 2.9. It is an inorganic siloxane polymer named Accuspin T-23. When cured at 380–400°C, it can tolerate temperatures as high as 450°C before decomposing.

15.10.3 Second-Generation Low- k Dielectrics ($2.5 < k < 2.8$)

The largest variety low- k dielectrics is found in the second generation of such substances, which exhibit k values between 2.5 and 2.8. It is believed that these materials will begin to be incorporated into ICs at the 0.18 μm generation. Both *spin-on* and *CVD* low- k materials have been developed within this range of permittivities. Both types are described here.

15.10.3.1 2nd-Gen Spin-On Dielectrics with $2.5 < k < 2.8$: Most spin-on organic low- k polymers¹¹⁷ are significantly different from their conventional SOG counterparts. That is, they do not evolve moisture during curing. Because they exhibit better crack resistance they can be applied in a single pass to a greater thickness. Three organic spin-on low- k dielectrics being considered are: *poly(ethylene) ethers (PAE)*, *fluoro-polyimides*, and *bis-benzocyclobutenes (BCB)*. In addition, there is an organo-siloxane polymer type of film. Spin-on dielectric films are deposited without the use of an expensive CVD tool (they need a track cluster tool with spinners and hot plates instead). The cost of the spin-on liquid, however, is typically much higher per wafer than the cost of the CVD reactant gases.¹³⁸ In addition, a 100-nm-thick capping SiO_2 layer (deposited by CVD) will probably still be needed with spin-on low- k layers. Finally, spin-on films reduce the topology of any previous steps by more than 50%.

The *PAE materials* are aromatic structures which have k values of about 2.5 and also have the ability to withstand high temperatures. A fluorinated PAE product from Allied-Signal is named FLARE, and has a k value of ~ 2.8 . It was designed to have low outgassing characteristics and high thermal and mechanical stability.¹¹⁹ Dow Chemical has a similar material named SiLK. A report by Fujitsu detailed their experience with integrating FLARE and SiLK into a two-level Cu damascene structure. They showed little change in k value after a five-hour anneal at 425°C.¹²⁰ Schumacher has pioneered the use of non-fluorinated PAE materials, with their latest such product being VELOX. This material can be deposited without an adhesion promoter and does not require a capping layer. A *fluoro-polyimide* named FPI-136M with $k = 2.7$ is sold by DuPont.

BCB is a co-polymer of divinyl-siloxane and bis-*benzocyclobutene* (which is often referred to as *DVS-BCB*, or simply *BCB*). It is a thermoset silicon-containing resin that exhibits a k value of 2.7, but which has limited thermal stability even at temperatures of 350°C. Nevertheless, there are no by-products to the polymerization reaction and no corrosive or volatile by-products are evolved during the cure. BCB demonstrates excellent gap-fill, adhesion, and planarization capabilities. Dow Chemical offers a BCB product for IC ILD applications called Cyclotene. A study in 1998 indicated that Cu drifts much more slowly through BCB films than through PAE.¹²¹ Thus, use of BCB films could relax the requirements of metal diffusion barriers. But its limited thermal stability constrains its use to applications with a reduced thermal ceiling.

The *SOG-like material* in this class is Accuspin T-24 from Allied-Signal. It is a hydrido-organosiloxane polymer (HOSP) which has a k value of 2.5 after curing. This HOSP has a 12%

carbon content, which allows it to be formed into layers up to $1.2\text{ }\mu\text{m}$ thick without cracking. Its k value remains stable at temperatures up to 380°C . HOSP demonstrates good gap-fill and planarization capabilities.

15.10.3.2 2nd Gen-CVD Dielectrics with $2.5 < k < 2.8$: Two dielectric materials with a value of k in the range of 2.5–3.0 and deposited by CVD have been reported: *amorphous fluorocarbon* (a-C:F), which has a k value of 2.5, and *fluorinated parylene* (*parylene-F*). The benefit of CVD deposition of low- k films is that the deposition process can be carried out in a multi-chamber tool that allows liner and cap layers to be CVD deposited sequentially in a single tool.

Dielectric films based on amorphous fluorocarbon (a-C:F) can be deposited by PECVD or HDP-CVD, using C_4F_8 and CH_4 as the reactants.¹²² They contain about 50% F and also have a unique amorphous C–C cross-linked structure and the same C–F bonds found in low- k polymer PTFE (see Sect. 15.10.4). The strong C–F bonds decrease the dielectric constant and the C–C cross-linked structure maintains the thermal stability. Another advantage is they have similar values of thermal conductivity as existing oxides. As high-speed devices will generate more heat, such relatively high thermal conductivity is a benefit. A process is available for the deposition of a-C:F films by HDP-CVD from Applied Materials.¹²³ They refer to such deposited films as Black Diamond™ films. They claim that the same tool set used to deposit the first generation FSG films can also be used to deposit a-C:F films. This will allow the same equipment base to be used through several generations. To protect the a-C:F layer from attack during resist removal, a capping layer of SiO_2 is CVD deposited onto it. If the a-C:F layer is deposited onto Al, a thin (50-nm thick) layer of CVD SiO_2 is deposited onto the Al prior to the a-C:F deposition step. This protects the Al from reacting with the F in the a-C:F film. The RC-delay of a ring oscillator fabricated in $0.18\text{-}\mu\text{m}$ CMOS was reduced by 22% when an interlevel dielectric film of a-C:F/ SiO_2 was used compared to one with just an SiO_2 film.

Parylene (poly-*p*-xylylene) has a dielectric constant of about 2.6. It does not absorb moisture, but is stable only up to about 300°C . Fluorinated parylene (*parylene-F*) exhibits a higher thermal stability (400°C). Parylene aliphatic tetrafluorinated poly-*p*-xylylene (*parylene-AF4*) demonstrates a k value of 2.25 and even higher thermal stability (greater than 450°C). Also it does not absorb moisture. A CVD process for parylene AF4 is commercially available from Novellus.¹²⁴

In a subtractive Al-based interconnect structure, the parylene-AF4 layer is deposited first to conformally cover the Al lines. It is then etched back until the tops of the metal lines are exposed, at which point an oxide capping layer is CVD deposited (Fig. 15-56). The parylene-AF4 remains in the spaces between the metal lines, where it does the most good in reducing side-by-side capacitance. The oxide above the metal lines is available as a layer through which via etching takes place (a process that is mature). The vapor deposition of parylene materials is a well-established technology, but the mechanical strength and applicability to via and damascene processes is still being researched. Adhesion is reportedly adequate for ULSI interconnect applications.¹²⁴

15.10.4 Ultra-Low- k Dielectrics ($k < 2.0$)

At the time of publication, ultra-low k dielectric materials were still being investigated for their suitability as IC interconnect layers. Several different types were being considered, and are briefly mentioned here.

