design of high-temperature furnaces used to anneal the SIMOX material has also been optimized in order to minimize the diffusion of metallic contaminants from the furnace heating-elements into the furnace tubes.

3. Pinholes in the buried-oxide-layer will occur if a particle is present on the wafer during the oxygen-implantation-process. Such particles screen the oxygen implant at that point. The result is a breach in the buried-oxide-layer that represents a flaw in the insulating barrier at that point. Careful wafer cleaning and attention to particulate control in the implant process can control this problem.

4. Despite the high-temperature anneal, the surface silicon may retain a dislocation density on the order of $10^3/cm^2$. The presence of such defects can cause reduced oxide-integrity and increased junction-leakage. However, devices fabricated in the best SIMOX material have been shown to be able to build excellent quality CMOS and bipolar ULSI circuits.[30]

5. SIMOX-wafers are more costly than bulk-Si-wafers, although the cost is expected to sharply decrease as the popularity of this technique expands. The cost reduction will be partly achieved by improvement in the implant equipment used to fabricate SIMOX-wafers. For example, SIMOX-dedicated ion-implanters are now available from IBIS Technology. Such implanters can furnish adequate throughput and acceptably-low levels of metallic-contamination. These production tools have an output of 100-mA of O^+ at an acceleration voltage 210-keV. The total transition-metal contamination on the wafers (from Ti to Zn) has been measured to be less than $5x10^{10}$ atoms/cm^2.

11.6.2.2 Low-Dose SIMOX

A recent trend with SIMOX fabrication is to use a lower oxygen-implant-dose than is used in the high-dose SIMOX approach. With the low-dose technique, improved, lower-cost SOI substrates are obtained. For example, a typical modern SIMOX implanter with a nominal 50-mA of O^+ beam current has a maximum output of only about 25 wafers of 200-mm diameter per day, round the clock operation. The low-dose SIMOX process also drastically increases the crystal quality of the silicon layer above the stoichiometric BOX layer. The oxygen implant dose is reduced to about $4x10^{17}$-$1.0x10^{18}$ ions/cm^2, and at an energy around 120-keV. Note that only a very-narrow process window close to the dose of $4.0x10^{17}$ ions/cm^2 will produce a continuous, precipitate-free BOX layer.

TSMC-1013
U.S. Patent 8,884,373

Table 11-3 compares the process conditions used to produce SIMOX with high-, medium-, and low-doses of oxygen.

A silicon-layer between 150-200-nm deep is produced, but the lower oxygen-implant-dose also reduces the BOX layer to thicknesses of only 90-150-nm (depending on the selected oxygen dose).[31] A high-temperature anneal in Ar (i.e., at 1350°C) is then carried out, just as is done in the high-dose SIMOX process. However, this high-temperature-anneal in an inert ambient is followed by a subsequent high-temperature anneal in a gas mixture of inert gas and 30-60% oxygen (at ~1350°C).

The oxygen in the furnace-ambient at this high-temperature both oxidizes some of the superficial silicon layer, but also diffuses through the surface silicon layer and reacts at the buried-oxide/Si-interface to form new oxide within the Si (Fig. 11-19). This has the effect of increasing the thickness of the BOX-layer.[31] This process of growing the BOX-layer to a greater thickness by annealing the low-dose SIMOX wafers in oxygen is called an *internal thermal oxidation process* (*ITOX*). Note that as long as the oxide grown on the surface of the silicon during ITOX is thinner than 500-nm, there exists a linear-relationship between the thickness of this oxide layer (t_{ox}) and the thickness increase in the buried

TABLE 11-3 Physical and Electrical Properties of Commercial SIMOX

Parameter	Units	Standard Dose SIMOX	Medium Dose SIMOX	Low Dose SIMOX
Silicon Thickness	A°	1500 - 2000	1500 - 2200	1500 - 1800
BOX Thickness	A°	370 - 400	170 - 190	90 - 150
HF-Defects Density	cm^{-2}	< 1	< 1	< 1
Dislocation Density	cm^{-2}	$< 10^7$	$< 10^6$	$< 10^5$
Stacking Fault Tetrahedara	cm^{-2}	$< 10^5$	$10^5 - 10^6$	$10^5 - 10^6$
Surface Roughness (10um x 10um)	A°	7 - 10	5 - 7	3 - 5
Metallics (TXRF)	cm^{-2}	$< 10^{11}$	$< 10^{11}$	$< 10^{11}$
Short Density	cm^{-2}	< 0.1	< 2	< 2
Electron Mobility	$cm^2/V\text{-}s$	850 - 1000	850 - 1000	850 - 1000
Hole Mobility	$cm^2/V\text{-}s$	200 - 325	200 - 325	200 - 325
Oxide Charge	C/cm^{-2}	$2 - 3 \times 10^{11}$	$1.5 - 3 \times 10^{11}$	$1 - 2.5 \times 10^{11}$
Interface State Density	$cm^{-2}eV^{-1}$	$2\text{-}3 \times 10^{11}$	$1 - 3 \times 10^{11}$	$1 - 2.5 \times 10^{11}$
Transconductance (Gmsat)	mS	400 - 440	900 - 1200	1000 - 1300

oxide Δt_{BOX}): $t_{ox} = 0.06\ \Delta t_{BOX}$ (Fig. 11-19c). The internal-oxide also grows at the expense of the bottom of the silicon overlayer.

In the low-dose SIMOX process, the BOX formation consists of oxygen precipitation, growth of the precipitates, and their coalescence during the high-temperature-anneal process. This sequence of events occurs because the oxygen implantation dose is less than the SiO_2 stoichiometry, even at the depth of the peak-oxygen-concentration, and a stoichiometric-SiO_2-buried-layer is required to be formed to produce effective dielectric isolation. By carrying out the ITOX process, the BOX in the low-dose-SIMOX-process is made thicker and exhibits better isolation characteristics. The upper part of the BOX after the ITOX process is essentially a thermal-oxide-layer. Since the upper segment of an ITOX-SIMOX layer is essentially a thermal oxide, such important BOX properties as the breakdown-field, show a marked improvement (i.e., the breakdown-field is increased from 3-4-MV/cm for the case of conventional low-dose SIMOX to more than 5-MV/cm for ITOX-SIMOX, see Fig. 11-19d).[49] This improvement can be attributed to the BOX-thickness incremental-increase due to ITOX. In addition, the ITOX reduces the size and density of silicon pipes in the BOX, which is the main cause of BOX breakdown.[50] In 2001 IBIS Technology licensed a low-dose-SIMOX-technology from IBM called Advantox® MLD.

In 2000, IBM said recent advances in the production of low-dose-SIMOX-wafers will cut incremental processing costs of SOI to only 3% over the costs of producing ICs on bulk-Si-wafers. Low-dose-SIMOX-wafers are commercially available from such vendors as Ibis, Komatsu, and Nippon Steel. For example, the Ibis-1000 model is currently available. The next generation oxygen implanter, the Ibis-2000 (which can accommodate both 200-mm and 300-mm wafers) was under development in 2001, and is forecast to be ready for delivery in the first half of 2002. Figure 11-20 is a photograph of a Power4™ processor made by IBM using a SIMOX-SOI-based process.

11.6.3 Wafer Bonding

Another popular technology for producing SOI wafers is *wafer bonding*.[32] Wafer bonding is a simple concept. Basically, wafer bonding refers to the phenomenon that mirror-polished, flat and clean wafers of almost any material, when brought into contact at room temperature, are locally attracted to each other by van der Waals forces and adhere, or *bond* to each other. The bonding process is started by locally applying a slight pressure to the wafer pair, and the bonded area spreads laterally over the whole wafer area in a few seconds. Because the bonding achieved at room temperature is typically relatively weak, wafer pairs bonded

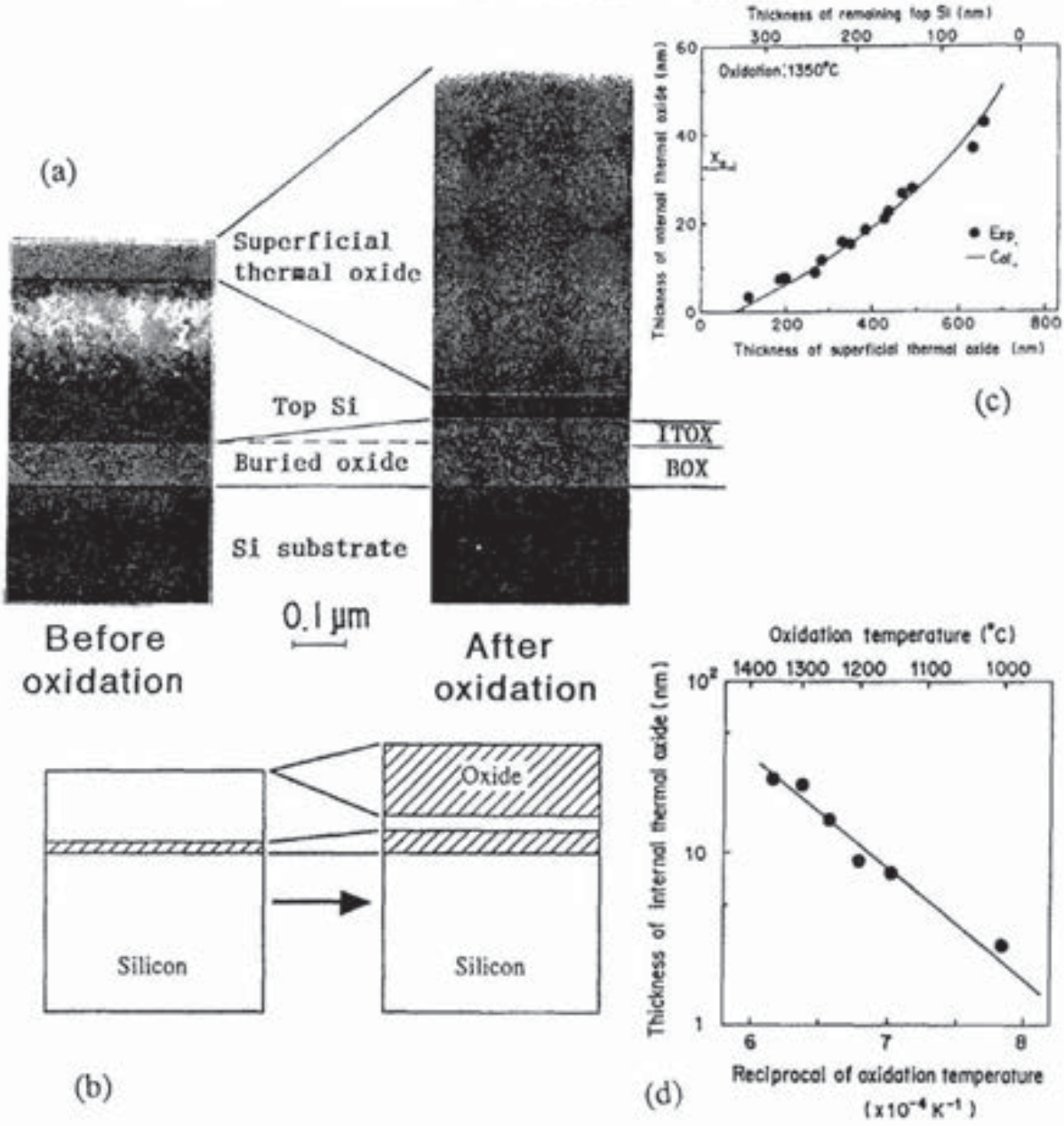


Fig. 11-19 a) An XTEM of SIMOX wafers before and after high-temperature-oxidation at 1350°C; b) A schematic drawing of SIMOX wafers before and after high-temperature oxidation at 1350°C; c) Relationship between the thickness of the ITOX layer and the superficial thermal-oxide-layer formed by high-temperature-oxidation at 1350°C; d) Relationship between the thickness of the ITOX-layer and the reciprocal of the oxidation-temperature.

at room temperature are usually subjected to a heat treatment to strengthen the bonding interface.

In the application of the bonding-process to form SOI-substrates, two thermally-oxidized-Si wafers (or one oxidized-wafer and one bare-Si-wafer) are first

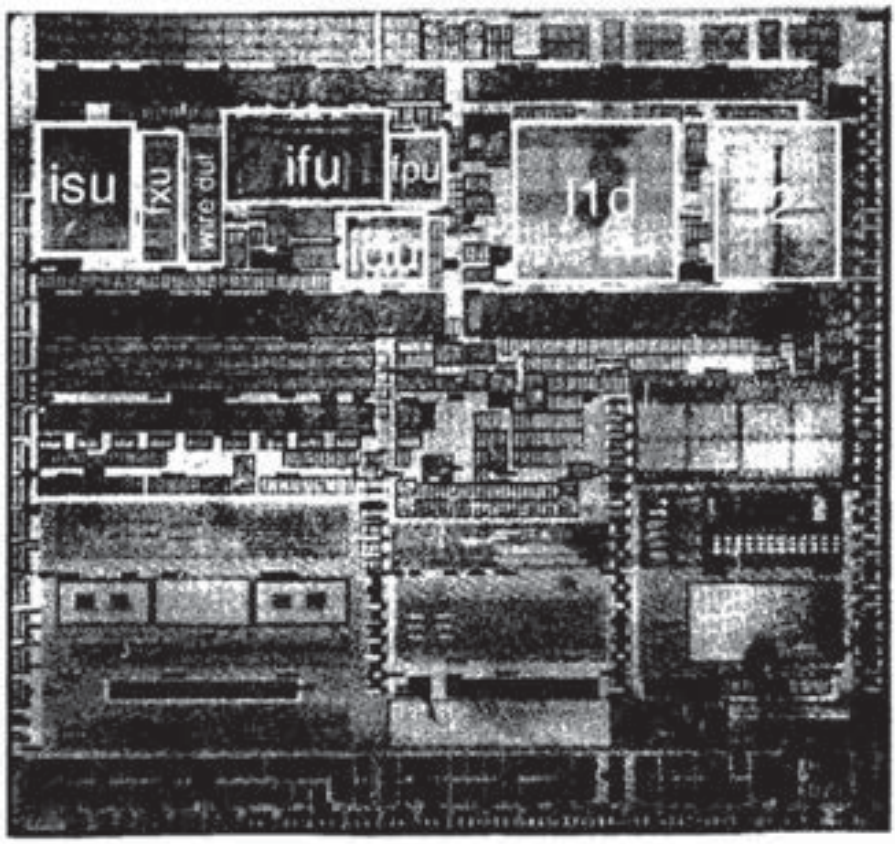


Fig. 11-20 Photograph of an IBM Power4™ processor using a SIMOX-SOI-based process.

first bonded by applying pressure to the wafers at room temperature (Fig. 11-21). A subsequent furnace anneal is used to increase the bonding strength. After such bonding, most of the silicon from one of the wafers is removed, leaving a thinned

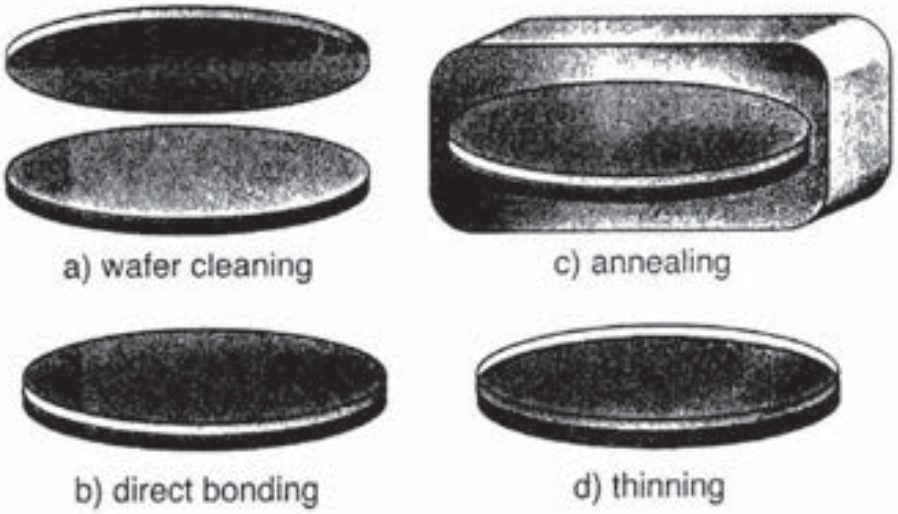


Fig. 11-21 Generic direct-wafer-bonding process flow.

layer of silicon from one of the wafers. The IC devices are built in this remaining thinned-silicon-layer, which now exists over the oxide present at the interface of the bonded wafers (i.e., a buried-oxide). The wide-availability of chemical-mechanical polishing (CMP) in IC processing, as well as a variety of other thinning approaches has led to wide and diverse use of wafer bonding. Unlike SIMOX, the wafer-bonding technique can form BOX-layers up to 50-μm thick. A book has recently become available dealing with the topic of wafer bonding.[33]

There are several approaches for achieving bonding. In one method, two oxidized-silicon-wafers are held together (face-to-face) by applying external pressure to the wafers while they are exposed to an oxidizing-ambient at about 700°C.[34] Bonding is believed to take place by the conversion of oxygen trapped between the wafers, into SiO_2. The conversion consumes the oxygen and creates a partial-vacuum that forces the wafers into intimate contact, allowing the chemical-bonding-reaction to proceed. The chemical-reaction is believed to be the polymerization of silanol bonds to form siloxane networks.

A second approach applies a moderate voltage across two oxidized-wafers placed face-to-face at temperatures in the range of 1100–1200°C.[35] Since no mechanical force is required when the wafers are first placed in the furnace, this method is less likely to introduce crystalline-defects. A third approach uses one oxidized-wafer and one bare-wafer. Both are treated with an acid solution that forms O-H groups on the surface of the wafers. After drying the wafers are placed face-to-face at room-temperature in clean air. After this treatment a self-adhesive contact is formed which holds the wafers together during the bonding process. The bonding is completed by annealing the two wafers at 1100°C in a nitrogen-ambient for 4 hours.[36]

Recently it has been demonstrated that bonding can occur at temperatures as low as 150°C, if the process is performed in a moderate vacuum (e.g., several torr).[37] It is believed that the vacuum significantly reduces the amount of trapped nitrogen at the interface. Trapped nitrogen can prevent an intimate contact of the bonding-surfaces during annealing, and hinders formation of the required covalent-bonds. It is also possible to achieve full-bonding-energy directly at room-temperature under ultra-high vacuum conditions if the wafer surfaces are free of adsorbates.[38]

It should be mentioned that particles or protrusions on either surface being bonded, do prevent bonding, (at least locally) around these disturbances. These lead to locally-unbonded-areas that are frequently termed *interface bubbles* or *voids*. Such voids are, of course, undesirable, since they can lead to delamination of devices located in imperfectly-bonded-areas during chip-assembly and

packaging. Voids with a diameter of several millimeters are readily created by particles 1-μm or less in size (and these are easily revealed by infrared or ultrasonic imaging). Therefore particles on wafers have to be avoided by proper cleaning and by carrying out all bonding procedures in an ultra-clean environment (Fig. 11-22).

11.6.3.1 Bond-and-Etch-back SOI (BESOI): Once the wafers are bonded it is necessary to remove most of the silicon from one of the wafers to produce the SOI structure. There are several proposals being evaluated for this removal process. The *bond and etch-back SOI* (BESOI) technique is used to produce relatively thick-films of both oxide and silicon. The BESOI wafers typically have thicker, yet better-quality silicon and buried-oxide-layers than do SIMOX wafers.

The silicon-layer on BESOI substrates can be thinned in one of several ways. Mechanical grinding and polishing can be used to thin the Si to a thickness down to ~1-μm, but only to within 10-30% uniformity. Better uniformity can be achieved by using an etchstop. Several etch-stoplayers have been developed, including: a boron-doped-layer, a silicon-germanium-layer, a carbon-implanted-layer, and a porous-silicon-layer.

In the first of these etchstop methods, a lightly-doped B-region can act as an etchstop-layer when it is formed on a heavily boron-doped-substrate (Fig. 11-23). That is, the wafer that will ultimately contain the device consists of a lightly-B-doped epitaxial-layer on a heavily-B-doped-wafer. After oxidation and bonding of the lightly-doped surface of this wafer to another wafer, the heavily doped

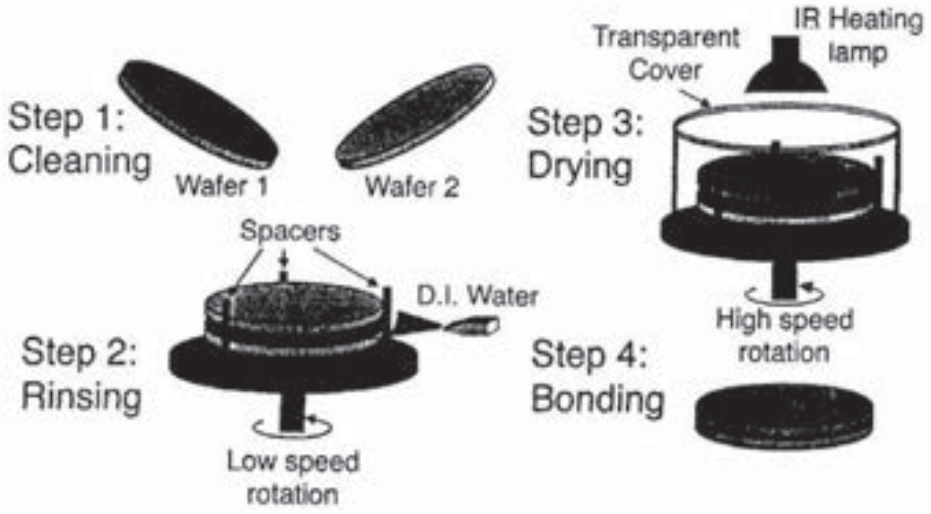


Fig. 11-22 Schematic of the microcleanroom setup and procedure.

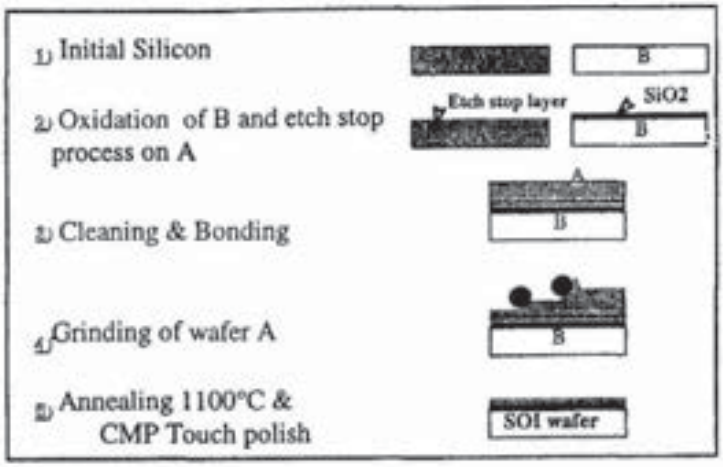


Fig. 11-23 Schematic of the Bonded-and-Etched-Back SOI technology (*BESOI*).

substrate is removed by an etch process that preferentially etches such heavily doped silicon. The epi-layer that remains bonded to the second wafer determines the thickness of the active-device-layer and the sum of the thickness of the oxides on the two wafers determines the thickness of the buried-oxide. One of the issues with this approach is that two silicon-wafers are required to make one device-SOI- wafer. Another is that the uniformity of the thickness of the silicon-layer depends on the etch-stop-selectivity (which can vary, depending on the etch-stop method used, from 10 to 10^5).

Such BESOI methods are widely used for "thick" SOI applications (i.e., > 1-μm), such as power devices (high-voltage and high-temperature devices) and micromechanical machines (MEMS). The power-devices are used for military, industrial, and aerospace applications. The share of the thick-film segment of the SOI market dropped from 80% in 1995 to 53% of the market in 1999 and was taken over by the thin-film segment. Future thick-film SOI applications will be driven by industrial and automotive applications.[54]

11.6.3.2 Hydrogen-Implantation-Induced Layer Splitting to Create SOI Substrates (SMART-CUT®): A recent (and elegant) wafer-bonding-based method has been developed for forming SOI substrates using a technique to obtain very-uniform and thin regions of single-crystal-silicon. This technique employs a process called *hydrogen-implantation-induced layer splitting*.[39] The process is also referred to by its inventor M. Bruel (and by the company that sells such SOI

substrates [SOITEC SA, France]), as *Smart-Cut.*® The Smart-Cut process also allows the sacrificial-wafer to be reused and only one wafer is therefore required to make a bonded SOI substrate. (The SOI material made with the Smart-Cut process is termed UNIBOND.®) There are two reasons why this process may be significantly less-expensive than the process used to fabricate SIMOX-wafers. First, the sacrificial wafer can be re-used. Second, because the implant-dose is an order of magnitude lower than even one in the low-dose SIMOX-process (and the implant is done at room temperature in a conventional implanter), the cost of the implant step is significantly lower, compared to those used in SIMOX.

Figure 11-24 shows how the process works. Two device-quality-wafers (labeled A and B) start the process. Wafer-A is oxidized (to a thickness of about 0.5-μm) to form what will become the buried-oxide of the SOI-structure. Then this wafer is ion-implanted with a high-dose of hydrogen ions ($\approx 5 \times 10^{16}$ H^+ ions/cm^2) through the oxide, using a conventional high-current-implanter. The peak-concentration of this implanted hydrogen-dose is at a projected range R_p of about 0.2-μm below the oxide-surface. At this point, only micro-cavities or micro-bubbles are formed at the R_p of the implantation. For reasons explained next, the depth of the H^+ ion implantation determines the thickness of silicon above the BOX.

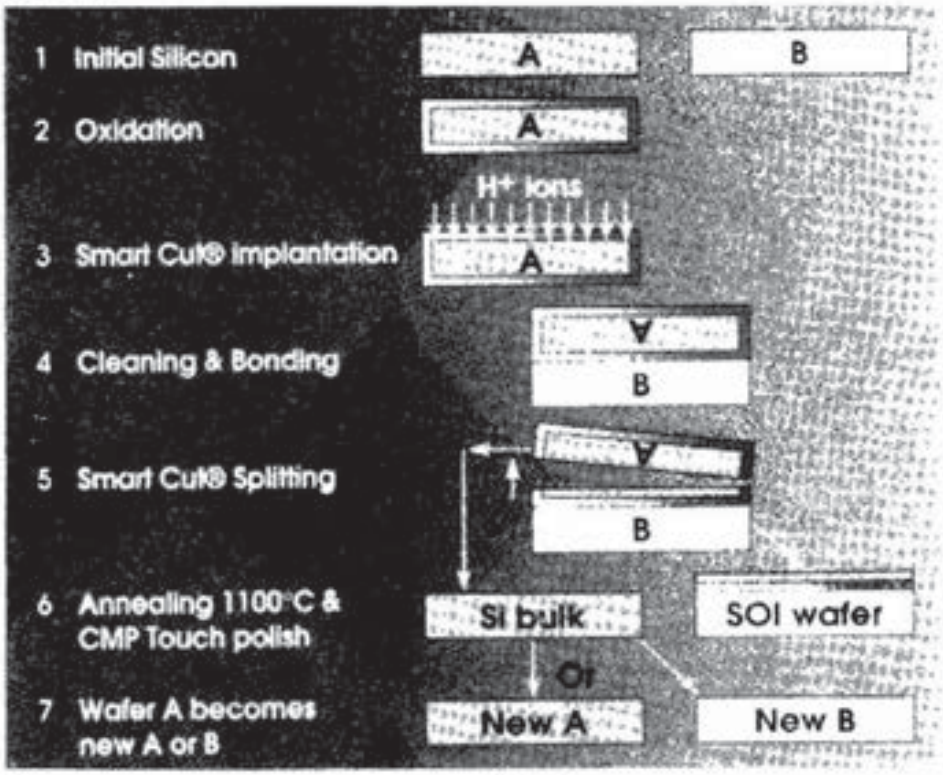


Fig. 11-24 Process flow for fabricating Smart-Cut® wafers.

After implanting the hydrogen into Wafer-A, both Wafers A and B are cleaned and low-temperature bonded face-to-face. The bonded wafers are next annealed (at 400-600°C). During this anneal the implanted-hydrogen precipitates, and this precipitation-process merges the buried network of bubbles and voids at the vicinity of the depth of R_p. As these hydrogen-filled microcavities continue to grow and merge, they eventually form a continuous microcrack along the line corresponding to the peak of the implanted-hydrogen. At that point Wafer-A delaminates from along that line, and Wafer-B (which now has a BOX-layer covered by the silicon layer from Wafer-A) becomes the SOI wafer. Meanwhile, Wafer-A remains available to be used as the sacrificial wafer of another UNIBOND substrate. Both wafers are then annealed at 1100°C to remove defects. This is followed by a "light" CMP (or "touch-polish") to erase the surface-roughness resulting from the layer-splitting process. Wafer-A, which is now only a fraction of a micron thinner than it was before the start of the process is used as Wafer-B when the process starts over. The use of ion implantation for layer separation improves the layer-thickness-uniformity over that of bond-and-etch-back SOI-wafers. The method is used to prepare commercially available 200-mm SOI wafers with a silicon thickness above the BOX in the 100-nm range and a uniformity of better than 5%. The dislocation density is $< 10^2/cm^3$, the surface roughness is < 2 Å (rms), and metal contamination is $< 5x10^{10}/cm^3$.[40]

11.6.3.3 Nanocleave® Wafer-Bonding SOI: The fabrication of starting wafers for SOI can also be carried out in other ways. One of these processes, named *Nanocleave®* was introduced by Silicon Genesis Inc. in 1998.[43] The process combines growing a SiGe/Si epitaxial layer, low-temperature wafer-bonding, and a room-temperature cleaving-process to produce sub-100-nm-thick layers of single-crystal-silicon over a buried-oxide substrate. One of the key advantages of the Nanocleave process is that it can produce such thin SOI-layers with a uniformity of several percent over an entire wafer. This type of uniformity is necessary for the successful fabrication of fully-depleted CMOS-SOI-FETs. Other thinning techniques (such as etch-back or CMP have increasing difficulty maintaining such SOI-layer thickness-control when silicon layers of sub-100-nm thickness are being produced. The Nanocleave process uses an atomic-scale cleaving-process that produces cleaved surfaces with an as-cleaved surface roughness less than 5-Å (as measured by AFM). In addition, the Nanocleave process potentially offers a method that is lower in cost than any of the other fabrication processes used to fabricate SOI-substrates. The Nanocleave technique also offers another path toward the capability of fabricating both "ultrathin" Si-

SOI-layers (10-to-50-nm-thick Si) and "nanothin" Si-SOI (5-to-10-nm-thick Si) required for the MOSFETs with gate-lengths smaller than 50-nm that are scheduled to begin production by 2010.

The Nanocleave process begins with the formation of a *cleave-plane*, the growth of a thermal-oxide-layer and Si-device layer on a high-quality "donor" wafer (Fig. 11-25a).[44] This "donor" wafer is then bonded to an inexpensive oxidized "handle" wafer (on which the device layer will ultimately reside).

The cleave-plane is created by first epitaxially growing a layer of Si-Ge on the "handle wafer, and then continuing the growth of this layer with an epitaxial-Si device layer (Fig. 11-26a).[63] The buried Si-Ge-layer is a compressive-strain-layer of a non-dislocated Si-Ge alloy, which contains no porous-regions, inclusions, or voids, as shown in Fig. 11-26a. When device wafer is separated, the resulting separated surface has a surface roughness, as measured by AFM profiling of 0.2-nm, or less.

After the cleave plane has been formed and a thermal-oxide is grown on the surface of the wafer with the cleave-plane (which will become part of the buried-oxide-layer in the finished SOI-device), the "donor" wafer is bonded to the

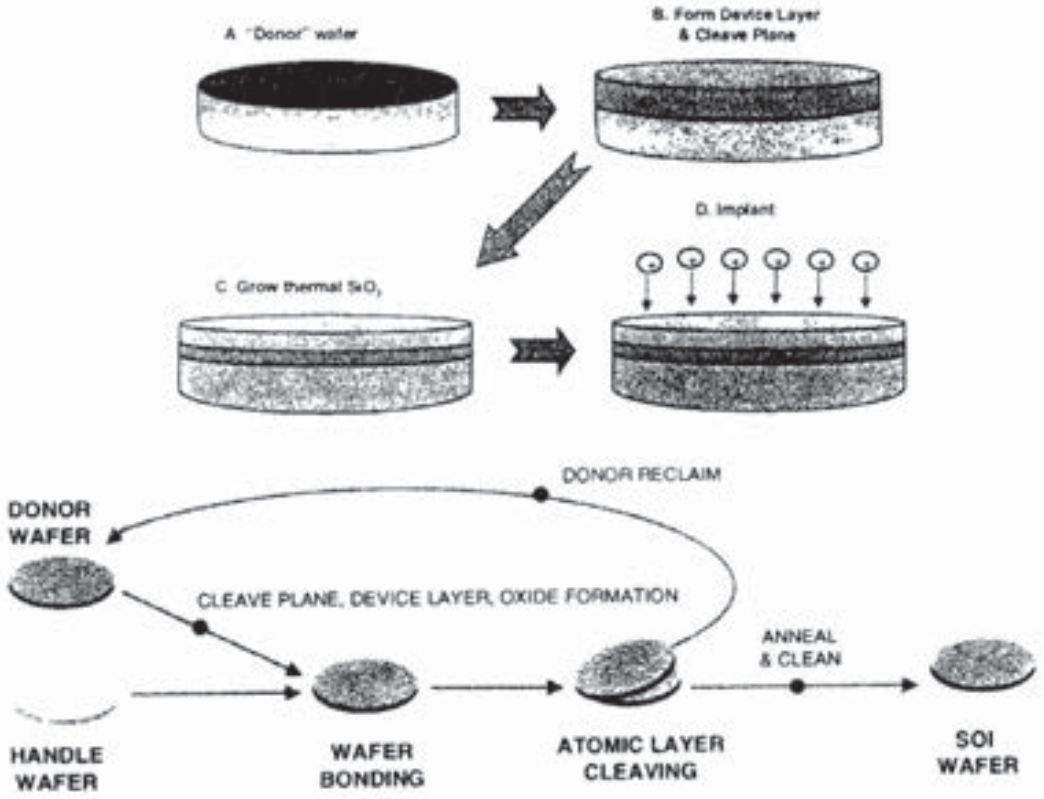


Fig. 11-25 Process flow of the Nanocleave® SOI technology.

"handle" wafer (which also contains a surface-oxide layer). Plasma-bonding techniques are employed to increase the initial bond-strength to sufficient levels for the room-temperature process used to transfer the silicon-layer from the "donor" to the "handle" wafer. This transfer process (by cleaving the device layer from the "donor" wafer from the "handle" wafer) is carried out next (Fig. 11-25b). Since the bond-strength is significantly greater than the interface-strength along the cleave plane, a novel cleaving process is used to separate the Si device layer from the "donor" wafer (which results in the creation of an SOI structure on the "handle" wafer). In the Nanocleave process, the donor wafer cleaving is done at room temperature (unlike in the Smart-Cut-process that uses an elevated temperature anneal for the cleaving step). In the Nanocleave process, the drive force is a static pressure of N_2 gas. A mechanical fixture secures the bonded wafer pair as a high-pressure nitrogen-gas is introduced along the strained Si-Ge layer. That is, cleaving is controlled by fracture along a cleave-front within the Ge-rich Si-layer. A thin-film of material from the "donor" wafer remains bonded to the substrate "handle" wafer, producing an SOI-structure. The result is an SOI-wafer with a silicon-layer whose thickness uniformity is within 1% across a 200-mm wafer. Furthermore, the atomic-layer cleaving-process is claimed to produce an as-cleaved surface roughness of less than 1-nm (typically 2-5 Å).[63] This is an order-of-magnitude smoother than the 80 Å of hydrogen-induced thermal cleaving obtained with the Smart-Cut technique. The smoother surface in the Nanocleave process is acceptable for many IC applications with no additional

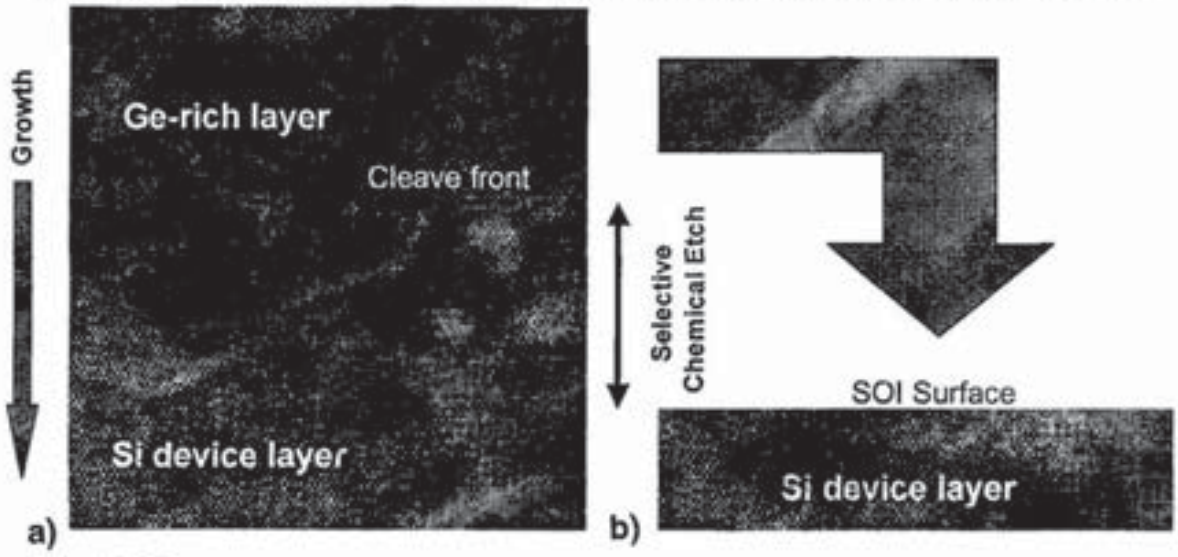


Fig. 11-26 a) HR-TEM image of the boundary area of a Ge-rich, epitaxial strain-layer and a crystalline-Si-device-layer in a state prior to atomic-layer-cleaving; b) The residual Ge-rich layer on the as-cleaved-wafers are subsequently removed through a selective etch step.

surface polishing. The edge of the SOI-layer has a smooth and regular character without the need for edge polishing. The residual Ge-rich layer on the as-cleaved wafers is subsequently removed through a selective etch step (11-26b).

A fully automated tool that separates bonded donor-handle wafer-pairs is available for this purpose. The high-quality wafer can be used to produce many thin-film transfers, reducing the process cost. Another factor that contributes to further cost-reduction arises from the fact that the "handle"-wafer is separated from the device-layer by the buried-oxide-layer. Consequently, a "handle" wafer with considerably relaxed electrical and chemical specifications (and therefore lower cost) can be used in this process.

Post-cleave-processes for layer-thickness-adjustment include: **1**) Si layer thinning using etching processes; and **2**) layer-thickening by growth of additional Si or Si-Ge multilayers. The layer-thickness-control for thickened wafers is better than 0.5% with a within-wafer thickness-uniformity of 2% or better (Fig. 11-27).

11.6.3.4 ELTRAN® Wafer Bonded SOI

A relatively new SOI structure (originated and developed by CANON Inc.), which is also based on wafer bonding, is called *ELTRAN.*® The name ELTRAN comes from *Epitaxial LAyer TRANsfer by BESOI* (Bond and Etch-back SOI method) with selective-etching of porous-silicon (Fig. 11-28).[45,46] ELTRAN wafers became commercially available in 1997 for 200-mm substrates.

The porous-silicon surface on an otherwise bare-silicon-wafer is formed by an anodizing treatment of the silicon surface. The ELTRAN method is characterized as BESOI implemented by growing an epitaxial-layer on a layer of porous-silicon. (It has been demonstrated that high-quality epitaxial-layers can be grown on porous-silicon.) The surface of the epitaxial-layer is then thermally oxidized, and this layer will eventually become the buried-oxide-layer of the SOI-structure. This wafer (with a thermally-oxidized epitaxial-film grown on porous-silicon) is then bonded to a "handle" wafer as the "donor" wafer. In early ELTRAN wafer fabrication process, the "donor" wafer material was removed after bonding by a grinding process (down to the porous-silicon-layer). The exposed porous-silicon that remains after the grinding-process is then removed by a wet-chemical etch. A mixture of HF and H_2O_2 is used for this etching step, since it provides an etch selectivity as high as 105:1 between the porous-silicon and the epitaxial-layer. At the end of the process, only the epi-layer remains on the top of the oxide layer and handle-wafer. The ELTRAN process has been shown to produce 470-nm-thick SOI-layers with a 14-nm thickness uniformity.

In the more recent ELTRAN process (Fig. 11-29), a double porous-silicon

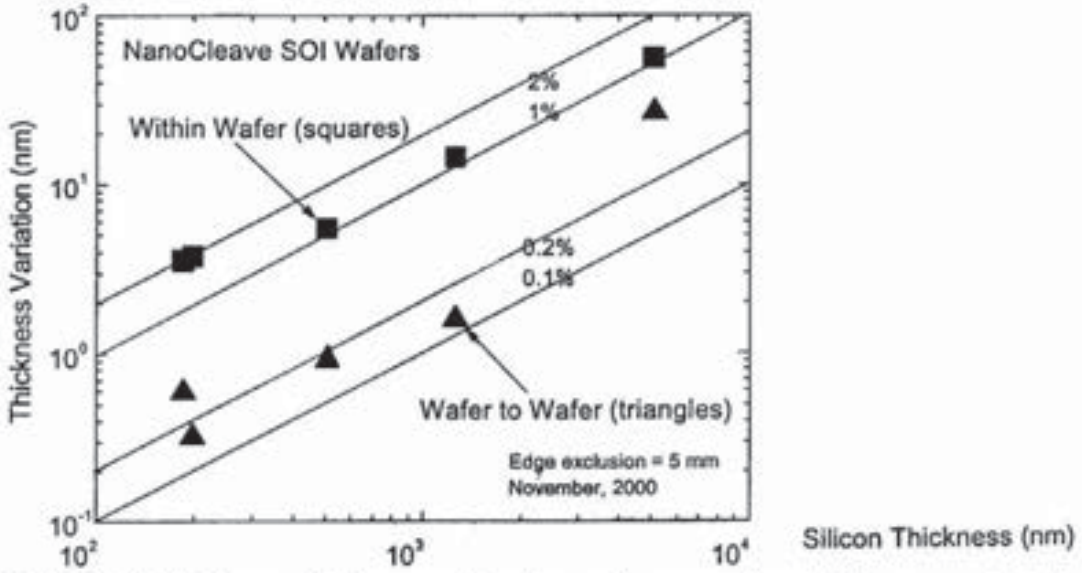


Fig. 11-27 Si-SOI layer-thickness-uniformity and average layer-thickness-control for post-cleave thickening of as-cleaved SOI-wafers with cleaved 200-nm Si-SOI-layers.[63]

layer is grown on the "donor" wafer, with each of these porous-silicon-layers having a different porosity. The porosity of the porous-silicon-films on the wafer surface can be adjusted by changing the value of the anodization current. After bonding it becomes possible to physically separate the "donor"-wafer from the

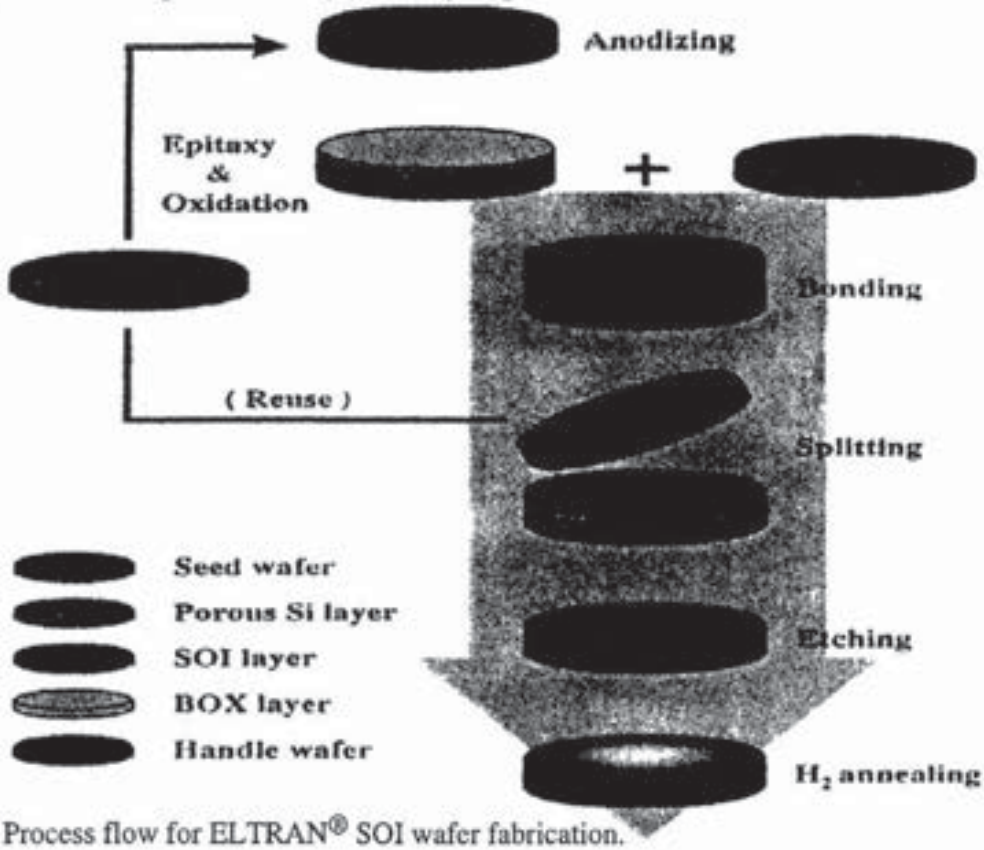


Fig. 11-28 Process flow for ELTRAN® SOI wafer fabrication.

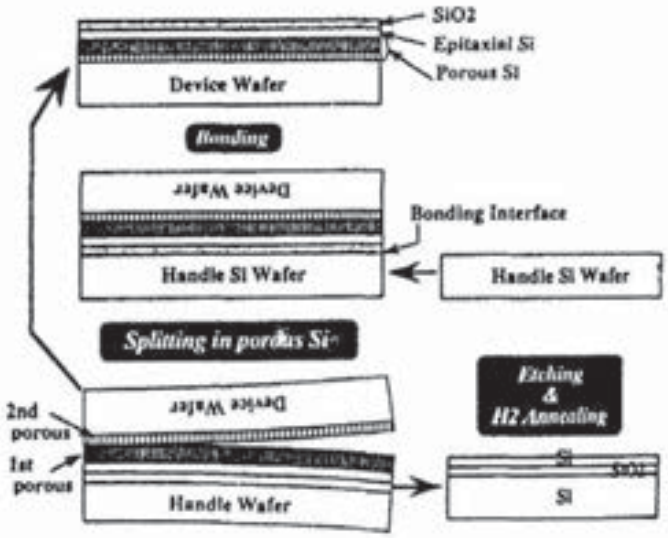


Fig. 11-29 The ELTRAN process flow for device wafer re-usage. Bonded pairs are split within the porous-Si layer. The device wafer sides are used as a starting device wafer. The handle sides result in ELTRAN wafers after etching the porous-Si and H_2-annealing.

"handle" wafer using a physical splitting process (i.e., a water-jet slices the SOI-layer from the seed layer).[62] The splitting occurs along the interface between the two porous-silicon layers. That is, since the porous-layers are a double-stacked structure with different porosities, the strain-energy accumulated at the interfacial area limits splitting to an extremely narrow range (because the interface between the first and second porous layers is mechanically the weakest point in the structure). Therefore, the stress-configuration induced by the double porous-layers is responsible for the splitting mechanism.

For splitting a "water-jet" technology is used. The water jet – a small collimated water stream – is injected at the edge of the bonded-pair and splits the two wafers at the thin porous-silicon layer. The double-layered porous silicon structure contains the stress concentration, with the result that both of the split wafers (SOI and seed wafer) are covered by the porous silicon protective layers. The water-jet drives high-pressure water into narrow gaps by fluid wedging and applies uniform pressure, reducing the chance of damage. The splitting yield is reported to be over 90%, and automatic water-jet splitting-machines are in production use.