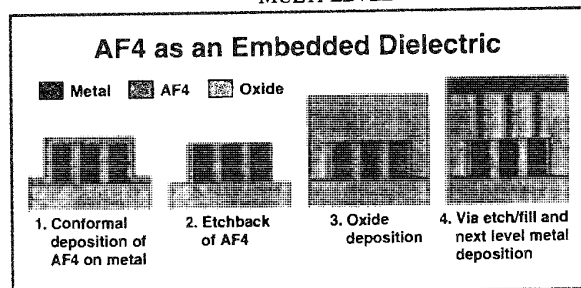


Fig. 15-56 In this integration scheme, the *low-k* material is encapsulated between the closely spaced metal lines where the *low-k* dielectric has its greatest influence on RC delay. The interlevel dielectric is oxide, which has superior mechanical and thermal transport properties.¹²⁴ Reprinted with permission of Semiconductor International.

The first of these are the nanoporous silica (SiO_2) xerogel materials whose k value can be tuned from $k = 1.3$ to $k = 2.5$ (by varying the film porosity). Since the k of air is 1.0, by making a dielectric layer porous, more of it consists of air, thus decreasing its overall k value. That is, in xerogels air is trapped as bubbles in a solidified gel. Xerogels are formed by evaporation drying, with some shrinkage. One commercial version is offered by Allied Signal under the name of Nanoglass. In temperature testing of nanoporous silica, feasibility studies indicated that such films are stable up to 500°C . Porous organic polymers are also being examined as *low-k* layers. Note that all such porous materials (oxide- and organic-based), must withstand such process steps as CMP, etching, and heat treatments without degradation of the pore structure. A second type of ultra-*low-k* dielectric is formed with polyfluorotetraethylene (PTFE), a chemical related to Teflon (a DuPont tradename). PTFE is a chemical compound consisting of carbon and fluorine, made up of uncross-linked $-\text{CF}_2-$ polymers (and so contains 67-at% F). A prototype spin-on version of a PTFE dielectric with a k value of 1.9 was being evaluated in 1998 (i.e., it was not yet being offered for sale). The developer, W.L. Gore and Associates gave it the name of SPEEDFILM. The material consists of an aqueous emulsion containing sub-20 nm PTFE particles and a surfactant. Films made with this material have good thermal stability up to 400°C . Although Teflon does not adhere well to metals SPEEDFILM has undergone adhesion testing, as well as CMP and etching studies.

Research continues on other ultra-*low-k* film materials. One possible class of candidates are porous organic polymer materials. This is a fruitful pursuit, because if one starts with a k value of 2.65 and makes it 20% porous, one can get a film with a k value of 1.9. With an inorganic material like SiO_2 ($k = 3.9$), the film has to be 70% air to get k down to 2.0.

15.11 HIGH-DENSITY-PLASMA CVD (HDP-CVD) OF DIELECTRIC FILMS

As the spaces between polySi and metal lines decrease, their aspect ratio increases. It becomes increasingly more difficult to deposit dielectric films into such high-aspect-ratio, narrow spaces without forming voids in the dielectric film. The previous gap-filling methods (CVD dep/etch; SOG sandwich films; and sandwich O_3 :TEOS films) for one reason or another, all eventually fail to adequately fill such gaps. Recent advances in high-density-plasma (HDP) sources (see

Chap. 14) have made them candidates for depositing dielectrics with HDP-CVD. They yield adequate gap-filling of such high-aspect ratio spaces.¹⁴ HDP-CVD dielectric films are also being studied for trench filling in shallow-trench-isolation structures (STI, Chap. 16), and for depositing low-hydrogen-content silicon nitride etch-stop layers (containing only 3–4% hydrogen, compared to 7–8% in PECVD nitrides) in damascene structures (Sect. 15.12). The basis for the improved gap-filling capability is that HDP sources provide a high-density of low-energy ions to the wafer surface. These ions simultaneously sputter the film being deposited by CVD. The sputtering removes film material from some regions on the wafer surface more rapidly than other regions. This effect is exploited to yield films that fill narrow gaps with no voids. Another benefit of HDP-CVD is that the films can be deposited at low temperatures.

This mechanism is described in detail here. The film being deposited by CVD initially covers all the surfaces of the wafer conformally (including the sidewalls and bottom of any trenches or contact holes). As the film grows thicker, an overhang develops at the top of the trench. This will close off the opening of the trench before the dielectric material completely fills it, leaving a void. If the overhang can be prevented from forming, it may be possible to fill the trench without void formation. This is accomplished in an HDP source by the bombardment of the wafer surface with a high density of ions. For depositing oxides by HDP-CVD a mixture of SiH_4 , O_2 , and Ar (or He) gases is most commonly used. The HDP source excites the oxygen and argon gas molecules. A bias applied to the wafer pulls the energetic ions out of the plasma and directs them toward the wafer surface. The oxygen reacts with silane to form SiO_2 while the Ar ions simultaneously sputter deposited material away. A high density plasma source is needed to maintain a high ion density at the relatively low pressures (a few mTorr) needed to obtain the directionality required for good gap fill. The good gap fill is achieved as described below.

As discussed in Chap. 11, the sputter yield is higher for surfaces which face the incoming ion beams at an angle, than at those surfaces perpendicular to the incoming ion directions. The maximum sputter etching rate occurs on surfaces at about 45° to the normal. Thus, a deposition:etch ratio exists that varies with location on the wafer surface. The net deposition rate is higher on the flat surfaces (the field regions of the wafer and the trench bottoms) than it is on the trench corners. (These angular surface regions are the 45° facets formed by depositing dielectric film at the top of the trenches.) By adjusting the dep/etch ratio, the trench openings can be kept from closing up, while the dielectric film fills the trenches from the bottom up (Fig. 15-57). The sputtering from the via bottoms also causes re-deposition of some material onto the sidewalls. This helps re-distribute the film so the gap is filled without forming a void. To achieve good filling, the ratio of deposition rate to removal rate should be no larger than 3.0. The wafers are also heated by the ion bombardment. Thus, the deposition rate is limited by the highest sputter etch rate that can be tolerated. The process is run at low pressures (<10 mtorr), which allows the ions striking the wafer surface to cross the sheath with no collisions. They arrive with the full energy they receive from the voltage drop across the sheath.

At the conclusion of the deposition process, the surface of the dielectric will generally still contain sharp ridges over the underlying metal lines (Fig. 15-12). If the metal lines are small enough and the deposition is continued, these ridges will eventually be sputtered away. But even in this situation, only local planarization is achieved. In most cases, an additional capping layer is also deposited by a non-sputtering CVD process. In all three cases, CMP is still needed to

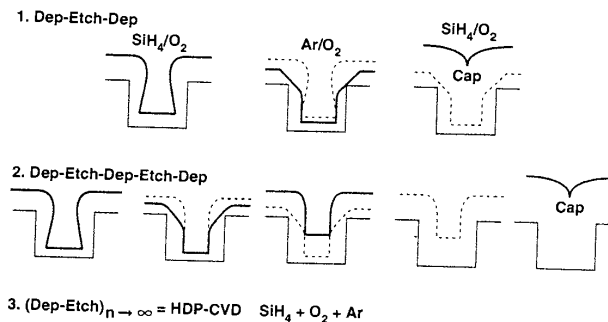


Fig. 15-57 High-density plasma CVD achieves good gap filling through simultaneous deposition and etching. The etch:dep ratio is usually kept between 0.14 and 0.33.¹⁴³ Reprinted with permission of Semiconductor International.

restore the wafer surface to a planar condition (although the amount of CMP required is reduced from that needed for a dielectric film deposited with a non-HDP-CVD process).

HDP-CVD tools have been developed by several vendors, including Applied Materials, Novellus, and Lam Research. Their widespread adoption, however, has been restrained by several considerations. The main one is that alternative dielectric processes that do not require HDP may be needed for sub-0.18 μm technologies. In one such scenario, low- k dielectrics applied by spin-on techniques may also be able to fill narrow, high-aspect-ratio features. Thus the expensive HDP tools would have to be replaced within a relatively short period of time. Another possibility is that widespread use of Cu dual-damascene interconnects would eliminate the need for dielectric layer gap-fill.

Another drawback of the HDP-CVD approach occurs when the space between metal lines must be filled. Here the aspect ratio of the inter-metal line space is higher than for a PMD application (poly-to-poly space). Therefore, the relative sputtering rate must be higher to prevent void formation. This slows down the net deposition rate, making the throughput low. (Note: the spacing between metal lines is generally larger than the critical dimensions that define a particular device generation. For example, the gap size for 0.35 μm is typically about 0.5 μm , and for the 0.25 μm device generation, it is 0.35 μm .) A protective dielectric layer must also be deposited without any sputtering at the outset of the dielectric deposition process. This is needed to keep the corners of the metal lines from being clipped during the dep/etch process. Such clipping of the could result in the incorporation of metal into the growing oxide film, could potentially lead to metal-to-metal leakage.¹²⁵ HDP-CVD processes also require vacuum pumps with high pumping speeds. For 200-mm wafers 2000 L/sec turbopumps are needed. For 2300-mm wafers 4000-5000 L/sec turbopumps will be required.

15.12 DAMASCENE AND DUAL-DAMASCENE INTERCONNECT STRUCTURES

As described in Sects. 15.4 and 15.7, the conventional way of fabricating the interconnect structure of an IC is to first blanket deposit a metal film onto the wafer surface. Next this film is patterned by etching, and a dielectric film is deposited over the etched metal lines. Vias are then

etched in the dielectric film to allow connection passages for the next layer of metal that is deposited onto the dielectric film. This approach requires etching the metal and having deposition processes with good gap-fill capability for both the dielectric and the metal layers. An alternative interconnect fabrication technology has been developed in which no metal etching step is required. Also the gap-filling capability is only required of the metal (but not for the intermetal dielectric). Since metal etch (especially the etching of Cu) and dielectric gap fill are viewed as two of the industry's greatest challenges in the drive to smaller dimensions, the elimination of these steps could represent a major benefit. The approach that offers this benefit is called the *damascene structure*, and it was pioneered by IBM.¹²⁶ Note that a third advantage of damascene interconnects is that they circumvent some problems associated with lithographic overlay tolerance, making it possible to achieve higher interconnect packing density.

The damascene interconnect structure is named after the practice of creating metal inlays, as developed by the artisans of Damascus in the Middle Ages. The process of creating the *single damascene structure* consists of first etching a trench or groove in a planarized dielectric layer, and then filling that trench with a metal, such as aluminum or copper (Fig. 15-58).¹²⁷ In a *dual-damascene (DD) structure*, a second level is involved where a series of holes (i.e., contacts or vias) are etched and filled, in addition to the trench. The vias and trenches are defined by using two lithography steps. The via is filled with metal in the same step as the metal line. After filling, the excess metal that is deposited outside the trench is removed by a CMP process, and a planar structure with metal inlays is achieved. Once a planarized surface is achieved, it is no longer necessary to perform CMP on the dielectric layers. Thus, one CMP step is also eliminated. Another simplification is that the plug-formation process is also eliminated.

Although the filling of narrow spaces by the dielectric layer is eliminated, the filling of narrow, high-aspect-ratio holes with metal is still a formidable challenge. In conventional interconnects, such holes are adequately filled with CVD-W. This can also be done in the single

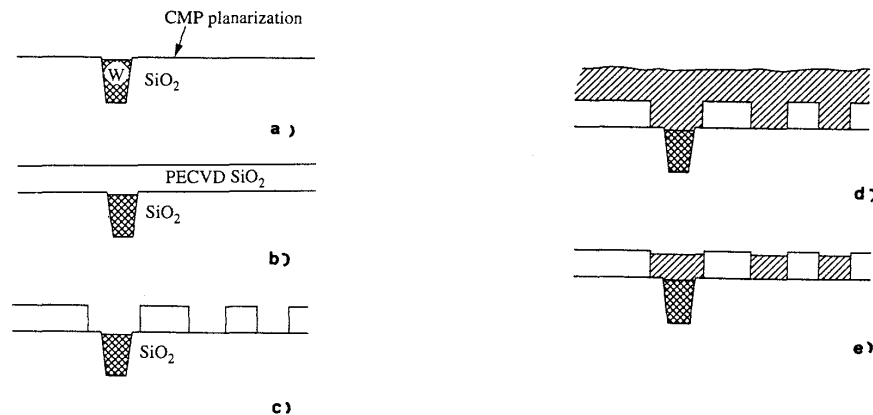


Fig. 15-58 Single Damascene process: (a) W plugs formed after the SiO₂ is planarized; (b) IMD is deposited by CVD; (c) Trench is etched in the IMD; (d) Metal is deposited to fill trenches; (e) CMP metal.