The device layer over a buried-oxide is then formed by the etchback of the porous-silicon. After the porous-silicon-film is removed, the surface of the device silicon-layer is flattened by hydrogen-annealing (see Chap. 2, for a discussion of hydrogen-annealing of Si wafers).

There is another economic benefit of the more recent, double-porous-silicon-layer process used to form ELTRAN wafers. In this process (and like in the Smart-Cut and Nanocleave processes discussed in the previous sections), after the splitting process, the "donor"-wafer becomes available to be used as another a "handle"-wafer (instead of being ground away, as in the early ELTRAN process). Canon announced that it has expanded the production of ELTRAN wafers from 2000-per-month to 10,000-per-month in November 2000. This production expansion is being supported by a greater marketing of ELTRAN wafers. To this effect Canon opened a marketing office in the U.S. to achieve this goal.[53]

11.7 LATERAL ISOLATION IN SOI

SOI-CMOS-structures eliminate all inter-device leakage-currents by building *n*- and *p*-type islands on an insulating substrate for *p*- and *n*-channel transistors, respectively. Field-inversion and device-to-device-punchthrough do not exist in SOI-technology, but new parasitic transistors in an SOI-structure can still give rise to leakage-currents. SOI-CMOS-technology depends on high-quality isolation. The methods used to laterally isolate SOI structures will be described here, as well as a description of the leakage paths that arise as a result of each of them. The three techniques for providing lateral isolation in SOI are: **1**) MESA isolation; **2**) LOCOS isolation; and **3**) shallow-trench isolation (STI).

11.7.1 MESA Isolation

MESA isolation was the early method used to create islands of silicon on an insulating substrate (Fig. 11-30). The technique is attractive because of its simplicity. That is, in MESA isolation the active-device-regions (the silicon islands) are formed into a "mesa" by using a mask-step and a silicon-etch step to remove the silicon in the field-areas down to the insulating-substrate. The SOI oxide helps as the etch-stop-layer, while anisotropic-etching allows for minimization of the distance between adjacent mesas. Passivation of the islands (mesas) is performed at the gate-oxidation step, where the gate-dielectric is grown not only on top of the silicon islands, but on their edges as well.

Although the MESA isolation method is simple, it also has some inherent problems. The first involves the etching of the mesas and the thermal-growth of the gate-oxide that is subsequently grown over the entire mesa. That is, the etching of the mesas produces silicon islands with sharp upper-corners. When the gate oxide is grown on these islands, the oxide at the top corner is thinner than on the flat areas of the mesa (typically 30-50% thinner than on the top surface - see

Vols. 1 and 2 for a discussion of oxide-thinning at corners during thermal-oxidation processes.) When the gate-material is run over the mesa and voltage is applied to turn-on the MOSFETs built in the mesa, the thin-oxide at the corner can cause premature gate-oxide-breakdown. The gate-oxide-breakdown-voltage may also be reduced on the sidewalls, since the crystal orientation may be different from that of the top surface of the silicon wafer. The growth-rate of SiO_2 is dependent on the silicon-crystal-orientation of the surface on which it is being grown. Furthermore, crystal-orientations other than (100)-Si generally have larger fixed-oxide-charge-densities and surface-state-charge-densities than (100)-Si, and this can decrease the threshold-voltage of MOSFETs built on such mesas.

The second problem relates to the steep sidewalls produced by an anisotropic-etch of the silicon-mesas. When a polysilicon-layer (used to form the MOSFET gate) is deposited on the mesa (typically to a thickness of 400-nm), an anisotropic-etching of the poly will leave stringers along the mesas, and this may short the gates of devices within the same mesa. Overetching is therefore needed to remove such stringers, but this can lead to poor dimensional-control of the poly gate-lengths.

The third problem involves early inversion along the top of the mesa sidewalls, which behave like lateral (edge) transistors in parallel with the top (main) device. This is due to charge-sharing of the depletion-regions caused by the poly running over the top and sides of the mesa-sidewalls. Thus, the corners show a lower threshold-voltage than the main channel of the MOSFET. This manifests itself as a kink in the subthreshold-characteristic of the drain-current. Such a deterioration of the subthreshold behavior is most pronounced in NMOSFETs in SOI-mesas, and can be alleviated by performing a boron-implant into the field-regions of the silicon wafer before etching the mesas. This boron-implant is then driven-in laterally into the active-area. The mesas are then etched, leaving the sidewalls more highly doped than the active-channel. Figure 11-31 shows a perspective drawing of a laterally-isolated MESA SOI-NMOSFET with a T-gate and a body-substrate contact.

11.7.2 LOCOS-Isolation for SOI

LOCOS-isolation is the second technique that has been used to produce lateral-isolation among active-silicon-islands in SOI. LOCOS is more complicated than MESA isolation, but it has the advantage of already being in wide use in bulk-CMOS technology. LOCOS-isolation in SOI is much the same as LOCOS in bulk-CMOS (see Vols. 1, 2, or 3 of *Silicon Processing for the VLSI Era* for more details on the LOCOS process.) However, there are some issues that do arise

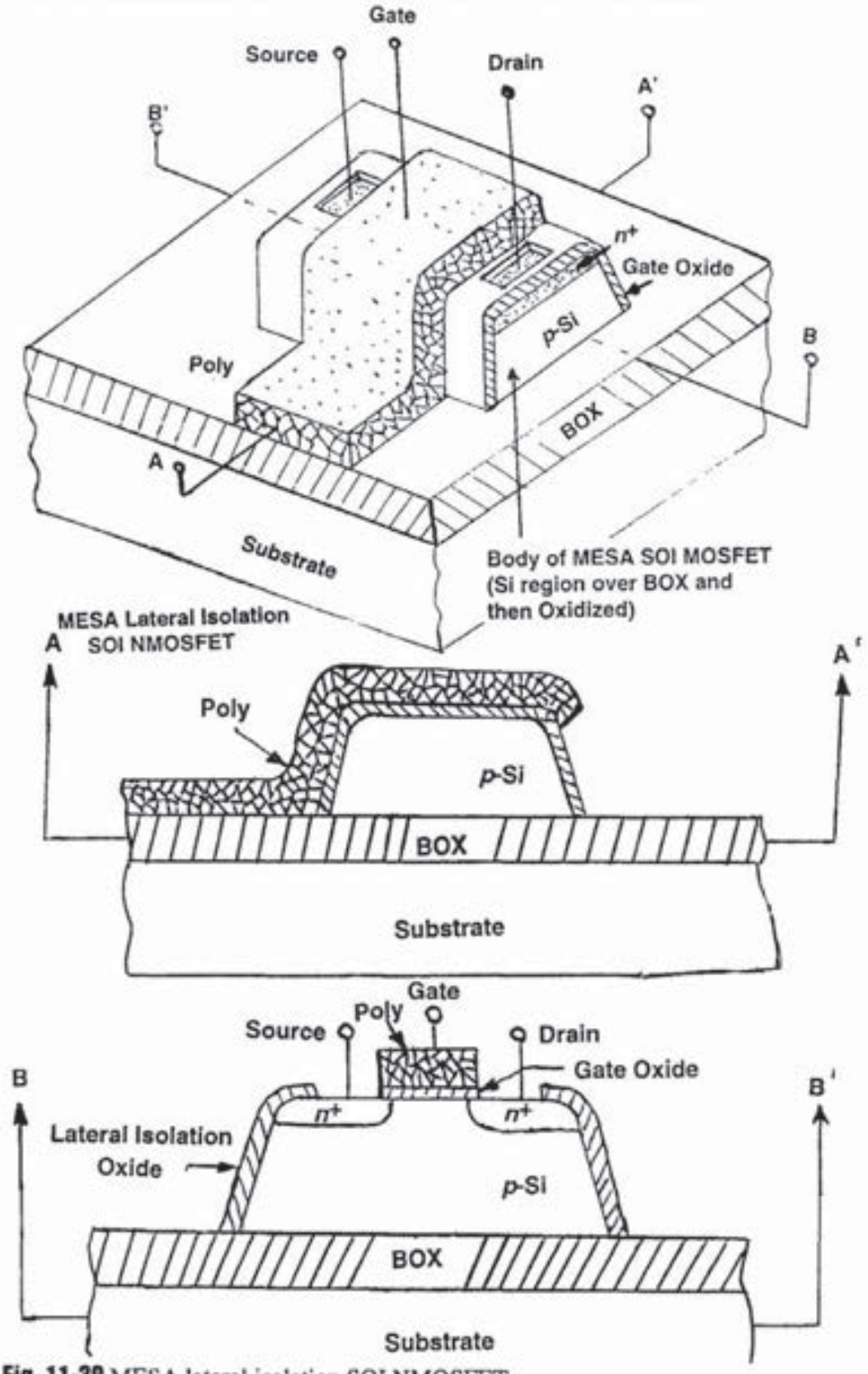
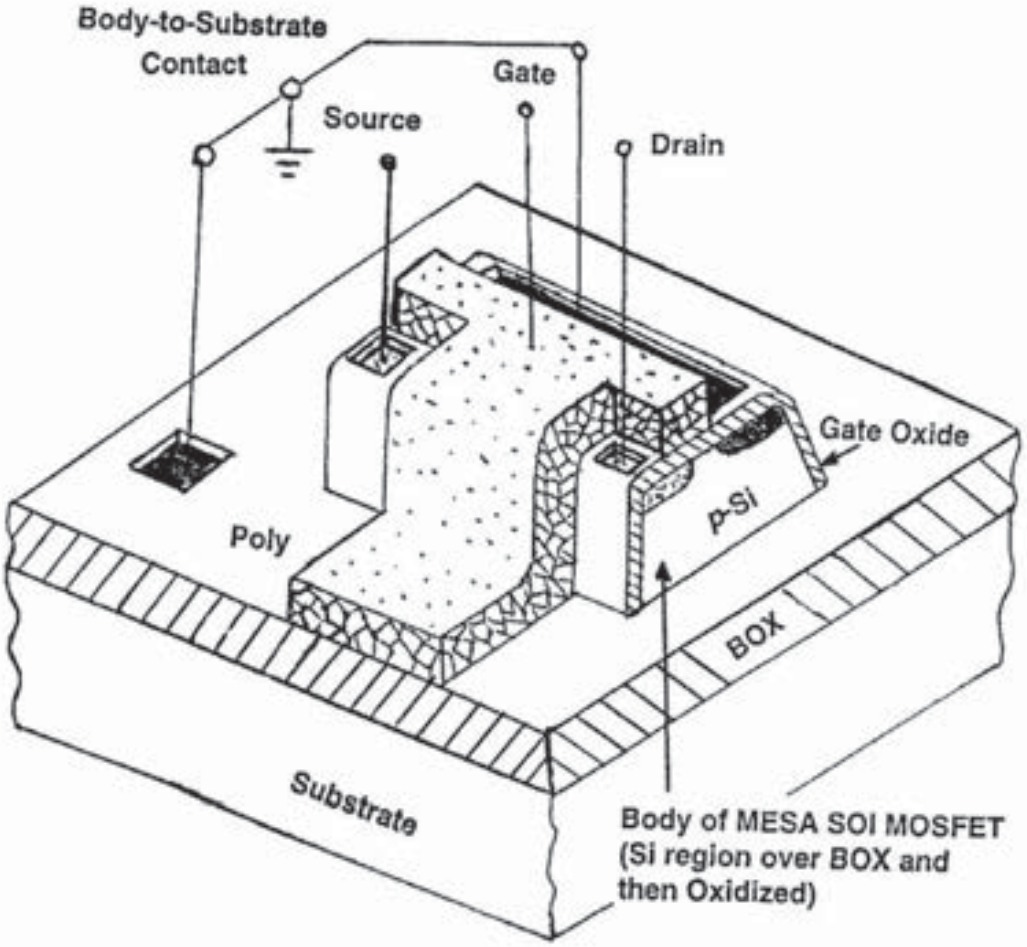


Fig. 11-30 MESA lateral isolation SOI NMOSFET.



11-31 Mesa lateral-isolation SOI NMOSFET with a T-gate and body-to-substrate contact.

when LOCOS is applied as the technique for producing lateral-isolation in SOI. These issues are discussed in this section.

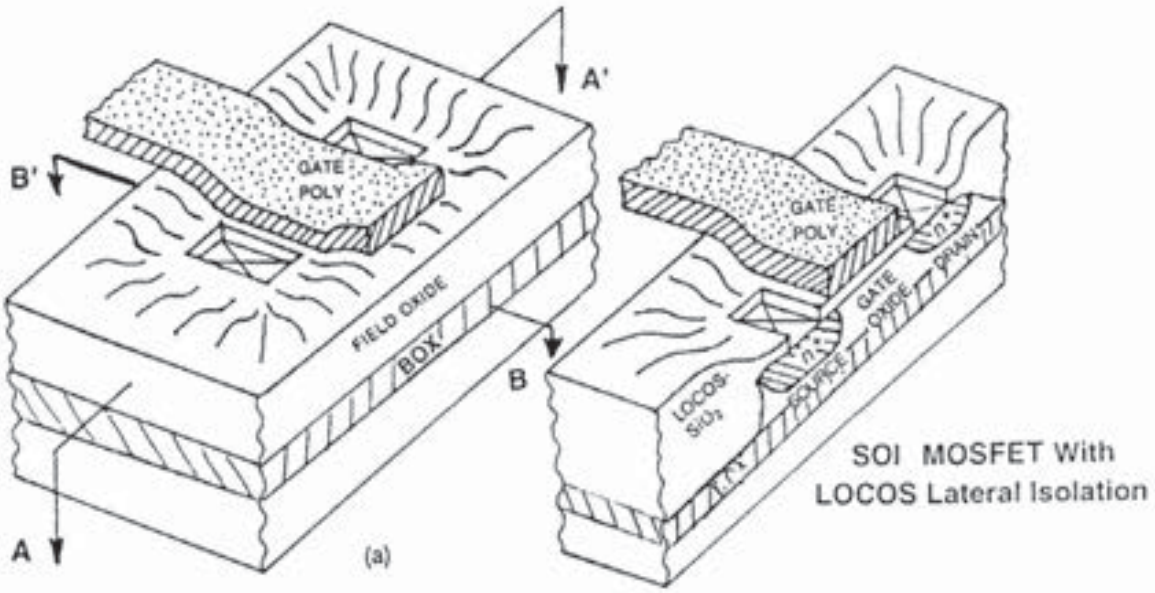
The LOCOS lateral-isolation-technique in SOI consists of the following steps (for producing NMOSFETs): a thin pad-oxide is grown, and a silicon-nitride-layer is then deposited by LPCVD onto the pad-oxide. Using a masking step, the nitride is patterned to define the active-silicon-areas. Boron is then implanted around the islands of *p*-type silicon that will contain the NMOSFETs (i.e., a channel-stop-implant). Next, the field-oxide is thermally grown. No oxide growth occurs over the active-areas which are protected by the silicon-nitride. A nitride and pad-oxide stripping step completes the process. In SOI the field-oxide-growth is carried out so that the thermally-grown field-oxide completely

consumes the silicon in the regions not covered by the nitride, with the thermally-grown field-oxide reaching through to the buried oxide (see Fig. 11-32).

The lateral-encroachment of the field-oxide (called the "bird's beak") is proportional to the thickness of the grown field-oxide. This encroachment has typical values of a fraction of a micron, and this hampers the use of the LOCOS technique for submicron bulk-CMOS processes (where field-oxide-thicknesses of 0.3-05 μm are grown). The problem of "bird's-beak"-encroachment in SOI-CMOS is even worse than in bulk-CMOS processes. This is because the oxidation-kinetics in SOI are somewhat different, particularly when the growing field-oxide reaches the buried-oxide. That is, consumption of silicon during oxide-growth on a bulk-silicon-wafer is 0.45-t_{ox}, where t_{ox} is the oxide-thickness. That means that a thermal-field-oxidation process that produces a silicon-dioxide-layer about twice as thick as the consumed silicon-layer must be used in a bulk-CMOS. For example, a silicon-layer of about 125-nm is consumed when a 250-nm SiO_2 field-oxide-layer is grown.

By comparison, a 125-nm silicon-layer above the buried-oxide of an SOI-wafer is not fully oxidized if only such a 250-nm silicon-oxide growth-process is used. That is, unoxidized-filaments of silicon remain in SiO_2-layers grown by the 250-nm SiO_2-process on an SOI-wafer with a 125-nm silicon-layer above the buried-oxide. To ensure that such silicon-filaments no longer remain in the silicon-dioxide-layer, an oxidation-process equivalent to one that would grow a 350-nm-thick SiO_2-field-oxide on a bulk-wafer must be used. In other words, a process that grows an oxide 3-times-as-thick as the silicon-layer consumed has to be used in SOI to completely laterally-isolate the active-regions with a LOCOS process. The longer field-oxidation-process produces even more "bird's-beak" encroachment (and more lateral-boron-diffusion from the channel-stop-implant into the active-region) than a LOCOS-process in a bulk-CMOS process.

Another problem of the LOCOS method of lateral-isolation is that source-to-drain leakage-current in the subthreshold-region of operation can also be increased by the parasitic-edge-transistor that exists along the sidewalls of the silicon-active-regions in parallel with the main-channel of the SOI-MOSFETs. The problem is most severe in LOCOS-isolated SOI-NMOSFETs for the following reason: A p^+ channel-stop-implant is carried out before the field-oxidation to prevent such edge-transistors from being turned-on prematurely and causing excess subthreshold-leakage-current. However, during the field-oxidation step, such implanted-boron segregates into the growing-oxide on the top of the silicon-surface - and into the buried-oxide as well. As a result, the boron doping at the tip of the bottom corners of the active-device islands ends up being



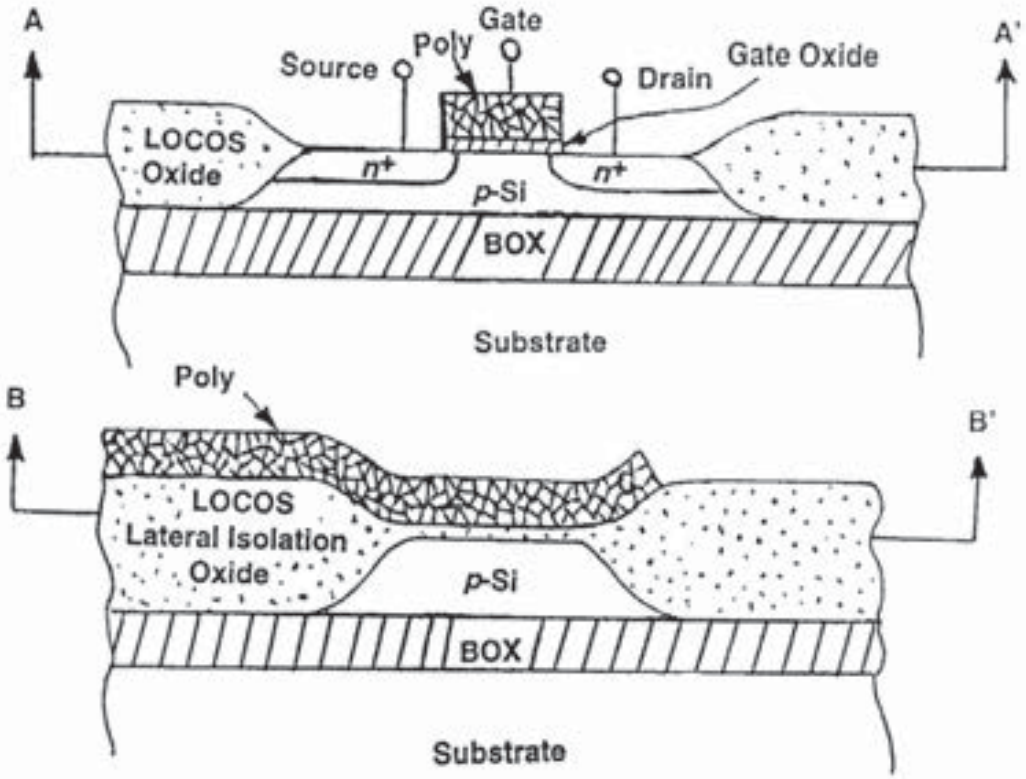


Fig. 11-32 SOI-MOSFET that uses LOCOS lateral isolation.

quite low, allowing edge-transistor-leakage to arise. One method suggested to alleviate this problem is to perform a second boron channel-stop-implant after the LOCOS field-oxide-step has been completed. Since the implant step is carried out after the field-oxide step, such B will be implanted into both active-regions and field-regions of the wafer. The energy of the implant can be adjusted to place the peak of the B-implant-dose close to the meeting point of the grown- and buried-oxides. Thus, this implant step will not impact the doping-concentration of the main MOSFET channels near the surface of the active-regions.

11.7.3 Shallow-Trench Isolation (STI) for SOI

Because of the problems associated with the MESA and LOCOS isolation techniques for providing lateral isolation in SOI technology, the more advanced SOI technologies have adopted shallow-trench-isolation as the lateral-isolation method of preference (Figs. 11-1 and 11-33). That is, the active-regions of the silicon are formed as in the MESA isolation method, but in STI the regions between the mesas are filled with a CVD-dielectric-material and then the surface is planarized by CMP to yield a flat wafer surface. Oxide remains in the trenches between the mesas, and its top surface lies in the same plane as that of the silicon-islands. STI eliminates the problems created by the sidewalls of the mesas, and the issues associated with the LOCOS-isolation method for SOI. (That is, the problem of "bird's-beak" encroachment does not occur when STI is used. Also, anomalies in the oxidation kinetics associated with field-oxide growth in SOI do not have to be addressed since the field-regions between the active devices are filled using CVD rather than a thermal-oxidation process). However, the same general issues involved with shallow-trench-isolation methods in bulk-CMOS still have to be addressed when applying an STI approach to SOI. Chapter 9 of this text provides a detailed discussion of STI technology.

11.8 SOI PRODUCT EVOLUTION

Table 11-4 presents a list of the SOI circuits that have been introduced during the International Electron Devices Meeting (IEDM) from 1989 to 1998.[64] It provides an indicator of SOI R&D efforts. The initial publication of the gain in the speed of a relatively small IC fabricated in SOI has been followed by a series of studies of performance-enhancements of larger ICs, mostly using SIMOX wafers. In general, these ICs were fabricated using partially-depleted SOI-MOSFET technologies, and silicon- and BOX-thicknesses were all comparable (and fairly thick). In 1998, IBM announced it would apply its leading-edge SOI-CMOS technology (including copper interconnects), to commercialize SOI-CMOS prod-

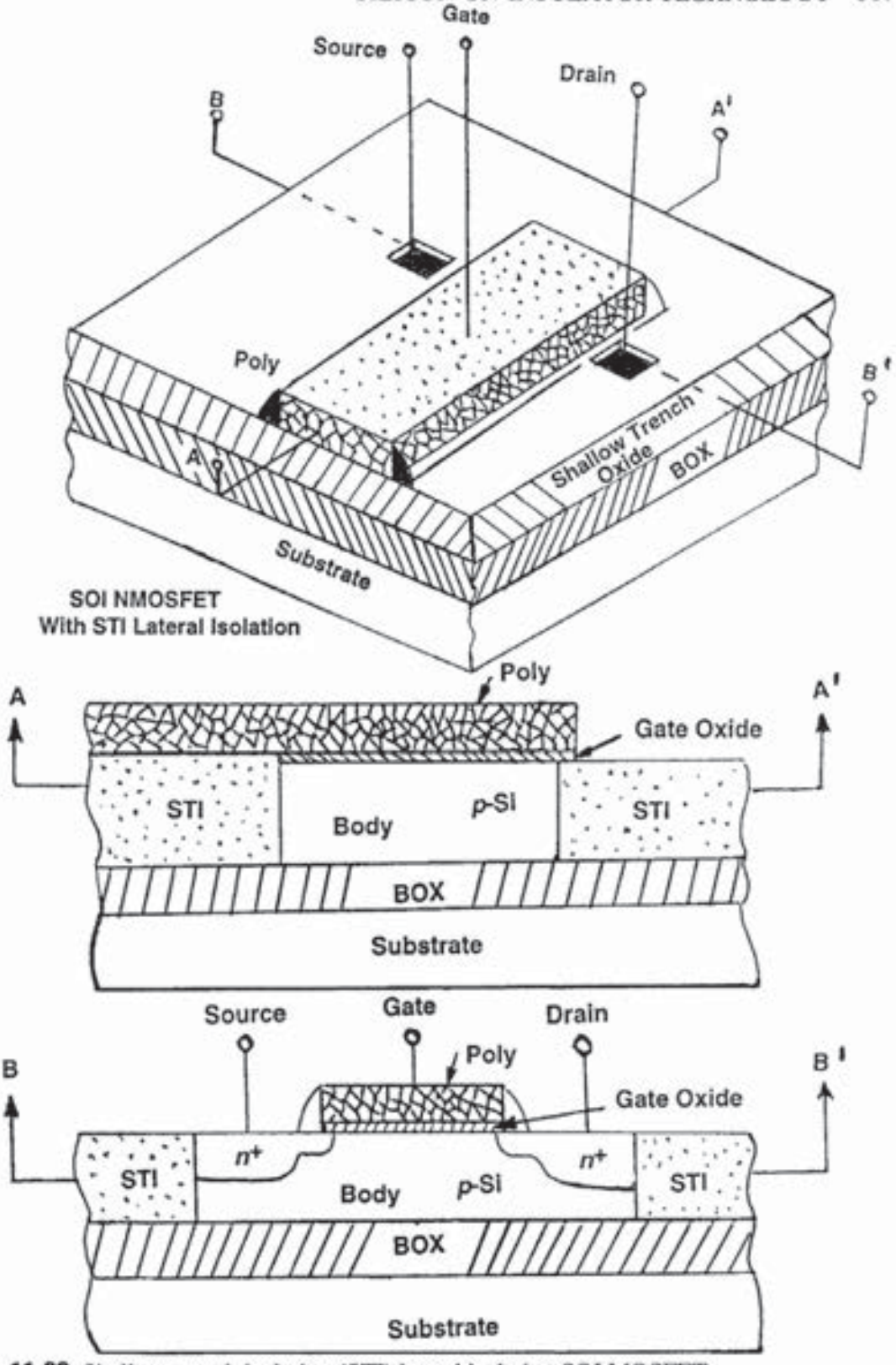


Fig. 11-33 Shallow-trench isolation (STI) lateral isolation SOI MOSFET.

Table 11-4 Summary of SOI Circuits Presented Annually at IEDM[64]

	Bell Labs	IBM	Motorola	Mitsubishi	Hyundai	DEC	IBM	Intel
	1989	1993	1995	1996	1997	1997	1997	1997
Circuit	Prescaler	0.5Mb SRAM	CPU	16M DRAM	1G DRAM	1M SRAM	4 Mb SRAM	4 Mb SRAM
Technology (μm)	0.50	0.50	0.50	0.50	0.17	0.35	0.25	Sub-0.25
V_{DD}		1.0		1.0		2.0	1.8	1.8
Silicon thickness	134		100	100	200	200	200?	150
BOX (nm)	235		400	400	250	200	400	400
Lateral isolation	LOCOS			MESA	LOCOS	STI	STI	STI
Silicon layer characteristics	PD/FD	Non-depleted	PD		Body-cont., thin Si film (75nm) beneath FOX	PD	Non-depleted SOI	PD
Metal levels	1		2			4	6	5
Performance improvement	50%						25–35%	
Power reduction						50%	170–300%	18%
Yield (if known)	Low		Same as bulk			Approaching bulk		Low
Conclusions and claims			Bulk CMOS design, performance 2 × bulk	Both body-tied and floating body MOSFETs	Not clear if circuit is functional!	20% performance improvement	Highest-performance 0.25μm CMOS technology reported FB effects negligible	No speed advantage over bulk for sub-0.25μm logic applications

ucts based on SIMOX. Figure 11-16 shows a SEM of a 0.18-μm IBM IC structure built using SOI. The company believes SOI processors are 30% faster than bulk-CMOS versions, and that they provide comparable circuit yields and add less than 10% to wafer-processing costs. Other companies have also announced the transition to SOI for high-speed microprocessors or low-power logic circuits. This is a positive development that supports IBM's claim that CMOS could move more towards SOI in the next few years (see Fig. 11-5).

In 1999, the presentation of new SOI devices occurred primarily at the *International Solid-State Circuits Conference* (instead of at the IEDM). Announcements were made concerning the successful fabrication of a number of SOI circuits in this ISSCC 1999 Conference. These circuits included 32- and 64-bit microprocessors, and 16-Mbit DRAMS on SOI. Depending on the technology and circuit layout performance, 10-40% enhancement was reported. A 0.18-μm SOI-CMOS technology for high-performance logic applications, however, was described by IBM in the 1999 IEDM. This SOI technology used a low-dose SIMOX substrate, shallow-trench-isolation, dual gate-oxide-thickness, dual threshold-voltage (V_T), dual-doped-polysilicon, cobalt-silicide, low-k dielectric

in their interconnect scheme, and seven layers of copper metallization. The SOI circuit devices were partially-depleted MOSFETs, and the test chip built using this technology contained a microprocessor-circuit that demonstrated performance above 1-GHz.[58,59] These reports indicate a new-degree of maturity in SOI technology.

The 0.18-μm SOI process of IBM was qualified by that firm for production in the summer of 2000, and work was continuing to qualify a 0.13-μm SOI process in 2001. The 0.13-μm SOI technology would allow production of high-performance devices with 44-million transistors. IBM also said that it had begun shipping its AS400 servers with 0.22-μm SOI devices early in 2000, and that their newer "Turbo" Unix machines would employ 0.18-μm SOI-processors running at well over 800-MHz. The 0.18-μm designs utilize shallow-trench-isolation, dual-work-function-poly, local-interconnect, and borderless tungsten-interlayer interconnect.

At the end of 2001, IBM reported on their 2nd generation 0.13-μm SOI process. The power supply voltage was 1.2 V, and a thinner SOI-layer and a gate-dielectric material with a higher concentration of nitrogen were used. The interconnect structure can have as many as 8 levels of metal. A completely low-k dielectric stack (i.e., with SiLK for the main dielectric material and SiC as the hardmask material and the post-CMP cap-dielectric). The Cu wires are formed using a dual-damascene process.

11.9 PROBLEMS ASSOCIATED WITH SOI

11.9.1 Floating Body Effects in Partially-Depleted SOI MOSFETs

In partially-depleted SOI MOSFETs, the depletion-region under the gate when the MOSFET is operated in inversion does not extend all the way to the buried-oxide interface. Instead, a neutral-region still exists. When the body is grounded (say, with an independent body-contact), a partially-depleted SOI-MOSFET behaves very much like a bulk-MOSFET, and most standard $I_D(V_{GS},V_{DS})$ equations and design concepts still apply. However, if a body-contact (connected by a conducting-line to ground) is not made to the SOI-MOSFET, so called *floating-body-effects* arise, some of which have a detrimental impact on the SOI MOSFET behavior. These are discussed in more detail in the following sections.

11.9.1.1 Kink Effect in Partially Depleted SOI NMOSFETs: The *kink effect* manifests itself as a "kink" in the output characteristics of partially-depleted SOI MOSFETs (Fig. 11-34). The kink appears in the graph of the drain current of

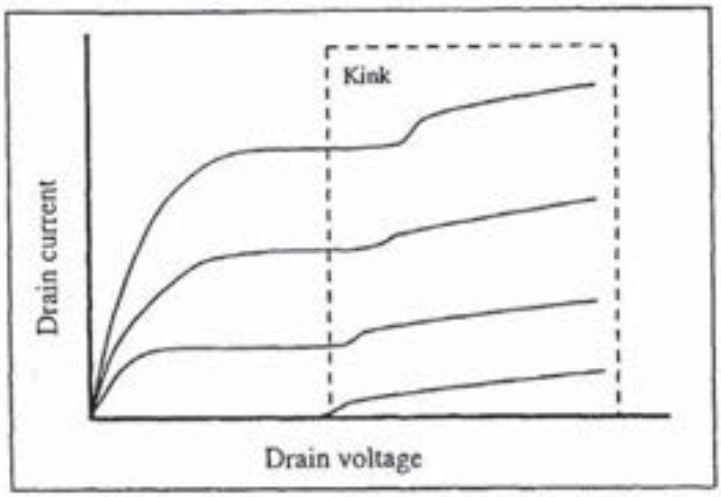


Fig. 11-34 Illustration of the kink-effect in the output characteristics of a partially-depleted SOI-MOSFET.

such devices above some value of the drain voltage. It is often quite pronounced in SOI-NMOSFETs, but is usually absent in SOI-PMOSFETs. The kink-effect is also not observed in any bulk-MOSFETs operated at room temperature if a substrate or well contact exists. It may be observed in bulk-NMOSFETs if they are operated a low temperatures or in NMOSFETs fabricated in a floating *p*-well.

The kink-effect occurs for the following reason. When the drain voltage is large enough in an NMOSFET, the electrons near the drain are accelerated across the channel-drain depletion-region by an electric-field that is strong enough to cause electron-hole pair generation by impact-ionization (in PMOSFETs the coefficient of electron-hole pair generation by energetic-holes is much lower than that of energetic-electrons - which are the carriers in NMOSFETs). Drain voltages typically get this high when the MOSFET is operating in the saturation region (Fig. 11-34).

In a bulk-NMOSFET, electrons generated by this effect rapidly move into the drain, adding to the drain-current. The generated holes exit the device through the body-contact where they become substrate-current. In partially-depleted SOI-NMOSFETs, the generated electrons still move out of the drain and add to the drain-current. The generated holes, however, collect in the NMOSFET body, where they raise the potential of the body, with respect to the source (which is grounded). The amount by which the potential of the body is raised can be calculated, but this requires iterative solving of a set of complex equations, and

will not be described here. However, when the body-potential rises about 700-800 mV, it starts to significantly impact the output-characteristics of the NMOSFET.

That is, an increase in the body-potential has the same effect as biasing the substrate (body-effect). In other words, when the body of an NMOSFET is positively biased, it will give rise to a decrease in the threshold-voltage, V_T, of the NMOSFET (this is termed the *body-effect*). The decrease in V_T causes an increase in the drain-current, and this increase shows up as a "kink" in the drain-current curve of the SOI-NMOSFET. The kink represents an annoyance when analog circuits are designed with SOI NMOSFETs, but can actually lead to the fabrication of faster digital-circuits.

The situation regarding floating-body-effects in fully-depleted SOI-MOSFETs is somewhat different than in partially-depleted devices for several reasons. First, in the fully-depleted SOI-MOSFET, the body of the MOSFET at threshold has the same potential as the surface of the silicon at the top Si/SiO_2 interface. Hence, any holes generated by impact ionization at the drain will migrate out of the source and not collect in the body (which, in a partially-depleted SOI MOSFET, will have a lower potential than the source, and will thus attract the generated holes). Hence, there is no collection of holes in the body region, and the body-effect is also insignificant. In other words, the generation of holes by impact-ionization in fully-depleted SOI-MOSFETs does not decrease V_T, and hence no kink-effect is observed in fully-depleted SOI-MOSFETs, even in the absence of a body contact (i.e., body contacts are thus no longer needed). Second, since the potential of the body-region is higher, the electric-field at the drain will be smaller, reducing the number of electron-hole-pairs generated by impact-ionization.

There is, however, some effect on the subthreshold-slope of both partially-depleted and fully-depleted SOI-MOSFETs that is produced by the electron-hole pairs generated by impact-ionization at the drain. That is, as the MOSFET is being turned-on (but is still in the subthreshold-regime), the generated-holes will collect in the body (which has a lower potential than the source in both partially- and fully-depleted SOI-NMOSFETs if it is being operated in the subthreshold-regime). These injected-holes charge the body and produce a forward-bias on the body-source junction that reduces the threshold-voltage. As a result, the subthreshold-current increases, and the subthreshold-slope of the $I_D(V_{GS})$ curve is not as steep (i.e., the turn-off behavior of the device is degraded).

11.9.1.2 Transient I_D Effects: The floating-body may also produce transient-effects in the drain-current of a partially-depleted SOI-MOSFET. For example, an *overshoot* in the drain-current occurs when the MOSFET is turned-on (e.g., by increasing the gate-voltage from 0 to V_{DD}). What is happening is that majority-carriers are being expelled from the volume of the silicon of the floating-body under the gate that is becoming a depletion-region, and these carriers collect in the remaining neutral-region of the floating-body. This causes the potential of the floating-body to rise, and (as is described in the section on the kink-effect), this causes an increase in the drain- current. However, these excess-majority-carriers recombine and eventually equilibrium is re-established. At that point, the potential of the floating-body is returned to the value it would have without the presence of the excess injected- majority-carriers. Thus, the drain-current gradually subsides with time to a lower value. A reciprocal transient *undershoot* in the drain-current is observed when the gate voltage is switched off (e.g., if the gate-voltage decreases from V_{DD} to 0).

Such short-duration transient spikes in the drain-current can impact such SOI-circuits as SRAMs and DRAMs, which use pass-gate configurations. The amount of charge that is passed by the transient floating-body-effects must be accounted for in the design of such circuits.

11.9.1.3 Effect of Building Partially Depleted SOI-MOSFETs Without Body-Contacts: As discussed in the previous sections, the floating-body of partially-depleted SOI- MOSFETs has an impact on various MOSFET performance characteristics, most of which appear to be undesirable. However, Shahidi made a radical proposal in 1994. He suggested that in many important commercial circuit applications, partially-depleted SOI-MOSFETs could be used without contacts to the floating-body, and that this could be accomplished by making some modifications to the circuit design. Furthermore, by leaving the body floating, not only would the density be increased, but circuit-speed would be improved as well.

Early applications of SOI were limited to space and military programs where radiation-hardness was more important than performance and cost. These products were built with partially-depleted SOI, and fixed potentials to the body were provided by well- and substrate-contacts, just as is done in bulk-CMOS technology. In commercial applications, the considerations of performance and cost become paramount, and radiation-hardness is almost never an issue. Thus, Shahidi suggested that the floating-body-effects would not significantly degrade the performance of many SOI-ICs tailored for specific applications. If that was the case, the floating-body-contacts could be left out. In fact, several bulk-SRAM

circuits and microprocessors were implemented in a partially-depleted SOI-technology that did not employ floating-body-contacts, and these SOI-ICs did indeed exhibit high-performance. In addition, they met all other commercial product specifications (such as stress, reliability, burn-in, and yield). Although some circuit-modifications to the bulk-CMOS circuit design had to be made when implementing them in SOI, these modifications resulted in a higher performance gain than would have been obtained by merely re-mapping the bulk-CMOS design into SOI.[51] The floating-body-effects on the change in drain-current were compensated for in the circuit designs by a number of solutions, and these allowed the SOI-ICs to be successfully designed and fabricated using partially-depleted SOI-MOSFETs without contacts to the floating-body.[52]

Another floating-body issue arises when designing 0.13-μm SOI-CMOS circuits (and beyond), as a result of the very thin gate-oxide-thicknesses needed for these generations of devices. That is, such thin-gate-oxides give rise to a significant amount of gate-to-body tunneling-current. The charge introduced into the floating-body modulates the body-voltage and therefore causes a so-called *history-effect*, which can become excessive on 0.1-μm SOI-CMOS circuits.[60]

This history-effect can be explained as follows: The gate-current in these circuits can change the device "ON" V_{bs}, but does not usually change the "OFF" V_{bs}. Therefore, the gate current will not impact the delay when the input signal switches from "0" to "1" for the first time, but it will impact the delay when the input signal then falls from "1" to "0" (1^{st} switch V_{bs}) - and the impact of this delay will persist when the input signal switches the second time from "0" to "1" (2^{nd} switch V_{bs} – Fig. 11-35). This history-effect must be minimized in 0.1-μm SOI-CMOS by appropriate device-design.

Minimizing the history-effect can accomplished by making two modifications to the SOI-MOSFET device-design: **1**) the silicon-dioxide gate-dielectric is replaced with a dielectric having a higher *k*-value (which reduces the gate-current for an equivalent-oxide-thickness); and **2**) the value of gate-side junction capacitance C_j is adjusted.[60] (That is, C_j controls the drain-capacitance coupling to the body, and hence determines the 2^{nd} switch V_{bs}.) By adjusting C_j at a given level of gate leakage-current, the 1^{st} and 2^{nd} switch delays can be balanced (thereby eliminating the history effect in the switching delay). From a performance point of view, a low-value of C_j is beneficial. C_j can be reduced effectively by scaling the SOI silicon-layer thickness.

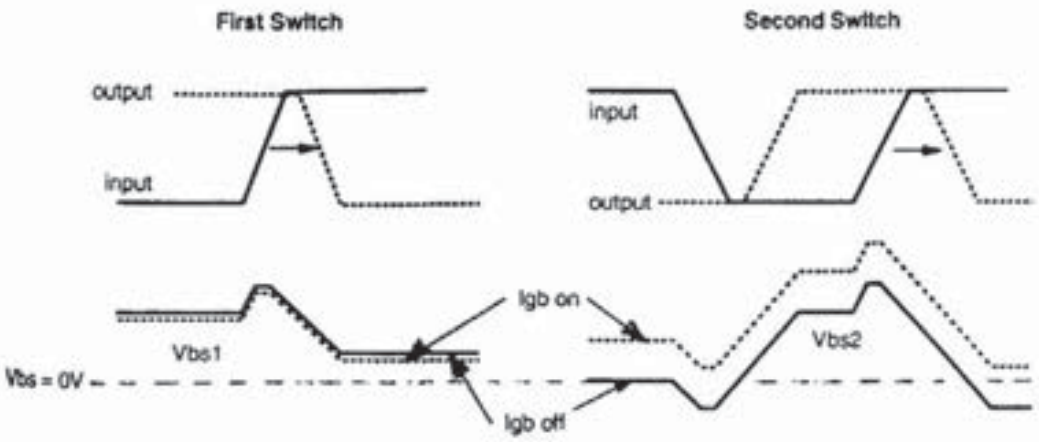


Fig. 11-35 Definition of 1st and 2nd switch rise-to-fall delay for the NMOSFET in an SOI-CMOS inverter. (© IEEE 2000).[60]

11.9.2 Short-Channel-Effects in SOI-MOSFETs

SOI-MOSFETs offer good short-channel behavior, especially for devices with channel lengths smaller than 0.1-μm. Fully-depleted SOI-devices are being studied for low-power/low-voltage applications because they appear to exhibit the best short-channel behavior. However, the silicon-thickness and the buried-oxide-thickness have to be made ultrathin in single-gate fully-depleted SOI-MOSFETs to achieve the best short-channel behavior (i.e., the silicon-thickness must be considerably smaller than the junction-depth of a comparable bulk technology). Even in such ultra-thin SOI-MOSFETs, back-channel-potential control-issues can exacerbate such short-channel-effects as threshold-voltage-rolloff and this issue must be addressed.[54,55]

The problem of threshold-voltage-rolloff arises in deep-submicron single-gated fully-depleted SOI-devices because the depletion-region-charge of the source/drain-regions cannot terminate in the buried-oxide region, and thus must terminate in the substrate below the buried-oxide or in the channel-region (Fig. 11-36).[55] Because the buried-oxide is thick, a portion of the depletion-charge will terminate in the channel. Thus, the source and drain behave like fringing back-gates. If the channel is short, the fringing-fields from these parasitic back-gates can induce back-channel-conduction, which enhances the problem of threshold-voltage-rolloff with decreasing channel-length. This phenomenon has been

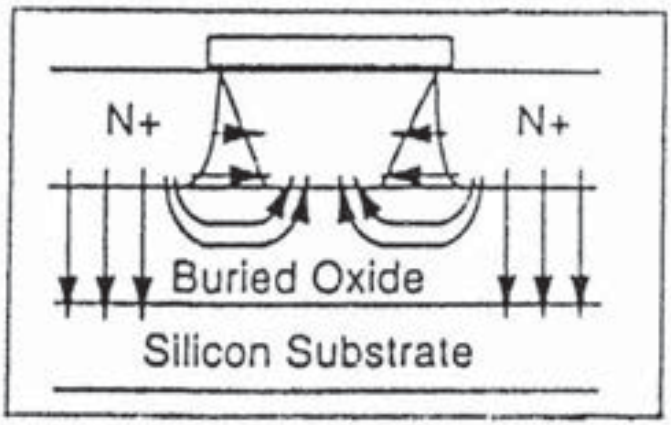


Fig. 11-36 Electric-field lines in the source/drain and body *pn*-junctions in a short-channel SOI-MOSFET. Some of these lines terminate in the substrate beneath the buried-oxide (BOX), while others terminate in the body, producing fringing-electric-fields.[55]

studied by simulation and empirically modeled.[55] Reducing the buried-oxide thickness can also help alleviate the 2-D charge-sharing through the buried-oxide, because the bottom substrate is brought closer to the source and drain regions, provided there is no significant depletion in the substrate. Simulation studies have been presented that predict that single-gate SOI-MOSFETs, even with ultra-thin silicon channels, do not provide adequate immunity against short-channel-effects in 25-nm channel-length devices.[57]

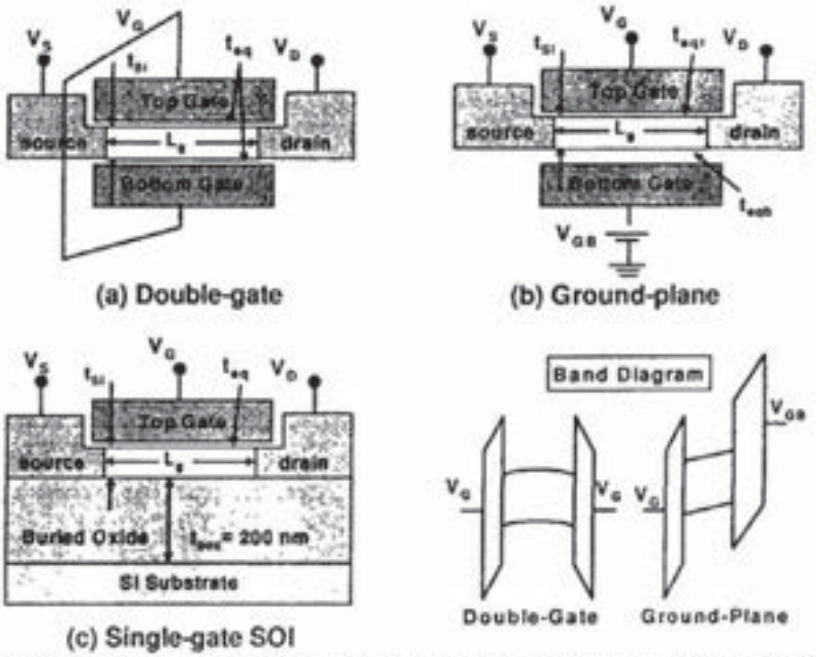


Fig. 11-37 a) Double-gate; b) Ground-plane; c) Single-gated SOI-MOSFET. Gate material work-function is set at the mid-gap of the silicon band gap. On-chip biasing of the ground plane is assumed. The source/drain junction is abrupt. The channel length (L_g) is the metallurgical channel length. (© 1998 IEEE).[57]

As a result, two alternative approaches have been proposed to reduce the threshold-voltage-rolloff problem described above. That is, it has been suggested that either a *double-gate SOI-device*, or a *ground-plane SOI-device*, be used instead of a single-gate SOI-device (Fig. 11-37). Simulation studies show that double-gate SOI-MOSFETs give the best control over the threshold voltage rolloff problem, but the difficulties in fabricating double-gate SOI MOSFETs are daunting. Instead, a compromise structure (namely an SOI device with a low-resistance ground plane) has been proposed.[56,57] In the ground-plane SOI MOSFET, the low resistance ground plane can be biased with a power supply separate from that used to control the gate. This offers the flexibility of tuning the threshold voltage by means of the bottom gate bias.[57] Both double-gate and ground-plane SOI-MOSFETs are shown to able to meet the electrostatic design requirements.