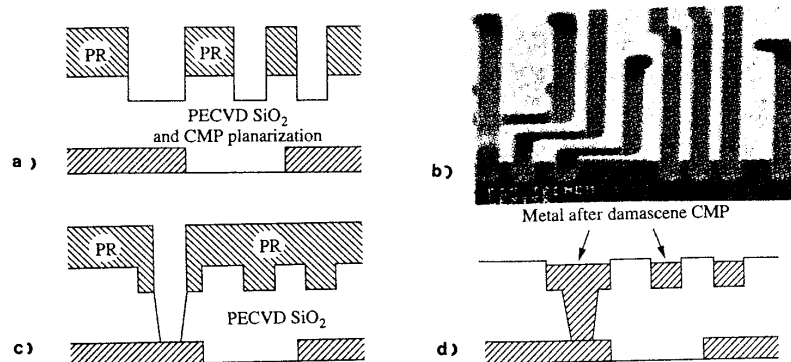


Fig. 15-59 Dual Damascene process: (a) IMD is deposited by CVD and planarized by CMP. Trench is defined by PR #1 and then etched; (b) SEM of the etched trench after PR #1 has been stripped; (c) Vias are defined by PR #2 and then etched, using PR #2 to protect other regions of the IMD from etching; (d) Metal is deposited to simultaneously fill trenches and vias. CMP to remove excess metal. Repeat the sequence for the next level of metal.

damascene structure. But in dual-damascene interconnects, the metal that fills the holes is Al or Cu. (Although it should be noted that the contact holes will likely be filled with CVD-W plugs in many dual-damascene structures, to keep Cu away from the Si substrate.) Aluminum via filling was discussed in Sect. 15.6, and the filling of vias with Cu in Sects. 15.6 and 15.8. A dual-damascene process using Al was described by workers at IBM in 1998.¹²⁸

The dual-damascene structure can be fabricated in a number of different ways. In one approach (Fig. 15-59), the trench patterns are defined in the ILD after resist is spun on and the trench pattern mask is used to expose the resist. After the trench is etched, the first resist layer is stripped. A second resist layer is then spun on, and the via mask is used to create openings in this resist layer so that the vias can be etched without further etching the dielectric in the trenches. After the resist is removed, the metal that fills both the vias and the trenches can be deposited, and polished back to create the dual-damascene structure. A second approach to fabricating the dual-damascene structure is shown in Fig. 15-60. It is described in Sect. 15.13.

15.13 MULTI-LEVEL METAL DAMASCENE INTERCONNECT STRUCTURE FOR 0.25 μm CMOS (AND BELOW)

Here we present an example of a multi-level-metal damascene interconnect structure that could be used in a 0.25 μm (or smaller) CMOS process. We assume that Cu is the main interconnect metal. We also stipulate that an STI isolation structure is used, so that the substrate is flat up to the time of the gate stack formation. We will see that both single-damascene and dual-damascene processing are actually used in this particular interconnect structure example. Many variations of this example damascene interconnect structure have been published.^{128,130,131,132,133} Some advanced interconnect structures may elect to use both conventional plug and etched Al-wire interconnects for some metal levels, and damascene for others. One such example is given in Ref. 134.

The fabrication sequence of this interconnect structure is begun by carrying out a single-damascene process sequence to form Metal-1 and CVD-W plugs in the contact holes. The

-
- Diagram illustrating the cross-sectional structure of a TSV (Through-Silicon Via) structure. The structure shows a central via filled with Cu (Copper) and lined with a Seed layer and Barrier layer. The via is surrounded by an Etch stop layer. The top of the structure is covered by a Passivation (and etch stop) layer. The bottom of the structure is covered by an Oxide CMP layer. The structure is labeled with various layers and dimensions:
- Top layer: Passivation (and etch stop)
 - Layer below top: Cu CMP
 - Layer below Cu CMP: M4
 - Layer below M4: Cu
 - Layer below Cu: Oxide CMP
 - Layer below Oxide CMP: Via
 - Layer below Via: Etch stop
 - Layer below Etch stop: Seed layer
 - Layer below Seed layer: Barrier layer
 - Bottom layer: M3
- Dimensions are indicated:
- T_m : Thickness of the top passivation layer.
 - IMO : Thickness of the oxide CMP layer.

process sequence is started by depositing a dielectric layer over the gate stack. This film serves as the PMD layer in this structure, and it is polished flat with a CMP step. Vertical contact holes are etched into the PMD layer. A CVD-W plug process is then used to fill the contact holes, and the W-plugs are planarized so that their top surface is flush with the top surface of the PMD. Next, an additional dielectric film is deposited onto the PMD layer, and trenches are etched into it, deep enough that the tops of the W-plugs are cleared of dielectric. A liner film (e.g., TaN) and a Cu seed layer are sequentially deposited by PVD as blanket layers over the entire wafer surface. A Cu film is then blanket deposited by electroplating. The Cu film (and the TaN liner layer) are polished off of the top surface of the dielectric film, leaving the Metal-1 Cu lines inlaid in the grooves. The main challenge here is to obtain the same polishing rate on the tantalum (which is a hard refractory metal), and the copper (which is comparatively soft). This completes the formation of the Metal-1 and contact plugs by the single-damascene process. The surface of the wafer is flat for the next dielectric deposition.

Micron Ex. 1023, p. 86
Micron v. YMTC
IPR2024-00790

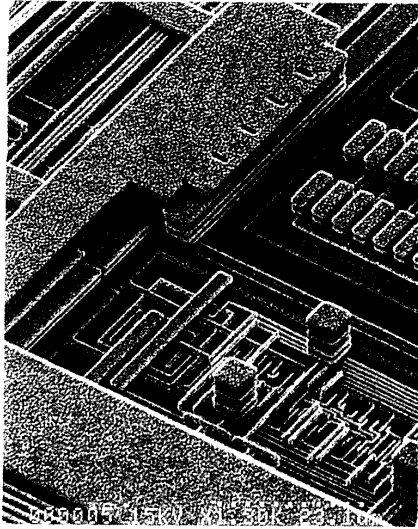


Fig. 15-61 SEM of an IBM Cu-based dual-damascene interconnect structure.⁷³ Reprinted with permission of Semiconductor International.

Note that in practice the via width may only be slightly smaller than the trench width.*

After the second resist layer is stripped, the formation of the metal line is begun. Again, a TiN or TaN liner and a Cu seed layer are sequentially deposited. This may require an ionized-metal-plasma sputtering process as the via hole is quite deep and has a high aspect ratio. The main part of the Cu interconnect is then deposited by electroplating. Note that both the via and the trench are filled simultaneously. A metal CMP process is used to remove the excess metal. A passivation/etch stop layer (silicon nitride) is deposited onto the top of the layer containing the grooves in the dielectric. This completely encapsulates the Cu with diffusion barriers. The entire interconnect structure is finally covered with an SiO₂/SiN passivation layer. Figure 15-61 shows a Cu-based dual-damascene interconnect structure fabricated by IBM.¹⁴⁷

* Note that the via pattern may be created in the etch stop layer by doing the via lithography step *before* depositing the second low-*k* film on the etch stop layer. Only a single trench/via etch process is then required, as the etching will continue in the areas from which the hard mask was removed.¹³¹ A third variation is to pattern the vias first with resist #1 and the via mask, and following this with an etch through the upper low-*k* layer and the etchstop material. Then, using resist #2 and the trench mask, the trench pattern is formed. The final etch creates the trenches and also etches the vias.¹³²

REFERENCES

1. C.T. Sah, "Evolution of the MOS Transistor - From Conception to VLSI," *Proceedings of the IEEE*, October 1988, p. 1280.
2. W. Shockley, *Bell Systems Technical Journal*, **30**, No. 10, p. 990, October 1951.
3. A.K. Sinha, *J. Vac. Sci. Technol.*, **19**, 778 (1981).
4. R.W. Keyes, *Proceedings of the IEEE*, **69**, 267 (1981).

5. A. H. Dansky, *Electronics*, October 9, 1980, p. 46.
6. D. J. McGreivy in "VLSI Technologies", D. J. McGreivy and K. A. Pickar, eds., IEEE Computer Society Press, Los Angeles (1982), p. 185.
7. D. Pramanik, "CVD Dielectric Films for VLSI," *Semiconductor International*, June 1988, 94.
8. L. Rothman, *J. Electrochem. Soc.*, **127**, 2216 (1980).
9. G. W. Ray and P. W. Marcoux, *Proceedings 2nd Internatl IEEE VMIC Conf.*, Santa Clara, CA, p. 52, 1985.
10. P. B. Johnson & P. Sethna "Using BPSG as an Interlayer Dielectric" *Semicond. Interl.*, Oct. 1987, p. 80
11. T. Abraham, *Proceedings 4th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 115, 1987.
12. B. Lee, A. Pierfederici, and E. C. Douglas, *Proceedings 4th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 85, 1987.
13. G. C. Smith and A. J. Purdes, *J. Electrochem. Soc.*, Nov. 1985, p. 2721.
14. D. Moy *et al.*, *Proceedings 6th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 26, 1989.
15. E. J. McInerney and S. C. Avazino, *IEEE Trans. Electron Dev.*, March, 1987, p. 615.
16. S. Mehta and G. Sharma, *Proceedings 6th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 80, 1989.
17. L. Koyama and M. Thomas, *Proceedings 2nd Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 45, 1985.
18. C. Kaanta *et al.*, *Proceedings 5th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 21, 1988.
19. S. Morimoto *et al.*, "Characterization of the Chemical-Mechanical Polishing of Intermetal Dielectric Films," *J. Electrochem. Soc.* 1993, p. 449.
20. J.M. Steigerwald, S.P. Murarka, and R.J. Gutmann, *Chemical Mechanical Planarization of Microelectronic Materials*, John Wiley & Sons, New York, 1996.
21. M.A. Fury, "The Early Days of CMP," *Solid State Technology*, May 1997, p. 81.
22. B. Davari *et al.*, "A New Planarization Technique, Using a Combination of RIE and Chemical Mechanical Polishing (CMP)," *Tech. Dig. IEDM*, 1989, p. 61.
23. R. DeJule, "CMP Grows in Sophistication," *Semiconductor International*, November 1998, p. 56.
24. A.R. Seturaman *et al.*, "Copper vs Aluminum: A Planarization Perspective," *Semiconductor International*, June 1996, p. 177.
25. F. Kaufman *et al.*, "Chemical Mechanical Polishing for Fabricating Patterned W Metal Features as Chip Interconnects," *J. Electrochemical Society*, Vol. 138, November 1993, p. 3460.
26. K. Wijekoon *et al.*, "Tungsten CMP Process Developed," *Solid State Technology*, April 1998, p. 53.
27. I. Kim *et al.*, "Optimized Process Developed for Tungsten CMP," *Semicond Internatl.*, Nov. '96, p. 119.
28. S. Roy, "Advanced Metallization and Interconnect Systems for ULSI Applications," 1995, p. 733.
29. C.C. Yu *et al.*, *Advanced Metallization for ULSI Applications Proceedings*, October, 1993.
30. J.-F. Wang *et al.* "CMP of Copper Multilevel Interconnects," *Proc. VMIC* 1995, p. 505.
31. R. DeJule, "CMP Challenges Below a Quarter Micron," *Semiconductor International*, Nov. 1997, p. 54.
32. D.R. Evans, "Electrochemical Interaction Between Copper and Barrier Materials During CMP," *Abs. of the Electrochem. Soc. Mtg.*, Fall, 1996, San Antonio, TX, Abs. No. 504, p. 612.
33. M. Tomozawa, "Oxide CMP Mechanisms," *Solid State Technology*, July 1997, p. 169
34. F. Preston, "Theory & Design of Plate Glass Polishing Machines" *J. Soc. Glass Tech.*, **11**, 214 (1927).
35. T. Izumitani, S. Harada, "Polishing Mechanism of Optical Glasses," *Glass Technol.* **12**, 131 (1971).
36. L. Cook, "Chemical Processes in Glass Polishing," *J. Non-Cryst., Solids*, **126**, 160 (1990).
37. J. Olsen and F. Moghadam, "Planarization Techniques," in *Handbook of Metallization for Integrated Circuits*, Eds. S.R. Wilson, C.J. Tracy, and J.L. Freeman, Noyes Pubs., Park Ridge, NJ, 1993, Chap. 6, pp. 426-438.
38. W.J. Patrick *et al.*, "Application of CMP to the Fabrication of VLSI Circuit Interconnections," *J. Electrochem. Soc.*, **138**, June 1991, p. 1778.
39. P.A. Burke, "Semi-Empirical Modeling of SiO₂ CMP Planarization," *Proc. VMIC* 1991, p. 379.