11.9.3 Self-Heating Effects in SOI MOSFETs

SOI MOSFETs are thermally insulated from the substrate by the poorly heat-conducting buried-oxide layer. Consequently, removal of excess heat generated by Joule heating within the transistors is less efficient than in bulk-CMOS technology. As a result, the device-temperature in SOI technologies can rise substantially above what it would be in a comparable bulk technology. Although the heat transfer rate is lower in an SOI device, heat is nevertheless still transported out through a number of paths, namely: **1**) through the buried oxide layer; and **2**) laterally through the silicon-island into the contacts and metallization. As we shall see below, however, such heating effects have not proved to be a significant problem in SOI CMOS technology.

The increased operating temperature within an SOI-MOSFET can have an impact on the device-performance. The most significant effect is that the mobility of the carriers in the channel is reduced. That is, if a MOSFET is being operated in the saturation-region, and the drain voltage is increased, there will be a larger voltage drop at the drain-junction. This will increase the power dissipated in the device and its temperature will rise. Once the device temperature is increased by this effect by 50°C (to 150°C), a significant reduction in carrier-mobility will occur. This effect will manifest itself in the output characteristics of the device in the form of a drop in the drain-current with increasing drain voltage (Fig. 11-38).

However, in CMOS circuits, all of the MOSFETs normally operate in a standby-mode. Large drain-currents in any device occur only during switching, and these cease after a very brief period of time (< 1-ns). Since the thermal time constants of SOI MOSFETs are on the order of several tens of nanoseconds, it

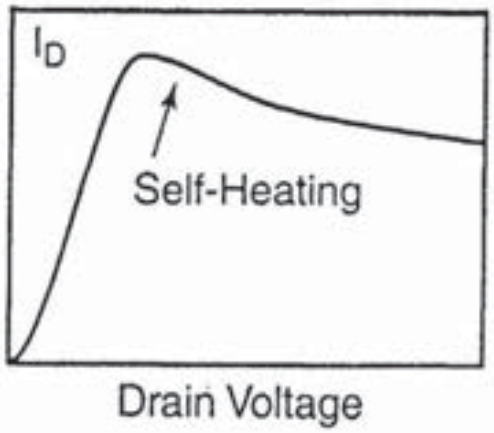


Fig. 11-38 The effect of self-heating in SOI-MOSFETs on output characteristics of the device.

seems that this self- heating effect generally will cause no problems in CMOS digital circuits, with a few exceptions. That is, in logic-circuits that exhibit short transition times, temperature rises of only 2-3 degrees are observed compared to bulk-CMOS. However, in such circuits as off-chip drivers (that can be ON for a significant portion of the switching cycle), the temperature rise is observed to be higher (e.g., ~10 degrees).

11.9.4 Reliability Issues for SOI

There are also some reliability issues unique to SOI-technology that should be addressed when considering SOI for ULSI applications.[61] Such an analysis indicates that SOI-circuits do not exhibit significantly worse reliability behavior than comparable bulk-CMOS-circuits. Some of the reliability issues associated with silicon-on-insulator ICs are reviewed here.

First, bulk- and SOI-CMOS circuits appear to react comparably with respect to *off-state current degradation* when subjected to bias-temperature stress-testing. Hence, this apparently does not represent a special reliability-problem for SOI. Second, channel hot-electron-degradation (CHE) in SOI-devices is more complex than in bulk-MOSFETs because charge-trapping and interface-state-generation can occur at both the gate-oxide-interface and at the buried-oxide-interface. Third, the defects responsible for gate-oxide yield and reliability have been minimized by 2000 in state–of-the-art SOI-materials to acceptable levels. This claim is supported by TDDB tests on gate-oxides on both bulk- and SOI-materials (i.e., the oxides grown on both the bulk- and SOI- materials show

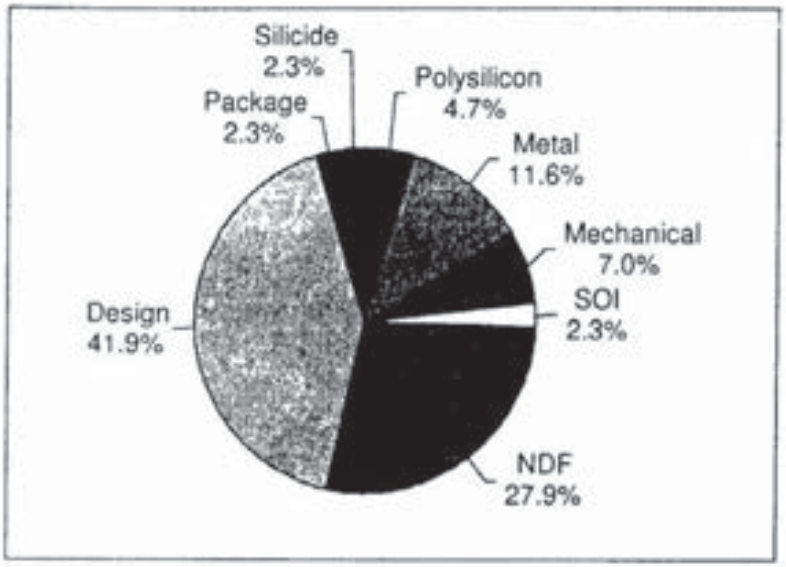


Fig. 11-39 SOI unique defect fails (induced by burn-in stress), are negligible part of the overall failure population.[61] (© 2000 IEEE)

similar failure distributions). Fourth, SOI-devices appear to be less susceptible to charging damage than bulk-CMOS devices. Finally, when failure mechanisms in SOI-microprocessor-circuits are catalogued, the percentage of fails due to SOI-unique defects represent a negligible portion of the failure distribution (Fig. 11-39). In summary, these reports indicate that by using state-of-the-art SOI material, highly reliable products can be produced.

REFERENCES

1. International Technology Roadmap for Semiconductors, 1999 Edition, Semiconductor Industry Association, San Jose, CA.

2. "Proc. 5th Inter. Workshop on Measurement, Characterization, and Modeling of Ultrashallow Doping Profiles in Semiconductors," *J. Vac. Sci. Technol.*, **B-18**, 2000, p. 337-604.

3. G. Shahidi, *et al.*, "Mainstreaming of the SOI Technology," *Proc. IEEE Inter. SOI Conf.*, p. 1-4, 1999.

4. A. Agarwal, "Ultra-Shallow Junction Formation by Spike Annealing in a Lamp-Based or Hot-Walled Rapid Thermal Annealing System: Effect of Ramp-Up Rate," *Materials Sci. in Semiconductor Processing*, 1, p. 237, 1998.

5. K.K. Ng and J. Lynch, "The Impact of Intrinsic Series Resistance on MOSFET Scaling," *IEEE Trans. Electron Dev.*, **ED-34**, p. 503, 1987.

6. M.I. Current *et. al.,* "Process Integration Issues for Doping of Ultrashallow Junctions," *J. Vac. Sci. Technol.* **B-18,** p. 468, 2000.
7. C.-T. Chaung *et al.*, "SOI for Digital CMOS VLSI: Design Considerations and Advances," *Proc. IEEE,* **86**, April 1998, p. 689.
8. S.T. Liu *et al.*, "Radiation Response of SOI Materials," *Mtg. Abs. Electrochem. Soc,.* Spring 1999, Abs. 414.
9. J.-P. Colinge, "Subthreshold Slope of Thin-Film SOI MOSFETs," *IEEE Electron Dev. Letts.,* **EDL-7**, pp. 244, 1986.
10. C.L Chen *et al.*, "High-Frequency Characterization of Sub-0.25-μm Fully-Depleted Silicon-on-Insulator MOSFETs," *IEEE Electron Dev. Letts.,* **EDL-21,** October 2000, p. 497.
11. *ibid.* Ref. 7.
12. *ibid. Ref. 9.*
13. J.-P. Colinge, "Fully-Depleted SOI CMOS for Analog Applications," *IEEE Trans. Electron Dev.,* **45**, May 1998, p. 1010.
14. G.G. Shahidi *et al.*, "A Room-Temperature 0.1-μm CMOS on SOI," *IEEE Trans. Electron Dev.,* **41,** December 1994, p. 2405.
15. N. Kistler *et al.*, "Sub-Quarter Micron CMOS on Ultrathin SOI, *IEEE Electron Dev. Lett.*, **EDL-12**, May 1992, p. 235.
16. M. Chan *et al.*, "Recessed-Channel Structure for Fabricating Ultrathin SOI MOSFET with Low Series Resistance," *IEEE Electron Dev. Lett,* **EDL-14**, January 1994, p. 22.
17. A. Toriumi *et al.*, "Performance and Reliability Concerns of Ultra-Thin SOI and Ultra-Thin Gate Oxide MOSFETs," *Tech. Dig. IEDM,* 1995, p. 847.
18. G.W. Cullen and C.C. Wang, Eds., *Heteroepitaxial Semiconductors for Electronic Devices,* Springer-Verlag, 1978.
19. G.W. Cullen *et al.*, *J. Crystal Growth*, **56**, 287 (1982).
20. M.S. Abrahams and C.J. Buiocchi, "Cross-Sectional Electron Microscopy of Silicon on Sapphire", *Appl. Phys. Lett.*, **27**, 325 (1975).
21. K.E. Violette, P.A. O'Neill, M.C. Ozturk, K. Christensen, and D. Maher, "Low- Temperature Selective Silicon Epitaxy by Ultra High Vacuum Rapid Thermal Chemical Vapor Deposition Using Si_2H_6, H_2 and HCl," *Appl. Phys. Lett.,* **68**, 66 (1996).
22. A. Gupta and P.K. Vasudev, "Recent Advances in Hetero-Epitaxial Silicon-on-Insulator Technology, Part II," *Solid State Technology,* **26**, 129 (1983).
23. R. Reedy, T.W. Sigmon, and L.A. Christer, "Suppressing Al Outdiffusion in Ion Implantation Amorphized and Recrystallized Silicon-on-Sapphire Films," *Appl. Phys. Lett.*, **22**, 537 (1983).
24. G.A. Garcia, and R.E. Reedy, "Electron Mobility Within 100 nm of the Si/Sapphire Interface in Double-Solid Phase Epitaxially Regrown SOS," *Electron. Lett.,* **22**, 537 (1986).

25. G.W. Cullen, M.T. Duffy, and R.K. Smeltzer, "Recent Advances in Heteroepitaxial Si on Sapphire Technology," *VLSI Science and Technology/1984*, Electrochemical Society, N.J., 1984, p. 230.
26. D. Munteanu *et al.*, "Silicon-on-Sapphire Material and Device Properties at High and Low Temperature," *Mtg. Abs. Electrochem. Soc.*, Spring 1999, Abs. 435.
27. K. Izumi *et al.*, *Electron Lett.*, **14**, (1978) 593.
28. J. Liu *et al.*, "Formation of Buried Oxide in Silicon Using Separation by Plasma Implantation of Oxygen," *Appl. Phys. Letts.*, **67**, p. 2361, 1995.
29. One supplier of SIMOX wafers is Ibis Technology Corp, Danvers, MA.
30. C.E. Chen *et al.*, *1984 IEDM Tech. Digest*, 701 (1984).
31. S. Nakashima *et al.*, "Investigations on High-Temperature Thermal Oxidation Process at Top and Bottom Interfaces of Top Si of SIMOX Wafers," *J. Electrochem. Soc.*, **143**, p. 244, January 1996.
32. *Wafer Bonding and SOI Technologies*, Abstracts of the Fall 1997 Meeting of the Electrochemical Society, Paris, France.
33. Q.-Y. Tong and U. Gosele, *Semiconductor Wafer Bonding*, VCH-Wiley, New York (1999).
34. J.B. Lasky *et al.*, "SOI by Bonding and Etchback," *Tech. Dig. IEDM*, 1985 p. 684.
35. R.C. Frye *et al*, "A Field-Assisted Bonding Process for Silicon Dielectric Isolation," *J. Electrochem. Soc*, **133**, 1673 (1986).
36. H. Ohashi, "Improved Dielectrically Isolated Device Integration by Silicon-Wafer Direct Bonding Technique," *Tech. Dig. IEDM*, 1986 p. 210.
37. Q.-Y. Tong, W.J. Kim, T.-H. Lee, and U. Gosele, "Low-Vacuum Wafer Bonding," *Electrochemical and Solid State Letter*, **1**, 52 (1998).
38. A. Plossl and G. Krauter, *Mater. Sci. and Engr Reports*, **R25**, 1 (1999).
39. A.J. Auberton-Herve, "SOI: Materials to systems," *Tech. Dig. IEDM* 1996, p. 3.
40. Unibond Wafer Specification Sheet. SOITEC USA Inc. Peabody, MA, (1998).
41. IBM Microelectronics
42. H.W. Lam, A.F. Tasch, Jr., and R.F. Pinzzotto, "Silicon-on-Insulator for VLSI and VHSIC," in *VLSI Electronics*, Ed. N.G. Einspruch, Vol. 4, p. 1, Academic Press, NY (1982).
43. S. Farrens *et al.*, "High-Quality SOI Structures by the Genesis Process," *Mtg. Abs. Electrochem. Soc.*, Spring 1999, Abs 396.
44. M.I. Current *et al.*, "Ultrashallow Junctions or Ultrathin SOI?" *Solid State Technology*, September 2000, p. 66.
45. K. Sakaguchi *et al.*, "ELTRAN by Splitting Porous Si Wafers," *Mtg. Abs. Electrochem. Soc* , Spring 1999, Abs. 395.

46. T. Yonehara *et al.*, "ELTRAN: Epitaxial Layer Transfer," *Mtg. Abs. Electrochem. Soc.*, Spring 1999, Abs. 394.
47. K. Suzuki *et al.*, "Scaling Theory for Double-Gate SOI MOSFETs," *IEEE Trans. Electron Dev.*, **40**, 2326 (1993).
48. C.H. Wann *et al.*, "A Comparative Study of Advanced MOSFET Concepts," *IEEE Trans. Electron Dev.*," **43**, 1742 (1996).
49. M. Tachimori *et al., Proc. Electroche,m. Soc. on SOI Tech & Dev.* Vol. 96-3, p. 53 (1996).
50. A. Matsamura *et al.*, "Recent Progress in Low-dose SIMOX Wafers Fabricated with Internal-Thermal-Oxidation (ITOX) Process," *Mtg. Abs. Electrochem. Soc.*, Spring 1999, Abs. 391.
51. D.H. Allen *et al.*, ISSCC, 1999.
52. F. Assaderaghi and G. Shahidi, "Advanced SOI Circuits for the Microelectronics Market," *Mtg. Abs. Electrochem. Soc.*, Spring 1999, Abs. 383.
53. "Canon Brings Big Company Muscle to the SOI Market With Yet Another Technology," *Solid State Technology*, June 2001, p. 50 (and 167, an Ad which announces the expanded production of ELTRAN wafers).
54. T. Ogawa, "At Long Last, SOI Market is On the Move," *Solid State Technology*, February 2001, p. 62.
55. L. Su *et al.*, "Deep-Submicrometer Channel Design in Silicon-On-Insulator (SOI) MOSFETs," *IEEE Electron Dev. Letts.*, Vol. 15, September 1994, p. 366.
56. T. Ernst and S. Cristoloveanu, "Ground Plane Concept for Reduction of Short-Channel Effects in Fully-Depleted SOI Devices," *Mtg. Abs. Electrochem. Soc.*, Spring 1999, Abs. 431.
57. H.-S. P. Wong, D.J. Frank, and P.M. Solomon, "Device Design Considerations for Double-Gate, Ground-Plane, and Single-Gated Ultra-Thin SOI MOSFETs at the 25 nm Channel Length Generation," *Tech. Dig. IEDM* 1998, p. 407.
58. E. Leobandung, *et al.*, "Scalability of SOI Technology into 0.13 μm 1.2 V CMOS Generation," *Tech. Dig. IEDM* 1998, p. 403.
59. E. Leobandung *et al.*, "High-Performance 0.18 μm SOI CMOS Technology," *Tech. Dig. IEDM* 1999, p. 679.
60. S.K.H. Fung *et al.*, "Controlling Floating-Body Effects for 0.13 μm and 0.1 μm SOI CMOS," *Tech. Dig. IEDM* 2000, p. 231.
61. R. Bolam *et al.*, "Reliability Issues for Silicon-on-Insulator," *Tech. Dig. IEDM* 2000, p. 131.
62. K. Sakaguchi and T. Yonehara, "SOI Wafers Based on Epitaxial Technology," *Solid State Technology*, June 2000, p. 88.
63. M. Current *et al.*, "Atomic-Layer Cleaving and Non-Contact Thinning and Thickening for Fabrication of Laminated Electronic and Photonic Materials," *Proceedings 2001 Materials Research Society Spring Meeting*, Paper 18.3.

64. A. Kamgar and J.T. Clemens, "Isolation's Path to SOI Technology," *Solid State Technology*, October 1999, p. 109.
65. J.W. Sleight *et al.*, "A High Performance 0.13 μm SOI CMOS Technology with a 70 nm Silicon film and with a Second Generation Low-k Cu BEOL," *Tech. Dig. IEDM*, 2001, p. 621.
66. R. Chau *et al.*, "A 50-nm Depleted Substrate CMOS Transistor (DST)," *Tech. Dig. IEDM*, 2001, p. 621.

Chapter 12

MULTILEVEL INTERCONNECTS FOR ULSI

The semiconductor industry continues to pursue the development of ever more powerful ICs. This effort basically involves a two-pronged approach. The first approach seeks to increase the speed of individual transistors through a continual reduction in the minimum size of device features. Such scaling also provides an increase in the device density on the chip. The second approach entails developing increasingly complex interconnect systems, which now use multilayer structures of metal wiring separated by interlayer dielectrics (ILDs). This enables the higher-speed devices to be interconnected (resulting in ICs with enhanced performance and system functionality). Efforts in both of these development arenas are supported by continually more sophisticated device and circuit designs.

The tradeoffs between individual device and interconnect performance depend on the details of the circuit architecture. When the IC industry was in its early stages (i.e., from the mid-1970s to about 1990), circuit design rules were in the 1-5-μm range. During that time, interconnect delays could generally be ignored because they were typically much smaller than the device switching times (which were ~50-Mz). However, by the mid-1990s, microprocessors and other large area ULSI devices with millions of circuit elements and speeds greater than 1-GHz were envisioned. Designers then began to realize that interconnect delays would become comparable to (and could eventually exceed) the intrinsic delays of the transistors. It was generally agreed that beyond the 0.25-μm technology node, the overall circuit performance would be dominated by the efficiency with which devices are connected, rather than by the speed of individual devices (Fig. 12-1).[1]

In addition, by the mid-1990s, two different technological domains had also started to collide – namely circuits designed for purely for radio-frequency (rf) applications and others designed merely for digital applications. That is, rf circuits were typically low-density and high-frequency (GHz) designs used in avionics and wireless communication, whereas digital circuits strove for high-density and covered a wide performance spectrum (1-100 MHz). The merger of performance requirements (or *mixed-signal products*), reflects a growing trend,

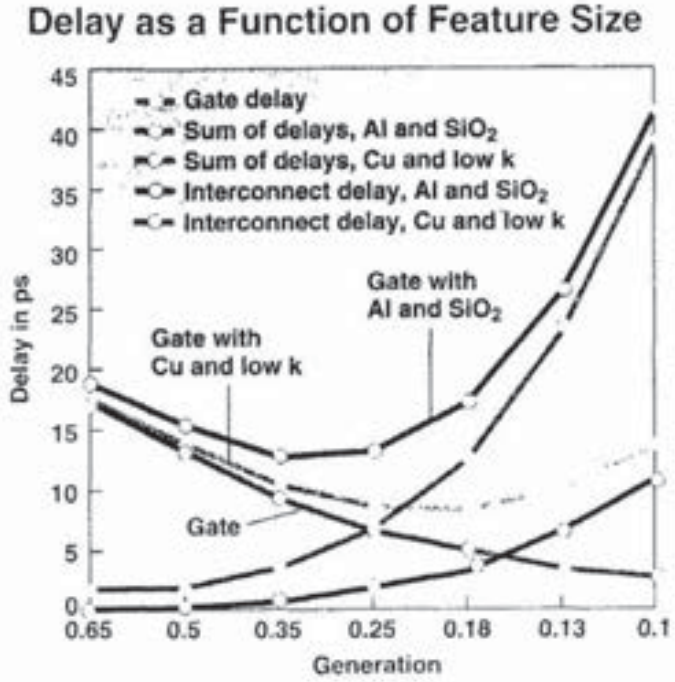


Fig. 12-1 RC-delay as a function of feature size.[1] (© IEEE 1997)

and places additional demands on interconnect systems as performance requirements continue to escalate.

In this chapter, some of the challenges related to the performance and density issues of ULSI interconnects will be discussed. The discussion will be divided into two parts:

1. The performance requirements of ULSI interconnect systems;

2. The evolution of the materials and architecture of interconnects for ICs.

The exact feature size at which the interconnect RC delay equals and then surpasses the intrinsic transistor delay depends on the details of the circuit design. Furthermore, the total overall interconnect delay t_{int} is composed of four components and can be approximated by:[2]

$$t_{int} = R_0 C + 0.4\,[(RC)^{1.6} + (tof)^{1.6}]^{1/1.6} + 0.7\,RC_{in} \qquad \textbf{(12.1)}$$

where R_0 is the output resistance of the driver device, tof is the time of flight of the electromagnetic wave in the interconnect, and C_{in} is the input capacitance of the device in the next gate. These four terms contribute differently to the overall delay, depending on the gate-length and gate-width, the interconnect line-length, and other geometrical parameters. However, most signs point to the feature size of the order of 0.5-μm at which the interconnect RC-delay starts to become a significant part of the overall signal delay.

12.1 EVOLUTION OF THE STRUCTURE AND MATERIALS USED IN IC INTERCONNECTS

Interconnects for ICs have evolved from the single-level-metal system on the original IC patented by Noyce, to the multi-level-metal interconnect systems used in today's ULSI. In this section we will describe the evolution of these interconnect systems. This will point out the limitations that existed for previous generations of IC interconnect technologies, and how they were overcome by learning how to build more advanced interconnect systems. By describing the state-of-the-art of current interconnects, we will also be able to see the limitations they face, and what kinds of techniques are being explored to overcome them.

12.1.1 Single-Level-Metal Interconnects in Early ICs

In 1959 Robert N. Noyce invented the concept of the monolithic-integrated-circuit (IC) while at Fairchild Semiconductor. His patent application described how planar silicon-bipolar-transistors and resistors could be interconnected with thin and narrow aluminum lines adherent to an underlying silicon-dioxide film (SiO_2) that covered the surface of the silicon chip. Contact holes were etched in the oxide film, and these allowed the aluminum lines to make contact to the circuit devices fabricated within the silicon substrate. After the aluminum lines were patterned, a passivation layer (e.g., a CVD P-doped SiO_2 film) would be deposited over the entire wafer. Figure 12-2 shows a schematic of how such a single-level-metal interconnect system was fabricated. The materials used in these early ICs (Al and SiO_2) were selected because they had many unique advantages. These advantages proved to be so compelling that this same combination of materials was used for many generations of IC interconnect

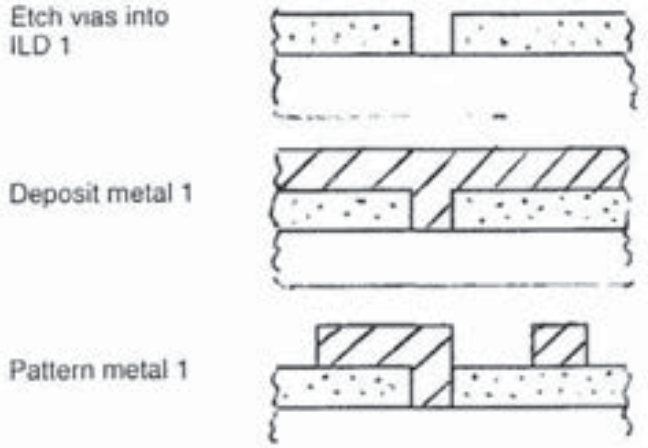


Fig. 12-2 Schematic of process steps used to produce a single-level-metal interconnect.

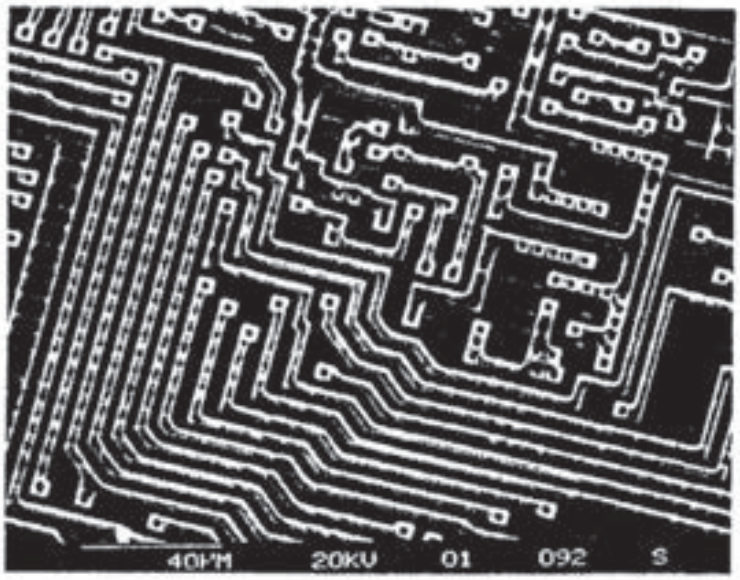


Fig. 12-3 SEM showing the patterned metal layer on a circuit with only one layer of metal.

technology. As a point of reference for materials parameters, the resistivity of pure Al is $\rho_{Al} = 2.7\text{-}\mu\Omega\text{-cm}$, and the dielectric constant of SiO_2 is $k_{SiO_2} = 4.2$ (for the type of SiO_2 used in dielectric films under the aluminum). Later versions of this interconnect system used Al films alloyed with Si to improve contact stability. Figure 12-3 shows an SEM of a circuit using only one-level of metal.

Note that the metal wires made in this traditional way of creating the IC interconnect structure are formed by: **1**) depositing a blanket layer of the metal over the entire wafer; and **2**) selectively removing the metal from regions on the wafer where the metal is not needed (i.e., with a *subtractive* etching process).

12.1.2 Double-Level-Metal Interconnects in Early Bipolar ICs

As the density of ICs increased with decreasing feature size, it became necessary to have more than one level of interconnect. In NMOS ICs, the polysilicon-gate material could serve as a local-interconnect level (eventually covered with a silicide to form a polycide), and this level could be covered with a single-level of metal (Al-Si) to form a two-level interconnect system.

In bipolar technology there is no MOSFET poly-gate structure, and thus to get two-levels of interconnect, it is necessary to use two-levels of metal (Fig. 12-4). Thus, the first two-level-metal interconnect systems for ICs were used in bipolar technology. (Figure 12-5 shows the steps used to fabricate such a 2-level-metal

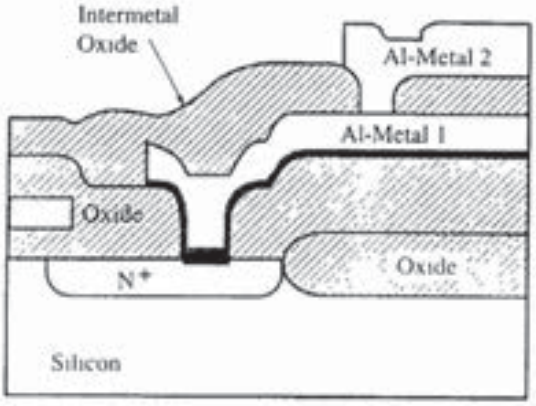


Fig. 12-4 Schematic of a double-level-metal structure used in early bipolar-ICs.

interconnect structure.) In these 2-level-metal bipolar-ICs there was no attempt made to planarize the dielectric layer that separated the 2-levels of metal (i.e., the *interlevel dielectric*, or *ILD*). The holes opened in the ILD to allow Metal-2 to contact Metal-1 were termed *vias*. While the two-level-metal process could be successfully implemented on bipolar-ICs, the step coverage of Metal-2 as it ran over Metal-1 lines was marginally acceptable. Thus, it would not have been possible to continue to use this process to form a three-level-metal interconnect, even in bipolar-ICs.

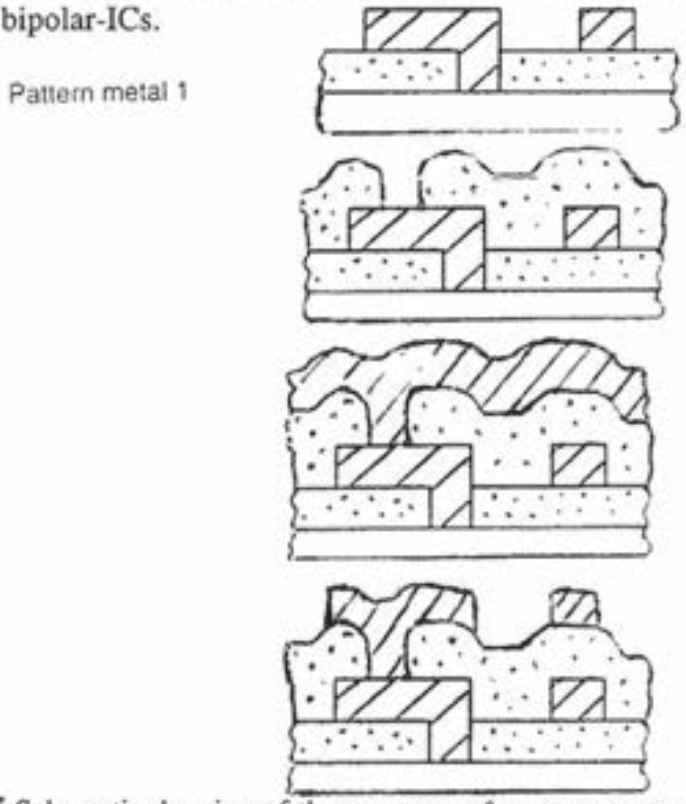


Fig. 12-5 Schematic drawing of the sequence of process steps used to produce a double-level-metal interconnect structure as shown in Fig. 12-4.

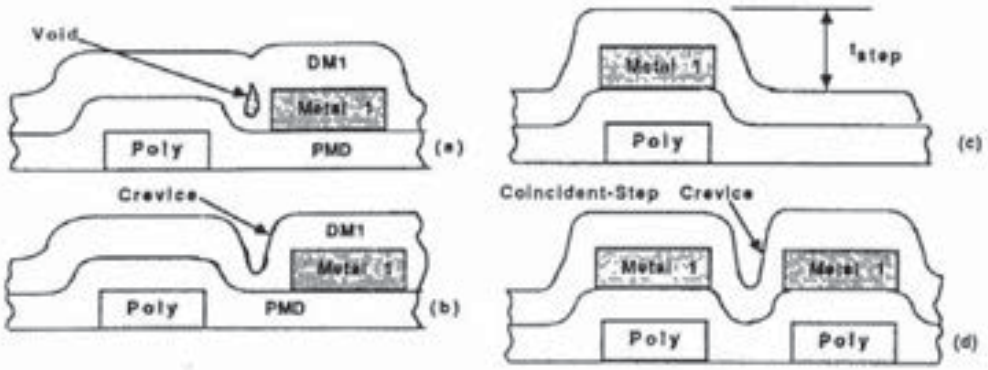


Fig. 12-6 Topography conditions that arise in MOSFETs that make it difficult to fabricate a double-level-metal structure without some kind of planarization method.

Likewise, when CMOS ICs came to replace NMOS ICs (in the mid-1980's), it was no longer possible to use the polysilicon as a local interconnect layer in the same manner that it had been used in NMOS. This was because polysilicon in early CMOS technologies was *n*-doped, and therefore it could not be used to make contact to PMOS source/drain regions. Instead, it also became necessary to use a double-level-metal interconnect structure in these CMOS ICs. Unfortunately, the same approach to fabricating a double-level-metal structure used in the early bipolar ICs could not be used in CMOS ICs. That is, the steps created by the polysilicon layer (together with the steps caused by Metal-1) were too high to allow adequate step coverage when Metal-2 lines had to run over such (poly + Metal-1 hills), as is shown in Fig. 12-6.

12.1.3 Double-Level-Metal Interconnects in CMOS ICs (Partial Planarization of the ILD)

To enable double-level-metal interconnect structures to be fabricated in CMOS ICs, the fabrication sequence used for making double-level-metal interconnects in bipolar ICs had to be modified. That is, it was necessary to be able to partially planarize the ILD prior to depositing Metal-2. The partial-planarization technologies used in the CMOS ICs of the late 1980's included *resist etchback* and the use of *spin-on glass* (SOG). Details concerning these partial-planarization methods can be found in Vol. 2, Chap. 4. Furthermore, the contact holes and vias were opened with a two-step etch process. The first half (which opened the top half of the via) used an *isotropic* etch process, while the second etch step (that opened the remainder of the via) used an *anisotropic* process. This

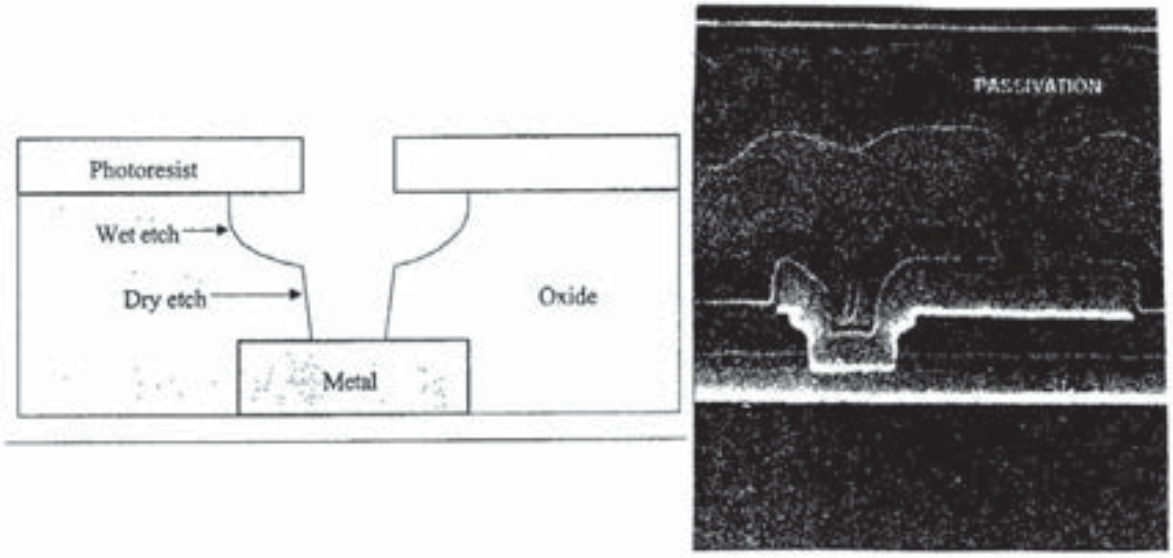


Fig. 12-7 a) Schematic of a via with a "champagne-glass" shape; b) SEM of actual via.

two-step via-opening process produced via profiles that had a "champagne-glass" shape (Fig. 12-7). Such vias were easier to adequately fill with Metal-2. Note that Al-alloy films were typically used (e.g., Al:2%Cu) because they exhibited better electromigration characteristics than did pure-Al films. Figure 12-8 depicts a schematic of a double-level-metal structure for CMOS technology, and Fig. 12-9 shows a SEM of an actual CMOS IC with such a two-level-metal interconnect-structure. Three-level-metal structures in bipolar-ICs could also be fabricated with this approach.

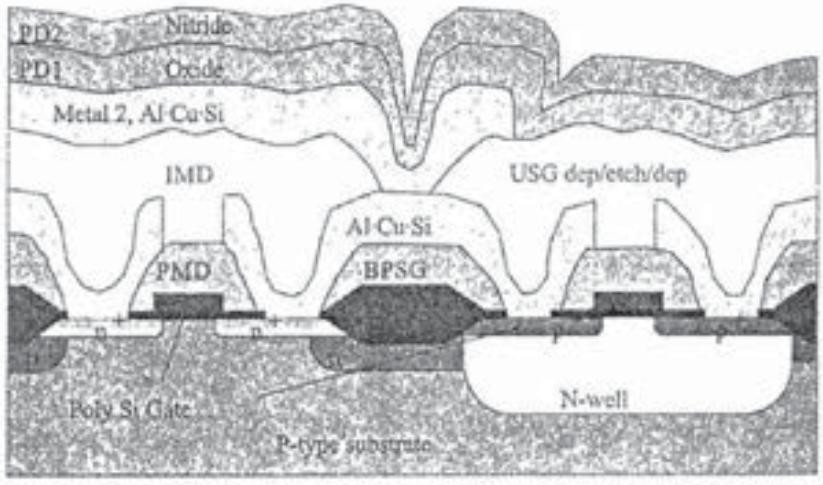


Fig. 12-8 Schematic of a double-level-metal structure in CMOS with partial-planarization.

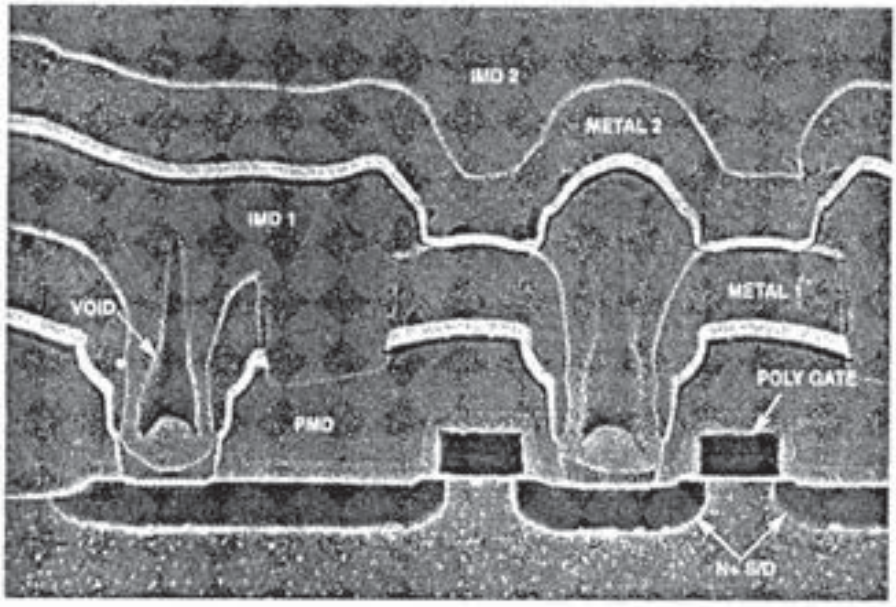


Fig. 12-9 SEM of an actual double-level-metal structure in CMOS with partial planarization.

The partial-planarization of the ILD allowed the step-coverage of Metal-2 to be sufficiently improved, and the "champagne-glass" via profile allowed Metal-2 to adequately fill the vias. Nevertheless, the topography of the IC remained too rough to allow this technology to be extended for producing three-level-metal interconnects in CMOS-ICs.

12.1.4 Three-Level Metal Interconnects in CMOS ICs: (CMP of the ILD & W-Plugs)

To enable three-level-metal interconnects (or more) to be fabricated in CMOS-ICs, a technique for globally-planarizing the ILD after it was deposited was needed. The technology developed for this purpose was *chemical-mechanical-polishing (CMP)* of the CVD-SiO_2 ILD. With this technique, the next level of metal would not have to cover any steps, but could be deposited on a completely planar surface (Fig. 12-10).

While CMP of the ILD solved the problem of metal lines having to cross high steps, it created a new difficulty. That is, as shown in Fig. 12-10, the vias that have to be opened in the planarized ILD are of varying depths (with some being quite deep). Such deep vias could no longer be adequately filled by the traditional PVD process used to deposit metal layers (sputtering), even if a two-step via-etch process was used. To address this issue, a CVD (instead of PVD) process was

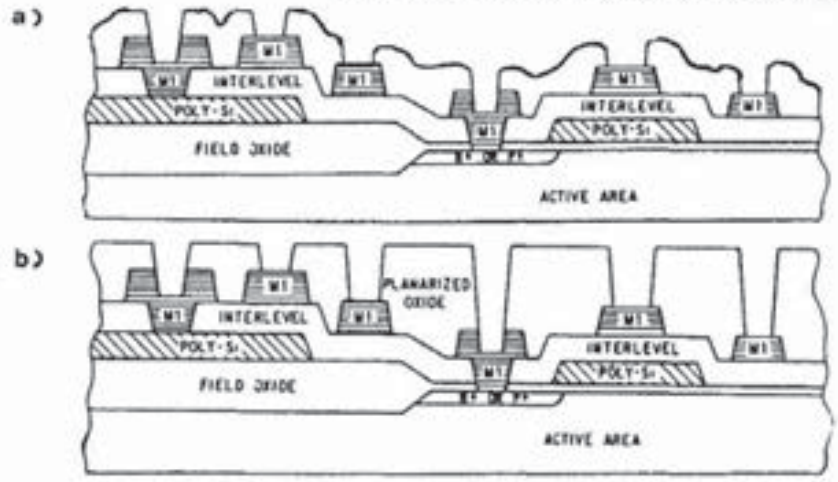


Fig. 12-10 a) If no planarization is used, the via and contact holes between adjacent conductor levels all have approximately the same depth; b) If planarization of the intermetal dielectric layer is achieved, the vias will vary widely in depth.

used to deposit the metal to fill the vias. The metal chosen to fill the vias in this manner was tungsten (CVD-W). A blanket CVD-W process was developed that could deposit a layer of W over the entire wafer covered with a planarized ILD (Fig. 12-11), such that variable-depth-vias across the wafer could also be completely filled. This blanket-W film would then removed from the planar wafer surface (by etching or CMP), leaving W only in the vias (i.e., *W-plugs* were created). The process of filling recesses in a dielectric layer with a metal film and removing the overburden with etchback or CMP (rather than depositing, patterning, and etching the metal on top of a dielectric layer) is called a

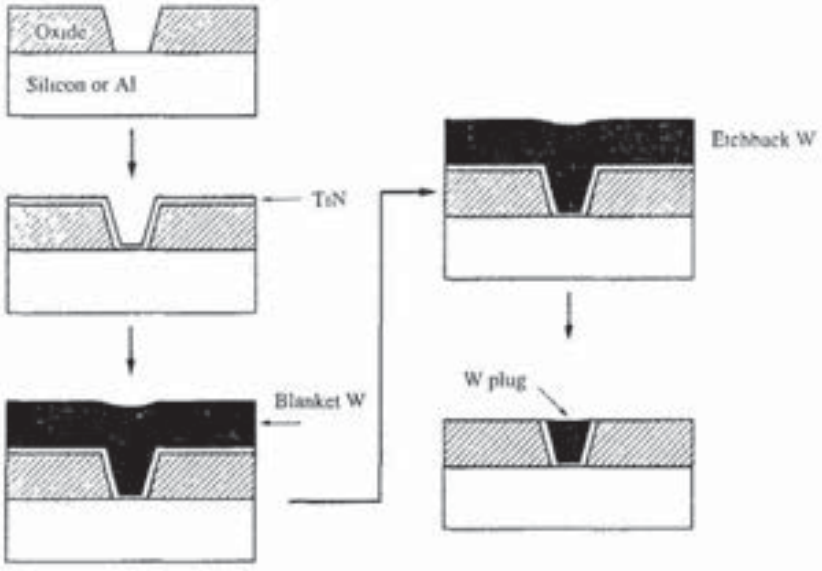


Fig. 12-11 Schematic of the process used to form W-plugs with CVD-W and etchback.

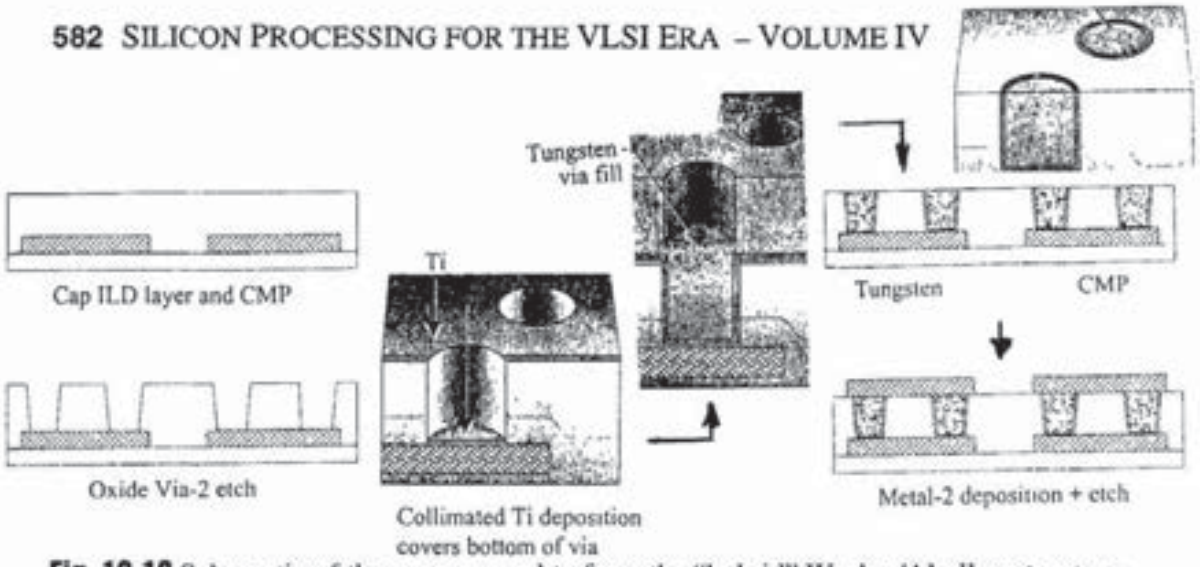


Fig. 12-12 Schematic of the process used to form the "hybrid" W-plug/Al-alloy structure.

"*damascene*" method. Its name comes from the ancient decorative practice of inlaying metal in wood or ceramics that originated in Damascus.

The layer of aluminum alloy (e.g., Al:2%Cu) from which the interconnect lines are patterned is then deposited and etched by the usual subtractive process. Fig. 12-12 shows a schematic of the steps used to form one-level of this type of interconnect structure. Figure 12-13 shows a schematic drawing and a SEM of such an interconnect structure with stacked vias.

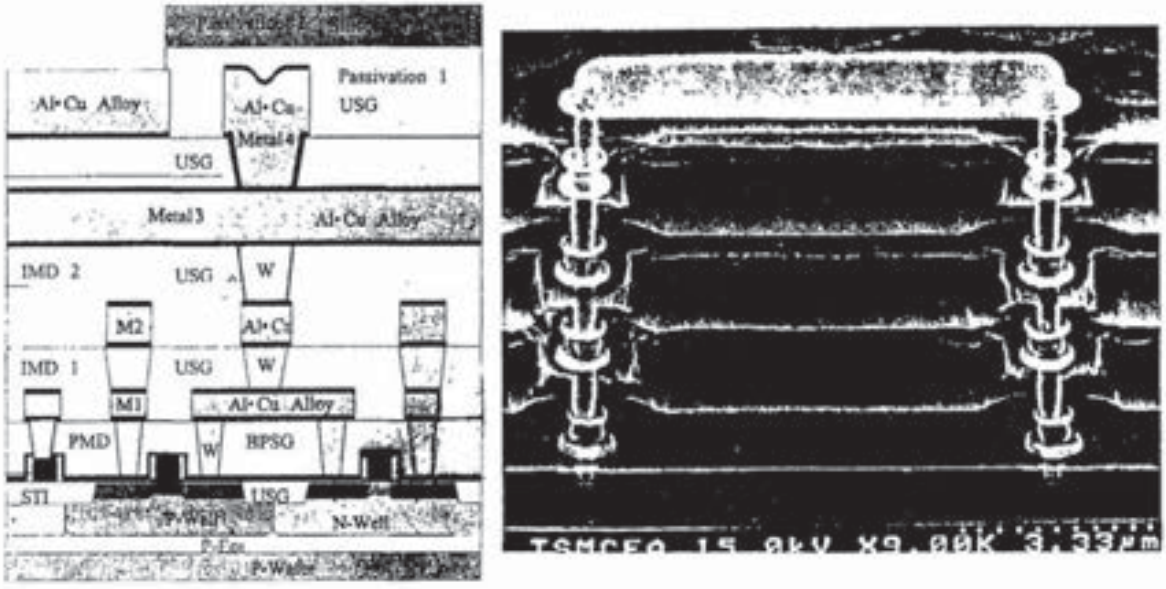


Fig. 12-13 a) Schematic of a W-plug/Al-alloy interconnect structure with stacked vias; b) SEM of an interconnect structure with stacked vias and five-levels of metal.