40. J. Schlueter, "Controlling Contamination During CMP Processes," *Sol. State Technol.*, May '96, p. S8.
41. W.-T. Tseng, "CMP of Fluorinated Silicon Dioxide: Is it Necessary and Feasible?" *Solid State Technology*, February, 1997, p. 61.
42. A.E. Braun, "Slurries and Pads Face 2001 Challenges," *Semiconductor International*, Nov. 1998, p. 65.
43. I. Ali, S.R. Roy, "Pad Conditioning in Interlayer Dielectric CMP" *Sol. State Technol.*, June 97 p. 185
44. J. McCann and C. Smith, "Bulk Slurry Distribution for Integrated Manufacture," *Semiconductor International*, November 1996, p. 94.
45. I. Ali *et al.*, "Chemical-Mechanical Polishing of Interlevel Dielectric: A Review," *Solid State Technology*, October 1994, p. 63.
46. W.L. Smith *et al.*, "Film Thickness Measurements for Chemical Mechanical Planarization," *Solid State Technology*, January 1996, p. 77.
47. E. Worthington, "New CMP Architecture Addresses Key Process Issues," *Solid State Technology*, January, 1996, p. 61.
48. R. DeJule, "Advances in CMP," *Semiconductor International*, November 1996, p. 88.
49. H.E. Litvak and H.-M. Tzeng, "Implementing Real-Time Endpoint Control in CMP," *Semiconductor International*, July 1996, p. 259.
50. H.E. Litvak, *U.S. Patent No. 5,499,733*. Issued March 19, 1996.
51. H. Litvak, *Proc of the SEMI Taiwan CMP Seminar*, Sept. 12, 1995, Taipei, p. 123.
52. W. Krussell *et al.*, "Mechanical Brush Scrubbing for Post-CMP Clean," *Solid State Technology*, June 1995, p. 109.
53. R. Davenport, B. Brown, "Integrated CMP Clean Methods," *European Semiconductor*, July '99, p. 36.
54. T.L. Myers *et al.*, "Post-Tungsten CMP Cleaning: Issues and Solutions," *Solid State Technology*, October 1995, p. 59.
55. J.M. de Larios, *et al.*, "Post-CMP Cleaning of Oxide and Tungsten Applications," *Semiconductor International*, May 1996, p. 121.
56. D. Hymes *et al.*, "The Challenges of the Copper CMP Clean," *Semicond. Internatl*, June 1998, p. 117.
57. Y. Zhang *et al.*, "Wafer dimensional analysis for Chemical Mechanical Planarization," *Solid State Technology*, July 1997, p. 179.
58. J. Reynolds *et al.*, "Defect Inspection Technologies to Meet the Challenges of Advanced CMP Processes," *Solid State Technology*, March 1998, p. 39.
59. W. O'Mara, "The Inevitable Shakeout in the CMP Equipment Market," *Solid State Technology*, May 1998, p. 84.
60. P.C. Fazan and V.K. Mathews, "A Highly Manufacturable Trench Isolation Process for Deep Submicron DRAMs," *Tech. Dig. IEDM*, 1993, p. 57.
61. K. Smekalin, "CMP Dishing Effects in Trench Isolation," *Solid State Technology*, July 1997, p. 187.
62. B. Withers *et al.*, "Wide margin CMP for STI," *Solid State Technology*, July 1998, p. 173.
63. L. Peters, "Choices & Challenges for Shallow Trench Isolation," *Semicond. Internatl.*, Apr., '99, p. 69.
64. P. Burggraaf, "CMP: Suppliers Integrating, Applications Spreading," *Semiconductor International*, November 1995, p. 74.
65. R.C. Ross and J. L. Vossen, *Appl. Phys. Letts.*, A3, (1984), p. 239.
66. D. W. Skelly and A. Gruenke, *J. Vac. Sci. Technol.*, A4, (1986), p. 457.
67. P.-L. Pai and C. H. Ting, *Ext. Abs. Electrochem. Soc. Conf.*, Fall, 1988, p. 364, Abs. No. 256.
68. C. Kaanta *et al.*, *Tech. Dig. IEDM*, 1987, p. 209.
69. N. S. Tsai *et al.*, *Tech. Dig. IEDM*, 1988, p. 462.
70. G.A. Dixit *et al.*, "A Novel High-Pressure, Low-Temperature Aluminum Plug Technology for Sub-0.5 μ m Contact/Via Geometries," *Tech. Dig. IEDM*, 1994, p. 105.
71. F.S. Chen *et al.*, "Planarized Aluminum Metalization for Sub-0.5 μ m CMOS Technology," *Tech.*

- Dig. IEDM*, 1990, p. 51.
72. H. Talieh, *et al.*, "Novel Method for Aluminum Planarization for Submicron High Aspect Ratio Contacts," *Proc. of the 10th International VLSI Multilevel Interconnect Conference*, 1993, p. 211.
 73. P. Singer, "Tantalum, Copper, and Damascene: The Future of Interconnects," *Semiconductor International*, June 1998, p. 90.
 74. T. Takewaki *et al.*, *Symp. on VLSI Technology, Dig. of Tech. Papers*, (1995), p. 31.
 75. D.S. Armbrust *et al.*, "Integration of HDPCVD FSG for 0.25 μm CMOS Technologies," *DUMIC Proceedings*, 1998, p. 90.
 76. J. Mendonca *et al.*, "TaN as a Cu Diffusion Barrier in Metallization Structures," *Proc. of the 1998 Advanced Metallization Conf. For ULSI Appl.* P. 741.
 77. P. Ding *et al.*, "Copper Barrier, Seed Layer and Planarization Technologies," *Proceedings of the 1997 VMIC Conf.*, p. 87.
 78. D. Edelstein *et al.*, "Full Cu Wiring in Sub-0.25 μm CMOS Technol.," *Tech. Dig. IEDM*, 1997, p. 773.
 79. S. Venkatesan *et al.*, "A High Performance 1.8 V, 0.20 μm CMOS Technology with Copper Metallization," *Tech. Dig. IEDM*, 1997, p. 769.
 80. J.A. Norman *et al.*, "New OMCVD Precursors for Selective Copper Metallization," *Proc. VLSI Multilevel Interconnect Conf.*, 1991, p. 123.
 81. J.D. Reid, "Fundamentals of Copper Electroplating," *Semiconductor International*, June 1998, p. 94.
 82. A. Contolini *et al.*, "A Copper Via Plug Process by Electrochemical Planarization," *Proc. VLSI Multilevel Interconnect Conf.*, 1993, p. 470.
 83. V.M. Dubin *et al.*, "Selective and Blanket Electroless Cu Plating by Contact Displacement for Deep Submicron Via and Contact Filling," *Proc. VLSI Multilevel Interconnect Conf.*, 1995, p. 315.
 84. J. Van der Meerakker, "On the Mechanism of Electroless Plating - I - Oxidation of Formaldehyde at Different Electrode Surfaces," *J. Appl. Electrochem.*, **11**, p. 387 (1981).
 85. A. Contolini *et al.*, "Copper Electroplating Process for Sub-Half Micron ULSI Structures," *Proc. VLSI Multilevel Interconnect Conf.*, 1995, p. 322.
 86. T. Nakano *et al.*, "Relationship Between Chemical Composition and Film Properties of Organic Spin-On Glass," *J. Electrochem. Soc.*, March 1995, p. 918.
 87. D. L. W. Yen and G. K. Gopal, *Proceedings 5th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 85, 1988.
 88. S. N. Chen *et al.*, *Proceedings 5th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 306, 1988.
 89. S. Morimoto and S. Q. Grant, *Proceedings 5th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 411, 1988.
 90. J. Chu *et al.*, *Proceedings 5th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 474, 1988.
 91. M. Loboda and G.A. Toskey, "Understanding Hydrogen Silsesquioxane-Based Dielectric Film Processing," *Solid State Technology*, May 1998, p. 99.
 92. P. E. Riley and A. Shelley, *J. Electrochem. Soc.*, May 1988, p. 1207.
 93. P.-L. Pai *et al.*, *J. Electrochem. Soc.*, Nov. 1987, p. 2829.
 94. C. H. Ting *et al.*, *Proceedings 4th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 61, 1987.
 95. S. Ito *et al.*, *Ext. Abs. Electrochem. Soc. Conf.*, Spring, 1989, p. 258, Abs. No. 178.
 96. M. Nakamura and R. Kanzawa, *J. Electrochem. Soc.*, **133**, p. 1167, (1986).
 97. R. J. Hopkins *et al.*, *Ext. Abs. Electrochem. Soc. Conf.*, Spring, 1989, p. 257, Abs. No. 177.
 98. H. M. Naguib *et al.*, *Proceedings 4th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 93, 1987.
 99. S. Morimoto and S. Q. Grant, *Proceedings 5th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 411, 1988.
 100. J. K. Chu *et al.*, *Proceedings 3rd Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 474, 1986.
 101. P. Elkins *et al.*, *Proceedings 3rd Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 101, 1986.
 102. L. B. Vines and S. K. Gupta, *Proceedings 3rd Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 507, 1986.
 103. V. Grewal, A. Gschwandtner & G. Higelin, *Proc. 3rd IEEE VMIC Conf*, Santa Clara CA, p. 107 1986.
 104. A. Rey *et al.*, *Proceedings 3rd Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 491, 1986.
 105. D. L. W. Yen and G. K. Rao, *Proceedings 5th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 85, 1988.
 106. S. N. Chen *et al.*, *Proceedings 5th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 306, 1988.

107. C. H. Ting *et al.*, *Ext. Abs. of Electrochem. Soc. Meeting*, Fall 1988, Chicago, IL, Abs. No. 257, p. 366.
108. L.D. Molnar, *Semiconductor International*, August 1989, p. 92.
109. M. Kawai *et al.*, *Proceedings 5th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 419, 1988.
110. C. Chiang *et al.*, *Proceedings 4th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 404, 1987.
111. G. Hausmann and P. Mokrisch, *Proceedings 5th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 293, 1988.
112. H.G. Tompkins, C. Tracy "Desorption from Spin-On Glass," *J. Electrochem. Soc.*, Aug. '89, p. 2328.
113. H. Kojima *et al.*, *Proceedings 5th Intl. IEEE VMIC Conf.*, Santa Clara, CA, p. 411, 1988.
114. L. Forester, A. L. Butler, and G. Schets, *Proc. 6th IEEE VMIC Conf.*, Santa Clara, CA, p. 72, 1989.
115. H. W. M. Chung, S. K. Gupta, & T. A. Baldwin, *Proc. 6th VMIC Conf.*, Santa Clara CA, p. 373, 1989.
116. L. Peters, "Pursuing the Perfect Low-k Dielectric," *Semiconductor International*, Sept. 1998, p. 64.
117. S. Murarka, "Low Dielectric Constant Materials for Interlayer Dielectric Applications," *Solid State Technology*, March 1996, p. 83.
118. H. Meynen *et al.*, "The Integration of Low k HSQ in Sub-0.35 μm Processes," *Proc. of the Electrochem. Soc., Symp. on Low and High k Dielectric Constant Mats.*, May 1998, Abs. 157.
119. K.S. Lau *et al.*, "Characterization and Thin-Film Properties of FLARE," *Proc. 1996 VMIC Conf.*, p. 92.
120. M. Ikeda *et al.*, "Integration of Organic Low-k Material with Cu-Damascene Employing Novel Process," *IEEE International Interconnect Conf., (IITC)*, June 1998, p. 131.
121. A.L.S. Loke *et al.*, "Copper Drift in Low-k Polymer Dielectrics for ULSI Metallization," *Dig. of Tech. Papers 1998 Symp. on VLSI Technology*, p. 26.
122. Y. Matsubara *et al.*, "Low-k Fluorinated Amorphous Carbon Interlayer Technology for Quarter Micron Devices," *Tech. Dig. IEDM*, 1996, p. 369.
123. Y. Matsubara *et al.*, "RC Delay Reduction of 0.18 μm CMOS Technology Using Low Dielectric Fluorinated Amorphous Carbon," *Tech. Dig. IEDM*, 1998, p. 841.
124. J. Gaynor and A. Ralston, "AF4 as a VLSI Interconnect Material," *Semiconductor International*, December 1997, p. 73.
125. R. Annapragada and S. Bothra, "Process Considerations in the Integration of HDP-CVD films," *Meeting Abs. of the Electrochem. Soc.*, Nov. 1998, Abs. No. 327.
126. P. Singer, "Making the Move to Dual Damascene Processing," *Semiconductor Internatl.* August 1997, p. 79.
127. B. Luther *et al.*, "Planar Copper-Polyimide Back End of the Line Interconnections for ULSI Devices," *Proceedings of the 1993 VMIC Conf.*, p. 15.
128. R. Iggulden *et al.*, "Dual-Damascene Aluminum for 1-Gbit DRAMs," *Solid State Tech.*, November 1998, p. 37.
129. X.W. Lin and D. Pramanik, "Future Interconnect Technologies and Copper Metallization," *Solid State Technology*, October 1998, p. 63.
130. B. Zhao *et al.*, "A Cu/Low-k Dual Damascene Interconnect for High Performance and Low Cost Integrated Circuits," *Dig. of Tech. Papers 1998 Symp. on VLSI Technology*, p. 23.
131. Y. Morand *et al.*, "Copper Integration in Self Aligned Dual Damascene Architecture," *Dig. of Tech. Papers 1997 Symp. on VLSI Technology*, p. 31.
132. S. Venkatesan *et al.*, "A High Performance 1.8V, 0.2 μm CMOS Technology with Copper Metallization," *Tech. Dig. IEDM*, 1998, p. 769.
133. D. Edelstein *et al.*, "Full Copper Wiring in a Sub-0.25 μm CMOS ULSI Technology," *Tech. Dig. IEDM*, 1998, p. 773.
134. M. Igarashi *et al.*, "The Best Combination of Aluminum and Copper Interconnects for a High-Performance 0.18 μm CMOS Logic Device," *Tech. Dig. IEDM*, 1998, p. 829.
135. A. Kamgar *et al.*, "Impact of Al Melting on Diode Integrity," *Proceedings of the 6th International VLSI Multilevel Interconnect Conference*, 1989, p. 190.
136. R.L. Jackson *et al.*, "Processing and Integration of Copper Interconnects," *Solid State Technology*, March 1998, p. 49.
137. T. Batchelder *et al.*, "In-line Cure of SOD Low-k Films," *Solid State Technology*, March 1988, p. 29.
138. E. Korczynski, "Low-k Dielectric Integration Cost Modelling," *Solid State Technology*, October 1997, p. 123.