Note that in several aspects this sequence involves a hybrid process-technology approach. First, two-types of metal are used, one for the plug that fills the via (W), and another to form the metal lines (Al:Cu). Second, a single-damascene process is used to remove unwanted W from atop the entire surface of the chip, while a subtractive etch-process is used to selectively remove Al:Cu from some regions of the chip surface, and thus form the interconnect lines.

The ILD-CMP/W-plug/Al:Cu-line interconnect-fabrication-approach solved some significant problems, allowing it to be used to form interconnect structures with an indefinite number of levels of metal. These solutions are: **1**) the metal lines on every level have no steps to cover; **2**) vias of varying depth can be completely filled using the CVD-W step; and **3**) the vias can be stacked on top of one another if a CMP process is used to planarize the surface of the W-plugs (which allows the density of the interconnect structure to be increased).

12.1.4.1 Problems of W-plug/Al-line Interconnect Structures: While this approach to fabricating multi-level interconnects represented a significant advance, it also introduced some drawbacks. First, it is a complex process, in that it uses two-types of metal, and two-different metal-deposition-processes (each with its own set of process tools). This fabrication complexity is compounded by the fact that two-CMP steps, one ILD-etch-step, and one metal-etch-step are required. Second, the overall resistance of the Al-line/W-plug-structure can grow to be too large for small sized interconnect paths (whereas this resistance value can be reduced by using a lower-resistivity conductor-material - e.g., Cu). Finally, there are some problems associated with W-contacts and vias. That is, electromigration problems in Al often occur in the vicinity W-contacts and vias. This may be partly due to mechanical stresses that can develop at W/Al interfaces (which may enhance electromigration, or promote void and hillock formation). But it is also because the different material in the contact or via serves as a point for flux-divergence. Al also has a lower electrical resistivity than W. This can be detrimental with respect to electromigration effects, especially in circuits where current flow between different materials is significant.

12.1.5 Dual Damascene Cu/Low-*k* Interconnects

An alternative approach to forming interconnects on ULSI ICs is the *dual-damascene approach*. The Cu/low-*k* dual-damascene interconnect scheme has been pursued because it has several advantages over the subtractive W-plug/Al-alloy interconnect approach. First, an all-Cu interconnect structure has a lower resistivity and better electromigration resistance characteristics than the W-

plug/Al-alloy system. Second, by incorporating low-*k* dielectrics, the parasitic capacitances of the interconnect structure are reduced, also improving the RC-delay. Third, only one CMP step is needed (i.e., CMP of the Cu/diffusion barrier film stack), compared to the two-CMP steps needed for the subtractive interconnect (i.e., CMP of the W-plug and the ILD dielectric layer). Fourth, the damascene approach avoids having to etch Cu lines, a difficult process to implement in manufacturing. Fifth, only a single interface exists for the via (i.e., the barrier metal in contact with the Cu film in the metal level below, versus the two interfaces at the top and bottom of the W-plug).

The Cu/low-*k* dual-damascene interconnect also has some difficult challenges, some of which are listed here. First, Cu is a fast-diffusing and device-failure-inducing contaminant, even when present in trace concentrations in the silicon. Therefore, it must be surrounded by a diffusion barrier on all sides to prevent it from entering the silicon substrate. Deposition of diffusion-barrier-materials into high-aspect-ratio trenches and vias with sufficient thickness everywhere becomes harder as the via and trench dimensions are scaled. Second, CMP of Cu (and the underlying diffusion-barrier-metal-layer) is also a complex process. Third, the deposition of Cu-layers by electroplating introduces another new, and potentially contaminant-producing process technology into the wafer fab. Fourth, developing low-*k* materials to replace CVD-SiO_2 as the ILD layer for the dual-damascene structure is not simple, and new materials will need to be found for several generations of low-*k* dielectric applications. These issues are discussed in detail in other chapters of this book, namely: low-*k* dielectrics (Chap. 14); issues of dual-damascene interconnects (Chap. 15); and copper deposition and CMP (Chap. 16).

The following sequence of steps is used to create a dual-damascene interconnect structure (see Fig. 15-3 in Chap. 15):

1. A low-*k* interlevel dielectric (ILD) film is first deposited;

2. Vias and trenches are next patterned and etched in the ILD (two masking and etch steps are needed to accomplish this result);

3. A multi-layer film (i.e., barrier-metal/Cu) is deposited into the vias and trenches of the ILD (the barrier-layer and Cu-seed-layers are deposited by PVD, and the main Cu-layer by electroplating);

4. The barrier-layer/Cu film is removed from regions above the surface of the ILD by CMP, leaving the barrier-layer/Cu recessed in trenches and vias;

5. The entire wafer surface is covered with a non-conductive diffusion-barrier-layer material (e.g., SiN) to complete the encapsulation of the Cu.

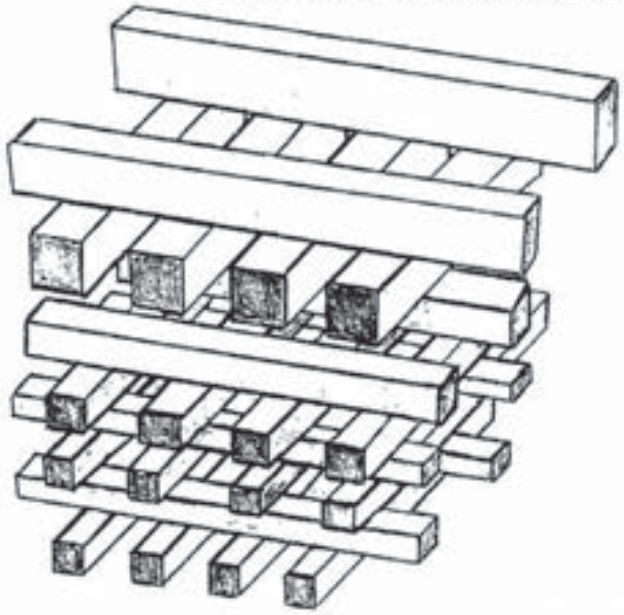


Fig. 12-14 Schematic showing the orthogonal architecture of multilevel interconnect lines.

12.1.6 Other Geometrical Aspects of Multilevel ULSI Interconnects

There are two other geometrical aspects of ULSI multi-level interconnects that should be discussed: **1**) orthogonal directions of alternating metal levels; and **2**) tiers of metal lines containing the same pitch.

In a multi-level-metal interconnect structure, alternating levels are run orthogonal to the levels above and below (Fig. 12-14). This serves three purposes. First, maximum flexibility of routing signals is made possible by using such an orthogonal multilevel structure. Second, the intralevel-capacitance between levels is minimized. Third, crosstalk can be reduced. Figures 12-15 and 12-16 show examples of such an orthogonal structure in a W-plug/Al-line interconnect structure, and in a Cu dual-damascene interconnect structure, respectively. Figure 12-17 shows a cross-sectional SEM of a six-level-metal interconnect structure that reveals the use of orthogonal metal-level layout as well.

Multilevel interconnect structures are also typically arranged in *tiers*. Tiers are pairs of orthogonal metal levels with equal pitch. In some interconnect systems, only two tiers are used (i.e., a *local interconnect-tier* and a *global interconnect-tier,* Fig. 12-18a). In a *local interconnect-tier,* the wires are only used within a cell or a block. In such local tiers the finest metal-pitch allowed by the design rules is employed. In a *global interconnect-tier,* the wires are used for inter-block communication. Thus, the lines may be quite long, and are thicker and wider than those in the local interconnect tier. In interconnect structures that contain six or more levels of metal, there may be a more optimized three-tier network (i.e., one

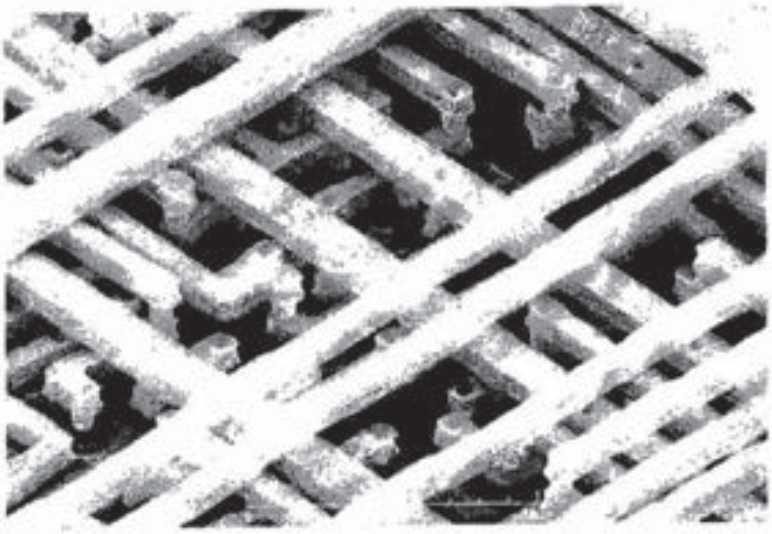


Fig. 12-15 SEM of orthogonal architecture of metal lines in a W-plug/Al interconnect.

that contains a third, *semi-global tier*, as shown in Fig. 12-18b).

12.2 PERFORMANCE REQUIREMENTS OF ULSI INTERCONNECT SYSTEMS

By 2001, microprocessors for consumer PCs were operating at speeds greater than 1-GHz (and even faster in the most advanced workstations). Data in Table

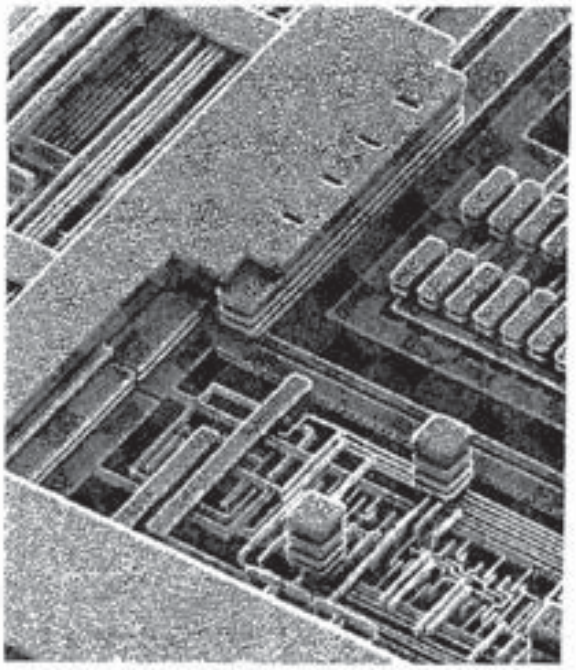


Fig. 12-16 SEM of orthogonal architecture of metal lines in a Cu interconnect.

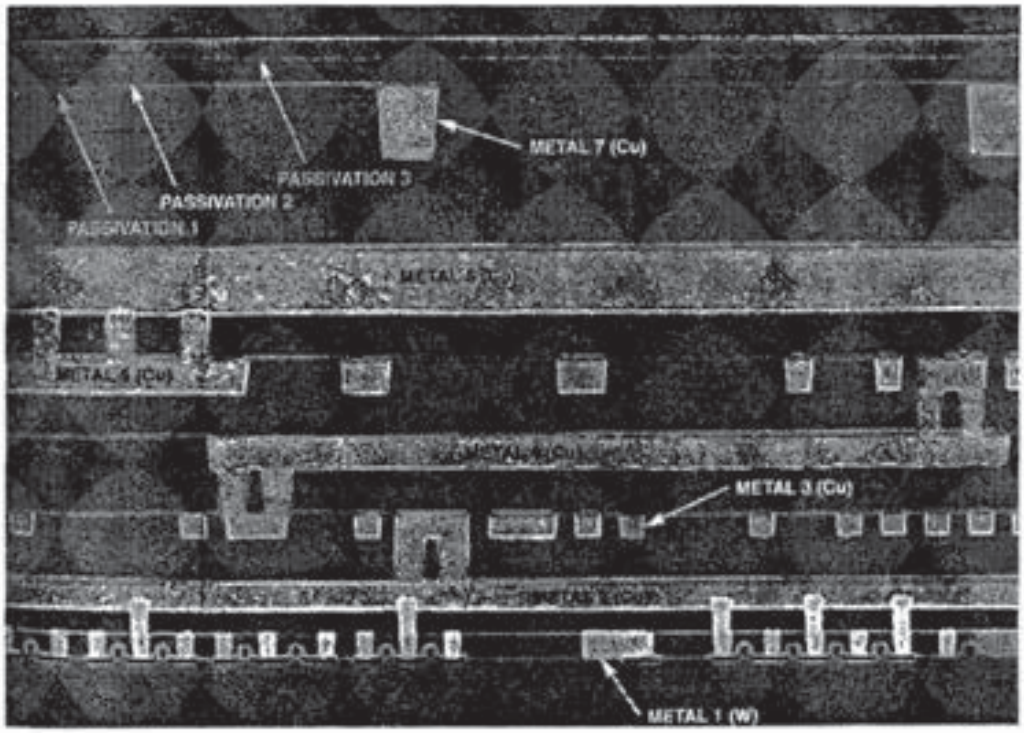


Fig. 12-17 SEM of a cross-section of a 6-metal-level interconnect.

12-1 taken from the 1999 ITRS,[3] summarizes the projected interconnect needs of microprocessors out to 2014, with respect to performance, density, feature size, and other critical technology parameters. Here we discuss those parameters and issues involved in allowing multilevel interconnect structures to achieve the

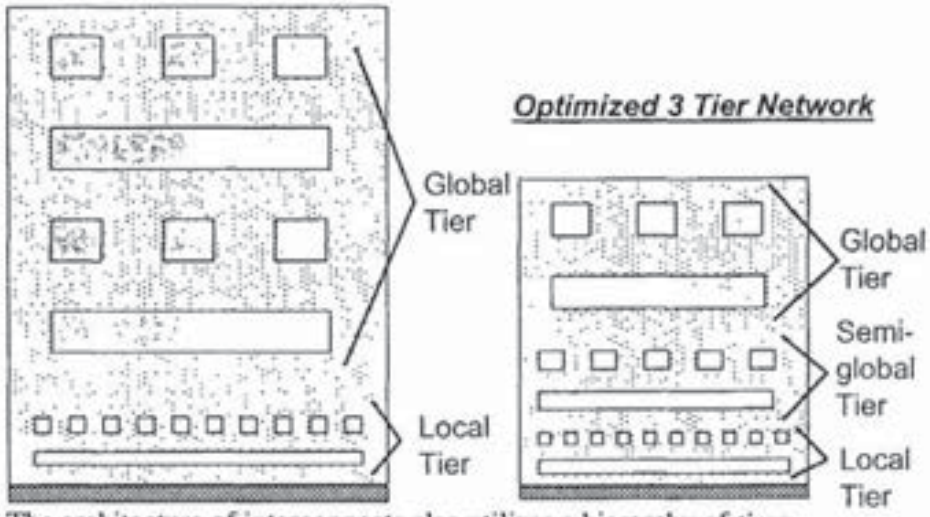


Fig. 12-18 The architecture of interconnects also utilizes a hierarchy of *tiers*.

TABLE 12-1 Selected Overall Technology Characteristics based on the 1999 International Technology Roadmap for Semiconductors.[3]

Year Technology node	1999 180 nm	2002 130 nm	2005 100 nm	2008 70 nm	2011 50 nm	2014 35 nm
MPU ½ pitch	230	160	115	80	55	40
MPU gate length (nm)	140	85–90	65	45	30–32	20–22
DRAM ½ Pitch (nm)	180	130	100	70	50	35
MPU chip frequency on-chip across chip clock–MHz (Cost-performance)	600	800	1100	1400	1800	2200
MPU chip frequency on-chip local clock–MHz (High performance)	1250	2100	3500	6000	10000	13000
High volume-MPU (Cost-performance) chip size at ramp (mm^2)	170	214	235	269	308	354
High volume-MPU (Cost-performance) Mtransistors/cm^2 at ramp (including SRAM)	7	18	41	100	247	609
Memory chip size at production (mm^2)	132	152	174	199	229	262
Memory chip Gbits/cm^2*	0.20	0.52	1.23	3.05	7.51	18.5
Maximum power w/high performance heat sink (Watts)	90	130	160	170	174	183
Maximum power (Watts) –Battery (hand-held)	1.4	2.4	2.0	2.0	2.2	2.4
Minimum logic Vdd–(Maximum performance) Volts	1.8	1.5	1.2	0.9	0.60	0.60
Minimum logic Vdd–(Lowest power) Volts (battery power)	1.5	1.2	0.90	0.60	0.50	0.30

performance goals listed in Table 12-1. The performance parameters that will be addressed are the following:

1. Propagation-delays *(RC*-delays) in digital circuits;
2. Power dissipation issues in ULSI interconnects;
3. Crosstalk in ULSI interconnects;
4. Contact resistance;
5. ESD protection.

12.2.1 Propagation-Delays in ULSI Multilevel Interconnects

As noted in the introduction of this chapter, for technology nodes beyond 0.25-μm the overall performance of ICs will generally become limited by the performance of the interconnect system on the chip. The most widely used figure of merit of an interconnect system is its $R_{int}C_{int}$-*delay*, which represents a measure of the time-delay for signal propagation. R_{int} denotes the resistance of the interconnect metal-line and C_{int} denotes the effective capacitance between the line and its surroundings. For the multi-level interconnect systems found in ULSI chips at the 0.25-μm technology nodes and beyond, the estimated effect of various geometrical and materials parameters on the $R_{int}C_{int}$ delay can be modeled using the test-structure shown in Fig. 12-19.[11] This test-structure consists of a set of metal lines between two grounded metal planes, all embedded in the interlayer-dielectric medium (ILD). Let us assign the following symbols to those aspects of the interconnect system depicted by this test-structure.

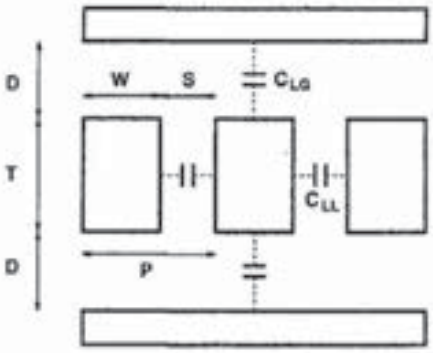


Fig. 12-19 Simplified *RC*-delay test-structure.[11] (© IEEE 1995)

That is, the metal-line *width* is w, the metal-line *thickness* is t, the *spacing* between the lines is s, the metal-line *pitch* is p ($p = w + s$), and the *thickness of the ILD between the top of the metal lines and the grounded planes* is d. Also let the *length* of the metal lines be l, and the *resistivity* of the metal be ρ Some equations can now be derived that allow the contributions made to R_{int} and C_{int} from: **1**) the materials properties; and **2**) the interconnect architecture, to be expressed. In these derivations the pitch p and the metal thickness t will serve serve as the parameters in the equations. In that way it will be possible to set $w = ap$, and $d = bt$, where a and b are constants appropriate for the given geometry.

Using this approach, the resistance of a metal line is calculated from:

$$R_{int} = \rho l/wt = \rho l/apt \qquad \textbf{(12.2)}$$

Next, the capacitance between the metal lines and the grounded planes C_{LG} (separated by an ILD whose relative dielectric constant is k), is calculated from:

$$C_{LG} = k\varepsilon_o wl/d = k\varepsilon_o apl/bt \qquad \textbf{(12.3)}$$

and the intralevel capacitance C_{LL} (assuming the same ILD) is found from:

$$C_{LL} = k\varepsilon_o tl/s = k\varepsilon_o tl/(p - w) = k\varepsilon_o tl/(1 - a)p \qquad \textbf{(12.4)}$$

where ε_o is the permittivity of free space. In Eqs. 12-3 and 12-4, C_{LG} represents the "vertical" capacitance between a metal line and a ground plane above (or below), and C_{LL} represents the "horizontal" capacitance between two neighboring lines. The total capacitance C_{int} of a metal line relative to its surroundings is thus written as:

$$C_{int} = 2(C_{LG} + C_{LL}) = 2k\varepsilon_o l\{[ap/bt] + [t/(1 - a)p]\} \qquad \textbf{(12.5)}$$

For the $R_{int}C_{int}$ time delay constant, it follows that:

$$R_{int}C_{int} = 2\,R_{int}\,(C_{LG} + C_{LL}) \qquad \textbf{(12.6)}$$

By combining Eqs. 12-2, 12-3, 12-4, and 12-6, we get:

$$R_{int}C_{int} = 2\rho k\varepsilon_o l^2\{[1/bt^2] + [1/a(1 - a)p^2]\} \quad \textbf{(12.7)}$$

This equation can also be rewritten as:

$$R_{int}C_{int} = 2\rho k\varepsilon_o(l^2/p^2)\{[p^2/bt^2] + [1/a(1 - a)]\} \quad \textbf{(12.8)}$$

The right-hand side of Eq. 12-7 has a natural interpretation, in that the factor $2\rho k\varepsilon_o$ represents the materials contributions, and the factor $l^2\{[1/bt^2] + [1/a(1 - a)p^2]\}$ represents the geometrical contributions, to the interconnect time delay.

Equation 12-8 can also be examined from a slightly different perspective if one defines the aspect ratio A of the metal lines as the ratio of line-thickness to line-width; i.e., $A = t/w$, or equivalently $A = t/ap$. In terms of A, Eq. 12-8 can be expressed as:

$$R_{int}C_{int} = 2\rho k\varepsilon_o(l^2/p^2)\{[1/a^2bA^2] + [1/a(1 - a)]\} \quad \textbf{(12.9)}$$

It can be observed that the term $1/a^2bA^2$ in Eq. 12-9 represents the "vertical" contribution to the total capacitance, and $1/a(1 - a)$ represents the "horizontal" contribution to the total capacitance.

Several conclusions about the performance of a multilevel interconnect system can be drawn by further examination of Eqs. 12-7 to 12-9. Two separate cases can be considered: **1**) where the same ILD is used in the interlevel and intralevel spaces (i.e., the value of k is constant throughout); and; **2**) where the spaces between metal lines are assumed to be filled with a material of a lower k.

12.2.1.1 Case When the *k*-Value of the ILD is Constant Throughout the Interconnect System: In this case, from Eqs. 12-8 and 12-9, it can be seen that for a fixed length l and a fixed aspect-ratio A, the RC-delay increases quadratically with decreasing feature size (i.e., decreasing pitch p). These two equations also indicate that for $A << 1$, the interlayer "vertical" capacitance is more important to the RC-delay than the intralayer (or "horizontal") capacitance., whereas the opposite is true for $A >> 1$. Also note that for the simplest geometry (where $a = 0.5$ and $b = 1$ (i.e., $d = t$), Eq. 12-9 reduces to:

$$R_{int}C_{int} = 8\rho k\varepsilon_o(l^2/p^2)\{[1/A^2] + [1]\} \quad \textbf{(12.10)}$$

12.2.1.2 Case When the *k*-Values of the Interlevel-ILD and the Intralevel-ILD are Different: In this case, one ILD is used between the metal lines and ground planes (whose "vertical" [between levels] k-value is given by k_{LG}), and another ILD is used between the metal lines themselves (whose "horizontal" [intralevel] k-value is given by k_{LL}). In that case, Eq. 12-9 becomes:

$$R_{int}C_{int} = 2\rho\varepsilon_o(l^2/p^2)\{[\,k_{LG}/a^2bA^2] + k_{LL}[1/a(1 - a)]\} \quad \textbf{(12.11)}$$

Again, for the simplest geometry where $a = 0.5$ and $b = 1$ (i.e., $d = t$), Eq. 12-11 reduces to:

$$R_{int}C_{int} = 8\rho\varepsilon_o(l^2/p^2)\{[k_{LG}/A^2] + [k_{LL}]\} \quad \textbf{(12.12)}$$

Equation 12-12 demonstrates several important conclusions. First, if a high aspect ratio interconnect structure is used on a given level (i.e., A is large), the large value of A will mitigate the influence of k_{LG} (i.e., it doesn't matter as much if an ILD with a high or low value of k_{LG} is used between interconnect levels). However, the value of k_{LL} plays significant in determining $R_{int}C_{int}$ for this case. Second, Eqs. 12-11 and 12-12 indicate that the contribution of the intralayer capacitance to the total capacitance (namely, that due to the term $1/\{a[1-a]\}$), is minimized when $a = 0.5$; i.e., when w is one half the pitch p.

While Eqs. 12-11 to 12-12 offer some general insight for minimizing the $R_{int}C_{int}$-delay, the interconnect structure can also be treated as consisting of two tiers, termed: **1**) the *local interconnect tier*; and **2**) the *global interconnect tier* (see Sect. 12.1.6). There are some differences in the ways in which the minimum *RC*-delay can be achieved in each of these tiers. The issues of are considered next.

12.2.1.3 Minimizing the *RC*-Delay in Local-Interconnect Tiers: The local interconnect tiers are those used to route interconnect lines *within* a cell or block, and in which the finest metal pitch allowed by the design rules are employed. In these tiers, the routing lines are so short that it can be assumed that $R_{int} << R_{trans}$ and $C_{int} >> C_{trans}$ (i.e., the wire resistance of lines shorter than about 1-mm is much less than the transistor on-resistance R_{trans}). In addition, the lines are so close that the largest component of C_{int} is intralevel capacitance, C_{LL}. In this case, the *RC*-delay in local interconnects is not found from Eqs. 12-2 through 12-4 (which assume the RC is controlled by R_{int} and C_{int}) but instead is described by the following equation:

$$R_{trans}C_{int} \sim R_{trans}\, k\varepsilon_o tl/s \quad \textbf{(12.13)}$$

Thus, unlike in the global interconnect tiers where the *RC*-delay is controlled by $R_{int}C_{int}$ (and where the delay is thus dependent on the *square* of the interconnect length l), in the local interconnect tiers the delay is only *linearly-dependent* on the interconnect length. In addition, in this region C_{int} is dominated by the intralevel coupling between adjacent lines. Furthermore, in the local interconnect regions it is also clear that changes in the metal resistance will not play a big role in altering the RC delay, since $R_{trans} >> R_{int}$. This means that in this line-length regime, the resistance and capacitance are decoupled. Consequently, factors that reduce C_{int} (such as lowering the value of k between metal lines, and reducing

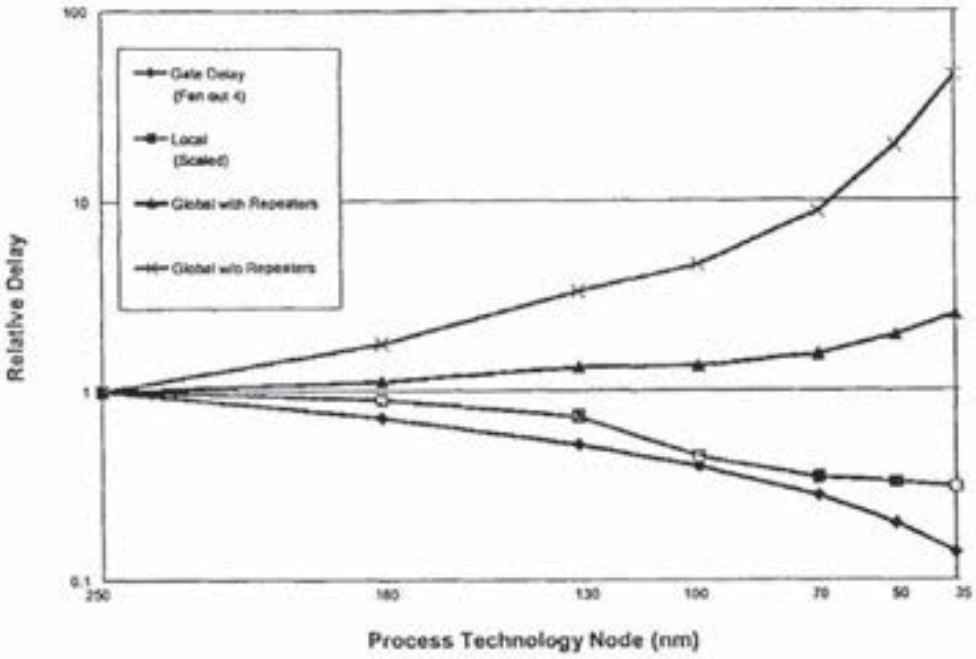
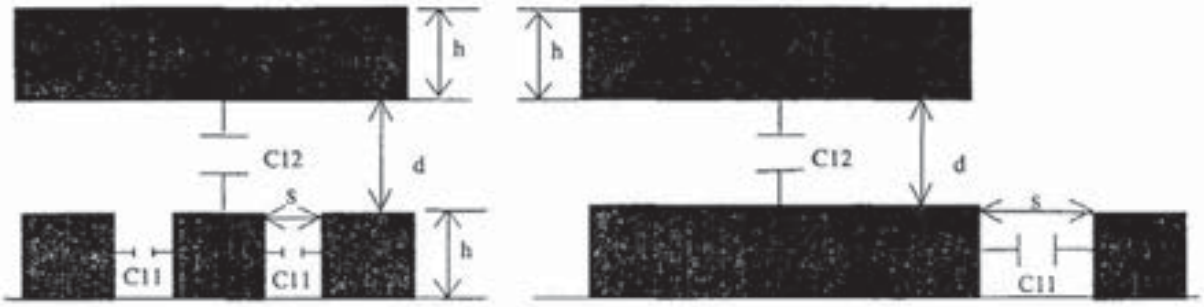


Fig. 12-20 Interconnect versus delay, from ITRS99.[3]

the height of the metal lines), will help decrease the *RC*-delay (provided that the interconnect resistance of the metal line does not exceed that of R_{trans} due to the reduction in the height of the metal lines). Figure 12-20 shows it is predicted that the relative delay of local interconnect tiers will remain roughly comparable to the gate delay as devices are scaled down toward 50-nm gate-lengths.

12.2.1.4 Minimizing the *RC*-Delay in the Global-Interconnect Tiers: The factors that minimize the *RC*-delay in the global tiers of the interconnect structure are somewhat different than the factors that govern minimizing *RC*-delay in the local interconnect tiers. (The *global interconnect tiers* are those used for inter-block communication employing the upper metal levels.) First, there are two coupling regimes that are encountered in global interconnect tiers – the *intralevel regime* and the *interlevel regime*. Figure 12-21a depicts the conditions that exist in the intralevel regime (which is typical of the denser interconnect lines found in the lower levels of global interconnect tiers). In this regime the intralevel capacitance dominates ($C_{LL} >> C_{LG}$). In the interlevel regime (which tends to exist in the upper levels of the global interconnect region, where wide wires are used to lower the resistance of long runs – Fig. 12-21b), the interlevel capacitance dominates ($C_{LG} >> C_{LL}$).

In the *intralevel global interconnect-regime*, $R_{int}C_{int}$ is calculated by combining Eq. 12-2 and 12-4, to get:



(a) Dense metal case where C11>C12
Major contribution in metal level

(b) Wide metal case where C11< C12
Major contribution between metal levels

Fig. 12-21 Capacitance coupling contributions for interconnect systems: a) Dense interconnect case where major coupling is between minimum-spaced lines; and b) Wide interconnect lines where the major coupling is between metal-levels.

$$R_{int}C_{int} = 2R_{int}C_{LL} = 2\,k\varepsilon_o\rho l^2/ws \qquad \textbf{(12.14)}$$

In the *interlevel global interconnect-regime* the interlevel-capacitance dominates, and hence $R_{int}C_{int}$ is calculated by combining Eqs. 12-2 and 12-3:

$$R_{int}C_{int} = 2\,k\varepsilon_o\rho l^2/td \qquad \textbf{(12.15)}$$

We see that in both cases the *RC* delay is strongly impacted by the length of the interconnect line (i.e., *RC* increases as the square of the line-length).*

In many circuits it is undesirable to operate under these conditions (i.e., in which the interconnect limits the propagation delay of the circuit). To overcome this limitation, some circuit designs include *repeaters* or *buffers* along the interconnect lines that exceed a critical length. The use of repeaters (while not desirable from a circuit-density standpoint), is nevertheless an essential technique for optimizing signal propagation. Implementing repeaters can even extend the usefulness of a given process technology. Their use will optimally halve the delay of a long line, with a small penalty of the repeater-delay and some device area.[13] Figure 12-22 shows how the addition of repeaters to an interconnect system can reduce the number of levels of metal as the functional density of ICs is increased.[16] In another example, Figure 12-23 depicts the interconnect delay contribution vs. feature size for varying repeater insertion lengths of 3000-μm

* Note that the intermediate situation (i.e., where both intralevel- and interlevel-capacitance effects are comparable in magnitude) is treated by Eqs. 12-8 and 12-9, since this situation lies between the two extremes and includes capacitive contributions both in- and between-metals-levels.

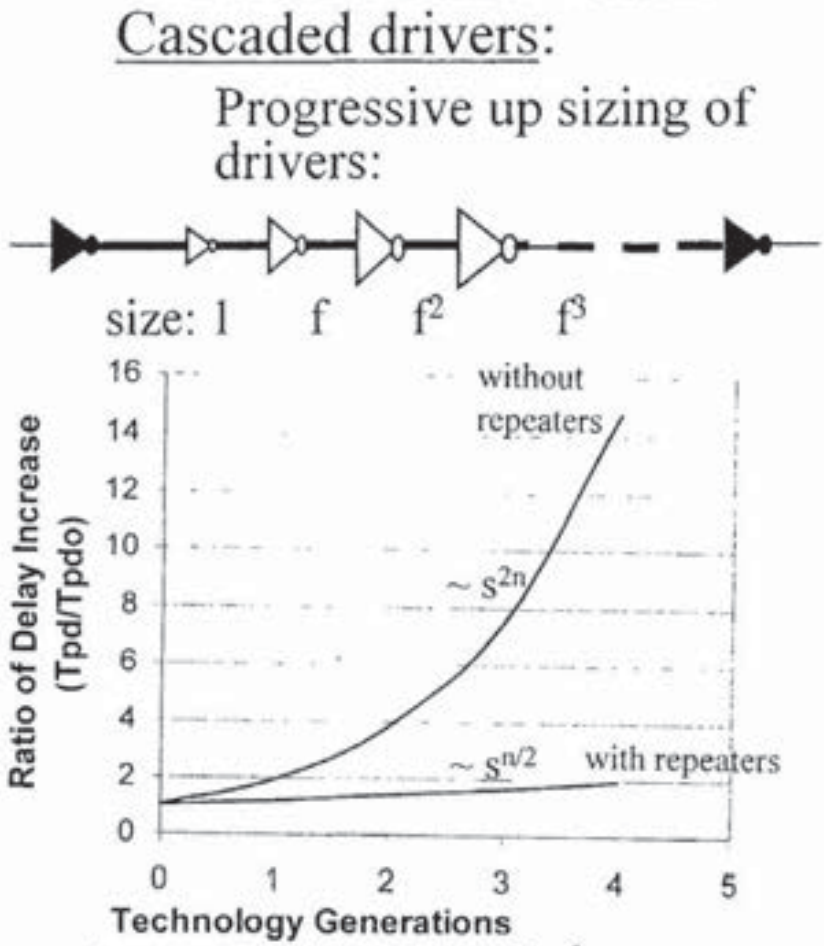


Fig. 12-22 Impact of repeaters on global propagation delay.[5]

and 5000-μm.[14] Note that reducing the spacing at which repeaters are placed can compensate for delays due to RC parasitics and circumvent the need for new materials (e.g., Cu and low-k dielectrics).

On the other hand, where long interconnect lines must be used, large pitches p must be employed, and the lines should be made wider and thicker. This implies that long interconnect lines should only be implemented in the upper interconnect metal layers where there is this room for increased pitch (wider lines and spaces), and where the lines can be made thicker. This is the approach that is commonly used today.

Thus, a multi-level interconnect system requires a hierarchy of scales, with the smallest lines in the first metal-level, and progressively wider (and perhaps thicker) and longer lines in the upper levels. One disadvantage of using thicker and wider lines in the upper levels is that fewer metal lines can be fabricated on any level. Consequently, to achieve the same functionality in the face of continued feature size scaling, the number of metal levels used in interconnect systems will also continue to grow.

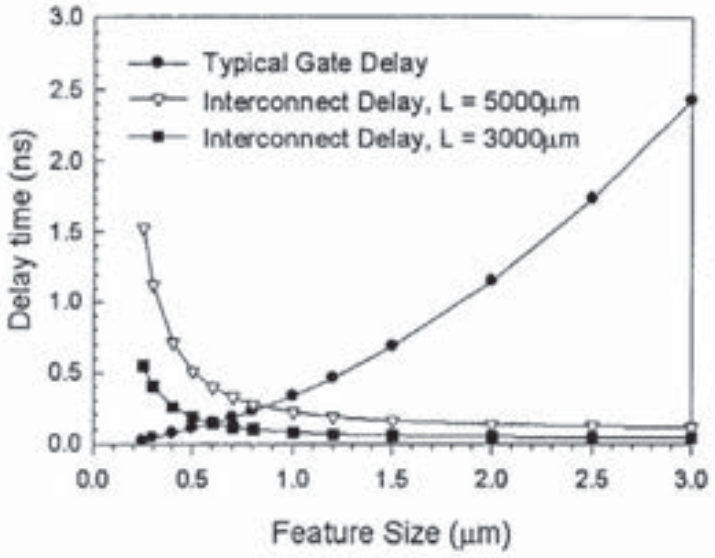


Fig. 12-23 Gate and interconnect delay versus feature size. Interconnect delay is shown for repeater spacings (L) of 3000-μm and 5000-μm.[14]

On the materials side of this issue, it is evident that the lowest $R_{int}C_{in}$-delay in global interconnect-layers will be achieved if a metal with the lowest possible ρ is used together with an ILD having the lowest possible k-value. With respect to the metal, this means using copper. (The details of fabricating metal interconnects with copper are described in Chap. 16). With regard to the ILD, it is especially beneficial to selectively reduce k_{LL} between the metal lines. We will explain how this can be accomplished in Chap. 14.

12.2.2 Power-Dissipation Issues in ULSI Interconnects

As the requirements for performance in ULSI continue to increase, the power demands will also arise unless the overall system capacitance, resistance and/or operating voltages can be reduced. The power-dissipation associated with a functioning IC can be calculated using either of the following two relationships:

$$P \sim f\, C_{total} V^2 \qquad \textbf{(12.16)}$$

or

$$P \sim I_{total}^2\, R_{total} \qquad \textbf{(12.17)}$$

where:

P = power dissipated (watts)
f = operating frequency (Hz)
C_{total} = total IC capacitance
= $C_{trans} + C_{int}$ (farads)
R_{total} = total IC resistance
= $R_{trans} + R_{int}$ (ohms)

V = IC operating voltage (volts)
I_{total} = IC operating current (amperes)
C_{trans}, C_{int} = transistor & interconnect capacitance contributions
R_{trans}, R_{int} = transistor & interconnect resistance contributions

Although Eqs. 12-16 and 12-17 indicate that the IC operating voltage plays the greatest role in determining the power dissipation, its selection is based on circuit and system design issues. On the other hand, the system resistance and capacitance are the chief parameters that can be controlled by the fabrication technology. Currently, the transistor resistive losses are responsible for most of the chip power dissipation (and these are discussed in Chaps. 1 and 5). However, as the ICs are being operated at higher and higher frequencies, an appreciable fraction of the power dissipation and voltage drops are associated with the resistive component of the interconnect. Due to the decreased size of interconnects on ULSI (~200-nm minimum feature sizes in 2000), interconnect resistance is increasing and losses can become significant if the wiring is not designed properly. The overall resistance of the interconnect includes both the line resistance and the resistance of one or more contacts and/or vias along the wiring path. Joule-heating caused by current flowing through the resistive interconnect enhances electromigration and represents a growing reliability concern. The power dissipated at the higher-resistance transition-regions of the contact or via interface is of special concern, especially for large pulsed currents. Since power-dissipation in interconnects is a growing problem, it is necessary to ensure interconnects are properly optimized to reduce such power-dissipation.

12.2.3 Crosstalk in ULSI Interconnects

Another issue that impacts the performance of ULSI interconnects is *crosstalk*. Interconnect lines of substantial length act as "antennae" which experience unwanted voltage pulses induced by the rapid voltage swings in neighboring wires of close proximity (which arise from the propagation of digital pulse waveforms along the driven lines). Induced voltages due to intralevel-coupling (the largest crosstalk component between dense wires in the same level) can easily achieve 20% of the supply voltage for very rapidly changing clock waveforms.[4] Figure 12-24 shows a simulation of the amount of crosstalk experienced by an undriven wire that is 1-mm long, separated by 100-nm from an adjacent interconnect line of the same length and driven with a very fast clock edge.[5] V1 is the driving waveform and V2 is the crosstalk voltage induced in the undriven line. The rate of rise of the clock edge imposes a large induced voltage

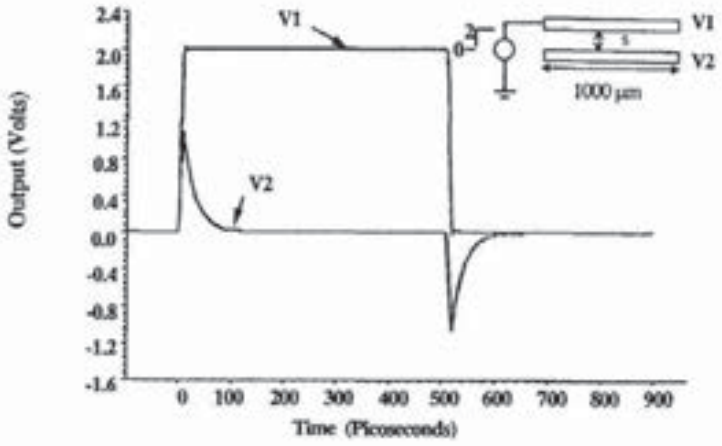


Fig. 12-24 Electrical modeling result showing the crosstalk voltage induced from a driven interconnect-line (V1) into an undriven line (V2) seeing a 20-psec waveform-edge. The lines were spaced at s = 1000-nm with a common length of 1-mm.[10]

which is related to the parasitics of the interconnect. In this example, where the clock-edge rise time is 20-psec, the magnitude of the imposed crosstalk voltage can be as much as 60% of the voltage waveform V1 (2 x 0.6 = 1.2V). As the intralevel-spacing between interconnects decreases, the magnitude of the crosstalk noise can become substantial, as is evident from the simulation data shown in Fig. 12-25.

It is more difficult to establish the level of crosstalk in dynamic systems than in the example given above. More sophisticated circuit simulators are required to

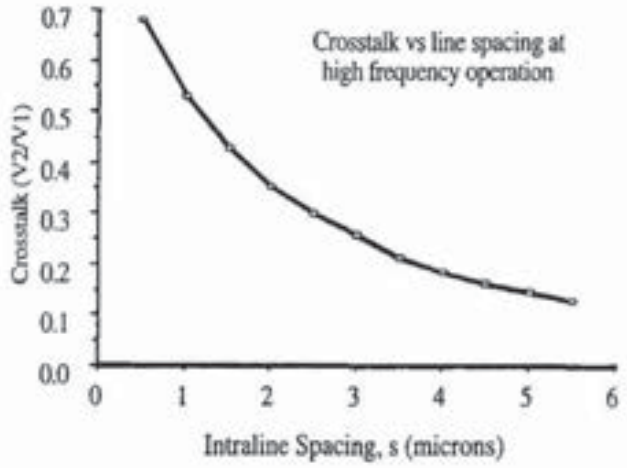


Fig. 12-25 Simulation results showing that the magnitude of the crosstalk between 2 interconnects separated by different line spacings, s.[10] Conditions described in Fig. 12-24.

estimate the impact of such crosstalk on circuit device performance. The potential of surrounding wires, wire resistance, wire length, and the capacitive load all affect the amplitude of crosstalk signals. Nevertheless, it is possible to reduce intralevel crosstalk by changing the relative dielectric environment. That is, by inserting an embedded low-k dielectric between dense intralevel wires, electric-field lines can be diverted to the short orthogonal lengths of underlying interlevel wiring (lowering capacitive-coupling between intralevel interconnects).

Crosstalk will become a significantly greater problem as ULSI mixed-signal devices begin to emerge in the marketplace. It is predicted that noisy digital wires and circuit elements will cause a number of problems for analog elements that are in close-proximity to such sources of noise. New techniques involving such novel electrical-isolation-methods as shielding and signal conditioning will be needed. This is because, in some circuits, analog devices must be immune to such "noise" down to levels that are three to six orders of magnitude smaller than the power-supply voltage-levels.

12.2.4 Contact Resistance in ULSI Interconnects

The magnitude and uniformity of contact and via resistance is another forthcoming challenge (see Vol. 2, Chap. 3 for more background information on contact and via resistance). Contact and via resistances can be substantial and have a significant impact on the overall wiring resistance of the interconnect system. These resistance contributions act in the same way as adding length to an interconnect line, and thus can factor heavily into voltage drops and delays in the wiring. For 0.25-μm contacts with an interfacial-contact-resistivity ρ_c to heavily doped silicon of $5x10^{-8}$-$1x10^{-7}$-Ω-cm^2, such contact-resistance values are approximately 100-200-ohms/contact.[6] Vias for systems using Al alloys with W-plugs have practical metal-to-metal contact resistivities of approximately $2x10^{-9}$ Ω-cm^2 and a resistance of about 5-Ω/via at the same geometry.[7] Vias which use copper metallization, however, are expected to show substantially lower contact-resistance because they have only one critical barrier interface per via (instead of two interfaces when W-plugs and Al alloy interconnect lines are used), and because the interfacial-contact-resistivity should be lower than $1x10^{-9}$-Ω-cm^2.[8]

Cleaning prior to the deposition of the contact- and via-filling conductor-films and control of dimensional tolerances are critical to the reproducibility of via or contact resistance. It is predicted that production variability will increase with the migration to smaller feature sizes and have a negative impact on resistance. This

resistance must be accounted for in the delay analysis when multiple levels with a large number of contacts and vias are used.

12.2.5 Electrostatic Discharge Protection Issues of Interconnects

Electrostatic discharge events (ESD) can cause large transient currents to flow in an IC. This can cause "fusing" in upper wiring levels, thus jeopardizing the long-term reliability of the IC. ESD robustness has been shown to be a severe reliability concern by several research groups, for interconnect generations with metal features smaller than 0.25-μm.[9,10] Since short term discharges can exceed 2000 volts for durations of hundreds of nanoseconds, it has been shown interconnects are becoming more prone to failure due to this type of electrical stress. Copper interconnects exhibit a substantially better resistance to this type of failure, when compared to aluminum interconnects, due to the higher melting point and thermal conductivity of copper.

12.3 SUMMARY OF USLI INTERCONNECT ISSUES AND FUTURE TRENDS

Based on the discussion of this chapter, the following trends are likely to be encountered in ULSI interconnect systems:

1. If the materials of an interconnect system are not changed (i.e., the metal resistivity ρ and ILD-k values remain the same), the RC-delay will continue to increase as device dimensions (and therefore interconnect pitch p) are reduced. There are two main reasons why this increase in RC occurs:

   a. Increasing the aspect ratio A of metal lines (to keep R constant) beyond about 2 produces diminished improvements in RC - even for a constant pitch p (because C_{LL} rises while R remains constant). However, if the pitch is reduced, this makes C_{LL} grow even larger, which, in turn causes an additional rise in RC.

   b. The metal line lengths tend to scale less strongly than does the metal pitch p. Hence, even as devices keep shrinking, the ICs will continue to contain some rather long lines (which will have larger C_{total}/unit length, for lines with the same length).