- March 1998, p. 49.
137. T. Batchelder *et al.*, "In-line Cure of SOD Low- k Films," *Solid State Technology*, March 1988, p. 29.
138. E. Korczynski, "Low- k Dielectric Integration Cost Modelling," *Solid State Technology*, October 1997, p. 123.
139. S. Saito *et al.*, *Tungsten and Other Refractory Metals for VLSI Applications II*, 1984, San Juan Batista, CA, R. Blewer Ed., p. 69.
140. S.C. Sun, "Process Technologies for Advanced Metallization and Interconnect Systems," *Tech. Dig. IEDM*, 1997, p. 765.
141. M. Bohr, "Interconnect Scaling: The Real Limiter to High Performance ULSI," *Tech. Dig. IEDM*, 1995, p. 241.
142. C. Jang *et al.*, *Proc. 4th Intl. IEEE VMIC Conf. Santa Clara Ca*, p. 357, 1987.
143. P. Singer, "The Future of Dielectric CVD: High-Density Plasmas?" *Semiconductor International*, July '97, p. 127.
144. P.C. Andricacos, "Copper On-Chip Interconnections," *The Electrochemical Society Interface*, Vol. 8, No. 1, Spring 1999, p. 32.
145. A. Denboer, "Inside Today's Microprocessors," *Semiconductor International*, February 1994, p. 64.
146. P. Singer, "Filling Contacts and Vias," *Semiconductor International*, February 1996, p. 89.

PROBLEMS

- (a) Calculate the RC time constant of a $0.5\text{ }\mu\text{m}$ thick Al interconnect line formed on a $0.5\text{ }\mu\text{m}$ thick thermally grown field oxide. The length and width of the Al line are 1 cm and $1\text{ }\mu\text{m}$, respectively. The resistivity of the Al is $10^{-5}\text{ }\Omega\text{-cm}$. (b) What would be the RC time constant for a polysilicon line which had a sheet resistance of $30\text{ }\Omega/\text{sq}$ of identical dimensions.
- Why is the parasitic capacitance per unit area for Metal-1 on field oxide smaller than the parasitic capacitance of polysilicon on field oxide?
- Metal lines of varying widths are to be fabricated. Before being patterned, the sheet resistance of the metal film is measured and found to be $2.5\text{ }\Omega/\text{sq}$. Find the total resistance of 0.5 , 1.0 , 2.0 , and $5.0\text{ }\mu\text{m}$ wide and 1 cm long lines. What other information is still needed before it can be determined what type of metal makes up these lines?
- A particular interconnect system exhibits a specific contact resistance between the metal and the silicon substrate of $10^{-6}\text{ }\Omega\text{-cm}^2$. Assume the contact hole in this process is $0.5\text{ }\mu\text{m}$ long and $10\text{ }\mu\text{m}$ wide. (a) If the current flows vertically from the metal into the Si, what is the contact resistance?; (b) If the edge of the contact is $1\text{ }\mu\text{m}$ from the edge of the gate, and the sheet resistance of the current path in the drain region of the Si substrate is $80\text{ }\Omega/\text{sq}$, how large is the sheet resistance between the gate edge and the contact?
- To improve the performance of the MOSFET described in Prob. 2, it is decided to form a self-aligned silicide on the source and drain. Assume that the resistivity is $10\text{ }\Omega/\text{sq}$ and that the silicide extends to $0.2\text{ }\mu\text{m}$ from the gate ($0.8\text{ }\mu\text{m}$ from the contact edge). Calculate the series resistance of this contact.
- Assume that a four level metal process flow is to be designed and the following are the costs for various processing steps: (a) lithography costs \$10 per step; (b) metal deposition costs \$5 per step; (c) dielectric deposition costs \$5 per step; (d) etchback SOG cost is \$2 per step; and (e) CMP of the dielectric layers is \$8 per step. If the lithography process has a depth of focus of $1.0\text{ }\mu\text{m}$ and the poly thickness (M_0) and metal thicknesses (M_1 , M_2 , M_3 , and M_4) are $M_0 = M_1 = M_2 = M_3 = M_4 = 0.6\text{ }\mu\text{m}$. What is the most economical four-level metal process? What is the most economical process if the CMP steps cost \$20? One may select any combination of local (i.e., SOG etchback) and global (i.e., CMP) planarization processes to design the process flow.
- The design of a circuit requires that a maximum permissible current density of $2 \times 10^5\text{ A/cm}^2$ through a conductor 1 mm long, $1\text{ }\mu\text{m}$ wide, and nominally $0.5\text{ }\mu\text{m}$ thick. Assume that 10% of the conductor length passes over steps and at these locations the film is 50% of the nominal metal film thickness. What maximum voltage may be used across the conductor if the sheet resistance is $5.6 \times 10^{-2}\text{ }\Omega/\text{sq}$?