   We noted earlier, however, that in short- and intermediate-length local interconnects, changes in the metal resistance will typically not play a major role in the RC-delay (since in this regime $R_{trans} >> R_{int}$). For such cases, the resistance and capacitance are decoupled. That is, factors that reduce interconnect capacitance (such as a reduction in the dielectric

constant of the ILD, and/or a reduction in the height of the metal lines), will help improve *RC* delay (as long as the interconnect resistance increase due to reducing the metal film thickness does not approach R_{trans}).

2. If ρ and/or k can be decreased (especially a reduction in the value of k of the ILD used in intralevel metal spaces k_{LL}) this will cause a decrease in *RC*. Hence, there will be continual effort expended to integrate Cu and low-k dielectrics into ULSI interconnect systems. The issues involved with incorporating low-k ILDs are covered in Chap. 14, and the challenges encountered when attempting to implement Cu-metal-interconnects are described in Chap. 16. Figure 12-26 shows a possible migration path for materials use in IC interconnect structures.[15]

3. More levels of metal will need to be used as the feature sizes are decreased.

4. Optimizing interconnect length is a powerful method for reducing *RC*-delay. This can be influenced by design philosophy and device architecture. Use of repeaters, while not desirable from a density standpoint, is another key approach for optimizing signal-propagation, and can extend the use of a given process technology. Since incorporation of repeaters into an interconnect system is a circuit-design issue, this approach not considered further in this text.

5. Increasing the metal line aspect-ratios and adding more levels of interconnect are two methods for improving interconnect performance and density. However, these approaches will reach practical limits in a few

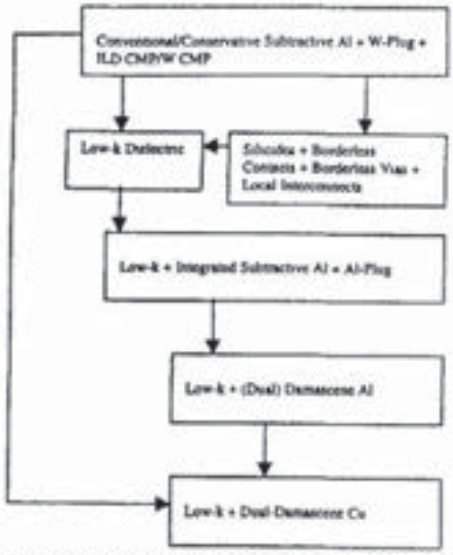


Fig. 12-26 Schematic illustrating an interconnect migration.[15]

generations.[11] (For example, continuing to increase the aspect ratio A above $A = 2$ will bring only diminished improvements in RC because the reduction in the metal resistance will be offset by an increase in intralevel capacitance.) New interconnect materials, improved circuit-design techniques, and possibly novel interconnect schemes will be needed to address the growing limitation that interconnects pose to ULSI performance. Circuit design needs to more accurately model interconnect performance to avoid overdesign or unexpected performance limitations, and automatic layout tools need to make more efficient use of existing interconnects. The novel interconnect approaches include RF-transmission and optical-interconnects. More information about these novel techniques can be found in Ref. 12.

REFERENCES

1. S.C. Sun, "Process Technologies for Advanced Metallization and Interconnect Systems," *Tech. Dig. IEDM*, 1997, p. 765.

2. R. Liu *et al., Proc. Int. Interconnect Technol. Conf.* (IEEE), 1999, p. 21.

3. International Technology Roadmap for Semiconductors, 1999 Edition, Semiconductor Industry Association, San Jose, CA.

4. C. Cregut *et al.*, "Low-*k* Dielectrics Influence on Crosstalk: EM Analysis and Characterization," *Proc. of the International Interconnect Technol. Conf.*, San Francisco CA, June 1-3, 1998, p. 59.

5. H.P. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*, Addison Wesley Publ., Reading MA, 1990.

6. I.S. Park, *et al.*, "Fabrication of 256 Mbit DRAM with a Fully Planarized Double Level Interconnection," *Proc. 12th VLSI Multilevel Interconnect Conf. (VMIC)*, Santa Clara CA, 1995, p. 45.

7. K. Kikuta *et al.*, "Comparitive Study of W-plug, Al-plug, and Al-Dual-Damascene for 0.18 μm ULSI Multilevel Interconnect Technologies," *Proc. of the Internat. Interconnect Technol. Conf.*, San Francisco CA, June 1-3, 1998, p. 140.

8. J. Zhang *et al.*, "CVD Cu Process Integration for sub-0.25 μm Technologies," *Proc. of the Internat. Interconnect Technol. Conf.*, San Francisco CA, June 1-3, 1998, p. 163.

9. S.H. Voldman, "ESD Robustness and Scaling Implications of Aluminum and Copper Interconnects in Advanced Semiconductor Technology," *IEEE EOS/ESD Symposium , 1997, p. 316.*

10. K. Banerjee *et al.*, "Characterization of VLSI Circuit Interconnect Heating and Failure under ESD Conditions," *1996 IEEE 34th International Reliability Physics Proceedings*, April 30-May 2, 1996, Dallas, TX, p. 237.
11. M.T. Bohr, "Interconnect Scaling – The Real Limiter to High Performance ULSI," *Tech. Dig. IEDM*, 1995, p. 241.
12. M.E. Thomas and R.H. Havemann, "Overview of Interconnect," Ch.12, in *Handbook of Semiconductor Manufacturing Technology*, Y. Nishi and R. Doering, Eds. Marcel Dekker, New York, 2000, pp. 303-4.
13. H.P. Bakoglu and J.D. Meindl, " Optimal Interconnection Circuits for VLSI," *IEEE Trans. Electron Dev.*, May 1985, p. 903.
14. J.C. Eble *et al*, "Minimum Repeater Count, Size, and Energy Dissipation for Gigascale Integration (GSI) Interconnects," *Proceedings of the International Interconnect Conf.*, San Francisco CA, 1998, p. 56.
15. J. Givens, "Engineering CMP for Manufacturability and Yield," *Abs. Mtg. Electrochem. Soc.*, Fall 1999, Abs. 1275.
16. S. Wills, "Interconnect Architecture," in *IEDM* 2000 *Short Course: Advanced Interconnects: Design, Process, Integration*, p. 27.

Chapter 13

POLYCIDES AND SALICIDES OF $TiSi_2$, $CoSi_2$, AND NiSi

Silicon forms many stable metallic and semiconducting compounds with metals, and such compounds are termed *silicides*. Several metal-silicides exhibit low-resistivity and high thermal-stability, making them suitable materials for several ULSI applications. The formation kinetics, film properties, and potential applications such silicides have been extensively examined by Murarka[1] and by Nicolet and Lau.[2] In this chapter the use of silicides for the purposes of fabricating polycides and salicides in ULSI applications is discussed.

13.1 POLYCIDES AND SALICIDES

The term *polycide* is a condensed form of the words polysilicon-silicide. Polycides are widely used for gate structures in MOSFETs that contain a non-self-aligned silicide. In this structure, a silicide is typically blanket-deposited over a doped polysilicon film and the composite film is patterned together (Fig. 13-1). Polycide gates can achieve a gate sheet-resistance smaller than 10-Ω/sq, compared to that of a polysilicon gate without a silicide of more than 40-Ω/sq. Even with its considerably lower-resistivity, the polysilicon gate with a silicide can only be used for local-routing over short-distances in deep-submicron ICs.

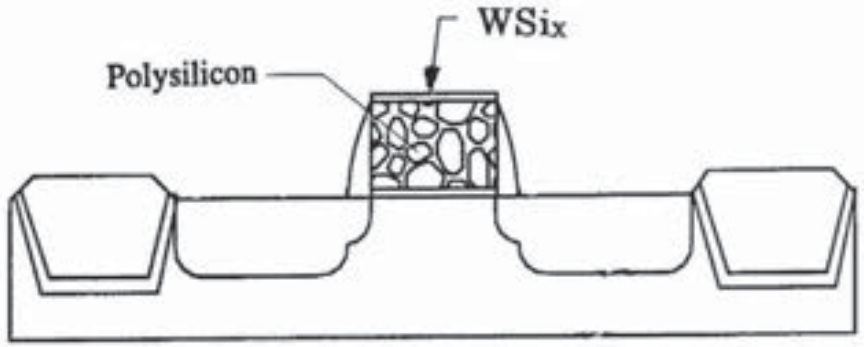


Fig. 13-1 Schematic cross-section of a polycide gate on a MOSFET.

Still, such a local-interconnect can be beneficial when only one or two levels of metal are used in a circuit - since the poly routing can help to reduce die size by as much as 20%. In fact, the incorporation of a polycide allowed the conductivity of poly lines serving as local interconnects to be so significantly reduced that the invention of the polycide-gate-structure was responsible for delaying implementation of multilevel-metallization in CMOS by several years.

Several types of polycides have been employed in IC technology, including WSi_2, $TaSi_2$, and $MoSi_2$. All are refractory-metal-based, and exhibit adequate thermal stability and resistance to processing chemicals. Since such polycides are formed as part of the gate fabrication sequence, they must be able to withstand the temperatures of any processes carried out after gate formation (typically 800-1000°C). Consequently, as WSi_2, $TaSi_2$, and $MoSi_2$ films remain stable at these elevated temperatures, they can be used for such gate polycides.

However, the silicides of Ti and Co (i.e., $TiSi_2$ and $CoSi_2$) are unstable above about 900°C (i.e., they agglomerate above ~900°C). Thus, it is not possible to use these materials as polycides if the gate formation sequence is followed by a thermal step that exceeds 900°C. Even if a process is used in which this temperature is not exceeded, neither $TiSi_2$ or $CoSi_2$ is a good candidate for a polycide that is patterned at the same time as the polysilicon gate. This is primarily due to an issue that arises during the patterning of the poly gate. That is, silicides composed of $TiSi_2$ are attacked by the Cl-based chemistry used for polysilicon dry-etching, and this makes linewidth control difficult if $TiSi_2$ must be etched during the patterning of the critical gate level. $CoSi_2$, on the other hand, lacks a volatile reaction by-product and is thus is difficult to pattern by dry etching. In addition, both $TiSi_2$ and $CoSi_2$ are vulnerable to processing chemicals, especially to HF solutions.

However, $TiSi_2$ and $CoSi_2$ are ideal candidates for *self-aligned-silicide* (also referred to as *salicide*) applications. As will be seen, such salicide processes are performed *after* the polysilicon lines for the gate structure have been patterned, and the source/drain regions of the MOSFET have been fabricated. Since the formation of the $TiSi_2$ or $CoSi_2$ in the salicide process are self-aligned techniques that require no etching, the etching difficulties mentioned above are avoided (as is exposure to temperatures > 900°C), permitting these materials to be used.

As the minimum dimensions of an IC approached 1-μm, the conventional contact structures used up to that point (see Chap. 3, *Silicon Processing for the VLSI Era*, Vol. 2 for details on contact formation in VLSI ICs) began to limit device performance in several ways. First, it was not possible to minimize the contact resistance if the contact hole was also of the minimum feature size, and

problems with cleaning the small contact holes became another concern. In addition, the area of the source/drain-regions could not be minimized because the contact-hole had to be aligned to these regions with a separate masking step, and extra area thus had to be allocated for misalignment. (The larger area that had to be used for such source/drain-regions also resulted in increased source/drain-to-substrate junction-capacitance, which slowed-down device speed).

Finally, when non-minimum-width MOSFETs were fabricated with conventional contacts, several small, uniform-sized contact holes were usually used rather than one wider contact hole. (The reason for this is that if all of the contact holes on an entire chip are of the same size, they are more likely to clear simultaneously during the etching process.) A problem with using either one wide contact hole (as well as if several contact holes of equal size are used) is that a metal film must be used to contact each contact hole. The result is that valuable routing area is thus removed from the chip surface (because it must be used for running metal connections to the contacts), and this results in an increase in the die size. The main approach used to alleviate this problem has been to employ a self-aligned silicide on the source/drain-regions (where these silicides are formed at the same time as the polycide structure is created). The process of simultaneously fabricating such self-aligned-silicides and polycides on the gate is referred to as a *salicide* (self-aligned silicide) process.

13.2 POLYCIDES FABRICATED BY CVD OF TUNGSTEN SILICIDE (WSi_X)

As mentioned earlier, various refractory metal silicides including WSi_X, $TaSi_2$, and $MoSi_2$ were explored in the 1980s as possible shunt materials in polycide structures. Tungsten silicide (WSi_X) films ended up being the most widely adopted for this application, with CVD being the chosen method of forming the WSi_X thin-films. Polycides with WSi_X films now find widespread use as word-line and bit-line interconnects in IC memory chips. Note that unlike CVD-W, which is used for plug/interconnect applications (and is thus never exposed to temperatures above about 500°C), these WSi_X films are deposited on polysilicon during the gate formation process. This means that such WSi_X films will have to have to be able to endure temperatures up to 800-1000°C during BPSG reflow, source/drain implant-activation, and poly reoxidation. WSi_X has also found some application as a stand-alone adhesion layer for blanket CVD-tungsten films.

A process sequence for producing the WSi_X polycide gate-structure is depicted in Fig. 13-2. It shows that WSi_X is blanket-deposited onto the doped polysilicon film, and then this WSi_X/poly multilayer is etched to form the polycide gate-structure. The CVD deposition of WSi_X won out over other candidate processes

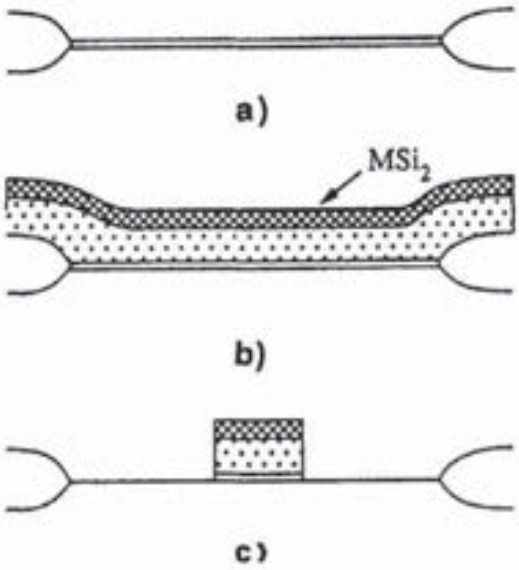


Fig. 13-2 Process sequence for producing a WSi_x-polycide-structure: a) Gate oxide growth; b) Polysilicon and WSi_x deposition by CVD; c) Patterning the polycide.

(i.e., evaporation and sputtering of WSi_x) for the following reasons: **a**) this process can produce high-purity WSi_x films without the need for high-vacuum deposition equipment; **b**) its throughput is acceptable; **c**) better step-coverage is obtained than with PVD; and **d**) good wafer-to-wafer and run-to-run uniformity is attainable. The first chemical reaction used to deposit WSi_x was the following, and it is capable of producing films with the benefits listed above:

$$WF_6\ (gas) + 2\ SiH_4\ (gas) \rightarrow WSi_2\ (solid) + 6\ HF\ (gas) + H_2\ (gas) \qquad \textbf{(13-1)}$$

Tungsten-silicide film deposition using this reaction is carried out at pressures between 50-300 mtorr (LPCVD) and at deposition temperatures between 300-400°C.[3] This reaction is similar to the silane reduction of WF_6 used to produce CVD-W films. As mentioned earlier, if it is desired that WSi_x be deposited, a higher flow rate of SiH_4 is needed. This ensures that the deposited film is WSi_x and not W. During film deposition, WSi_2 forms with excess Si collecting in the grain boundaries. Since an excess of silicon exists in such films, the chemical formula to denote them is given as WSi_x. However, unless $x \geq 2.0$, WSi_x films formed with this reaction are prone to cracking and peeling away from the underlying poly during later high-temperature steps. That is, the excess Si in the as-deposited silicide film avoids the possibility of consuming some of the underlying poly material, which could cause the above problems. Thus, in practice a SiH_4/WF_6 flow-ratio of more than 10:1 is used to ensure a deposition

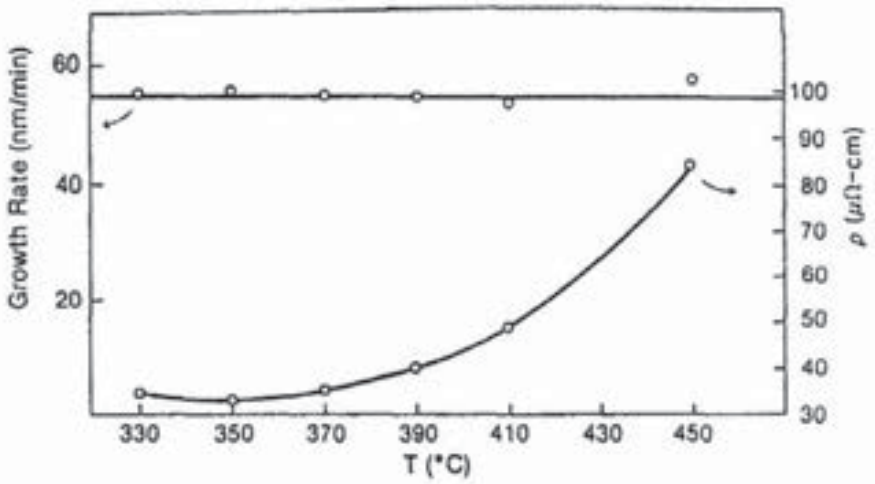


Fig. 13-3 Resistivity of CVD-WSi_x versus x, where x varies from 2.2 to 2.6.

with x ~2.2–2.6. As-deposited WSi_x-films exhibit high resistivity (~500-$\mu\Omega$-cm), but this drops to about 50-$\mu\Omega$-cm after an RTP anneal at 900°C. Note that the resistivity of CVD-WSi_x also depends on its stoichiometry as is shown in Fig. 13-3, where x varies from 2.2 to 2.6. The resistivity of the film increases as it becomes progressively richer in Si.

It was found that CVD-WSi_x films formed by the reaction of Eq. 13-1 had high fluorine-concentrations (~$10^{20}/cm^3$). This creates problems when such films are used with gate oxides thinner than about 20-nm, because low-field oxide breakdown and threshold-voltage shifts are observed.[4] Such problems arise because some of the fluorine gets incorporated into the gate oxide during anneal of the polycide stack. As a result, an alternative chemistry using dichlorosilane (DCS - SiH_2Cl_2) instead of SiH_4 was developed for CVD of WSi_x:[4]

$$WF_6(gas) + 3.5SiH_2Cl_2(gas) \rightarrow WSi_2(sol) + 1.5SiCl_4(gas) + 6HF(gas) + HCl(gas) \quad \textbf{(13-2)}$$

This DCS process is also carried at low pressure (LPCVD) in the temperature range of 570–600°C. The fluorine content is much lower than in films formed with the silane reduction of Eq. 13-1, and the chlorine content is also quite low. The resistivity is comparable in both types of films. However, the step-coverage is better in DCS films, and its deposition-rate is three to five times faster. Peeling and cracking is also less severe in WSi_x films formed from the reaction of WF_6 and DCS. DCS has thus largely replaced silane for CVD-WSi_x processes.

Exposure of the WSi_x films to high-temperature O_2-ambients (900°C) will cause a dense adherent layer of SiO_2 to grow on the silicide while the polycide structure underneath remains intact. The oxide forms by reaction of oxygen with the excess-silicon dissolved in the silicide film, and continues after this silicon is

consumed by diffusion of additional silicon from the polysilicon to the surface of the WSi_X. The oxide-formation reaction occurs at the oxide/silicide interface.

CVD-WSi_X films are deposited in cold-wall reactors. Batch processes were initially used but single-wafer systems have also been introduced. A process that integrates polysilicon deposition, polysilicon doping, and WSi_X deposition in a single cluster tool is available.[6] Advantages of such a configuration include cleaner processing and less cleanroom floor space.

13.3 SALICIDE (SELF-ALIGNED-SILICIDE) STRUCTURES

In the salicide process, the entire surface area of the source and drain regions of a MOSFET are covered with a conductor film (i.e., a metal silicide). This approach becomes even more attractive if such a film can be formed using a self-aligned process, in that a self-aligned process does not entail any additional masking steps. The process sequence performed to fabricate such a self-aligned conducting silicide on the source/drain-regions (and to simultaneously form a conducting silicide on gate-polysilicon structure [polycide]) is as follows:

1. After the source- and drain-regions have been implanted and annealed to form the source/drain-junctions (Fig. 13-4a), CVD-SiO_2 sidewall-spacers on the polysilicon gate are fabricated (i.e., by CVD deposition of an oxide, followed by an etchback step which leaves spacers on the poly sidewalls - Fig. 13-4b).

2. The metal film (of either Ti or Co) used to form the silicide is deposited (typically by sputtering).

3. Next, the wafer is heated, which causes the silicide reaction to occur wherever the metal is in contact with the silicon. Everywhere else, the metal remains unreacted (Fig. 13-4c). Note that RTP is extensively used in this first silicidation step (1st RTP-step). At the conclusion of this step, the silicide is typically in a high-resistance phase. A second RTP-step (carried out after the unreacted metal is removed in step **4**), is performed at a higher temperature than the 1st RTP step. This transforms the silicide into the desired low-resistivity phase.

4. The unreacted-metal is selectively removed through the use of a wet-etchant that does not attack the silicide, the silicon-substrate, or the SiO_2 (Fig. 13-4d). As a result, each exposed source and drain region, and each exposed polysilicon gate region is now completely covered by a silicide film, but no metal or silicide film remains elsewhere.

5. As noted in step 3, after the unreacted metal is stripped, a 2nd RTP step is performed at a higher temperature than the 1st RTP step. This higher temperature process transforms the high-resistance phase of the silicide remaining on the source/drain- and poly-regions to the desired low-resistance silicide-phase.

6. A dielectric layer is deposited onto the silicide, and contact holes are opened in it down to the silicide layer.

7. Metal is deposited into the contact holes to make contact with the silicide (Fig. 13-4e).

The goals of the above salicide process are as follows:

1. The silicide should meet the required low sheet-resistivity specifications called for by the particular technology node being fabricated.

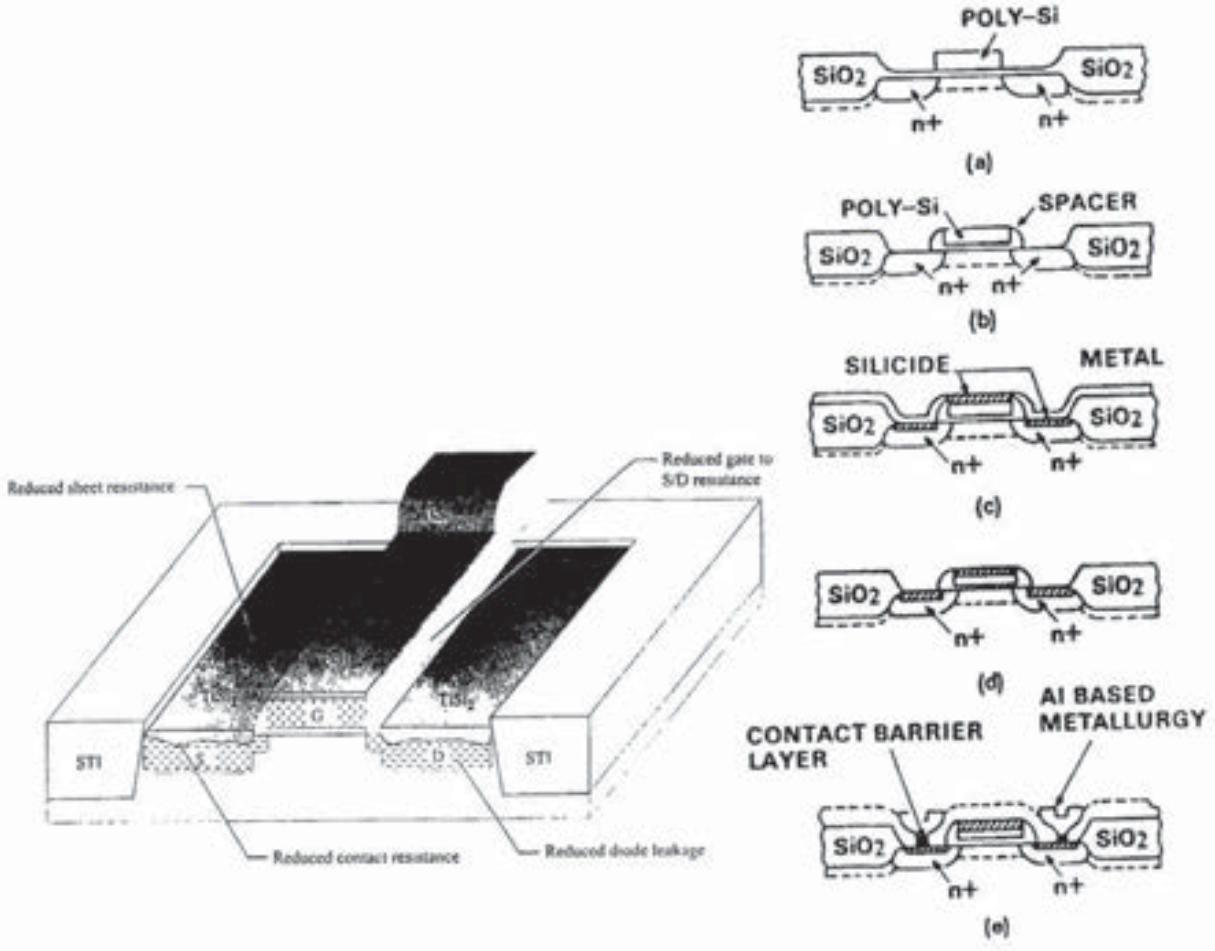


Fig. 13-4 Salicide processing steps and final structure.

2. Low junction leakage should exist on all source/drain geometries.

3. No bridging should occur across isolation regions or across gate sidewall spacers.

4. A low specific-contact-resistivity should exist for the silicide/Si contact.

The above contact-formation-method offers several advantages over the conventional contact structure:

1. The value of the parasitic-series-resistance component due to the bulk sheet-resistance of the source- and drain-regions R_{sh} becomes negligibly small because the ρ_{sh} value of the silicide is typically 1-2-Ω/sq, whereas that of the diffused junction region alone is 40-120 Ω/sq.

2. The contact area of the silicide and the Si is much larger than that of the metal-Si area in a conventional contact in a device of the same total area. Thus, the value of the contact resistance, R_{co} arising from the contact between the silicide and the Si substrate is smaller than in a conventional contact. That is, R_{co} is equal to the specific contact resistivity ρ_c divided by the contact area.

3. The contact hole is now used to make contact between the metal and the silicide, but the value of ρ_c exhibited by silicide-metal interfaces is typically $\leq 1 \times 10^{-9}$-Ω-cm^2. This is two orders of magnitude lower than nominal metal-to-silicon specific contact resistivities, ~1×10^{-7}-Ω-cm^2, obtained at or near solid-solubility silicon-doping-limits. Hence, the contribution to R_{co} due to the metal-silicide contact resistance at the contact hole is also usually negligibly small.

As it became apparent that a self-aligned contact-structure would be beneficial for reducing the values of the parasitic-series-resistances of the conventional contact structure, some other processing trends also allowed a new group of materials to be considered as contact metals to Si. These materials were $TiSi_2$ and the group-VIII silicides $CoSi_2$ and NiSi. When it also became necessary to fabricate shallower junctions, processing technology had to be performed at lower temperatures to avoid the formation of deep junctions. When the maximum temperature after the silicide formation was thus restricted to less than 900°C, the door was opened to the use of these silicide films (since all of them are unstable at temperatures higher than 900°C).

There are several important reasons why $TiSi_2$ and the group-VIII metal silicides are especially attractive as candidates for self-aligned ohmic contacts and local interconnects to silicon. These are as follows:

1. Such silicides exhibit lower resistivities than the refractory-metal silicides of W, Ta, and Mo. That is, the respective resistivities are 13-20-$\mu\Omega$-cm for $TiSi_2$ (with C-54-phase crystal structure), 16-18-$\mu\Omega$-cm for $CoSi_2$, and 15-18-$\mu\Omega$-cm for NiSi.

2. All of the group-VIII metals react with Si at 600°C or less, and with the exception of Ti, there is no reaction of the metal with the masking oxide during the silicide formation. Even with Ti, the reaction is minimal below 600°C, and the reaction products are soluble in the Ti-removal etch.[8]

3. When the silicides are formed at such low temperatures, the metal atoms diffuse into the Si, and react with it. As the metal atoms diffuse into the SiO_2, no Si is available for reaction, and hence, the metal over device areas covered with SiO_2 remains unreacted. This also prevents lateral formation of the silicide over the narrow oxide-spacer-region during the silicide formation-step.*

4. For the cases of Co and Ni (but not for Ti), dopant atoms in the silicon are "snowplowed" into the silicon substrate as the $CoSi_2$ and NiSi are formed. Such snowplowing maintains a high doping concentration at the silicide-Si interface, which helps produce low-resistance contacts.

5. In all cases, after the silicide has been formed, the unreacted metal can be etched off from the oxide surface using chemical etchants that do not attack the oxide, the silicide, or the substrate silicon. This last characteristic allows self-aligned contacts to be formed through the use of these silicides.

The focus of most development for the salicide process has revolved around $TiSi_2$ and $CoSi_2$. Both of these silicides exhibit lower resistivities than silicides formed with the refractory metals, W, Ta, and Mo and they can withstand process temperatures in excess of 800°C (such temperatures are frequently used to reflow the doped glasses deposited over the salicided source/drain-regions).

The salicide process exhibits a limitation related to the fact that the gate and the source/drain-silicides are formed at the same time. On the gate, it is desirable for the silicide to have the lowest possible sheet resistance (so that the gate electrode will also possess a low interconnect-resistance). To achieve this, a thick

* There is an exception to this statement. During $TiSi_2$ formation at temperatures above 600°C, Si diffuses into the Ti. The diffused Si then reacts to form $TiSi_2$ over the oxide-spacer regions, resulting in shorting of the gate and source/drain regions. Procedures for avoiding this effect are discussed later.

Silicide-layer on the gate is needed. Over the source/drain-regions, however, the silicide can be only of limited thickness, in order to prevent excess consumption of the substrate-silicon by silicide formation. Thus, a thicker silicide, though favorable at the gate level, is detrimental to contact formation (and vice-versa). Going forward, this tradeoff between sheet-resistance and silicon-consumption represents a key scaling issue, in that thicker salicide films are needed to achieve low sheet-resistance-values, while shallower junction-depths require thinner silicide films.

Another problem is that it is difficult to metallize shallow junctions (≤ 200-nm-thick), because of the junction leakage that is observed when the silicidation process is used. The implant damage (caused during the source/drain junction-formation) may not be sufficiently annealed-out by the lower-temperature silicide-formation step. This damaged region may remain within the depletion-region of the shallow junction, giving rise to higher leakage. In addition, the RIE-process used to open the contact-holes in the doped oxide (deposited after silicide formation), may damage the silicide.[9] Techniques to reduce or eliminate this last set of problems are discussed in later sections. A final problem of salicide technology is that it degrades the effectiveness of some electrostatic-discharge-protection devices used with CMOS circuits.

13.4 THE EFFECTS OF SALICIDES ON MOSFET PERFORMANCE

A self-aligned-silicide on the source/drain-region lowers the contact-resistance because the metal-Si contact area is increased from the area of the contact window(s) to the entire source/drain area. For example, consider the contact resistance R_{co} (Ω) of an *n*MOSFET with a source/drain area of 1-μm x 1-μm and a contact window the size of 0.4-μm, with and without a silicide on the source/drain-region. Assume that the specific-contact-resistivity is identical for a silicide/n^+-Si contact and a metal/n^+-Si contact (e.g., in this case 1×10^{-7}-Ω-cm^2), and 0.2×10^{-8}-Ω-cm^2 for the metal/silicide contact. (Note that the value of the specific-contact-resistivity of metal-to-silicide is much smaller than the specific contact resistivity of silicide or metal to n^+-Si.)

Thus, if no silicide is formed on the source/drain region, then only the contact resistance R_{co} of the metal/n^+-Si contact exists, and its value is calculated from:

$$R_{co}([\text{metal}/n^+\text{-Si}]\ \text{contact}) = (\text{specific contact resistivity})/\text{contact area}$$
$$= (1\times10^{-7}\ \Omega\text{-cm}^2)/(0.16\times10^{-8}\ \text{cm}^2) = 65\ \Omega \qquad \textbf{(13.3)}$$

On the other hand, if a silicide has been formed that covers the entire source/drain-regions, the contact-resistance will be the sum of the two contacts in series. That is, it will be sum of the contact resistance R_{co}([metal/silicide] contact), and R_{co}([silicide/n^+-Si] contact). The value of R_{co}([metal/silicide] contact) is found from:

$$R_{co}([\text{metal/silicide}]\ \text{contact}) = (1\text{x}10^{-7}\text{-}\Omega\text{-cm}^2)/(1\text{x}10^{-8}\ \text{cm}^2) = 10\ \Omega \quad \textbf{(13.4)}$$

The value of R_{co}([silicide/n^+-Si] contact), is found from:

$$R_{co}([\text{silicide/}n^+\text{-Si}]\ \text{contact}) = (0.2\text{x}10^{-8}\text{-}\Omega\text{-cm}^2)/(0.16\text{x}10^{-8}\ \text{cm}^2) = 1.2\ \Omega \quad \textbf{(13.5)}$$

The sum of these two contact resistances is:

$$R_{co}(\text{total}) = R_{co}([\text{metal/silicide}]\ \text{contact}) + R_{co}([\text{silicide/}n^+\text{Si}]\ \text{contact})$$
$$= 11.2\ \Omega \quad \textbf{(13.6)}$$

Therefore, even for a nominal-sized MOSFET (i.e., in the case above, having a width of 1.0-μm), the self-aligned-silicide significantly decreases the contact resistance.

However, even a contact-resistance of 65-Ω is only a small fraction of the channel-resistance. Reducing this quantity even further does not significantly increase device performance of the nominally-sized MOSFETs of a circuit. On the other hand, for wider MOSFETs (used, for example, to supply more current to subsequent stages of the circuit, or to provide drive-current for off-circuit loads), the benefit can much more significant. Assume that a transistor is 10-μm wide (instead of being only 1.0-μm wide, as in the above example), and that this MOSFET is capable of delivering 4-mA of current when V_{GS} is 3.3 V. Let us see how the use of a self-aligned-silicide on the source- and drain-regions impacts the performance of such a device (compared to one that does not). We will do this by considering the effect of the silicide on each side of the device separately, namely first on the drain-side, and then on the source-side.

The effect of the contact-resistance R_{co} on the *drain-side* of such a wider MOSFET device is to reduce the drain voltage V_{DS} by the amount of $R_{co}I_D$. If the MOSFET is operating in the saturation region (as is the case in most digital circuits, when the MOSFET is ON), then the reduction in V_{DS} has little effect. If the transistor is operating in the linear region (as in analog circuits), then the output current I_D is reduced from:

$$I_D = V_{DS}/R_{channel} \quad \textbf{(13.7)}$$

to:

$$I_D = (V_{DS} - R_{co}I_D)/R_{channel} \quad \textbf{(13.8)}$$

or:

$$I_D = V_{DS}/(R_{co} + R_{channel}) \quad (13.9)$$

Therefore, the effective resistance in the linear channel has been increased by the amount R_{co}. For this MOSFET, the channel resistance in the linear region, $R_{channel}$, can be calculated as V_{Dsat}/I_D, or approximately 250-Ω. If no silicide exists, and only a single metal/n^+-Si, 0.4-μm contact is made to this drain region, the contribution of R_{co}, is about 25%.

The effect on the performance of the device due to the R_{co} on the *source-side* of the device is more significant. That is, a resistance on the source-side of the MOSFET not only increases the drain-voltage by the same amount as on the drain-side, but it also raises the potential at the source by $R_{co}I_D$. The effective gate voltage V_{GS} is then reduced by the same amount:

$$V'_{GS} = V_{GS} - R_{co}I_D \quad (13.10)$$

A reduction in effective gate voltage is equivalent to a reduction in transconductance:

$$g_m = \Delta I_D/\Delta V_{GS} \quad (13.11)$$

$$g'_m = g_m/(1 + R_{co}g_m) \quad (13.12)$$

A MOSFET with a g_m = 300-μS/μm and an R_{co} = 65-Ω, has a $g'_m = 0.98\text{-}g_m$, or only a 2% loss in drive current for a 1-μm-wide MOSFET. But for a 10-μm-wide MOSFET, a 20% loss for the drive current results. Therefore, for wide MOSFETs, either multiple contacts, or a self-aligned silicide is needed to reduce the parasitic resistance.

The contact-resistance will degrade the device performance more severely on MOSFETs with smaller minimum dimensions. At the 0.25-μm node, the contact window will be about 0.3-μm wide, making R_{co} ~140-Ω, compared to the 65-Ωin the example in which the contact window was 0.4-μm wide. At the 0.18-μm technology node, R_{co} will be about 250-Ωwithout a self-aligned silicide. Even for a 1-μm-wide MOSFET, the g'_m will be reduced to only about $0.9g_m$. If a self-aligned silicide is used, R_{co} drops to about 18-Ω in the 0.25-μm node, and to about 25-Ω in the 0.18-μm node.

13.5 SALICIDES FORMED WITH TITANIUM SILICIDE ($TiSi_2$)

Salicides formed with titanium silicide $TiSi_2$ are currently being used extensively in the production of 0.25-μm CMOS devices. $TiSi_2$ is attractive for this salicide application because it exhibits low resistivity, and because it can reduce native-

oxide layers (making it the only known refractory metal that can reliably form a silicide on both polycrystalline and single-crystal silicon through a thermal reaction). Furthermore, devices fabricated with titanium silicide on the gate electrode are more resistant to high-field-induced hot-electron degradation than are conventional poly-Si gate devices.[10] It is conjectured that the $TiSi_2$ is an effective getter for the hydrogen atoms introduced during the hydrogen anneal. Less hydrogen is thus incorporated into the gate oxide, and this improves the hot-electron reliability.

Some aspects of the $TiSi_2$ process are not as favorable, including:

1. The reactivity of Ti with SiO_2 can cause unwanted reactions with the oxide spacers during the silicide-formation process (a problem referred to as *bridging*).

2. $TiSi_2$ is less thermally stable than WSi_2 or $MoSi_2$ (i.e., it agglomerates at $T > 900°C$).

3. Because Ti films have a high propensity to oxidize, the silicide must be formed in ambients that are free of oxygen.

4. For line-widths smaller than about 0.25-μm it becomes difficult to form low-sheet-resistivity $TiSi_2$ films (and this phenomenon is termed the *$TiSi_2$ narrow-line effect*).

5. Silicon is consumed by the silicide reaction (i.e., about 30-nm of Si is consumed by 10-nm of Ti to form 35-nm of $TiSi_2$). Silicon consumption during silicide-formation becomes a major issue in the process of forming very shallow source/drain-junctions.

6. During silicide formation, dopants in the junctions redistribute themselves. This represents an especially severe problem with $TiSi_2$.

Each of these issues will be discussed in the upcoming sections.

13.5.1 Reaction Kinetics of $TiSi_2$ Formation (and the Problem of Bridging)

In practice, a two-step formation-process is typically used to form $TiSi_2$.[11] The first step (1st RTP step) is carried out at a temperature is between 620-680°C in nitrogen for 15-60 sec. This short RTP step transforms the Ti in contact with Si (or polysilicon) into the metastable, high-resistivity C-49 phase-$TiSi_2$:

$$Ti + 2Si \dashrightarrow TiSi_2 \ (C\text{-}49) \qquad \textbf{(13.13)}$$

Following this step, selective etching is used to remove the unreacted Ti (e.g., in a room-temperature mixture of DI H_2O, 30% H_2O_2, and NH_4OH [5:1:1]). Then a second RTP-step at a temperature of 800°C in nitrogen for 30-60 sec is used to

lower the $TiSi_2$ sheet-resistance by transforming the metastable C-49-phase into the low-energy, low-resistivity orthogonal C-54-$TiSi_2$ phase:[12]

$$TiSi_2 \text{ (C-49)} \dashrightarrow TiSi_2 \text{ (C-54)} \qquad \textbf{(13.14)}$$

The resistivity of C-49-phase-$TiSi_2$ is about 3-4 times higher than that for C-54-phase-$TiSi_2$ (60-70-$\mu\Omega$-cm vs 13-20-$\mu\Omega$-cm). Phase-transformation occurs at 700-750°C, but it will be seen why the process is not carried out in a single step.

The reason why a two-step process is used to form C-54-phase-$TiSi_2$ involves the problem termed *bridging*. That is, oxide-spacers at the edges of the poly-silicon are formed to separate the silicided gate and source/drain-regions. But these spacers are typically only 200-300-nm wide. As a result, any lateral-formation of silicide can easily bridge the separation and cause the gate to be come shorted to the source/drain. Figure 13-5 illustrates the formation of $TiSi_2$ along the sidewall of the gate, resulting in a short between the gate and the source/drain. Such lateral-formation of $TiSi_2$ along the gate-sidewall oxide-spacer occurs because silicon is the moving species of this reaction. Thus, it diffuses into the Ti-regions that cover the spacer-oxide, where it subsequently reacts with the Ti. However, when annealing is performed at a lower temperature (and for short times), the Si does not diffuse as far, and this inhibits the lateral $TiSi_2$ reaction along the spacer. However, at least 600°C is needed to cause Ti to react with Si, even to form the C-49 phase. At this temperature, the Si-diffusion will already produce bridging - even if RTP is used.

It has also been reported that the surface of the SiO_2 sidewall spacers is covered with CF_x polymers after the dry-etch step used to from the spacers (carried out in a CHF_3/C_2F_6 plasma). The presence of these polymers enhances the lateral Ti-silicide growth over oxide-spacers and can result in bridging. Therefore a cleaning procedure (not based on a fluorine-based chemistry) should be used remove these polymers. A suggested post-spacer-etch cleaning-process in this report was an O_2 plasma + 2% diluted-HF in H_2O + APM (ammonia-hydroxide/hydrogen-peroxide mixture). This process was able to remove the fluorocarbon polymers while consuming less than 2.8-nm of the oxide-spacers.[13]

As a result, in addition to using the low-temperature RTP process, the 1st RTP step is also carried out in an N_2 ambient. This causes the Ti to absorb a significant amount of N_2 (e.g., more than 20atom%) before extensive Si diffusion occurs. The N_2 is absorbed preferentially at the Ti grain-boundaries, which "stuffs" the grain-boundary diffusion paths. This further reduces the diffusivity of Si in the Ti, and essentially suppresses the lateral silicide-reaction. Annealing in pure N_2 or pure forming gas (95% N_2 + 5% H_2), allows C-49-phase-$TiSi_2$ form-

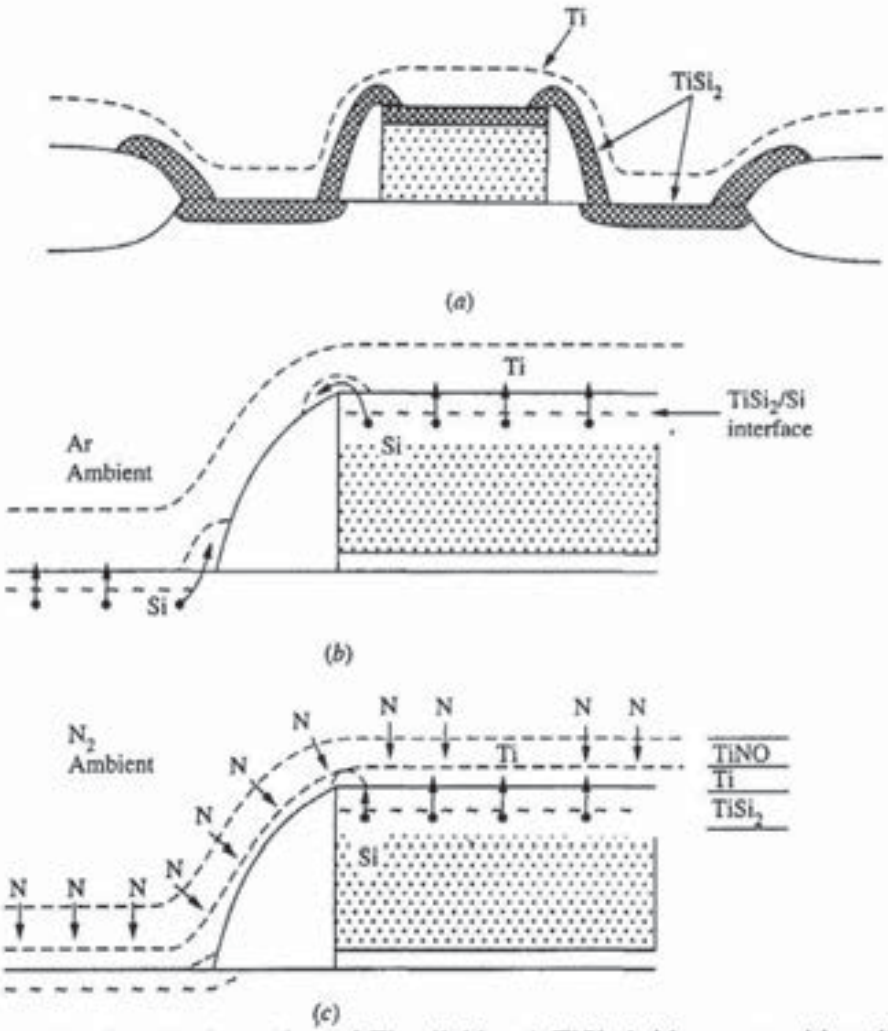


Fig. 13-5 "Bridging" during the formation of Ti salicide: a) $TiSi_2$ bridges over sidewall spacer and a short circuit is formed between the source-and-gate and the drain-and-gate; b) During $TiSi_2$ formation, Si diffuses into the Ti to form $TiSi_2$. Under favorable conditions, Ti can diffuse a long distance and form $TiSi_2$ on the sidewall spacer; c) In a nitrogen ambient, nitrogen diffuses into the Ti and blocks the diffusion of Si, reducing or eliminating bridging. TiN and/or Ti-O-N are formed on top of the $TiSi_2$.

ation without bridging (provided the temperature does not exceed ~680°C). It is also important to control the N_2 ambient so that it contains less than 5 ppma of oxygen or water to avoid unwanted oxidation of the Ti film.[15] The absorbed N_2 also reacts with the Ti to form TiN at the Ti surface.[14,15]

Another reason for performing the 1st RTP step at the temperatures below 700°C is that at higher-temperatures, Ti and the spacer SiO_2 will react rapidly enough to form significant amounts of titanium oxides. If any residues of this

reaction remain on the oxide spacer, they can degrade device performance by compromising the oxide integrity or by producing bridging.

13.5.2 The $TiSi_2$ Narrow-Line Effect

As the polysilicon lines onto which the $TiSi_2$ is formed are made narrower, the sheet resistivity of the $TiSi_2$ is found to increase above the 13-20-$\mu\Omega$-cm exhibited by C-54 phase-$TiSi_2$ (Fig. 13-6). As a result, the resistance of the polycide-coated gate-poly lines no longer meet the resistance specifications for the corresponding technology nodes. The factor that causes this increase in $TiSi_2$ sheet resistivity on narrow polysilicon-lines is that the transformation of the C-49-phase into the C-54-phase is incomplete. It was observed that after the 1st RTP-step, small-grained (0.2-μm) C-49-$TiSi_2$ was formed on both wide (> 10-μm) and narrow (0.4-μm) polysilicon lines. After a 725°C anneal, the $TiSi_2$ on the wide polysilicon-lines transformed into large-grained (1 to 10-μm) C-54 phase. However, on narrow-lines only smaller-grained (< 1-μm) C-54-phase-$TiSi_2$ was observed, and only on a few lines. At higher temperatures (775°C anneal), C-54-phase-$TiSi_2$ was formed on both wide- and narrow-lines. At 800°C, $TiSi_2$ started to agglomerate, and the line resistance rapidly degraded. The incomplete transformation of C-49 to C-54 on narrow lines causes the line to have some regions along the line consisting of C-49 and others of C-54 (Fig. 13-7), resulting in an overall increase (and variability) in the $TiSi_2$ sheet-resistance.

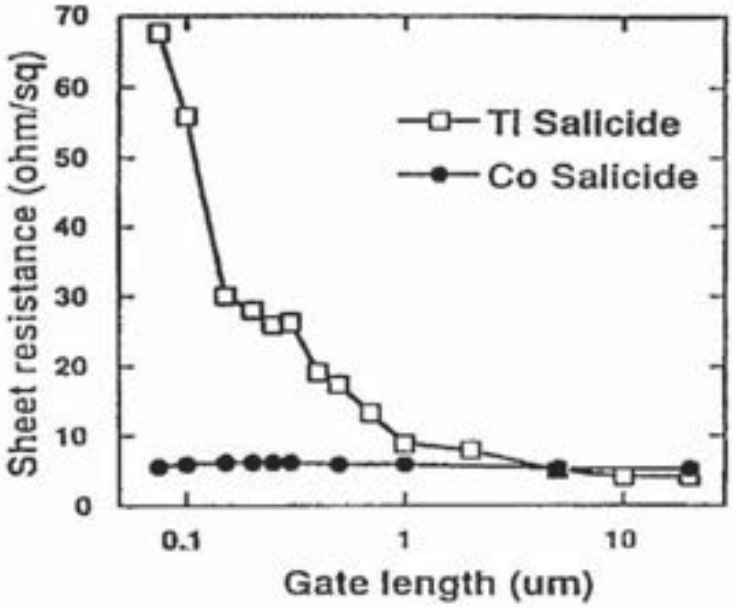


Fig. 13-6 The $TiSi_2$ narrow-width effect: As the polysilicon lines onto which the $TiSi_2$ is formed are made smaller (i.e., < 1 μm), the sheet resistivity of the $TiSi_2$ increases.

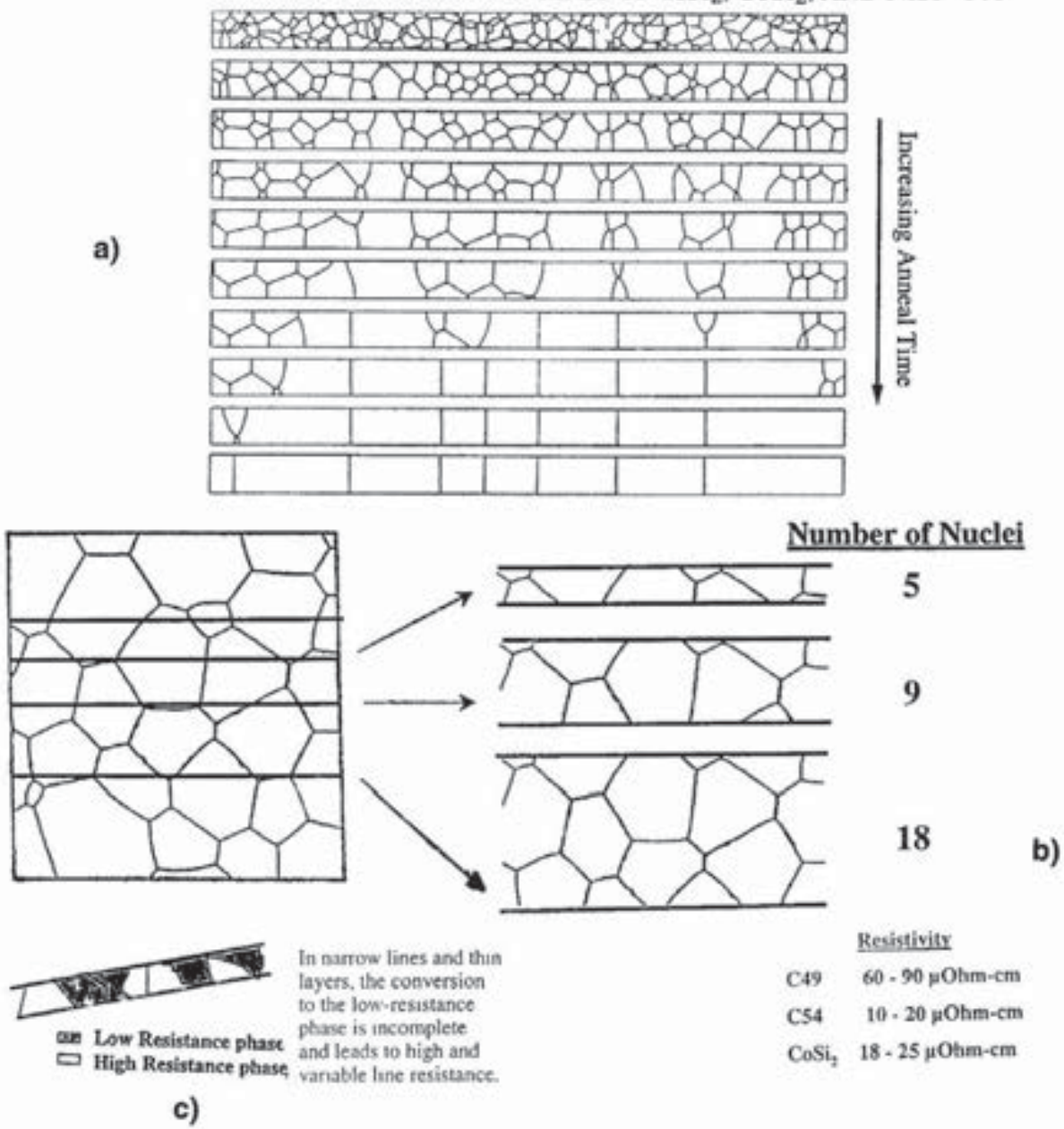


Fig. 13-7 a) Simulation of grain-growth within a line with increasing annealing time, starting with small grains; b) When the as-deposited grains are already large (relative to the line width) few nucleation sites for the C-54 phase exist. In fact, the narrower the line, the fewer the number of nuclei; c) Thus, the conversion to C-54-phase from C-49-phase is incomplete, and the final $TiSi_2$ sheet-resistance is a combination of low and high resistance phases, and is therefore higher (and variable).

The reason why there is difficulty in completely transforming C-49-phase into C-54-phase is based on the fact that on narrow polysilicon-lines the small grains of the C-49-phase are comparable to the width of the polysilicon lines.[16] As a

result, there is a low-density of grain boundaries - in particular triple-grain boundaries - to function as nucleation sites for C-54-phase-$TiSi_2$ grain growth. (It has been shown by that the nucleation of C-54-phase-$TiSi_2$ preferentially occurs at triple-grain junctions [or grain-boundaries].)[17] In addition, on narrow-lines, the grain growth is limited by the linewidth, and only a fraction of the line is covered by the C-54-phase grown from the few nuclei available. On the other hand, on wide-lines, there are an adequate number of such nucleation sites per unit length of the line to allow C-54 grains to nucleate and grow to a size that transforms the entire $TiSi_2$ along the line to the C-54-phase (see Fig. 13-7). One way to solve the problem is to heat the $TiSi_2$ lines to > 750°C, to accelerate the nucleation of the C-54-phase. Unfortunately, thin films of $TiSi_2$ agglomerate at such temperatures, rendering this solution unusable. The linewidth and C-49 grain-size-dependence of the C-49-to-C-54 transformation-kinetics are in excellent agreement with a nucleation-density-model based on these assumptions.[18]

As shown in Fig. 13-8, as the polysilicon linewidth and $TiSi_2$ thickness decreases, it takes higher temperatures to transform the C-49-phase entirely into C-54-phase. Unfortunately, as the thickness of the film is decreased, the temperature at which it begins to agglomerate also decreases. As a result, the breadth of the process window (here expressed as the allowed variation in the time at which the 2nd RTP step is performed at the appropriate temperature) is also reduced. Without modifying the process conditions in some way, at some point it becomes no longer possible to produce $TiSi_2$ films with the sheet-resistance values required to meet the process specifications.

13.5.3 Proposed Process Modifications to Mitigate the $TiSi_2$ Narrow-Line Effect

Several techniques have been pursued as solutions to the narrow-line effect. One method is to replace $TiSi_2$ with $CoSi_2$, as $CoSi_2$ does not exhibit this effect. (Salicide formation with $CoSi_2$ is covered in Sect. 13.6.) However, to allow $TiSi_2$ to continue to be used on polysilicon lines below 0.25-μm, it has been proposed that the grain-size of the C-49-phase in the $TiSi_2$ be decreased. This will increase the density of C-54 nucleation sites in the film, allowing the C-49-phase to be transformed into C-54-phase even on narrower lines (Fig. 13-7a). The method that has been studied to decrease the grain sizes of the $TiSi_2$ C-49-phase is to perform a *pre-amorphization implant* (or *PAI*). That is, a shallow implant is made into the Si-substrate and poly to amorphize the top 30-nm of the silicon before the Ti is deposited (using either Ge, Si, or Xe as the implant species, Fig. 13-9).

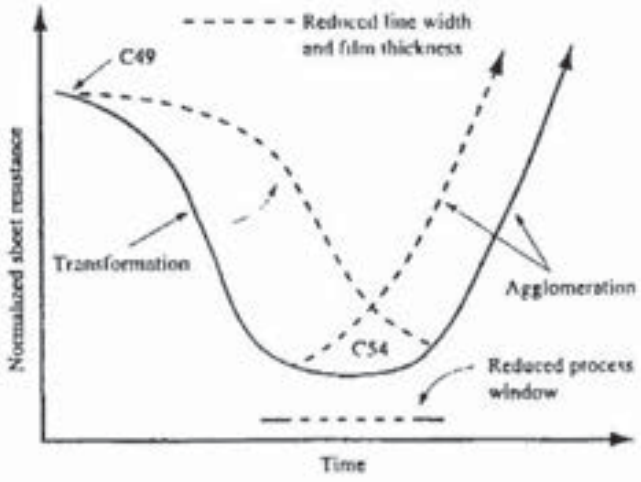


Fig. 13-8 For a thinner polysilicon line, the time required to transform the C-49-phase to the C-54-phase increases (i.e., at some chosen, constant temperature), but the time it takes for the $TiSi_2$ film on the line to undergo agglomeration decreases as the line width is decreased. Hence, the process-window for forming C-54-phase on thinner lines eventually closes.

This approach has been shown to induce a smaller C-49 grain size (~0.07-μm) compared to the 0.2-μm grain size found in the process without the PAI. This enables transformation of silicide films to the low-sheet-resistance-phase on structures down to 0.1-μm (Fig. 13-10).[19] It should be noted that such PAI steps can also cause increased junction leakage, due to damage by the implant of the Si substrate. Thus, the integration of PAI into a fabrication sequence involves careful design of the process conditions.[20]

An alternative approach to PAI which may allow the continued use of $TiSi_2$ on narrow polysilicon-lines is the implantation of refractory metal impurities into the Si surface before depositing the Ti, close to the Ti/Si interface.[21] Reports of using Mo implants for this purpose have been published.[22] This work indicates that the reaction between Ti and Mo-doped-polysilicon bypasses the C-49-phase, growing C-54-phase directly at low temperatures. On the contrary, however, the reaction between Ti and the Mo-doped single-crystal (100) silicon proceeds by the conventional sequence of C-49 growth first. To achieve a single-step growth of C-54 phase $TiSi_2$ on single-crystal-Si, it was found necessary to combine the Mo-implant into the Si-substrate with a PAI. Such a one-step Mo + PAI achieves low-sheet resistance on poly lines as small as 0.6-μm wide. Use of a single silicide-anneal step needed for the Mo + PAI process resulted in higher

concentrations of B at the silicide *p*MOS interface, and nearly an order of magnitude improvement in the *p*MOS silicide-to-source/drain contact-resistance.

13.5.4 Junction Dopant Redistribution and Silicon Consumption During $TiSi_2$ Formation

Note also that $TiSi_2$ in contact with Si was originally thought to be capable of being subjected to temperatures up to ~900°C. ($TiSi_2$ agglomerates at higher temperatures.) It was observed, however, that if the temperature exceeds 800°C, the value of ρ_c to p^+ Si becomes too large.[23] However, the ρ_c of $TiSi_2$-to-n^+ Si can be less than 10^{-6}-Ω-cm^2 for anneal temperatures of 800°C. Even at anneal temperatures of 800°C, the value of ρ_c of $TiSi_2$-to-p^+-Si is 10^{-5}-Ω-cm^2, while at temperatures of 900°C, it increases to 10^{-3}-Ω-cm^2.

The cause of this problem is the rapid diffusion of the boron from the PMOS source/drain regions into the overlying silicide. Since silicon is the moving species in the course of the silicide reaction, during the formation of silicide, a large number of vacancies are generated in the Si substrate near the Si surface. This enhances the diffusion of dopants in the junction. This enhanced

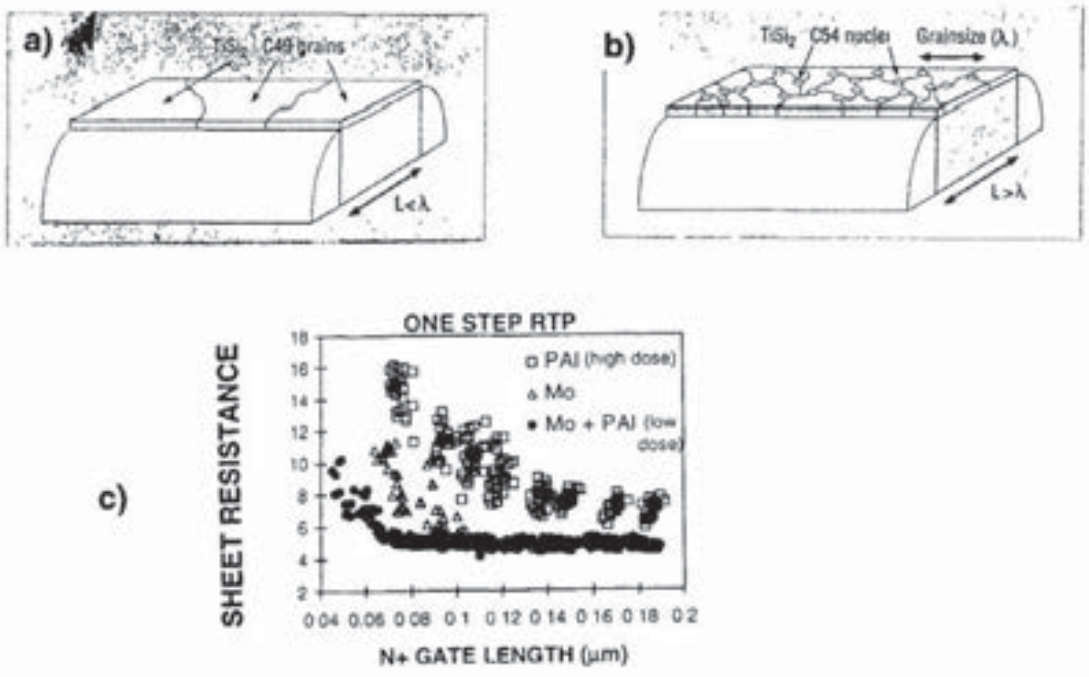


Fig. 13-9 The $TiSi_2$ C-49 (high resistivity) to C-54 (low resistivity) transformation-rate depends on the density of nucleation sites: a) Without a pre-amorphization implant (PAI) there are few nucleation sites on narrow polysilicon-lines; b) PAI reduces the C-49 grain size, increasing the number of triple-grain-boundaries and C-54 sites; c) By using PAI + MO the sheet–resistance of $TiSi_2$ lines remains low for narrower lines than when only PAI is used.

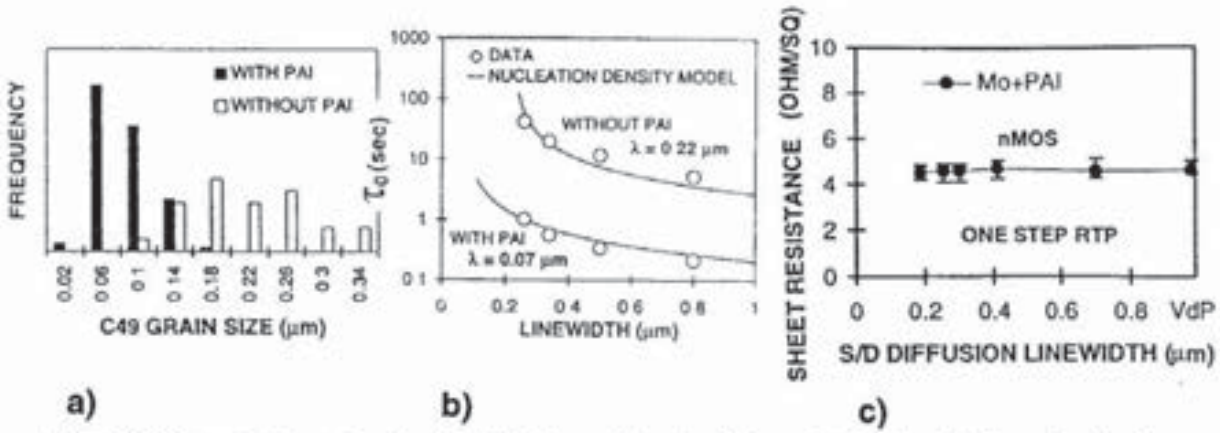


Fig. 13-10 a) C-49 grain size distribution with and without PAI. The C-49-grain size is significantly reduced by PAI, making it possible to achieve low sheet-resistance on 0.1-μm gates. b) Results and predictions of the nucleation-density-model for n^+ gates with- and without-PAI. The difference in C-49 grain size (λ) accounts for the large increase in transformation times. Transformation times τ_0 increase dramatically as the linewidth approaches λ, making the process without PAI incompatible with linewidths below 0.2-μm: c) Mo + PAI allows low-sheet resistance (4-Ω-sq) to be obtained for 0.2-μm-wide lines.

dopant diffusion has three deleterious effects. First, some of the dopants in the silicon diffuse deeper into the silicon, creating deeper junctions than would occur under normal diffusion. Second, the silicon containing the highest doping-concentration is consumed by the silicide. Third, the enhanced-diffusion causes more dopants from the silicon to diffuse into the silicide, further depleting the junction of dopants at the silicide/junction interface (and consequently, causing higher contact-resistance between the junction and silicide layer). Temperatures in excess of 800°C must thus be avoided following $TiSi_2$-formation in CMOS technologies. This can represent a problem if doped-glass-reflow is used in a CMOS process.[24] Reports have been published, however, that describe the use of nitrogen implants into the Si surface prior to the Ti-deposition to suppress the diffusion of B and As during the silicide formation.

Note, that one might think that a solution to the dopant-depletion problem would be to implant dopants into the *silicide* after it was formed, and then drive-out the dopants from the silicide to form the junction. Unfortunately, this solution is not effective for $TiSi_2$ because the dopants implanted into the $TiSi_2$ to produce the source/drain regions (e.g., B and As) form stable compounds such as TiB and TiAs.[25] Thus, the B and As species are tightly bound to the Ti and do not easily diffuse out from the silicide. This problem also contributes to the dopant depletion in the Si in contact with $TiSi_2$. That is, the B and As that diffuse into the silicide, do not back-diffuse into the Si, because once inside the silicide, the

dopants bind tightly with the Ti. An extensive study of the effects of $TiSi_2$-formation on dopant redistribution is presented in Refs. 26 and 27.

Another serious issue involving the use of silicides on shallow source/drain-regions is the junction-leakage caused by the silicide. To form the silicide, a substantial portion of the silicon from the junction is consumed. For example, to form 50-nm of $TiSi_2$ requires a 20-nm-thick layer of Ti and the consumption of approximately 50-nm of silicon from the junction region. For a shallow-junction of about 150-nm deep (0.15-μm), this represents one-third of the junction-depth (assuming the silicide is uniformly produced). However, the silicide film tends to be not totally uniform in depth, and even a minor non-uniformity (say 20-nm), means that in some places 70-nm out of the 150-nm-deep junction are consumed - leaving a very shallow effective junction. As the diffused junction-depth beneath the silicide will continue to decrease with technology scaling, the silicide must also decrease in thickness so that the amount of silicon consumed is not greater than the junction-depth. In fact, it is typically desirable to leave about 50-nm between the bottom of the silicide and the junction, or else leakage current through the junction can become too large.

13.6 SALICIDES FORMED WITH COBALT SILICIDE ($CoSi_2$)

Although $TiSi_2$ has been the most widely implemented material for the salicide process, $CoSi_2$ has also received significant attention as an alternative candidate material. In fact, many companies are switching to Co at the 0.25/0.18-μm nodes. The switch to $CoSi_2$ relieves some of the constraints that exist for many established $TiSi_2$ processes used in production, widening a process window with good margins. On the other hand, it shifts the concerns from issues of lateral scaling (i.e., the narrow-line effect that afflicts $TiSi_2$), to issues related to the vertical scaling of junctions (such as diode leakage in junctions with $CoSi_2$).

13.6.1 Advantages and Drawbacks of Cobalt Salicides

$CoSi_2$ exhibits the following attractive characteristics for application in the salicide process:

1. It offers the benefits of low resistivity (16-18-$\mu\Omega$-cm) and high temperature stability. For example, 300-nm-thick $CoSi_2$ films exhibit a ρ_{sh} of 0.5-Ω/sq and remain stable when in contact with Si up to 900°C.[28,29] This demonstrates the compatibility of $CoSi_2$ with a 900°C glass-reflow step.

2. Unlike a conventional $TiSi_2$ process, nucleation does not limit the formation or transformation of Co-silicide phases, even for polysilicon linewidths smaller than 0.1-μm (see Fig. 13-9). This is due to the fact that

$CoSi_2$ has grain sizes ten times smaller than those of $TiSi_2$ (i.e., ~0.05-μm versus ~0.5-μm).

3. $CoSi_2$ is normally formed with a two-step RTP process:

$$Co + Si \dashrightarrow CoSi \quad (450°C) \qquad \textbf{(13.15)}$$

$$CoSi + Si \dashrightarrow CoSi_2 \quad (700°C) \qquad \textbf{(13.16)}$$

$CoSi_2$ can also be formed by RTP directly in a single annealing step. A wet-etch is then applied directly to the $CoSi_2$. Note that while Co is the species that moves during several stages of the Co-Si reaction, Si is still the main moving species during the formation of CoSi. Thus, bridging across the sidewall-spacer or isolation can still occur. The use of a Ti or TiN capping layer effectively reduces lateral growth and eliminates bridging for $CoSi_2$. In contrast with Ti, Co does not react with oxide or nitride films, alleviating some of the issues found for Ti salicide processing. An RTP system optimized for the formation of $CoSi_2$ films is described in Ref. 30.

4. Under optimized process conditions, the interface of the silicide and silicon can be smooth, and low-resistivity contacts to the silicon can be formed (as low as 15-$\mu\Omega$-cm^2 to both n^+ and p^+Si). The conditions that give rise to such smooth $CoSi_2$/Si interfaces will be described later.

5. $CoSi_2$ processes (unlike those for conventional $TiSi_2$) generally do not exhibit the strong dopant interaction and redistribution issues that can lead to interfacial dopant-depletion and higher silicide-to-source/drain contact-resistance. Instead, low-resistance-contacts to shallow As and B junctions can be successfully fabricated, and the dopant profiles in such junctions appear to be essentially unchanged after the silicide has been formed.

6. A selective chemical etch exists that makes it possible to remove the Co without affecting the $CoSi_2$ (e.g., with either a single-etch step using a hot H_2SO_4:H_2O_2 solution,[31] or with a two-step wet-etch process, consisting of an NH_4OH:H_2O_2:H_2O-etch followed by an HCl:H_2O_2:H_2O-etch.[32])

7. $CoSi_2$ is far less susceptible to removal by plasma-etching than $TiSi_2$. Hence, there is negligible silicide loss during the overetch time that may be needed when dry-etching the contact-openings through the glass-overlayer.

8. No competing reaction to the $CoSi_2$ formation reaction occurs, whereas when $TiSi_2$ is formed in nitrogen, TiN is simultaneously formed. In addition, lower shear forces (important to device integrity) are present in $CoSi_2$ than in $TiSi_2$ of the same thickness.[33]

9. It may be possible to form shallow junctions by implanting dopants into $CoSi_2$ after it has been formed, and then drive-out the dopants to form the shallow-junction. Unlike for the case of $TiSi_2$, good junctions have been reported by using this technique. This process is described in more detail in Sect. 13.9.

The use of $CoSi_2$, however, also has some drawbacks:

1. Cobalt silicide is sensitive to the surface condition of the Si regions prior to the Co deposition. That is, the presence of a native oxide or other surface contaminants can impact the silicide-formation-process and the resultant $CoSi_2$ film-characteristics. For example, both the $CoSi_2$ sheet-resistance and the leakage of the *pn*-junctions existing beneath the $CoSi_2$ salicide-film are influenced by previous processing steps that alter Si surface conditions, such as the spacer-etch, the screen-oxide-etch, and the subsequent surface-clean-and-preparation steps.

Unlike Ti, Co cannot reduce native oxides on Si, and therefore the formation of $CoSi_2$ is more sensitive than Ti to residual contaminants or to a native oxide on the Si-surface. However, several methods have been devised to reduce the sensitivity of $CoSi_2$ to surface conditions. In one of them, advantage is taken of the reactivity of Ti (and its ability to reduce native oxides), by incorporating Ti at the deposition stage. That is, by applying a capping layer of Ti onto the Co film (i.e., deposited *in situ* prior to the silicide formation), the sensitivity of the $CoSi_2$ formation process to interfacial contaminants such as oxygen is significantly reduced.[34] This allows the formation of $CoSi_2$ through native oxide layers. A similar effect was reported for a Co-Ti (10-20 atom.%) alloy deposition.[35]

Further discussion of the effect of surface-contaminants present on the Si surface prior to Co-deposition and silicide-formation on the leakage-current of $CoSi_2$ salicided *pn*-junctions is given in Sect. 13.7.3.

2. The silicon consumption during the $CoSi_2$ formation is about 25% higher than for the formation of $TiSi_2$.

3. Gate-oxide integrity is a concern when Co salicides are used because Co contamination will degrade GOI, and both the diffusivity and solubility of cobalt are high (significantly higher than those of Ti) at high temperatures. Degradation of film and interface morphology at high temperatures, as well as process conditions that induce rough, non-uniform silicide-Si interfaces, can effect GOI. These issues could limit the compatibility of Co salicide

processes with aggressive scaling of polysilicon gate-thickness. GOI degradation can be avoided by a careful control of the thermal history.

4. $CoSi_2$ is sensitive to oxidation and contamination from the ambient. Measures must be taken to prevent this from occurring, and these are discussed in Sect. 13.7.4.

13.6.2 Details of the Cobalt Salicide Fabrication Process

The details of the process sequence for forming the $CoSi_2$ polycide and the self-aligned contacts to Si are as follows:

1. Since the formation of $CoSi_2$ is sensitive to the Si surface conditions, procedures must used to ensure that a clean Si surface exists prior to the Co deposition. One technique used to clean the silicon and polysilicon surfaces before the Co sputtering step is to dip the wafers in a dilute (100:1) solution of H_2O:HF just prior to loading them into the sputtering chamber. This is often followed by a short *in situ* rf-backsputtering procedure to remove ~10-nm of the silicon surface.[28,29] A prolonged rf-backsputtering process has the risk of allowing contaminant material sputtered from other parts of the wafer to redeposit onto the Si surfaces that are being cleaned. Nevertheless, the cleaning procedure is important, because cobalt is not able to reduce SiO_2 and because the presence of a native oxide can inhibit the silicidation reaction.

2. The cobalt is sputter-deposited to the desired thickness. The thickness is selected based on how much of the silicon in the source/drain regions can be consumed, as well as what value of silicide sheet-resistance is needed. For example, 35.9-nm of Si is consumed by 10-nm of Co to form 35.9-nm of $CoSi_2$. To form a $CoSi_2$ layer with ρ_{sh} = 1.5-Ω/sq, a 30-nm-thick layer of Co must be deposited. As noted earlier, the Si consumption during $CoSi_2$ formation is about 25% greater than that during $TiSi_2$ formation, constituting a disadvantage of $CoSi_2$ compared to $TiSi_2$.

3. The $CoSi_2$ is formed by reacting the Co with the Si and polysilicon. Although the reaction can be accomplished through either a furnace anneal or RTP, the latter has emerged as the more attractive approach.[36] Either a two-step anneal, or a single-step anneal can be employed, but the former is more commonly employed. An 80-nm-thick Co film will be completely converted to $CoSi_2$ at 700°C for 30 seconds. The gas ambient must be kept free of oxygen to ensure that oxide-free silicide films are formed.

4. The unreacted Co is etched away using the etchant mentioned earlier.

5. Finally, the wafer is covered with a CVD glass layer. Note that when Al is contacted to the $CoSi_2$, the maximum anneal temperature must thereafter be limited to 400°C (since Al will react with $CoSi_2$ at temperatures in excess of 400°C). To allow a higher anneal temperature to be used, a diffusion barrier layer (such as Ti:W), must be used between the $CoSi_2$ and the Al layers.[37]

13.6.3 Diode Leakage Issues Involving Cobalt Salicides

It was noted earlier that one of the major difficulties of using $CoSi_2$ is the issue of increased leakage of shallow source/drain-junctions when $CoSi_2$ is the salicide. A primary cause of such leakage is the formation of a rough interface between the $CoSi_2$ and the Si-substrate. Such interfacial roughness can arise if a native oxide or other contaminants are on the Si surface prior to the Co deposition. In that case, the silicidation reaction can be less uniform, proceeding faster in clean spots where the Co can diffuse faster, resulting in a rougher interface. As a result, an effective cleaning of the Si surface before sputtering the Co is mandatory for formation of smooth interfaces (and thus also low-leakage). As mentioned earlier, a dip in dilute-HF just before loading the wafer into the sputtering tool is thus normally performed. In addition, a brief *in situ* sputter-cleaning pre-Co-deposition-step is also used to remove contaminants from the Si surface.

The silicide interface is also rough if the silicidation is carried out at low temperatures. This, however, is the usual way of forming $CoSi_2$ films (i.e., a 1st RTP step is carried out at 450°C). At this temperature the formation of long 'needle-like' protrusions, or 'spikes' of some form of cobalt-silicide in the Si-substrate has been observed (with a length that is possibly great enough to penetrate shallow-junctions).[38] In contrast, high-temperature-processing promotes the formation of a smooth interface. At such higher temperatures the spikes formed at the lower temperatures are found to dissolve, and the $CoSi_2$/Si-interface becomes smoother. Such smoothing occurs because of the enhancement of diffusion kinetics and the increased solubility of Co in Si. Therefore, a very effective way to reduce diode leakage is the use of an optimized higher-temperature RTP, in the range of 800-850°C (Fig. 13-11). Because the growth of the low-resistivity Co-salicide-phase is easily achieved in deep-submicron structures, the second RTP step of $CoSi_2$ formation can be optimized in terms of its effects on diode-leakage, agglomeration, and source/drain-resistance, rather

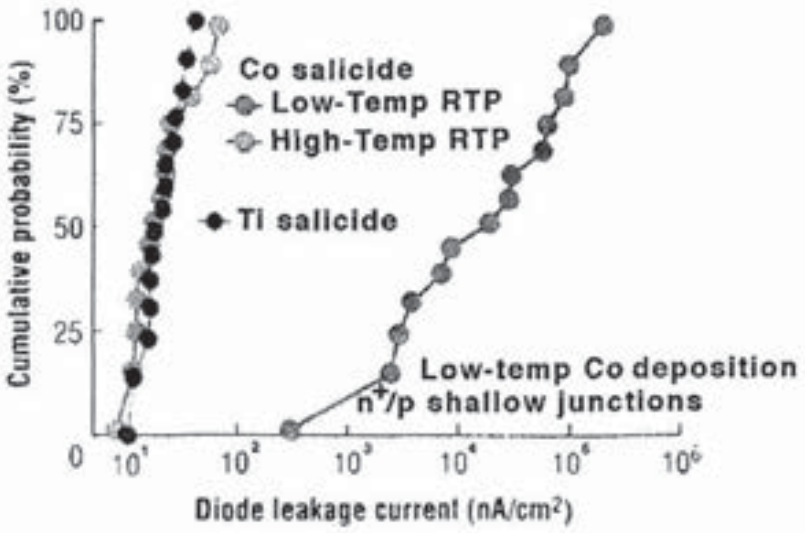


Fig. 13-11 High-temperature RTP is effective in improving diode-leakage in MOSFETs fabricated with $CoSi_2$ by eliminating high-leakage outliers.

than simply considering the growth of the low-resistivity phase (as must be done when using Ti- instead of Co-salicides). An upper-limit on the temperature that can be used during the second RTP-step is set by silicide agglomeration and the consequent degradation of the source/drain series-resistance. (An alternative technique for the reduction of diode-leakage is high-temperature cobalt deposition [i.e., the Co is sputter-deposited at 450°C].) In this process, the initial stages of the Co/Si reaction take place during the high-temperature cobalt-deposition-process.[39,40] $CoSi_2$-films deposited with such a high-temperature sputter-deposition-process were also reported to exhibit increased thermal stability, as compared to those sputter-deposited at room temperature.[41]

13.6.4 Sensitivity of Cobalt Salicide Films to Degradation by Oxygen Ambients: TiN-Capping of Cobalt-Salicide Films

$CoSi_2$ is also sensitive to oxidation and contamination due to the presence of oxygen or other contaminants in the ambient during the silicidation thermal steps. Such oxidation can cause the phase-transformation temperature from Co to $CoSi_2$ to change, and to produce films that exhibit higher sheet-resistance. A very effective approach to control ambient-contamination is the use of a capping layer, consisting of either Ti,[42] or TiN,[43] and deposited *in-situ* immediately following cobalt-deposition (Fig. 13-12). $CoSi_2$ processes with *in-situ* capping-layers have been shown to produce linewidth-independent, low-resistance films well below 0.1-μm dimensions. As also noted earlier, the use of a Ti or TiN capping layer

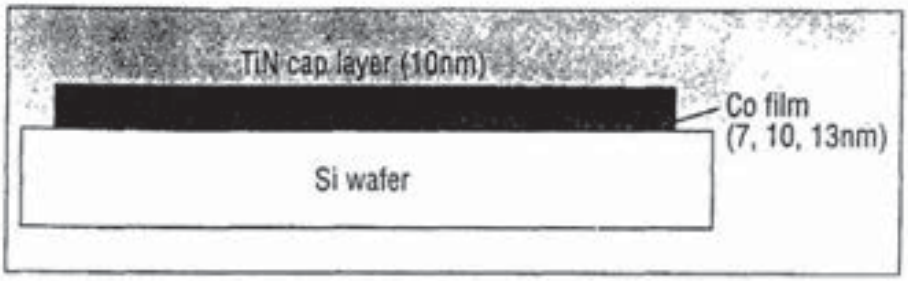


Fig. 13-12 Schematic cross-section of a $CoSi_2$ film with a TiN capping layer on a Si wafer.

also effectively reduces lateral growth of the $CoSi_2$ and thus eliminates bridging when carrying out the $CoSi_2$ salicide process.

For shallow junctions, the thickness of $CoSi_2$ films must also be scaled, to prevent excessive silicon-consumption of the substrate in which the shallow-junction is formed. The thermal-stability of $CoSi_2$ films decreases with thickness, and this becomes a major concern.[44] Deposition of a Ti capping-layer on the Co was found to increase the thermal stability of $CoSi_2$ when a Co-Ti-phase was formed on the surface.[45] According to another report, the Ti capping-layer also produced superior films than those having a TiN capping-layer. That is, the Ti-capped films formed on 0.15-μm-wide poly lines showed uniformly-low sheet-resistances and a thermal stability up to 900°C for 30 min. The TiN-capped films on the same lines showed a wide distribution of sheet-resistance values after annealing for 30 min at 800°C. The higher thermal-stability of the Ti-capped $CoSi_2$-layers was attributed to the fact that a large amount of Ti-atoms diffused into the $CoSi_2$ film along the grain boundaries, and that the presence of these Ti atoms slowed the agglomeration process.[46]

13.7 EPITAXIAL $CoSi_2$

The reaction mechanism that forms $CoSi_2$ layers on silicon-substrates that has been discussed to this point produces polycrystalline films of $CoSi_2$. However, it is also possible to nucleate and grow $CoSi_2$ epitaxially on (100)-Si. A potential advantage of epitaxial cobalt-silicide films is the achievement of smooth silicide-silicon interfaces that could result in lower diode-leakage-currents on ultra-shallow junctions. Thin epitaxial-silicide layers provide a shallow-junction contacting-technology that may open a path for further scaling of junction-depths.

Techniques have been developed for achieving epitaxial-$CoSi_2$-growth on (100)-Si that are compatible with salicide processing for IC manufacturing (Fig. 13-13).[47] One such approach consists of depositing a thin Ti-layer (~5-nm thick) before deposition of the Co film.[48] A different technique involves the growth of an ultrathin oxide layer (~0.5-nm-thick, which can be grown in a peroxide bath), and produces a similar effect.[49] In both cases, the mechanism appears to be solid-state epitaxy through a transport medium, with the main role of the interlayer being that of a diffusion barrier retarding the Co/Si reaction.

Several issues involved with the Ti-interlayer process (such as void formation at the silicide edges in patterned structures), significantly limits its applicability to salicide processing. The ability to control the characteristics of an ultrathin oxide layer may also limit the application of this process in a production environment. Epitaxial growth was also reported for a process that used a Co-Ti (10-20 atom%) alloy,[50] as well one that used a high-temperature sputter-deposition technique to deposit the Co-film.[51] An article comparing these techniques, and the influence of a BF_2 implant on the roughness of the $CoSi_2$/Si interfaces produced by each of these techniques, is listed in Ref. 52.

13.8 $CoSi_2$ AS A SORCE/DRAIN DIFFUSION-SOURCE

One technique for fabricating shallow-junctions that involves salicides is to implant the junction dopant after the salicide has been formed. To do this, the dopant would have to be implanted either: **1**) entirely into the silicide; or **2**) through the silicide. It is difficult to form very shallow junctions (< 0.15-μm) by

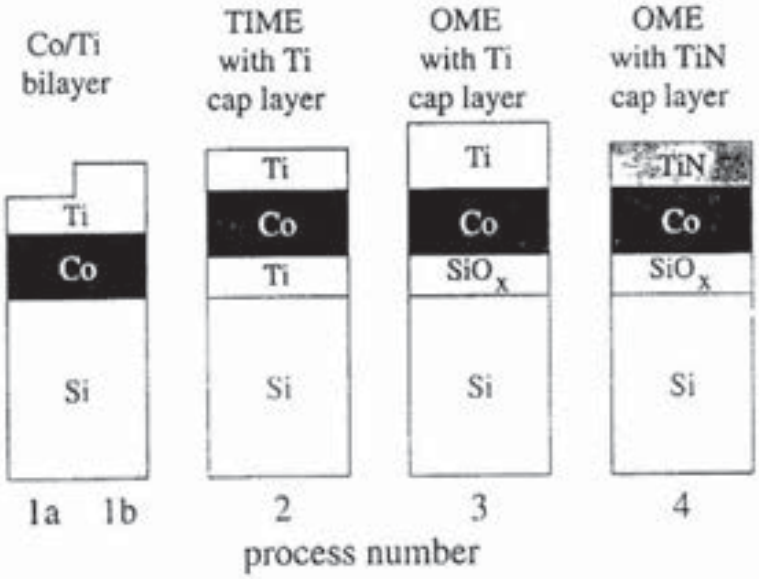


Fig. 13-13 Different silicide techniques; thickness are not scaled.

implanting through the silicide, since the peak of the implant has to be in the silicide and only the tail portion can be used to form the *pn*-junction in the Si. The junction-depth then becomes a sharp function of the silicide thickness, which is difficult to control precisely.

The second approach (implanting the junction-dopant into the silicide and driving it out) has been shown to successfully form good shallow-junctions with $CoSi_2$ but not with $TiSi_2$. This method has been dubbed the *Silicide as a Diffusion Source* (SADS) process.[53] As noted earlier, dopants such as B and As have high solubilities in Ti and also form stable compounds with Ti (i.e., TiB_2 and TiAs). As a result, if these dopants are implanted into Ti, it is very hard to get them to out-diffuse to form a junction.[54]

Forming junctions as shallow as 0.12-μm by implanting dopants into $CoSi_2$ were first reported in 1987.[55] In the SADS process that involves $CoSi_2$, the conventional S/D-implant-and-activation steps are skipped while a Co-salicide is formed, followed by S/D-implantation into the silicide-layer. A relatively low-temperature anneal is then performed to drive-out the dopants from the silicide film into the underlying silicon to form shallow junctions. Since B and As dopants diffuse rapidly in a $CoSi_2$ film, an ideal junction is expected to follow the silicide/Si interface-contour (i.e., conforming to any existing interface roughness), resulting in low-diode-leakage on the silicided ultra-shallow junctions. Two other benefits of the SADS process are that low source/drain series- resistance is achieved, due to the high dopant-concentrations at the silicide-Si interface, and that SADS is compatible with thin sidewall-spacers.

An example of an n^+/p-junction formed with the SADS process is shown in Fig. 13-14.[56] The p-doping concentration in the Si was $1x10^{18}$ cm^{-3}. It can be seen that a junction-depth of 40-nm beneath the $CoSi_2$/Si interface is obtained. Arsenic was implanted into the $CoSi_2$, followed by a low-temperature RTP-step. (Note that in Fig. 13-14, a baseline-temperature T_B is invoked (i.e., used in the conventional $CoSi_2$ process). Thus, the low-temperature-range referred to here is [T_B - 250°C] to [T_B - 50°C].) Junction-depths were observed to increase as the RTP-temperature was increased in this low-temperature-range. However, as the RTP-temperatures approached the baseline temperature ([T_B - 50°C] in Fig. 13-14), the SADS junction-depth became shallower (due to As evaporation from the $CoSi_2$ film). The minimum possible junction-depths in SADS process will probably thus be limited by the maximum temperature that can be endured by the thin $CoSi_2$ layers before they become thermally unstable (e.g., before they agglomerate). A capping-layer would aid in preventing evaporation at higher temperatures, allowing the formation of deeper junctions.

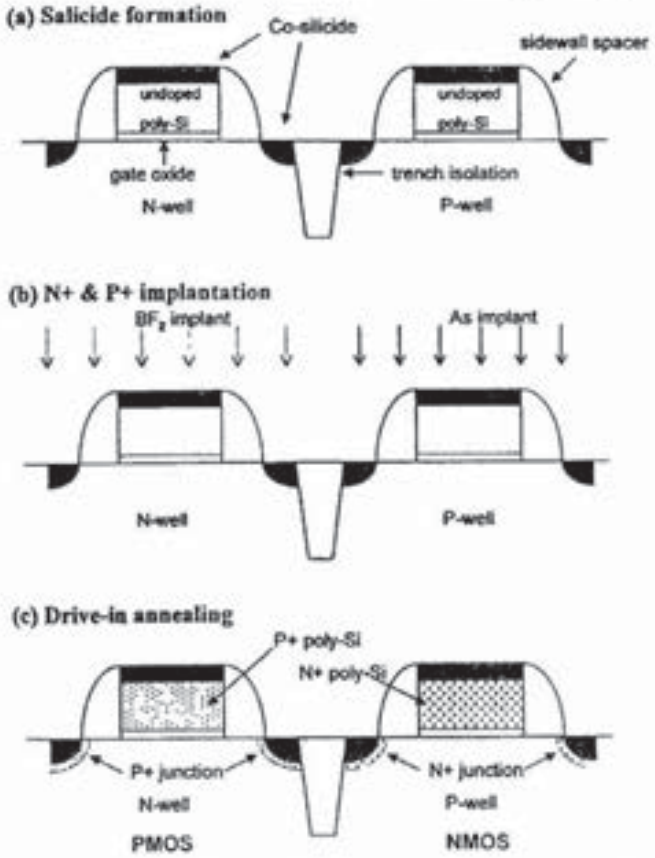


Fig. 13-14 Schematic of a process flow of 0.18-μm salicided dual-doped-gate CMOS using Co-silicide as the doping source.

Improvements in short-channel characteristics for thin (35-55-nm) spacers were verified for the SADS process in comparison with the conventional process, as better DIBL-values were observed (see Vol. 3, Chap. 5, *Silicon Processing for the VLSI Era*, for more information on DIBL).[57]

One of the main concerns associated with implementing silicided shallow junctions is diode-leakage. Good leakage characteristics, however, have been demonstrated with $CoSi_2$ as the diffusion source for both n^+/p and p^+/n junctions. However, most of the demonstrations were carried out in devices with low n-well and p-well concentrations that are not suitable for deep-submicron CMOS applications. For higher background-concentrations the depletion-region of the n^+/p-junction extends closer to the $CoSi_2/n^+$-Si interface. If the junction is too shallow, the depletion-region can extend through the whole n^+-Si-layer, and the shallow n^+/p-junction behaves as a Shannon-contact with high diode-leakage. Applicability of SADS is thus limited by the ability to obtain the appropriate junction-depths needed to achieve low diode-leakage, given the background-concentrations used in specific technologies.

A report that details the use of SADS for a full sub-micron, dual-doped-gate CMOS process was recently published.[58] As shown in Fig. 13-15, BF_2 and As were implanted separately into a Ti-capped, 35-nm-thick $CoSi_2$-layer that covered both the polysilicon and source/drain-regions of the MOSFETs. An As implant at 30-40-keV with a dose of $1x10^{16}$ cm^{-2}, was used to introduce the *n*-dopants, and a BF_2 implant at 30-keV with a dose of $5x10^{15}$ cm^{-2} was used to introduce the *p*-dopants. A single-step anneal at temperatures ranging from 800-1000°C for 30 seconds was used to drive-out the dopants to form the shallow junctions. These shallow-junctions exhibited good leakage-characteristics and low contact-resistance. The process also did not contribute to degradation of the gate-oxide, nor to B-penetration of the gate-oxide from the *p*-doped poly-gates.

13.9 NICKEL SILICIDE FILMS FOR SALICIDE APPLICATIONS

Ni silicides have also been investigated for use as salicide candidates, due to their low resistivity (~15-18-$\mu\Omega$-cm^2). A single RTP-step at temperatures in the 400-500°C range forms the NiSi film.[59] Like $CoSi_2$ the sheet resistance of NiSi films are not affected by the strong narrow-line-effects which plague $TiSi_2$ salicides. The slight increase in sheet resistance observed with decreasing linewidth has been attributed to thickening of NiSi films at gate edges. A potential advantage of NiSi over $CoSi_2$ is that silicon consumption is only 80% of that for $CoSi_2$.

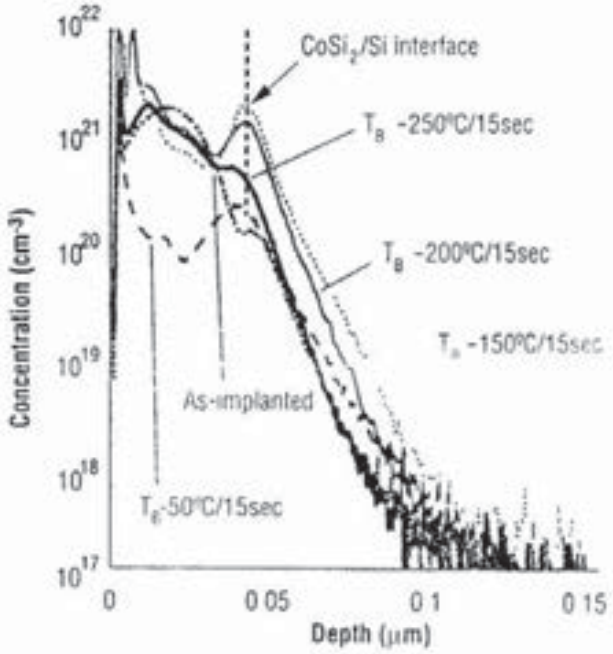


Fig. 13-15 Shallow junctions obtained using $CoSi_2$ as a diffusion source (SADS) process.

Several challenges, however, exist for implementing NiSi. First, NiSi (like $CoSi_2$) is sensitive to contamination on the Si surface, and this can lead to rough interfaces and higher-diode-leakage on shallow junctions. TiN capping-layers and nitrogen-doping can prevent oxidation from the ambient during silicide formation, and this reduces diode-leakage. High diode-leakage in the source/drain-edges of STI edge-intensive-structures has been reported, correlated to film-thickening and non-uniformity at these edges.[60] Furthermore, transformation into the high-resistivity $NiSi_2$-phase needs to be prevented during later processing steps. That is, NiSi films have poor thermal-stability, with an increase in sheet resistance observed at temperatures as low as 600°C.

REFERENCES

1. S.P. Murarka, *Sillicides for VLSI Applications,* Academic Pr., Orlando, FL 1983.

2. M.-A Nicolet and S. Lau, "Silicides," in N.G. Einspruch & G.B. Larabee,eds., *VLSI Microstructure Science*, Vol. 6, Academic Press, N.Y. 1983, p. 329.

3. K.C. Saraswat *et al.*, "Properties of Low Pressure CVD Tungsten Silicide for MOS VLSI Interconnection," *IEEE Trans. on Electron Dev.*, **ED-30**, No. 11, November 1983, p. 1497.

4. M. Venkatesan and I. Beinglass, "Single-Wafer Deposition of Polycrystalline Silicon," *Solid State Technology*, March 1993, p. 49.

5. T. Hara et al., "Composition of Tungsten Silicide Films Deposited by Dichlorosilane Reduction of WF_6," *J. Electrochem. Soc.* , **137**, p. 2955 (1990).

6. S.G. Telford *et al.*, "CVD WSi_x Films Using Dichlorosilane in a Single Wafer Reactor," *J. Electrochem. Soc.* , **140**, 3689 (1993).

7. C. Y. Ting, "Silicide for Contacts and Interconnects," *Tech. Dig. IEDM*, 1984, p. 110.

8. R. A. Haken, "Applications of the Self-Aligned $TiSi_2$ Process to VLSI MOS CMOS Technologies," Workshop on Refractory Metals and Silicides for VLSI-IIII, San Juan Batista, CA, May, 1985.

9. R. Liu, D. S. Williams, and W. T. Lynch, "Mechanisms for Process-Induced Leakage in Shallow Silicided Junctions," *Tech. Dig. IEDM*, 1986, p. 58.

10. S.-T. Chang and K. Y. Chiu, *IEEE Electron Dev. Letts.*, May, 1986, p. 244.

11. *ibid. Ref. 7.*

12. C. Y. Ting *et al.*, "Interaction Between Ti and SiO_2," *J. Electrochem. Soc.* **131**, p. 2934, 1984.

13. Y.-B Kim *et al.*, "Characterization of the Post Dry-Etch Cleaning of Silicon Ti-Self-Aligned Silicide Technology," *J. Electrochem. Soc.* April 1999, p. 1549.

14. *ibid. Ref. 8,*

15. R. M. Vadjikar and R. P. Roberge, *Ext. Abs. Electrochem. Soc. Meeting*, Fall 1988, Abs. No. 467, p. 681, 1988.

16. J.B. Lasky *et al.,* "Comparison of Transformation to Low-Resistance Phase and Agglomeration of $TiSi_2$ & $CoSi_2$," *IEEE Trans. Electron Dev.,* 1991, p. 262.

17. Z. Ma *J. Appl. Phys.* 77, 4384 (1995).

18. J.A. Kittl *et al.,* "Novel One-Step RTP Salicide Process with Low Sheet Resistance 0.06 μm Gates and High Drive Current," *Tech. Dig. IEDM,* 1997, p. 111.

19. J.A. Kittl *et al.,* "A Ti-Salicide Process for 0.10 μm Gate Length CMOS Technology," *Symp. VLSI Tech* 1996, p. 14..

20. J. Kittl *et al.,* "Salicides for 0.10 μm Gate Lengths: A Comparative Study of One-Step RTP Ti with Mo Doping,, Ti with Pre-Amorphization, and Co Processes," *Symp. VLSI Tech.* 1997, p. 103.

21. R.W. Mann *et al., Appl. Phys. Letts.* 67, p. 975 3725, 1996.

22. J. Kittl, *Appl. Phys. Lett.,* 73, p. 900, 1998.

23. D. B. Scott *et al.,* "Titanium Disilicide Contact Resistivity and Its Impact on 1-μm CMOS Circuit Performance," *IEEE Trans. Electron Dev.*, March 1987, p. 562.

24.R. K. Shukla and J. S. Multani, "Thermal Stabillity of Titanium Silicide Thin Films," *Proceedings of the 4th Intl VMIC Conf.,* Santa Clara, CA, 1987, p. 471.

25. V. Probst, *J. Appl. Phys.* 70, 693, p. 1991.

26.C. M. Osburn *et al.,*"The Effects of Titanium Silicide Formation on Dopant Redistribution," *J. Electrochem. Soc.*, June 1988, p. 1490.

27. M.E. Lunnon, J. Chen, J.E. Baker, "Furnace and Rapid Thermal Annealing of p^+/n Junctions in BF_2^+-Implanted Silicon," *J. Electrochem. Soc.,* **132**, 2473 (1985).

28. L. C. Van de Hove *et al.,* " A Self-Aligned $CoSi_2$ Interconnection and Contact Technology for VLSI Applications," *IEEE Trans. Electron Dev.*, March, 1987, p. 554.

29. S. P. Murarka *et al.,* "Self-Aligned Cobalt Disilicide for Gate and Interconnection and Contacts to Shallow Junctions," *IEEE Trans. Electron Dev.*, Oct. 1987, p. 2108.

30. W.S. Yoo *et al.,* "Cobalt Silicide Processing in a Susceptor-Based LR-RTP System," *Solid State Technology,* July 1999, p. 126.

31. J. Kittl *et al.,* "Salicides and Alternative Technologies for Future ICs: Part I," *Solid State Technology* June 1999, p. 86.

32. D.K. Sohn *et al.,* "Effects of Ti-Capping on Formation and Stability of Co Silicide," *J. Electrochem. Soc.*, January 2000, p. 373

33. L. Van den Hove et al., *Electrochem. Soc. Abstracts*, Vol. 88-1, Atlanta, GA, May 15- 20, 1988, p. 312.

34. Q.F. Wang, *Dig. Symp.VLSI Tech.,* 1995, p. 17.

35. Iinuma *et al.*, "Highly Uniform Heteroepitaxy of Cobalt Silicide by Using Co-Ti Alloy for Sub-Quarter Micron Devices," *Symp. VLSI Tech.*, 1998, p. 188.

36. A. R. Sitaram and S. P. Murarka, *Ext. Abs. Electrochem. Soc. Meeting*, Fall 1988, Abs. No. 263, p. 376, 1988.

37. E.K. Broadbent *et al.*, "Application of Self-Aligned $CoSi_2$ Interconnection in sub-Micron CMOS Transistors," *Proceedings of the 5th IEEE VMIC Conf.*, Santa Clara, CA, 1988, p. 175.

38. K.-I. Goto *et al.*, "A New Leakage Mechanism of Co Salicide and Optimized Process Conditions," *IEEE Trans Electron Devices*, January 1999, p. 117.

39. Q.Z. Hong, "$CoSi_2$ With Low Diode Leakage and Low Sheet Resistance at 0.065 μm Gate Length," *Tech. Dig. IEDM* 1997, p. 107.

40. K. Inoue *et al.*, "A New Cobalt Salicide Technology for 0.15-μm CMOS Devices," *IEEE Trans. Electron Dev.*, November 1998, p. 2312.

41. H.Y. Huang, "Thermal Stability of $CoSi_2$ Thin Films on (001)Si by High Temperature Sputtering Deposition," *Mtg. ECS Spring 1999*, Abs 516.

42. *ibid. Ref.* 34.

43. T. Yamazaki *t al.*, "21 psec Switching 0.1 μm-CMOS at Room Temperature Using High Performance Co Salicide Process," *Tech Dig. IEDM*, 1933, p. 906.

44. K.Maex, *Mat. Sci. Eng. Rep.*, R11, p. 53 1993.

45. R.T. Tung and F. Schrey, *Appl. Phys. Lett.*, 67, 2164 (1995).

46. ***ibid. Ref. 32.***

47. J. Hohaus *et al.*, "Influence of BF_2-Implantation on Roughness of Cobalt Silicide/Silicon Interfaces," *Electrochem. & Solid State Letts.* March 2000, p. 147.

48. L.A Dass, *Appl. Phys. Lett.*, 58, p. 1308, 1991.

49. R.T. Tung *Appl. Phys. Lett.*, 61, 3461, (1996

50. *ibid.* Ref. 35.

51. *ibid.* Ref. 40.

52. ***ibid.*** Ref. 47.

53. J.C. Hu, *SPIE* Vol. 3506, p. 103, 1998.

54. C.Y. Lu *et al.*, "Process Limitation and Device Design Tradeoffs of Self-Aligned $TiSi_2$ Junction Formation in Submicrometer CMOS Devices," *IEEE Trans. Electron Dev.* **ED-38**, p. 246 (1991).

55. R. Liu *et al.*, "Formation of n^+/p Junctions with $CoSi_2$," in *Proceedings of the 1st International Symposium on ULSI Science and Technology*, S. Broydo and C.M Osburn, eds. Vol. 87-11, The Electrochemical Society, Pennington, NJ 1987, pp. 446.

56. J. A. Kittl *et al.*, "Salicidies and Alternative Technologies for Future ICs,": Part 2," *Solid State Technology*, August 1999, p. 55.

57. J.C. Hu, *SPIE* Vol. 3506, p. 103, 1998.

58. J.-S. Park *et al.*, "Simultaneous Formation of Shallow Junctions and Gate Doping for Dual Gate Structures Using Cobalt Silicide as a Dopant Source," *J. Electrochem. Soc.*, October 1999, p. 3856.

59. T. Ohguro *et al.*, "Analysis of Resistance Behavior in Ti- and Ni-Salicided Polysilicon Films," *IEEE Trans. Electron Dev.*, p. 2305, 1994.

60. K. Goto, *Proc. 36th Intl. Reliability Phys. Symp., IEEE* p. 363, 1998.

Chapter 14

LOW-*k* DIELECTRICS

For the purposes of IC microelectronic applications, dielectrics are categorized into two classes: **1)** *low-k dielectrics* and; **2)** *high-k dielectrics.* Those having a dielectric-constant k lower than that of thermal silicon dioxide (i.e., $k < 3.9$) are classified as *low-k dielectrics*, while those having a value higher than that of silicon nitride ($k > 7$), are classified as *high-k dielectrics*. Note that the minimum value of k is 1.0 for air. In this chapter low-k dielectrics are discussed, and in Chap. 4 high-k dielectrics are covered.

14.1 INTRODUCTION TO LOW-*k* DIELECTRICS

Low-k dielectric films are important in advanced IC applications because their use can significantly improve circuit performance characteristics, primarily with respect to interconnect delay. As detailed in Chap. 12, 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 (Fig. 14-1). The value of C, in turn, can be reduced by using dielectric materials with smaller permittivity values (i.e., *low-k dielectric materials*).[1] As semiconductor device geometries shrink below 0.25-μm, the parasitic capacitance in the intermetal-dielectric between metal lines becomes more and more important, in terms of the *RC*-delay in switching.

The primary dielectric film material for interconnects has been SiO_2. The choice of SiO_2 was based on its good dielectric and mechanical strength, as well as its ease of processing. However, SiO_2 has a dielectric-constant ranging from k = 3.9 to 4.5, depending on the method of its formation. This value is too high to be applicable in devices below about 0.18-μm. That is, as indicated in Fig. 14-1 only by using Cu as a conductor together with a low-k dielectric as the insulator, can the sufficiently-low values of *RC*-delay required for deep-submicron ICs be obtained. Thus, there has been a push to integrate new dielectric materials with k-values smaller than that of SiO_2 into IC fabrication. It should be pointed out that it has become apparent that the tacitly-hoped-for dielectric with k ~1 is not realizable, and that some of the qualities of SiO_2 will also no longer be attainable in actual low-k dielectrics. Evaluations of various dielectric materials for low-k thin-film applications will provide the empirical data needed to determine the trade-off limits of k-value versus mechanical performance.

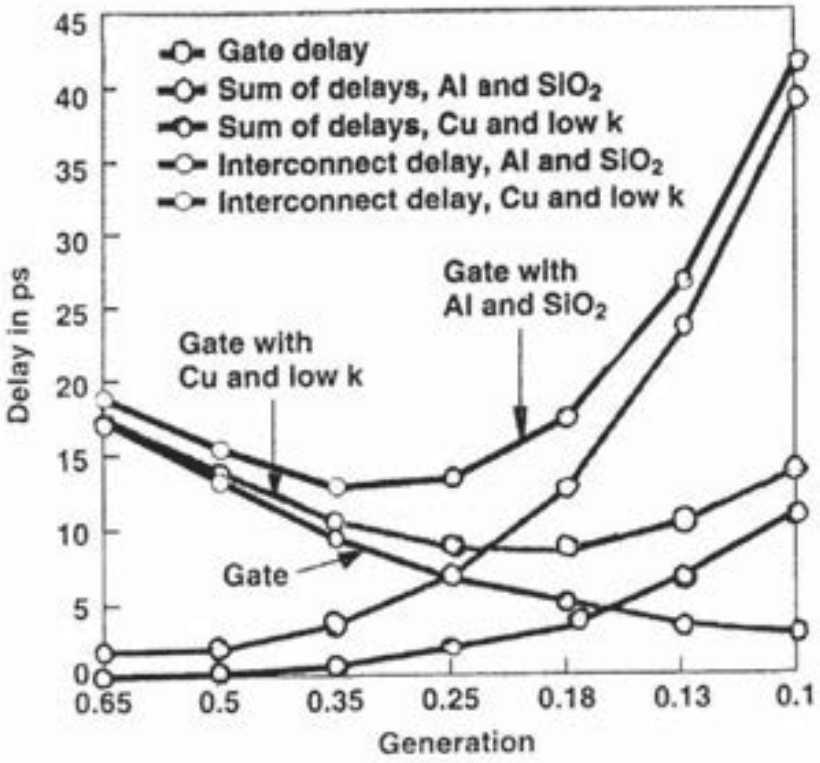


Fig. 14-1 *RC*-delays caused by interconnects are charted as a function of feature size for both Al/SiO_2 and Cu/low-*k* approaches. (© 1997 IEEE)

The conclusions from Fig. 14-1 can be justified by noting that the SiO_2 molecule is a pyramid-shaped tetrahedron (see Chap. 3), and its lattice structure thus fits together rather tightly. Although thermal SiO_2 does not possess long-range-order, it is nevertheless a relatively dense material (~2.3 g/cm^3). Since the goal is to find dielectric films with *k*-values smaller than that of oxide (~4.0), the factors that impact the dielectric constant should first be considered.

In general, no dielectric films as dense as thermal oxide can achieve *k*-values significantly smaller than 3.5. In fact, even CVD-oxides (which are also silica-based films) have about the same *k*-values, even though they are somewhat less dense. While CVD-oxides are more disordered than thermal-SiO_2 (which lowers their density), such disorder also makes them contain more silanol (OH) groups (i.e., at sites where Si-O-Si bonds are unable to form). Thus, they are prone to absorb more moisture than thermal oxides, and the absorbed water raises their *k*-value. Furthermore, materials that are less dense than thermal-oxide are also generally mechanically weaker. Nevertheless, to achieve a $k < 2$ in any dielectrics (including silica films), free space must be added, even though the strength of the material diminishes according to some power of the solids content.

The approaches used to produce dielectrics with k-values smaller than that of thermal-oxide and CVD-oxides therefore fall into three categories: **1**) reduction of the film density; **2**) reduction of ionization in the material, and **3**) reduction of the film polarization. Details of how these approaches are exploited in each low-k material-type are provided later, but a few general comments on them are given here.

First, the most dramatic technique for reducing the k-value is to *decrease the density* of the film material. Figure 14-2 shows the density of various low-k materials, as well as their k-values. It shows that the lower the material density, the lower the value of k. In all ultra-low-k dielectric-materials, the density is significantly lower than that of oxide films. This density reduction can occur either at the *microscopic level* (i.e., through the formation of a lattice structure that is more open than that of silicon oxide), or at the *macroscopic level* (i.e., through the formation of macropores in the film material - with sizes in the nm-range).

Second, *reduced ionization* and *reduced polarization* are only able to reduce the k-value to a much smaller extent. Reduced ionization is a common feature of the partially-organic low-k-materials. Such organic films contain carbon, often in the form of methyl groups (CH_3). These methyl groups replace oxygen atoms to create Si-CH_3 bonds in place of Si-O bonds. A smaller percentage of Si-CH_3 bonds ionize than do Si-O bonds (and they are less polarizable than are Si-O bonds). Reduced polarization is exploited in such dielectrics as fluorinated silicate glass (FSG). That is, the fluorine atoms introduced into the oxide lattice

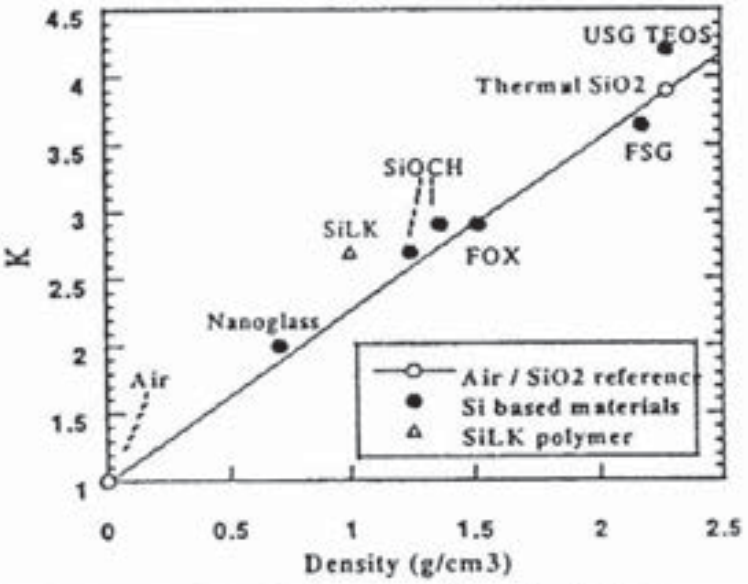


Fig. 14-2 Value of k is shown vs. density for various dielectrics in ULSI interconnects.

TABLE 14-1 ITRS Roadmap of the Effective Dielectric Constant Needed at Each Device Generation - Out to 2012

Table 1. SIA Roadmap								
1997								
Year of first product	1997	1999	2001	2003	2006	2009	2012	
Shipment technology node	250 nm	180 nm	150 nm	130 nm	100 nm	70 nm	50 nm	
Interlevel metal insulator effective k	3.0-4.1	2.5-3.0	2.0-2.5	1.5-2.0	1.5-2.0	<1.5	<1.5	
1999								
Year of first product		1999	2001	2002	2005	2008	2011	
Shipment technology node		180 nm	150 nm	130 nm	100 nm	70 nm	50 nm	
Interlevel metal insulator effective k		3.5-4.0	2.7-3.5	2.7-3.5	1.6-2.2	<1.5	<1.5	
2000?								
Year of first product			2001	2003	2005	2007	2009	2011
Shipment technology node			180 nm	150 nm	130 nm	100 nm	70 nm	50 nm
Interlevel metal insulator effective k			3.5-4.0	2.7-3.5	2.7-3.5	2.2-2.7	2.2-2.7	1.6-2.2

replace O atoms and form Si-F bonds instead of Si-O bonds. The shorter bond length of Si-F bonds makes the molecular structures containing a fraction of Si-F bonds less polarizable than those consisting of only Si-O bonds.

A few materials that were identified as good low-k on-chip insulators for ICs were first integrated into two-level-metal structures in the 1995-96 timeframe. These first-generation low-k films had k-values somewhat smaller than 4.0 ($2.8 < k < 3.5$). In 1999 the *International Technology Roadmap for Semiconductors* (ITRS-99) was published.[2] It urged 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 (Table 14-1). However, the industry proved to be more conservative. As Table 14-1 shows, there was a two-year delay in the implementation of the various low-k solutions in the years from 1997 to 1999. The delay in the adoption of low-k films was due to numerous technical challenges that had to be overcome before low-k materials could be successfully integrated into existing processes. As a result, the 0.25-μm and 0.18-μm technology generations were generally implemented with dielectrics having ranges of k = 3.0–4.0, and k = 2.5–3.0, respectively. As a result, ICs with low-k dielectrics did not appear in significant volume until about 1999. However, the consumption of materials used to produce low-k films is expected to gradually increase, and by 2003 the global low-k material market should reach $150 million and expand rapidly. In addition, it is quite possible that the expected

change from a k range of 2.5-3.0 to a k range of < 2.0 will be stretched out beyond the 70-nm generation in 2005, due to a significant increase in integration challenges for materials with dielectric constants below 2.0.

As mentioned, there are many process-integration issues that must be satisfied before a new material is successfully incorporated into an IC fabrication sequence.[3] Candidate low-k materials must also satisfy these requirements to be useful in IC fabrication. Here a number of low-k materials being considered for use in deep-submicron IC processes are described, as well as some of the process integration challenges they offer.

14.1.1 Spin-On versus CVD Methods for Forming Low-k Dielectrics

The two most common ways of depositing low-k dielectric films are: **1**) spin coating followed by furnace annealing (*spin-on deposition*); and **2**) chemical vapor deposition (*CVD deposition*). As will be seen, there are many varieties of low-k films within the spin-on and CVD-deposition groups being studied for use in advanced IC interconnects. Note that spin-on films are all applied in liquid solution form, and are then cured in nitrogen at temperatures as high as 450°C. Curing is necessary for: **a**) driving-off solvent; **b**) completing chemical reactions; **c**) driving-off unreacted monomers and oligamers; and **d**) densification.

Proponents of the CVD deposition approach point to the reuse of existing toolsets and simpler integration due to the silicon-dioxide-like structure of CVD siloxanes as advantages of their method. However, the cost of depositing films by CVD is generally higher than the cost of a spin-on deposition. Hence, this is a factor that must also be considered when a choosing a low-k material. Manufacturers of CVD-tools for depositing low-k films include Applied Materials, Novellus, ASMI, and Trikon Technologies.

Manufacturers of spin-on materials and spin-on dielectric (SOD) tracks contend that the spin-on dielectrics allow better extendibility to future generations (especially in the sub-0.25-μm regime, when porous, low-k materials are likely to be used). Manufacturers of *spin-on dielectric* (*SOD*) coaters include: Dainippon Screen (DNS, Tokyo), Tokyo Electron (TEL America, Austin, TX), FSI International (FSI, Chaska, MN), Semix (Fremont, CA), and SVG (San Jose, CA). Spin-on dielectric material vendors include: Honeywell, Dow Chemical, Dow Corning, Hitachi Chemical, and Tokyo Ohka Kogyo. Each SOD equipment vendor has optimized the coating and curing modules specifically for SODs, making them significantly different from the photoresist tracks on which they were originally based. Note that one cost factor in SOD use is the quantitative dispense of the SOD. Due to the high cost (> $2000/liter) and potential sensitivity

of some spin-on formulations to temperature, moisture, and oxygen, materials management could play a critical role in the successful integration of spin-on dielectrics. It should also be mentioned that use of fast heating and cooling rates in RTP reduces the spin-on-deposition processing time. Thus, implementation of rapid thermal processing tools into the spin-on low-k deposition process is being explored.

At present it is too early to tell which technology (spin-on or CVD), and which type(s) of low-k material will prevail. It seems that the whole industry will be characterized by uncertainty regarding this issue over the next few years. As one process integration engineer was said to remark in 1999, "My biggest problem right now is the confusion caused by the variety of material and process options that I face."[4]

14.1.2 Silicon-Based versus Carbon-Based Low-k Dielectrics

Low-k dielectrics are also categorized into two other groups: *silicon-based low-k dielectrics* (Si-based materials) and *carbon-based low-k dielectrics* (C-based materials). C-based materials typically have lower k-values, but some of their other characteristics may make Si-based materials the better overall choice, depending on the application. For example, Si-based materials usually have higher thermal stability and hardness compared to C-based materials, but Si-based materials tend to be more prone to moisture absorption.

Insofar as manufacturability issues are concerned, the Si-based materials are more compatible with existing device-fabrication flows. Adhesion to silicon, SiO_2, and silicon nitride (as well as to Al and Cu metallization) is better with Si-based materials. Also, the Si-based materials are easily etched in fluorine-based plasmas, whereas O_2-chemistries are needed for C-based materials. The selectivity of oxygen-based plasmas is not as high over organic photoresist. In addition, the Si-based materials are more compatible with CMP-processing because of their greater hardness.

14.1.3 Status of Low-k Dielectrics in 2001

Even by 2001 the news from the industry reflected the state of uncertainty as to which process will prevail. Early in 2000, IBM announced it would use SiLK (a spin-on film from Dow Corning with $k = 2.65$) in its 0.15- and 0.13-μm Cu/low-k technologies, while Texas Instruments stated it would use Black Diamond (a CVD process from Applied Materials which produces a film with $k = 2.7$) together with Cu, for its high-end DSP-products. Soon after, LSI Logic described how it would push the performance of its aluminum-based interconnects with

TABLE 14-2 List of Candidate Low-*k* Materials

Low-k Dielectric Candidates		
Material	Dielectric constant Low-k materials	Deposition method
Polyimides	3.0-3.6	Spin-on
Spin-on glasses	2.7-3.1	Spin-on
Fluorinated polyimides	2.6-2.9	Spin-on
DLC (diamond-like carbon)	2.8-3.0	CVD
Poly(arylene ethers)	2.6-2.9	Spin-on
Poly(arylenes)	2.6-2.8	Spin-on
Cyclotenes	2.6-2.8	Spin-on
Parylene N	2.6-2.8	CVD
Poly(norbornenes)	2.5-2.7	Spin-on
Polyimide-SSQ hybrids	2.7-3.0	Spin-on
Alkyl-silanes/N_2O	2.4-2.7	CVD
	Ultralow k	
Teflon-AF	1.9-2.1	Spin-on
Teflon microemulsion	1.9-2.1	Spin-on
	Porous dielectrics	
Polyimide nanofoams	2.2	Spin-on
Silica aerogels	1.1-2.2	Spin-on
Silica xerogels	1.5-2.2	Spin-on
Mesoporous silica	1.9-2.2	Spin-on

low-*k* FlowFill® CVD films from Trikon products on its G12 0.18-μm products. SONY said it had integrated FLARE® (a spin-on material from Honeywell) in its advanced devices. Yet, all of these announcements also appear to signal that the time of low-*k* dielectrics has come. In the discussion here, three ranges of low-*k* dielectric materials will be described: **1**) first generation (k = 2.8–3.5); **2**) second generation (k = 2.5–2.8); and **3**) ultra low-*k* (k = < 2.0). Table 14-2 is a list of various low-*k* dielectric material candidates.

As of mid-2000, the situation was described by Ku (of IBM), as[5] "Though leading candidates have emerged from among the dozen or so being considered a year ago, any choice appears to have downsides. No matter which material is selected, they are all fundamentally flawed in some major characteristic. Therefore, the criterion of low-*k* material selection for IC manufacturers is not purely technical; the question usually becomes which material shortcomings are we going to choose to integrate around." Table 14-3 shows the main commercial low-*k* products being used or evaluated on 2001.

14.2 DESIRED CHARACTERISTICS OF LOW-*k* DIELECTRIC FILMS

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 possess the following characteristics: **1)** It must exhibit adequate material properties (i.e., thermal, electrical, and mechanical); **2)** It must be able to work with the other materials of the interconnect structure; **3)** It must be compatible with the IC processes of cleaning, etching, CMP, and thermal treatments; **4)** It must be available in high purity form, and at low cost; and **5)** It must be able to operate reliably over the life of the product under the specified device operating conditions. Here the general characteristics of groups **1)** and **2)** are considered. Comments are made on how well the various dielectric materials meet these desired goals with respect to each type. The remaining issues are also discussed for each film in particular.

The most important material properties needed by low-*k* dielectric films are listed in order of importance in Table 14-4.[4] First, of course, is the *k*-value (in general, the smaller the better). However, next are the *thermal characteristics.* These include *thermal stability* (i.e., 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). Next, 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. It should be

TABLE 14-3 Some of the Main Low-*k* Products Being Used in 2001

Material	Manufacturer	Deposition	Category	k
FOx	Dow Corning	spin-on	inorganic (hydrogen silsesquioxane)	3.0
XLK	Dow Corning	spin-on	inorganic (porous hydrogen silsesquioxane)	2.0
SiLK	Dow Chemical	spin-on	organic (aromatic hydrocarbon polymer)	2.7
3MS-lowk	Dow Corning	CVD	C-doped oxide (Si-O-C)	2.7
Aurora	ASM-Japan	CVD	C-doped oxide (Si-O-C)	2.7
Black Diamond (BDII)	Applied Materials	CVD	C-doped oxide (Si-O-C)	2.0-2.7

noted, however, that the thermal-conductivity of typical low-k polymers is only about 2.4-3.7 mW/cm°C. This low thermal-conductivity is quite undesirable, since high-speed circuits generate significant amounts of heat that must be dissipated. The final thermal property a low-k dielectric should exhibit is a low thermal-expansion coefficient (to prevent cracking during thermal excursions (e.g., < 10 ppm/°C). It should be noted that k-value is generally traded off against thermal conductivity. For example, xerogel dielectrics ($k = 2.4$) possess thermal conductivity that is 10 times smaller than SiO_2 (which may therefore give rise to reliability problems from severe increases in circuit temperature).

The low-k films should also exhibit *little tendency to absorb moisture* and should also not exhibit an excessive permeability to moisture. Moisture can adversely impact the other materials of the interconnect structure, and can be outgassed during other processes (which can degrade the process ambient). The films should also have good *adhesion* properties. That is, these films must adhere well to the materials that are deposited on them, and they must also adhere tightly to the materials on which they themselves are deposited.

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. They should also exhibit high reliability under normal operating voltages for the design life of the IC in which they will be used.

As far as *mechanical properties* are concerned, the films should exhibit low stress (< 10^9 Pa) and have the capability of being deposited to a thickness greater than 2-μm without cracking. From a *chemical perspective* the films must also possess chemical compatibility with various reactants (e.g., they must demonstrate good resistance to such reactants as acids, bases, solvents, and resist strippers). They should also exhibit good etch selectivity with respect to the photomasking material (e.g., resist), and with respect to the substrate (e.g., underlying metal). Note that a variety of plasma treatments, thermal treatments, or other steps may be added to stabilize the film during processing. These become a large part of a chip manufacturers expertise in bringing low-k materials into the production environment.

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 also be important (e.g., no voids should occur when filling a space that is 0.35-μm wide and has an aspect-ratio = 2). However, although the gap-fill characteristics of the low-k films are not a consideration for damascene structures (since 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 become a

TABLE 14-4 List of Desirable Properties of Low-*k* Materials

1	Dielectric constant (out-of-plane, 1MHz @ room temperature)	<3.2
2	Thermal stability (C)	>425
3	Moisture absorption	<2.0
4	Adhesion (ILD/,metal, metal/ILD, ILD/ILD)	No peel after 450C cycles
5	Thermal expansion coefficient (ppm/C)	-10 to +50
6	Solvent resistance (acid, bases, solvents, PR strippers)	No weight change after dip
7	Etch rate (wet or dry)	
8	Etch selectivity (ILD/metal, metal/ILD, PR/ILD)	
9	Stress (Pa, tensile or compressive)	<1E9
10	Gap fill	No void @0.35μm and AR=2
11	Tensile strength (MPa)	<200
	Tensile modulus (GPa)	<1
	Elegation at break (%)	>5%
12	Thickness uniformity over topography (before and after curing)	
13	Cost	IBO
14	Corrosion to Al-Cu	
15	Environmental, health and safety	Pass

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, they will need to be compatible with CMP processing. Finally, use of spin-on dielectric films that evolve toxic vapors will be avoided.

From an economic perspective, the costs involved with both the materials needed to deposit the films and the CoO of the equipment required to fabricate the films should be low. In addition, the materials used to form the low-*k* films should be commercially available and have a reasonable shelf life.

14.3 GENERAL PROCESS INTEGRATION ISSUES INVOLVING LOW-*k* FILMS

There are several processes that need to be performed on the dielectric films after their deposition, and thus these films must be properly integrated into such processing. First, the films will need to undergo either etching or CMP.

As far as etching is concerned, 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. The resist-removal processing (and any resist-residue-removal steps) should not damage the dielectric layers, nor undercut the low-*k* dielectric layer beneath the hard-mask (if one is being used).

When etching *organic* low-*k* films, O_2-based plasmas are generally used. However, since an O_2-based plasma will also ash the resist masking-layer while the low-*k*-layer is being etched, an oxide or nitride hard-mask must be used to transfer the pattern to the low-*k* film. The etch-process-sequence thus consists of opening the hard-mask (using the PR to define the pattern), and then proceeding with the low-*k*-dielectric etch. Depending on the type of low-*k* film, a wet-chemistry or proprietary dry-chemistry may be used to subsequently remove the resist. Since use of a pure-O_2 plasma will tend to undercut the hard-mask, such undercutting must be suppressed through the use of additives to the O_2-plasma.

When etching *inorganic* low-*k* dielectrics such as *organosilicate glasses, OSG* (which are oxide-based materials that also contain hydrogen and carbon), a common requirement is that the etch-process be selective to PR and to a silicon-nitride etch-stop-layer. (Silicon-nitride layers are used as embedded hard-masks in dual-damascene processes - see Chap. 15.) In general, oxide-etch-processes that are selective to nitride do not etch the OSG films due to the presence of the organic component in the film. Therefore, alternate process chemistries that include appropriate gases to etch the organic component in the OSG films and that maintain selectivity to photoresist and silicon nitride must be developed.

With respect to films that must be planarized by CMP, such materials will have to be polishing-process-compatible. As an example of the issues involved with CMP, polishing xerogel films is difficult due to the poor mechanical strength of the film. This necessitates the deposition of conformal oxide liners (both beneath the xerogel layer, and also as a capping layer on top of the xerogel). The issues of CMP involving low-*k* films are also covered in Chap. 8.

Another issue involving the use of CVD-deposited organosilicate glasses (OSG) is the possibility that they will poison the DUV-resists. That is, if the OSG contains any nitrogen, it could enter the DUV-resist and cause a "foot" to be produced in the DUV-resist layer that contacts the nitrogen-containing OSG interface. There are a number of "fixes" for this problem, such as: **1**) the use of a bottom anti-reflective coating (BARC); **2**) the deposition of a spin-on or CVD capping-layer on the OSG layer; or **3**) treating the OSG layer in an nitrogen-containing plasma (to form a thin nitride barrier on the OSG surface). More information on the DUV-resist-poisoning problem is given in Chaps. 6 and 15.

Another issue of integrating a low-*k* film into an interconnect process is the

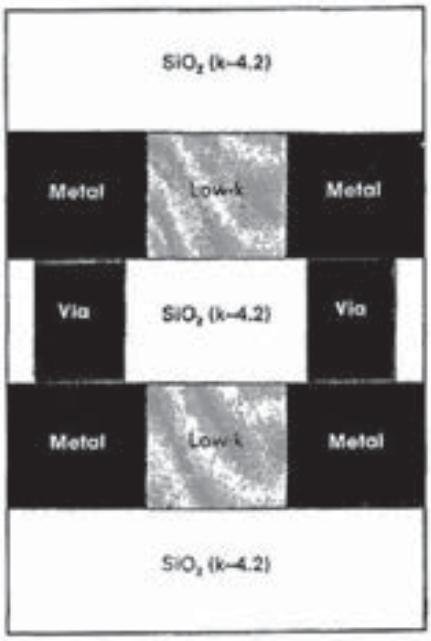


Fig. 14-3 Using low-k material between metal lines (but not at the via level) improves the mechanical strength and reliability of the stack at the expense of manufacturing complexity.

fact that such films lack the mechanical strength of oxides. One way to circumvent this difficulty is to place the low-k dielectric only *between* the metal lines, but not at the via levels (Fig. 14-3). Beyond strengthening the structure, this "embedded-approach" can reduce moisture-absorption, prevent via poisoning, and allow the fab to keep its conventional dry-etch and cleaning processes. The embedded-structure also makes for a more compatible CMP process, while allowing a company to learn from gradual integration of the low-k material. In addition, roughly 80% of the benefit is gained by using low-k material just between the metal lines. However, this approach requires that two ILD processes must be used. Figure 14-4 is a summary of the process integration issues involved with the use of low-k dielectrics as ILDs.

14.4 FIRST GENERATION LOW-k DIELECTRICS ($2.8 < k < 3.5$)

Dielectric materials with k-values moderately lower than SiO_2 were the first dielectric films to be adopted for this function. Three types of such moderately low-k films had found their way into IC production by 1999. The first two were the *spin-on dielectric* (SOD) films composed of *HSQ* and *MSQ*, with the third being CVD-deposited *fluorinated oxide films*,* Their main use was reported to be in the 0.5-μm to 0.25-μm CMOS generations.

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

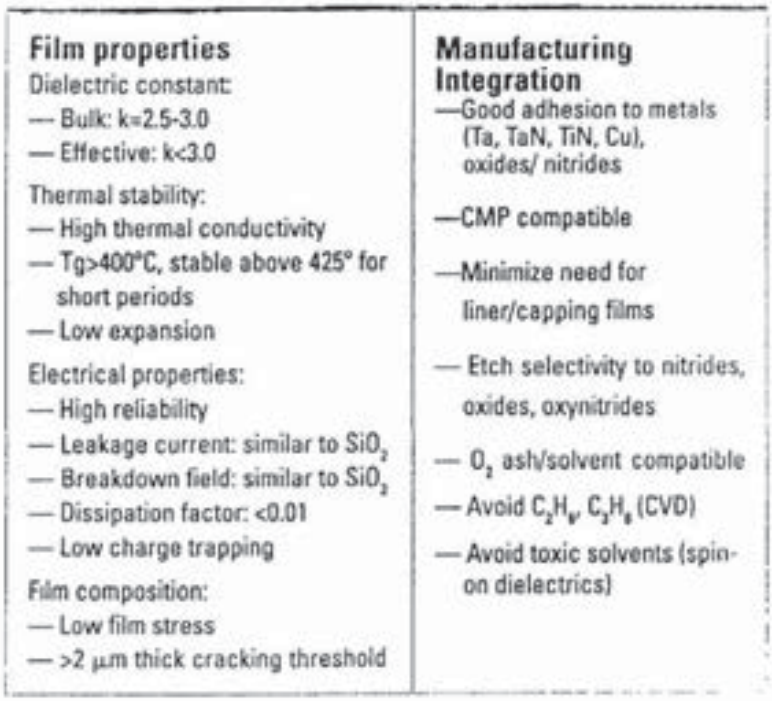


Fig. 14-4 Some of the many challenges of integrating low-k dielectrics in a copper dual-damascene process flow. Courtesy of Semiconductor International.

14.4.1 Hydrogen Silsesquioxanes (HSQ)

Films of the spin-on inorganic polymer, *hydrogen silsesquioxane* (*HSQ*) were initially investigated as replacements for siloxane spin-on-glass (SOG) films in non-etchback SOG planarization processes. When it was found that they also exhibit a lower k-value than do SiO_2 films (their k-value is < 2.9) as well as good gap-filling capability, they became candidates for low-k dielectric applications.[7]

HSQ is a subset of polyhedral oligomeric silsesquioxane macromers, a family of ordered three-dimensional polymers whose structure resembles a cage (Fig. 14-5a). HSQ can be considered a derivative of SiO_2, in which one of the four oxygen atoms bonded to every Si atom is replaced by hydrogen. That is, the caged structure of HSQ has an empirical chemical formula of $HSiO_{1.5}$. (The *sesqui* refers to the *one-and-a-half* in the stoichiometric bonding arrangement of $SiO_{1.5}$ to H.) In the SiO_2 lattice, all O-atoms act as bridges between the SiO_4 tetrahedra of the SiO_2 lattice (see Chap. 3 for details on the microscopic structure of SiO_2). However, when H (or a methyl group, CH_3), replaces an O-atom in the lattice, it no longer "bridges," but instead forms a terminal group. Thus, the tetra-

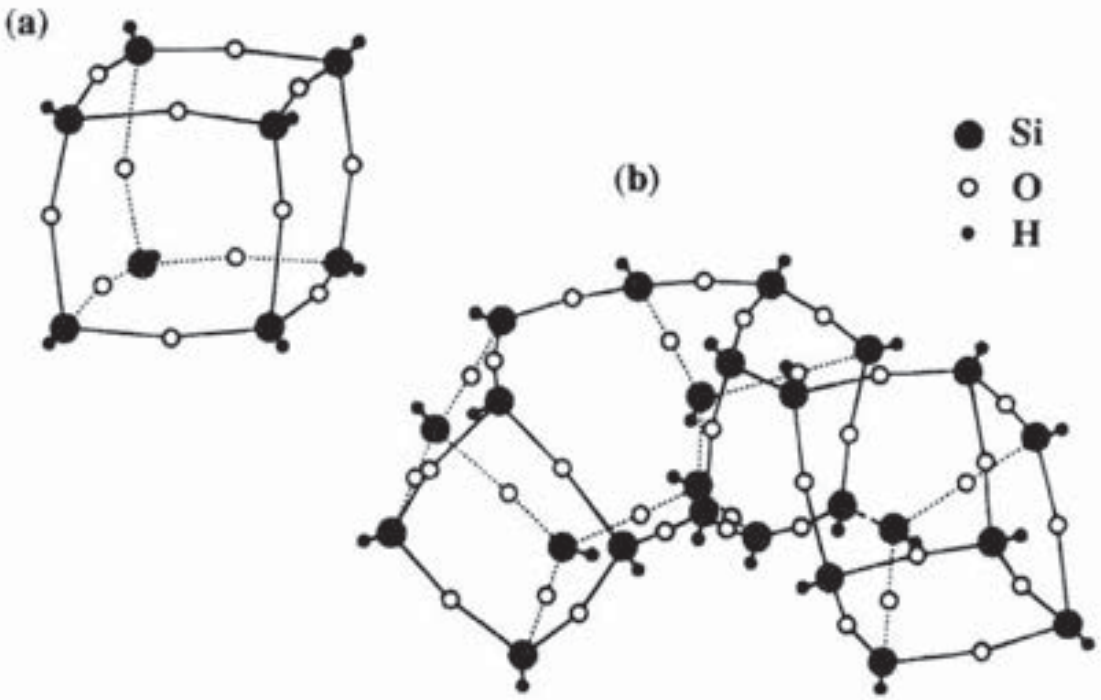


Fig. 14-5 Schematic representation of the molecular structure of HSQ: a) eight-corner oligomer; and b) resinous material.

hedron that contains the Si-H bond is then bonded only at three of its four corners. This results in a more open lattice structure (i.e.., it exhibits microporosity). Note that in HSQ films some of the Si atoms that would have bonded to O-atoms in SiO_2 are instead bonded to H-atoms. The incorporation of such Si-H bonds in spin-on dielectrics (SODs) is inherent to the spin-on material. In low-*k* Si-H bond-containing films formed by CVD, the Si-H bonds are formed by incorporating gases in a CVD-chamber that contain H.

In its uncured state, HSQ resin is composed of a mixture of Si-O networks and cage-like networks (Fig. 14-5b). This random mixture of networks is polymerized during a two-step thermal procedure consisting of : **1**) a bake at 150-250°C, and **2**) a cure at ~400°C. During these thermal treatments, a rearrangement occurs. That is, the cage structures are opened up, and there is an intermolecular redistribution of Si-O and S-H bonds. In other words, HSQ is cured from a near cage-like structure to a three-dimensional network which provides mechanical integrity to the film. The built-in microporosity of the cured HSQ network results in a reduced dielectric constant.

However, in addition to the rearrangement reaction during curing, other reactions which convert Si-H bonds to silica can also occur. If such conversion to silica is excessive, it will result in densification of porous HSQ (i.e., shrinkage of

also increase the dielectric constant.[8] That is, low-k properties in cured HSQ films can only be achieved if the density of Si-H bonding is maintained at a high level and the formation of Si-OH bonds and the absorption or creation of water in the film is minimized. Thus, when curing HSQ films, the process must balance between promoting the favorable rearrangement-reaction and inhibiting the Si-H conversion-reactions.[9] To achieve this balance, the curing-temperature must not exceed 450°C, and an optimum curing-temperature was reported to be in the vicinity of 400°C.[10]

The HSQ material is available in solution from Dow Corning under the trade name of FOx® (Flowable Oxide).[11] HSQ was reported to have been integrated into products by Texas Instruments in the 0.5-μm generation, and in 1997 an article describing its use in a five-layer Al-interconnect structure was published.[3] An oxide cap was used on the HSQ film to improve the stability of its low-k value during W-CVD and Al-fill. That is, such HSQ-films must be capped with oxide to prevent moisture uptake in the film, and the integration sequence must be optimized to prevent cracking.

The built-in microporosity of the HSQ films, nevertheless, also reduces their mechanical strength. To overcome this drawback, a plasma treatment in NH_3 at 350°C for 20 min was performed. This was found to incorporate N atoms into the film. With the incorporated nitrogen, the HSQ films were found to be mechanically stronger. Specifically, scratching that was observed during CMP of non-plasma-treated HSQ films is mitigated.[12] In addition, such films could now be cured at temperatures up to 500°C without triggering the Si-H conversion-reactions.[13] Another benefit of such a plasma-treatment is that the thin nitride film formed on the HSQ-surface (35-nm thick) reduces leakage current through the film and also provides a good barrier against Cu migration. Leakage current was reportedly reduced by three orders of magnitude below that of an as-cured HSQ film, and evidence that Cu-migration was also suppressed was presented.[14]

14.4.2 Spin-On Methyl-Silsesquioxanes (MSQ)

Spin-on *methyl-silsesquioxane* (*MSQ*) films are similar to the HSQ materials described in the previous section, except methyl groups (CH_3) are used instead of hydrogen to replace some of the Si-O bonds (i.e., Si-CH_3 bonds are formed instead of Si-H bonds). Therefore, MSQ is often referred to as an organo-silicate material. With an empirical chemical formula of $CH_3SiO_{1.5}$, there are one-and-one-half (*sesqui*) stoichiometry of oxygen bound to silicon in a caged structure of MSQ. MSQ films have the benefit exhibiting excellent crack-resistance, and can

thus be used to form thicker dielectric films. In addition, they show better moisture absorption properties than HSQ.

One example of a spin-on MSQ material is T-9,™ sold by Tokyo Ohka Kogyo. It's k-value is 2.7. The curing procedure is about the same as that for HSQ. Another spin-on film containing CH_3 is sold by Allied-Signal ($k = 2.9$). It's commercial name is Accuspin T-23.™ When cured at 380–400°C, it can tolerate temperatures as high as 450°C before decomposing. Note that development is being conducted to develop ultra-low-k dielectrics based on spin-on MSQs. One commercially available MSQ with sub-2.2 k value is Dendriglass.™

14.4.3 Fluorinated Silicate Glass (FSG) Films

FSG films are silicon oxyfluorides (F_xSiO_y) deposited by CVD. That is, FSG is produced by doping an oxide-film with F in an otherwise conventional CVD process. Fluorine is one of the most electronegative elements, and its presence in the oxide makes the oxide less polarizable (which causes the k-value to decrease). Specifically, fluorine atoms replace oxygen atoms in the oxide film to create some Si-F bonds instead of Si-O bonds. Because the attraction of fluorine to its valence electrons is greater, these Si-F bonds are less polarizable than the Si-O bonds they displaced. (Figure 14-6 shows the molecular building blocks of FSG.) In addition, fluorine-to-fluorine repulsion is stronger than oxygen-to-oxygen, and this makes oxide films containing fluorine less dense than those without incorporated fluorine. The FSG films used in ICs exhibit a minimum k value of ~3.5. This is because FSG films are formed in PECVD or HDP-CVD tools by adding a fluorine-source gas (e.g., SiF_4, CF_4, C_2F_6 , or NF_3, or other organo-silicate F sources), to the process-gas ambient used to deposit SiO_2 films by CVD (either silane and oxygen, or TEOS). Usually, SiF_4 or organo-silicate F sources are preferred because the incorporation of F into the silicate glass networks does not require breaking the strong Si-F bond (whose bond strength

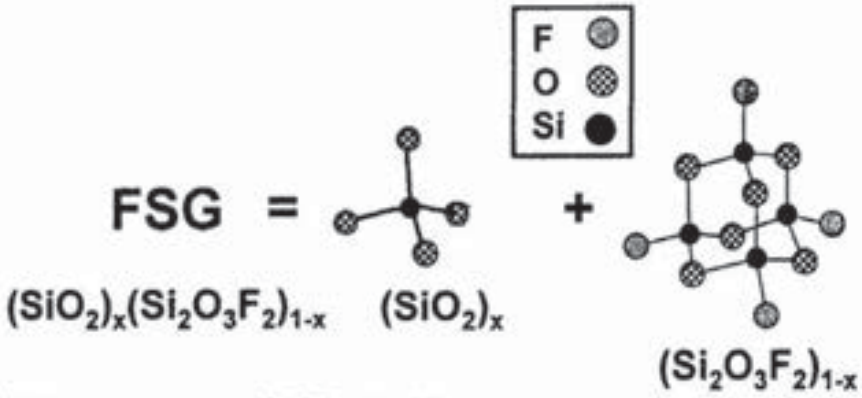


Fig. 14-6 Molecular structure of FSG materials.

is roughly 135 kcal/mole). By increasing the fluorine-source-gas flow-rate, more F is incorporated into the film. Higher concentrations of F cause the value of k to decrease, for the reasons explained above. However, a maximum of about 6% F is permitted to be incorporated into the FSG films because higher concentrations would make the FSG film unstable. That is, in such unstable films, F is evolved during RIE etching. This evolved F is highly reactive toward water. Aluminum interconnect lines embedded in this material would be corroded by the HF that is formed when the SiOF reacts with water. Because of this limitation, only a relatively small decrease in k values is possible with FSG. Thus, the benefits of replacing non-fluorinated CVD-oxides with FSG are modest.

However, FSG is relatively simple to integrate because of its similarity to the doped-CVD-oxides currently being used in the IC industry. In fact, FSG-films were reported to be successfully integrated into 0.35,- 0.25,- and 0.18-μm circuits by IBM.[15] The evolution of even small amounts of F from the FSG oxide has 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_2 formation, and the consequent loss of adhesion between the low-k film and the Ta. Despite such barrier stabilization problems, use of FSG has permitted extension of PECVD oxide processes to both subtractive-aluminum and copper-damascene interconnects. Note, however, that FSG solutions are different for the subtractive-aluminum process and the Cu-damascene process. That is, to obtain the gap-fill needed for the subtractive aluminum process, the sputter component encountered during the deposition of the FSG film bombards the depositing film. This breaks the bonds in the film and makes it less stable. Hence, even less F can be incorporated. In the damascene process, the dielectric film is deposited on a planar surface, and hence no gap-filling capability is needed. Thus, higher F-concentrations can be used. The k-value of FSG-films used with damascene interconnects is reduced by 0.2 to 0.3, compared to the k-values of FSG used with Al. As noted above, among the difficulties encountered when using FSG in a damascene process with Cu is the issue of keeping the fluorine tightly bonded in the CVD film (i.e., volatile TaF compounds can form at high temperatures).

Another possible application of FSG is replacing the conventional PECVD layer used to cap SOD low-k films. It has been demonstrated that use of FSG as a capping layer for an SOD resulted in 10% reduction in parasitic capacitance.[16]

14.5 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, and each of these types are described here.

As noted earlier, a heated battle at the $k = 2.8$ node between CVD- and spin-on-dielectric (SOD) films is taking place. It is staged on the one hand by the desire of wafer manufacturers to continue to use CVD-tool-sets for one more technology node, versus the potential lure of being able to extend the use of new SOD materials and process integration schemes for multiple generations.

That is, CVD-low-k processes can reuse existing tool sets and allow for simpler integration (due to the silicon-oxide-like structure of the CVD dielectrics). From a business perspective, CVD-tool-suppliers can also provide an integrated solution to the customer. For circumstances where there is a tight labor market for engineering talent (e.g., the Taiwan silicon foundries in 2000), the foundries need to rely on low-k solutions from their vendors that have been well characterized and can be easily introduced into production. Thus, adoption of CVD-low-k dielectrics is likely to be favored in such cases.

However, an alternative case for SOD-low-k dielectrics at the $k = 2.8$ node is made by the suppliers of SOD-materials and SOD-track-manufacturers. They claim that SOD processes are extendable to future technology nodes, especially those below 0.25-μm (when porous low-k materials will need to be used). Porous low-k SOD materials are available now for evaluation and testing. The SOD porous films may also produce films with smaller pore sizes than those deposited by CVD. This would make them mechanically stronger, allowing better film stability and better integration. The key for SOD processes is to develop an SOD solution that can be extended for multiple generations and can process 200-mm and 300-mm wafers on the same equipment. Some of the issues remaining to be resolved with organic SODs include poor adhesion and via-poisoning due to improper etch and resist removal. Spin-on-dielectrics also remain susceptible to cracking at thicknesses approaching one micron.

There is also a question at the 0.18-μm technology node as to which process should be implemented first - copper or low-k? It appears that the most cost-effective gain occurs with aluminum/low-k, particularly if $k < 3.0$ can be implemented. Hybrid structures with aluminum/low-k for the lower levels of metal, and copper at the very top levels of interconnect will allow manufacturers the ability to ship 0.15/0.13-μm products, forestalling the significant challenge of integrating copper and low-k. Some companies have elected to take this path, but others have decided to implement copper interconnects first, before incorporating a lower-k dielectric material.

14.5.1 2nd-Generation Spin-On-Dielectrics: (2.5 < k < 2.8)

Most spin-on *organic* low-k-polymers[17] are significantly different from their conventional SOG counterparts, primarily because they do not evolve moisture during curing. Since they exhibit better crack-resistance they can also be applied in a single pass to a greater thickness. Three organic spin-on low-k-dielectrics being considered are: **1**) *poly(alylene) ethers* (*PAE*); **2**) *fluoro-polyimides*; and **3**) *bis-benzocyclobutenes* (*BCB*). One other group of spin-on polymers (the *organo-siloxane polymers*), are more like their conventional SOG counterparts.

Spin-on-dielectric films are deposited without the use of an expensive CVD-tool. Instead, they are deposited with a less-costly track cluster-tool, consisting of spinners and hot plates. The cost of the spin-on liquid, however, is typically much higher per wafer than that of the CVD-reactant-gases.[18] In addition, a 100-nm-thick capping SiO_2 layer (deposited by CVD) will probably still be needed with spin-on low-k-layers. But an added benefit of using spin-on films is that they reduce the topology of any previous steps by more than 50%.

The *PAE materials* are aromatic structures with k-values of about 2.5 which can also withstand higher temperatures. A fluorinated PAE-product from Honeywell is named FLARE,™ and it has a k value of ~2.8. It was designed to have low-outgassing characteristics and high thermal and mechanical stability.[19] Dow Chemical has a similar material named SiLK (k = 2.62).™ Despite its name, SiLK is also a hydrocarbon that contains no silicon (and also no fluorine). A report by Fujitsu detailed their experience with integrating FLARE and SiLK into a two-level Cu-damascene-structure. These films showed little change in k-value after a five-hour anneal at 425°C.[20] Schumacher has also pioneered the use of non-fluorinated PAE-materials, with their latest such product being called 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.

In April 2000, IBM announced plans to integrate SiLK into its next-generation 0.15-μm and 0.13-μm Cu-11 process, intended to go into high-volume production in the first half of 2001. They claimed the interlevel dielectric layers made with SiLK demonstrated excellent resistance to Cu-migration, and the mechanical strength of SiLK films is high enough to make them compatible with CMP (and back-end processes such as wire-bonding). IBM said it chose to use a spin-on-dielectric film for two reasons. First, SiLK meets the critical thermal, mechanical and chemical requirementss. Second, it offers extendability to the next generation. "To get the effective dielectric constant down into the 2.2-2.3

range, there are no materials, and there is no way to get there with PVD or CVD. Our opinion is that the best way to achieve that is by the introduction of porosity. That would say, whether its an organic or an organosilicate, you're going to have to do a spin-on process."[21] One reported difficulty involving the use of SiLK is that if the film is immersed in HF solutions (even as dilute as 0.1%-HF), the HF will diffuse through the film and attack the oxide at the oxide/SiLK interface. This will lead to SiLK liftoff.[22]

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 called Cyclotene™ for IC ILD-applications. A study in 1998 indicated that Cu drifts much more slowly through BCB films than through PAE.[23] 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.

As noted above, there is also a class of SOG-like materials available in the 2nd-generation low-*k*-range. These are *hybrid-organo-siloxane polymers* (*HOSPs*) which have a *k*-value of 2.5 after curing. HOSPs have a 12% carbon-content, which allows them to be formed into layers up to 1.2-μm thick without cracking. Their *k*-value remains stable at temperatures up to 380°C. HOSPs demonstrate good gap-fill and planarization capabilities. However, the quality of HOSP films (and those of other carbon-containing spin-on-dielectrics) is degraded by damage from an oxygen plasma used when stripping resist. The conventional oxygen-based ashing-process can attack the low-*k*-dielectrics that contain any carbon, which will break down the film integrity. To overcome this limitation, a plasma treatment in H_2/N_2 or NH_3 (as discussed in previous sections), has proved to be effective. Honeywell markets such a low-*k* material under the name HOSP.™

Note that a plasma-treatment of organic SOD layers with an NH_3 gas reportedly improves some of the properties of these films (as was also discussed in the previous section the covered HSQ films). That is, such a treatment produces a thin nitride layer on the SOD surface, and this layer acts both as a barrier against Cu migration and a layer that improves the ashing resistance of such films.[24]

14.5.2 2nd Generation-CVD-Dielectrics: (2.5 < k < 2.8)

Three dielectric materials deposited by CVD and with a value of k in the range of 2.5–3.0 have been developed: **1)** *carbon-doped siloxanes,* also termed *organosilicates glasses* (*OSG*), which have a k value of 2.5; **2)**, *fluorinated parylene* (*parylene-F*); and **3)** an *inorganic low-k material using organosilane/ozone.* One 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. Another advantage is that the CVD processes use readily available and inexpensive gaseous precursors (in contrast to the expensive and proprietary spin-on liquid-precursors), which reduces the comparative chemical cost of the CVD approach. Furthermore, CVD does not require the handling of organic-solvent wastes.

14.5.2.1 2nd Generation Low-k-CVD Dielectrics: Organo-Silicate Glasses: The OSG dielectric films deposited by CVD basically have the same molecular structure as the HSQ- and MSQ-films deposited by spin-on methods. The CVD OSG films can be can be formed by PECVD or HDP-CVD, using a variety of precursors, including methylsilane, dimethylsilane, trimethylsilane (3MS), or tetramethylsilane, together with N_2O (or in some cases ozone). As noted earlier, in the discussion on HSQ- and MSQ-films, the methyl-groups form terminal Si-CH_3 bonds on the tetrahedral structures that make up the oxide lattice. These Si-CH_3 bonds contribute to reducing the k-value by increasing the microporosity of the film and by reducing its polarization tendency. In OSG-low-k-films formed by CVD, these methyl-groups come from the precursor gases used in the CVD process. In general, the CVD process must balance the film requirements for low-k with other desired film properties. That is, low methyl-content can lead to films that behave merely like standard oxide films (i.e., with k ~4.0). On the other hand, high methyl-content can lead to Si-C-Si bonding, which will degrade thermal stability and will hamper the etching of the film (i.e., through the production of carbon polymers during the etch process).

Precursors with more methyl-groups are capable of incorporating more carbon into the film, though the compounds are more difficult to dissociate. As a result, their deposition rates are smaller. For example, tetramethylsilane deposition rates are about 25% slower than trimethylsilane because it takes more energy to dissociate the tetramethyl molecule. Another significant difference lies in the delivery system, since trimethylsilane is a gas and tetramethylsilane is a liquid. (A trimethylsilane precursor gas is offered by Dow-Corning, which they have named Z3MS.™) An advantage of OSG films is that they have thermal

conductivity values similar to those of silicon oxides. As high-speed devices will generate more heat, such relatively-high thermal-conductivity represents a significant benefit. Manufacturers of OSG-CVD-systems include Applied Materials and Novellus.

A CVD-low-k film using organosilane and ozone was also recently reported.[25] This film can be easily integrated into an interconnect scheme that employs W- plugs and Al-lines. In the CVD process, the Si-H-bond of the organosilane precursor is selectively broken by an active O_3-molecule to form a highly mobile $Si(CH_3)_x$-intermediate to achieve good step-coverage and gap-filling between metal lines. Void-free filling was achieved at spacings of 0.12-μm and an aspect-ratio of 5:1. A post-deposition-anneal at 400°C polymerizes and forms a low-k film. A cap PECVD-oxide film is deposited over the low-k-layer to avoid direct CMP of the low-k-material. With the capping oxide, the normal ashing and cleaning process can be applied to remove resist-residue and etch by-products. The W-via-filling process is straightforward after standard de-gas and liner-deposition. No via-poisoning was observed after the W-CMP step.

14.5.2.2 2nd Generation Low-*k*-CVD Films: *FlowFill*® CVD Films: The *FlowFill*® films from Trikon Technologies are OSG-films (silsesquioxanes) deposited with a thermal-CVD process using a mixture of methylsilane and hydrogen peroxide (H_2O_2) gases.[26] These react in the gas-phase to form methylsilicic acid, which condenses on a cool wafer (i.e., the wafer is kept cooled to 0°C during the deposition process). The condensed material takes the form of a methyl-substituted oxide-gel that contains a considerable amount of water. The high flowability and surface-tension of this gel enables it to seamlessly fill gaps and provide a significant degree of local planarization. After deposition, this gel is converted into an amorphous methyl-doped silicon-oxide by a higher temperature cure in a separate chamber. That is, annealing at 400°C drives-off the water and densifies the film, leaving a high-quality methyl-incorporated silicon-oxide-film (which has a porous structure and a low dielectric constant). Figure 14-7 shows the process flow of this CVD approach.

The FlowFill-films exhibit a k-value of 2.8 and can be used for gap-filling applications in a subtractive-aluminum interconnect-process, or a Cu-damascene-process. For gap-fill, the deposition sequence involves first putting down a plasma-deposited base-layer, then depositing the low-temperature flow-layer and performing a cure step (and finally capping this layer with a standard PECVD film). The overall k-value of this composite-film is ~3.1. FlowFill layers can also be deposited thicker (to > 1.5-μm) with a modified cure step. This allows a

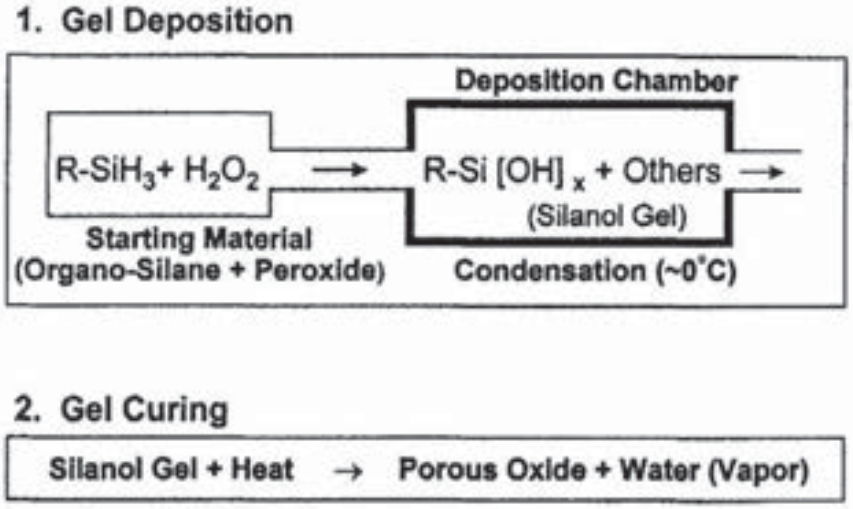


Fig. 14-7 Trikon *FlowFill*® CVD approach for depositing silsesquioxane films.

simple extension to timed-etch damascene structures using the same hardware and tool platform (i.e., the Trikon Planar® system). The film is attractive because its properties are very similar to spin-on glasses, whose process integration issues are well understood. However, its impact on circuit performance is limited by the need for encapsulation by the PECVD-SiO_2 capping layer.

14.5.2.3 2nd Generation Low-k-CVD Films: Black Diamond™ and Coral™ CVD Films: Applied Materials has developed a modification to the CVD OSG-films that they have termed Black Diamond.™ In this approach, a near-room-temperature PECVD-process is used to deposit a silsesquioxane film using the same organo-silane precursors, but with gaseous oxidizers like O_2 or N_2O instead of H_2O_2. By going from a cool deposition process (as is used in the *FlowFill* approach of Sect. 14.5.2.2) to a warmer one, Applied said it could eliminate the post-deposition furnace-cure step while positively influencing film-hardness and surface-roughness. This also produces an as-deposited film that already has a silsesquioxane type of porous structure, and which exhibits a k-value of about 2.7. Depending on the deposition temperature, curing may still be necessary to outgas the trapped reaction byproduct from the film. In addition, the high sticking-coefficients of the gaseous reactants result in poor step-coverage and gap-fill capability. Thus, Black Diamond films can be used as an inter-metal dielectric only in damascene interconnect schemes.

Applied Materials also claims that the same tool set used to deposit the first-generation FSG-films can also be used to deposit their Black Diamond films. This will allow the same equipment base to be used through several generations.

Applied Materials also announced that this material was validated for sub-0.18-μm devices by International SEMATECH.

Novellus calls their CVD OSG-films Coral.™ Their k-value is 2.7. Novellus claims that through process optimization, the hardness of the Coral film has been improved sufficiently so that it can be used in standard plastic packages. To protect the OSG-layer from attack during resist removal, a capping layer of SiO_2 is CVD typically deposited onto it (although SiC-layers are also being introduced).

14.5.2.4 2nd Generation Low-*k* -CVD Films: Parylene CVD-Films: Parylene (poly-*p*-xylylene)-based polymers are one of the most promising low-k CVD materials. Parylene polymers are high molecular weight (20,000-50,000 amu) linear chains of phenyl groups isolated by ethyl groups at the para positions. Figure 14-8 shows the different types of parylene polymers. The parylene precursors are low-melting-point solids that are composed of molecular dimers. The deposition consists of heating the solid in the presence of a carrier gas so that the dimers are carried to a cracking-furnace. The thermal energy in the furnace splits the dimers into the individual parylene monomers. The carrier gas then brings the monomers into a deposition chamber where they condense onto a wafer at temperatures below 0°C. Polymerization and cross-linking occur even at such low temperatures because the monomers are very reactive radicals. Parylene 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 depositing parylene AF4 is also commercially available from Novellus.[27]

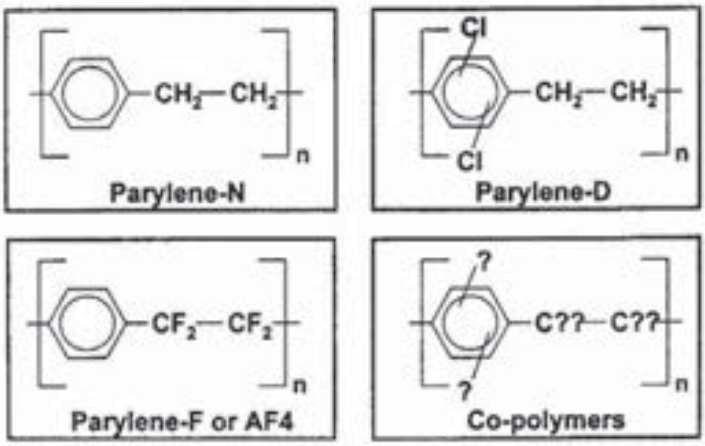


Fig. 14-8 Different types of parylene polymers.

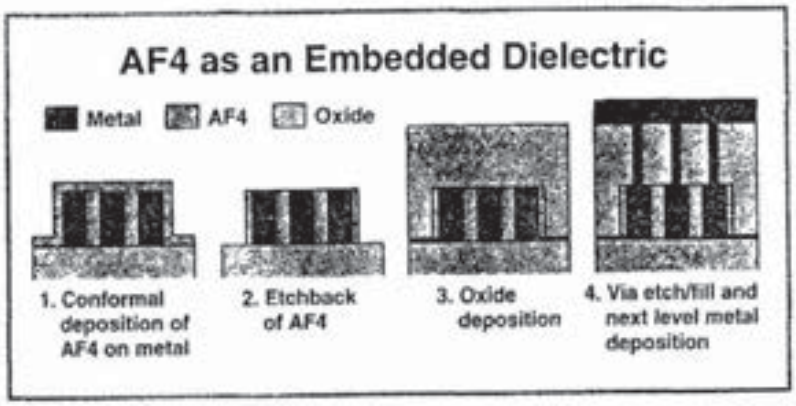


Fig. 14-9 In this integration scheme, the low-k material is encapsulated between the closely spaced metal lines where the low-k dielectric has its greatest influence on RC-delay. The interlevel dielectric is oxide, which has superior mechanical and thermal transport properties. Reprinted with permission of Semiconductor International.

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. 14-9). 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 chemical-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.[27]

14.6 ULTRA-LOW-k DIELECTRICS: (k < 2.3)

Ultra-low-k dielectric materials are also being investigated for their suitability as IC interconnect layers. Several different types were being considered, and are mentioned here. These types of ultra-low-k materials have not yet been proven for production-worthy devices. Specifically, according to the 2000 edition of the ITRS, it appears that such materials will not be included in commercial products until at least 2005.

One reason that ultra-low-k materials will be needed is that future high-performance circuits will use a copper damascene approach. Thus, the low-k dielectric likely be used in conjunction with an etch-stop (or polish-stop)

dielectric layer (which will have a higher dielectric constant). Consequently the *k*-value of the ultra-low-*k*-layers will need to be decreased so that the overall dielectric constant of such ILD-stacks can meet the desired *k*-value specifications. Furthermore, adhesion of the low-*k* material to the other materials in the structure may also involve use of *dielectric-liners* with higher *k*-values. Many of these liners will probably be new to semiconductor processing.

Note that numerous ways of producing films with $k < 2.2$ are available. However, all of these methods invariably rely on reducing the density of the dielectric material. The most commonly used approach is based on a porous-silica structure (although silsesquioxane and aromatic-organic-materials have also been formed into porous films). These types of films, however, exhibit deterioration of the mechanical properties with decreasing *k*-values, as shown in Table 14-4.[28] In order to improve the mechanical properties of the porous films, a number of methods are being pursued. For films produced by the standard sol-gel technique (see later paragraphs in this section for more on sol-gels), particular attention must be given to the gelation and aging procedure, and to the termination of bonds on the interior pore walls (which typically requires an additional process step of some kind (e.g., thermal or plasma). However, silsesquioxane-based precursors reportedly can be treated with a sol-gel technique without requiring additional pore-wall passivation. In order to produce sol-gel films with more order and uniform pore size, self-organization stimulated by surfactant templating has been utilized. (Note that the information contained in this paragraph was presented at the Ultra-Low-*k* Workshop sponsored by SEMATECH in March 1999.)

Besides the sol-gel techniques, porosity has been introduced into dielectric films by burning out sacrificial material.[29] Another porosity producing method involves the interpenetration of a low-*k* constituent with a silica structure that is formed by standard SOG coat and bake procedures. One such approach involves depositing a multiphase film by CVD (that contains one film material that has a thermally-unstable phase). This is followed by an anneal (400°C for 8 hours), which removes this labile phase from the material. The dual-phase materials used in this study were deposited by PECVD from mixtures of SiCOH precursors with gases containing mainly carbon- and hydrogen-atoms. Following the anneal, the *k*-value was found to be reduced to 2.4 from its pre-annealed value of 2.8. Efforts are also being undertaken to produce porous dielectrics by PECVD using organosilicate glasses from a number of silane-based precursors.

The challenges of integrating such porous ultra-low-*k*-materials in copper damascene are formidable. Not only must they meet the low-*k* requirements, but

the films should possess properties as close to the incumbent SiO_2-based materials as possible. These properties include low leakage-current, low thermal-coefficient-of-expansion, high dielectric-breakdown-voltage, low film-stress, low moisture-absorption, good adhesion to other materials, high thermal-conductivity, and mechanical-strength. Realistically, the materials properties of porous ultra-low-k-films will not match those of silicon dioxide, because they are compromised when pores are incorporated into a dielectric. Balancing low-k performance with sufficient mechanical properties (to survive mechanically intensive processes such as CMP), is one of the most significant hurdles facing ultra-low-k materials.

The first of the ultra-low-k materials considered here 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, and thus its overall k-value is lowered. 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 Honeywell (formerly Allied Signal) under the name of Nanoglass.™ It was originally developed in 1996. 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 (i.e., both oxide- and organic-based), must withstand such process steps as CMP, etching, and heat treatments without degradation of the pore structure. Other spin-on porous materials include porous-SiLK (from Dow Chemical), and porous versions of FLARE and HOSP (from Honeywell).

Dow-Corning has developed a spin-on film consisting of an interconnected pore structure based on HSQ-resin named XLK. In 2000 it could produce films with a k-value of 2.5, and the vendor claims it will be extendable to values of $k = 1.5$ in the future.[30,33] With a pore size of 2-3-nm, a 50% porosity, a tight pore-size-distribution, and a fully-interconnected network-structure, XLK appears to offer many of the characteristics needed for device integration. Pores are introduced through a three-step process on an SOD track system: **1**) gellation in the presence of a boiling-point solvent at ambient conditions; **2**) removal of the solvent without collapsing the resin structure; and **3**) formation of a strong network structure during cure near 400°C. Gellation occurs when the film is exposed to an NH_3 and H_2O environment for 30-120 seconds, which causes Si-H bonds to react with the H_2O to form SiOH. Then the SiOH bonds join together to form a SiOH gel. A plasma treatment in $H_2/N_2 + CF_4$ can be used to increase the mechanical properties of the XLK films.[31] In a single-level copper damascene



Fig. 14-10 Trench with 0.35-0.4-μm lines and spaces etched in 700-nm XLK film (ultra-low-k-dielectric) with 100-nm-thick oxide-cap-layer shows good sidewall etch-profiles.

integration study, 700-nm of XLK and 100-nm of an oxide-cap were deposited on a wafer with a 550-nm thermal oxide and a 100-nm nitride. A two-step recipe was used to etch the cap-oxide, then the low-k film, resulting in good etch-profiles (Fig 14-10). Following ashing in an HDP reactor, a 25-nm TaN barrier, with 100-nm Cu-seed and 1-μm of electroplated Cu were deposited. CMP of the structure showed no delamination. This work represented the first evidence of the feasibility of integrating an ultra-low-k film at the first-level of Cu-interconnect.

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-CF_2-polymers (and so contains 67-at% F). A prototype spin-on version of a PTFE dielectric with a $k = 1.9$ has been introduced. 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). A thermal-CVD PTFE film first developed by Watkins-Johnson (now SVG Thermco Group), with a precursor from DuPont Electronic Gases Group, is also being offered.

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

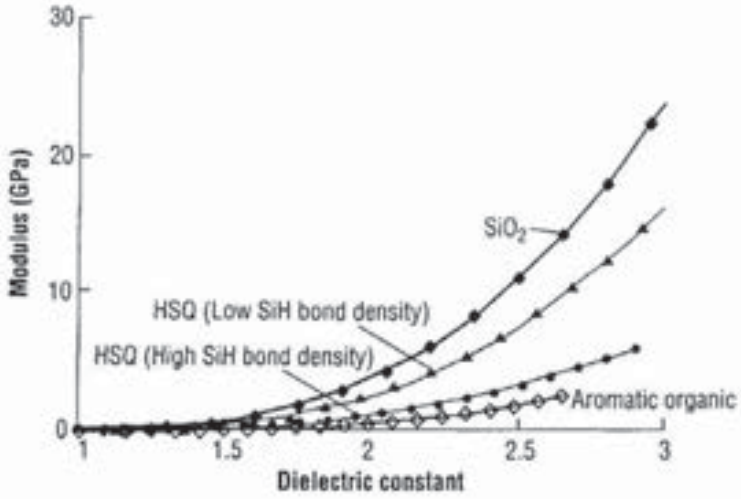


Fig. 14-11 Theoretical approximation of modulus of elasticity of porous ultra-low-k materials as a function of dielectric-constant indicates porous-SiO_2 offers the greatest mechanical stiffness.

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. Since the ultra-low-k films wont be commercially used until at least 2005, CVD companies still have time to develop a CVD film with a porous matrix.

In summary, the chief factor that determines the important characteristics of ultra-low-k materials is the pore structure. Each matrix material (whether it is silica-, silsesquioxane, or aromatic-organic based) has a unique pore structure.

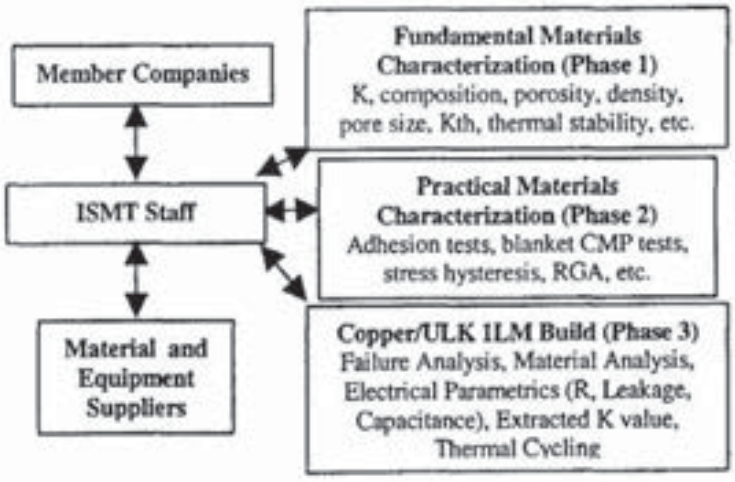


Fig. 14-12 Initial evaluation of candidate low-k materials at ISMT. The flow of information is meant to provide verification to material/equipment suppliers and probability of materials success in a two-level-metal integration.[32] (© IEEE 2001)

Material Property	Method	Specification	Guideline
K (relative permittivity)	MIS Capacitor 1MHz, RT, 40% RH	<2.5	<2.2
Thermal Stability (%loss/hr)	I-TGA In N2 and Ar, 425C	<1.0%	<0.5%
Film WIWNU, 1-sigma	Spectroscopic Ellipsometry	<3%	<2%
Young's Modulus (GPa)	Nanoindentation 1000nm thick film	>3	>4
Adhesion	ASTM Tape Pull	No Failures	No Failures
Maximum Pore Size	PALS/SANS+SXR	<25nm	<10nm
Film Quality	Visual Inspection Profilometry Light Scattering	No Striations, visual defects	Particles <100 @0.2um

Fig. 14-13 List of materials Specification and Guidelines. Specifications are the minimum expected material property, and Guidelines are the desired set of properties to ensure successful integration.[32] (© IEEE 2001)

For example, since air fills the pores, the k-value ultimately depends on that of the matrix material. In addition, the degradation of the modulus of elasticity E (which makes a material softer), is approximated by a power of the film porosity. Each type of porous dielectric-material exhibits a degradation of modulus of elasticity relative to fully dense silica (whose $E = 72$), as is shown in Fig. 14-11. Pore-size may also impact the continuity of Cu barrier-layer films, since these films are deposited with thicknesses comparable to the dimension of the pore size. Finally, compatibility with wet-chemical processing is influenced by pore structure. Open or connected pores are prone to absorb wet chemicals (such as water). Since the k-value of water is 72, any retained moisture in the pores degrades the dielectric constant. More details on this topic are given in Ref. 34.

One further change when ultra-low-k-dielectrics are used may involve going away from via-first dual-damascene interconnects (the approach most commonly used in 2000). With porous material, via-etching cannot be done first because after via opening the trenches are etched, and problems arise with resist absorption into the porous sidewalls of the via (see Chap. 15, Sect. 15.9.1.2).

Late in 2001 a report from Wetzel at International Sematech (ISMT) outlined a study that was being conducted there on candidate materials for ultra-low-k

ILDs.[32] Figure 14-12 shows the general scheme for the initial evaluation of such materials. The list of materials specifications sought for ultra-low-*k*-dielectrics is shown in Fig. 14-13. They suggested that integrating such low-*k* films into a dual-damascene process will face three main challenges: **1**) the ability to successfully survive CMP processing (especially with respect to the mechanical strength and adhesion of the films); **2**) the compatibility of ultra-low-*k* materials with DUV resist must be improved, especially when following a via-first dual-damascene approach; and **3**) porous-silica films must be able to maintain their initial *low-k* values through the processes that occur after the films are formed.

REFERENCES

1. L. Peters, "Pursuing the Perfect Low-*k* Dielectric," *Semiconductor International,* Sept. 1998, p. 64.

2. *International Technology Roadmap for Semiconductors* (ITRS-1999).

3. 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 Materials,* May 1998, Abs. 157.

4. S.K. Chiang and C.L. Lassen, "The Market for Low-*k* Dielectrics," *Solid State Technology,* October 1999, p. 42.

5. V. Ku, "Low-*k* Integration Challenges," *What's Up from SEMI,* June 27, 2000.

6. K.A. Monnig, "Low-*k* Dielectrics for Integrated Circuit On-Chip Insulators," *Vacuum & Thin Film,* June 1999, p. 27.

7. M.J. Loboda *et al.,*"Properties of α-SiO_x:H Thin Films Deposited from Hydrogen Silsesquioxane Resins," *J. Electrochem. Soc.,* Vol. **145**, p. 2861 (1998).

8. J. Bremmer *et al., Spring Mtg. Materials Res. Soc.,* April 1997.

9. Y.K. Siew *et al.,* "Thermal Curing of Hydrogen Silsesquioxane (HSQ)," *Abs. Mtg. Electrochem. Soc.,* Spring 1999, Abs. 203.

10. Y.K. Siew, *et al.,* "Thermal Curing of Hydrogen Silsesquioxane," *J. Electrochem. Soc.,* Vol. 147, January 2000, p. 335.

11. H. Kojima *et al., Proceedings 5th Intl. IEEE VMIC Conf.* CA, p. 411, 1988.

12. T.-C. Chang *et al.,* "Ammonia Plasma-Passivation Effects on Properties of Post-CMP Low-*k* Hydrogen Silsesquioxane (HSQ)," *Abs. Mtg. Electrochem. Soc.,* Spring 2000, Abs. 240.

13. D.J. Kim, "Effect of NH_3 Plasma Treatment on Improvement of Reliability of W-B-N/HSQ Thin Films," *Abs. Mtg. Electrochem. Soc.,* Spring 2000, Abs. 230.

14. K.-M. Chang *et al.,* "Using NH_3 Treatment to Improve the Characteristics of Hydrogen Silsesquioxane for Copper Interconnection Application," *J. Electrochem. Soc.,* Vol. **147**, May 2000, p. 1957.

15. L. Peters, "IBM Reveals Key Copper/Low-k Integration Issues for Production," *Semiconductor International,* August 2000, p. 25.

16. C.-F. Yeh, "Comprehensive Investigation on FSG Prepared by Temperature-Difference-Based Liquid Phase Dep.," *J. Electrochem. Soc.*, January 2000, p. 330.
17. S. Murarka, "Low Dielectric Constant Materials for Interlayer Dielectric Applications," *Solid State Technology*, March 1996, p. 83.
18. E. Korczynski, "Low-*k* Dielectric Integration Cost Modeling," *Solid State Technology,*, October 1997, p. 123.
19. K.S. Lau *et al.*, "Characterization and Thin-Film Properties of FLARE," *Proc. 1996 VMIC Conf.*, p. 92.
20. M. Ikeda *et al.*, "Integration of Organic Low-*k* Material with Cu-Damascene," *IEEE International. Interconnect Conf., (IITC)*, June 1998, p. 131.
21. P. Singer, "IBM to Use Dow's SiLK in Low-*k* Process," *Semiconductor International*, May 2000, p. 44.
22. G. Passemard and O. Demolliens, "Low-*k*: Damascene Integration Issues," *Abs. Mtg. Electrochem. Soc.*, Spring 2000, Abs. 228.
23. J. Gaynor and A. Ralston, "AF4 as a VLSI Interconnect Material," *Semiconductor International*, December 1997, p. 73.
24. L.-Q. Xia *et al.*, "Simplified Intermetal Dielectric Integration with Chemical Vapor Deposited Low-*k* Material Using Organosilane/Ozone Chemistry," *Abs. Mtg. Electrochem. Soc.*, Spring 2000, Abs. 229.
25. 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.
26. K.-M. Chang *et al.*, "A Novel Pretreatment Technology for Organic Low-Dielectric Material to Suppress Copper Diffusion and Improve Ashing Resistance," *J. Electrochem. Soc.*, Vol. **147**, June 2000, p. 2332.
27. M. Matsuura *et al.*, "A Highly Reliable Self-Planarizing Low-*k* ILD for Sub-Quarter Micron Interconnects," *Tech. Dig. IEDM*, 1997, p. 785.
28. B.T. Ahlburn, "Ultra Low-*k* Dielectrics, - Shortening the Time Line from R&D to Manufacturing ," *Abs. Mtg. Electrochem. Soc.*, Spring 2000, Abs. 218.
29. *Electrochemical and Solid-State Letts*, 2, p. 77, February 1999.
30. L. Peters, "IMEC, SEMATECH Show Feasibility of a Sub-2.0 Low-*k* Dielectric," *Semiconductor International*, July 2000, p. 33.
31. W. Chen *et al.*, "Plasma Treatment of Hydrogen Silsesquioxane (HSQ) Based Ultra-Low-*k* Films," *Abs. Mtg. Electrochem. Soc.*, Fall 2000, Abs. 535.
32. J. Wetzel *et al.*, "Evaluation of Material Properties & Performance of Ultra-Low-*k* Insulators for Inlaid Cu Metallization," *Tech. Dig. IEDM*, 2001, p. 73.
33. H.-J. Lee *et al.*, "Correlations Between Structural Characteristics and Process Conditions of HSQ-based Porous Low-*k* Thin Films," *Spring Materials Res. Conf., April 2001.*
34. J.N. Bremmer, "A New Class of Insulating Materials: The Emergence of Ultra-Low-*k*," *Solid State Technology*, September 2001, S3.

Chapter 15

DUAL-DAMASCENE INTERCONNECTS

As the submicron era evolved in the 1990's, several novel manufacturing methods were adopted to keep pace with Moore's law. One such approach, namely the formation of interconnect structures with *damascene processes* was chosen to fabricate IC interconnects. In this chapter we describe this technique.

15.1 DAMASCENE VERSUS SUBTRACTIVE INTERCONNECT STRUCTURES

The conventional way of fabricating each conducting layer of an interconnect structure of an IC is a so-called *subtractive process.* In such a subtractive process, a blanket metal film is deposited onto an insulator-covered wafer surface (with contacts - or vias - having been previously etched in the insulator layer - Fig. 15-1). Next this metal film (e.g., Metal-1) is patterned by etching - to create metal interconnect lines. Finally a dielectric film is deposited over these etched metal lines. Vias are then etched in the overlying dielectric film to allow connection passages for the next layer of metal that is to be deposited onto the dielectric film covering the Metal-1 lines. This subtractive approach requires several process capabilities: **1**) It must be possible to effectively etch the metal and; **2**) It must be possible to deposit both the dielectric and metal layers with a deposition process that has good gap-filling capability. In some interconnect schemes (especially as the ICs are scaled below 0.5-μm), it becomes difficult if not practically impossible, to meet these goals. For example, for technology nodes smaller than 0.5-μm it is desirable to replace Al with Cu as the intercon-

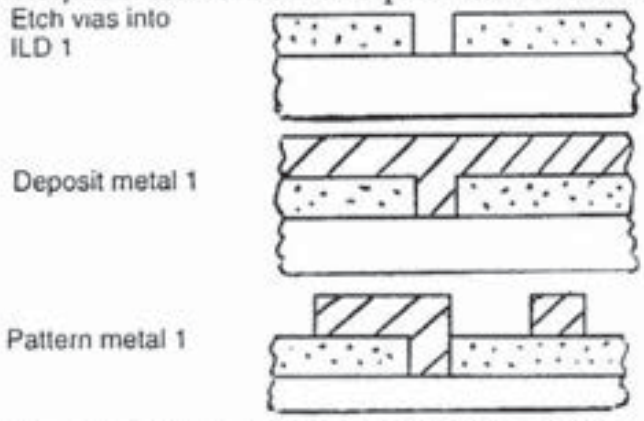


Fig. 15-1 Subtractive approach to forming interconnect structures for ICs.

nect conductor (i.e., because Cu has a higher conductivity than Al, and also exhibits superior electromigration characteristics). However, since it is difficult to etch Cu, an interconnect scheme that does not require Cu etching would be preferable. In addition, as the spaces between metal lines become smaller, it becomes progressively harder to fill such narrow gaps with an insulator film. Again, a significant advantage would be gained if it were possible to use a process that does not require gap-filling when insulator films are deposited. To achieve the above goals, an alternative interconnect fabrication technology to the subtractive method has thus been developed, called the *damascene method*.

15.2 DAMASCENE STRUCTURES AS ALTERNATIVE DEEP-SUBMICON INTERCONNECTS

In the damascene approach to forming interconnects on ICs, a trench is cut into the dielectric layer and then filled with metal. (The term *damascene* is derived from the decorative practice of creating metal inlays, as developed by the artisans of Damascus in the Middle Ages.) This method thus provides two major advantages over the subtractive method of forming interconnect structures:

1. In the damascene approach, no etching of the metal layers is required. The metal-etch process of the subtractive approach is replaced with a metal chemical-mechanical-polishing process (CMP). The metal-CMP step removes any metal deposited above the top surface of the trench (including the regions where the metal is deposited on the dielectric surface). This leaves a planarized dielectric surface with metal inlaid in trenches of the dielectric. Since the CMP process leaves a globally planarized surface, *the need for filling narrow gaps during the dielectric deposition process is eliminated*. The only gap-filling capability required by the damascene approach is during the deposition of metal into the recesses of the dielectric. Since metal etch (especially the etching of Cu) and dielectric gap-fill are viewed as two of the greatest challenges in the drive to smaller dimensions, elimination of these steps represents a major benefit. On the other hand, filling of even narrow, high-aspect-ratio trenches with metal (i.e., with W by CVD, or Cu by electroplating) is not as difficult.

2. There is only a bottom interface for Cu (versus top and bottom interfaces for Al-alloys with W plugs). This helps reduce overall via resistance.

We will also see that by using a *dual-damascene* process, the number of processing steps is smaller than for the subtractive approach. Hence, use of dual-damascene interconnects may also result in lower manufacturing costs. The

damascene approach was pioneered by IBM, and began to appear in commercial IC products at about the 0.25-μm node.[1]

15.3 SINGLE-DAMASCENE INTERCONNECT STRUCTURES

Damascene interconnects can be formed with either a *single-damascene* or *dual-damascene* process. The dual-damascene process offers the advantage of reducing the number of process steps, and thus most of our discussion focuses on it. However, the single-damascene process is always carried out on at least the first level of a multi-level interconnect structure of an IC. For this reason, it is useful to begin with a description of the single-damascene approach.

In the *single-damascene process* (see Fig. 15-2 [and Fig. 12-10 in Sect. 12.1.4]), the interlevel dielectric layer (ILD) is planarized after it is deposited. Next, a contact (or via) is patterned with a lithography step, and then the contact opening is etched into the ILD layer. After completion of the contact-hole etch step, a metal plug is formed in this opening - typically by blanket deposition of the metal, followed by an etchback or CMP step to form the plug (e.g., a W-plug). Thus, the surface of the dielectric containing the metal plug after such an etchback or polish step is globally planar (see Fig. 8-92). The metal lines of the interconnect level are formed next. This can be done using a subtractive process, as described in Sect. 15.1, or with another single-damascene process.

Assuming a single damascene process is applied twice (first to create plugs, and second to create metal lines), the second time it is employed involves the following process steps (Fig. 15-2). Another dielectric layer is deposited over the

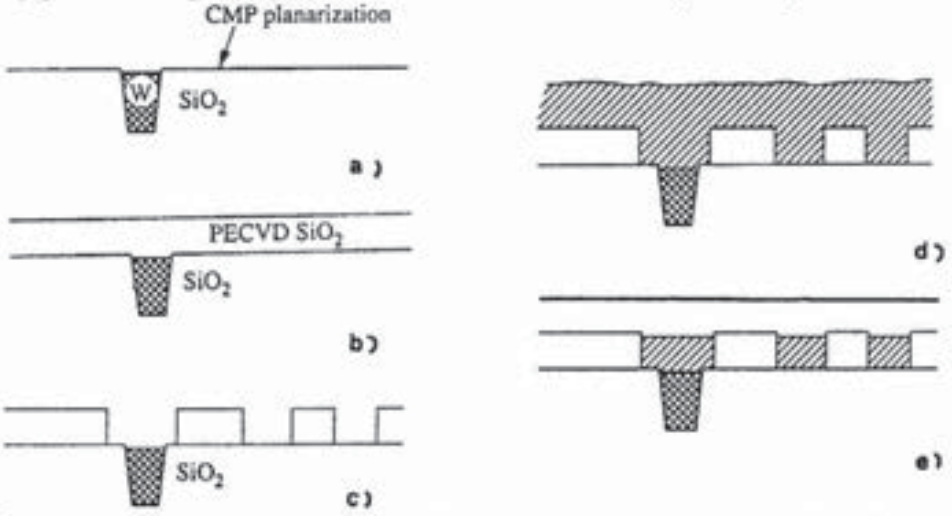


Fig. 15-2 A *single-damascene* process (applied twice) to form one level of an interconnect structure: (a) W-plugs formed after the SiO_2 is planarized; (b) IMD is deposited by CVD; (c) A trench is etched in the IMD; (d) Metal is deposited to fill trenches; (e) CMP metal.

underlying, plug-containing layer. This overlying dielectric film therefore also remains planarized. A second lithography step is then used to form the pattern of a trench, which is aligned to the previously fabricated via plug. With this aligned trench pattern acting as a template, a trench or groove is etched into the planarized overlying ILD. After the resist pattern is stripped, the trench is filled with a metal (such as aluminum or copper).[2] This metal is polished back by CMP to create a metal line embedded in a dielectric trench, and in contact with the underlying plugs. Finally, another dielectric layer is deposited onto the dielectric layer containing the metal-filled trenches. This dielectric layer will be the one through which the next level of vias (or holes to the contact pads) will be created.

It is worth emphasizing several important aspects of the single-damascene process: **1**) If CMP is used to form both the plug and the metal in the trench, no metal etch process is needed; **2**) No gaps have to be filled by the dielectric-film-deposition processes; **3**) The process still contains about the same number of steps as the subtractive interconnect process (i.e., the two polish steps replace the via and metal etch steps of the subtractive process, and the two metal-deposition steps replace the metal-plug-deposition step and main-metal-line deposition step of the plug/metal-line subtractive interconnect process); and **4**) Since the first-level interlevel dielectric layer is planarized before the contact holes are etched, *contact holes* will generally have different depths (e.g., those going to the substrate will be deeper than those going to the gates). Thus, to allow all contact holes to be opened, the shallower ones will expose the underlying material to the plasma etching reactants during the time the deeper contacts are being opened. To prevent the underlying material in the shallow contact holes from being attacked during this time, an etch-stop layer is often deposited after the salicide process is completed (typically using Si_3N_4 as such an etch-stop film). After the contact etch is completed, the nitride etch-stop layer is removed, Since the thickness of this layer is uniform on all regions, it can be etched away with a much shorter overetch time. This method allows all of the contact holes in a single-damascene process to be opened without damaging the underlying material on any part of the die.

15.4 INTRODUCTION TO DUAL-DAMASCENE INTERCONNECT PROCESSES

In a *dual-damascene (DD) structure*, only a single metal deposition step is used to simultaneously form the main metal lines and the metal in the vias. That is, both trenches and vias are formed in a single dielectric layer. The vias and trenches are defined by using two lithography steps (Fig. 15-3). Trenches are typically etched to a depth of 4000-5000-Å, and the vias are typically 5000-

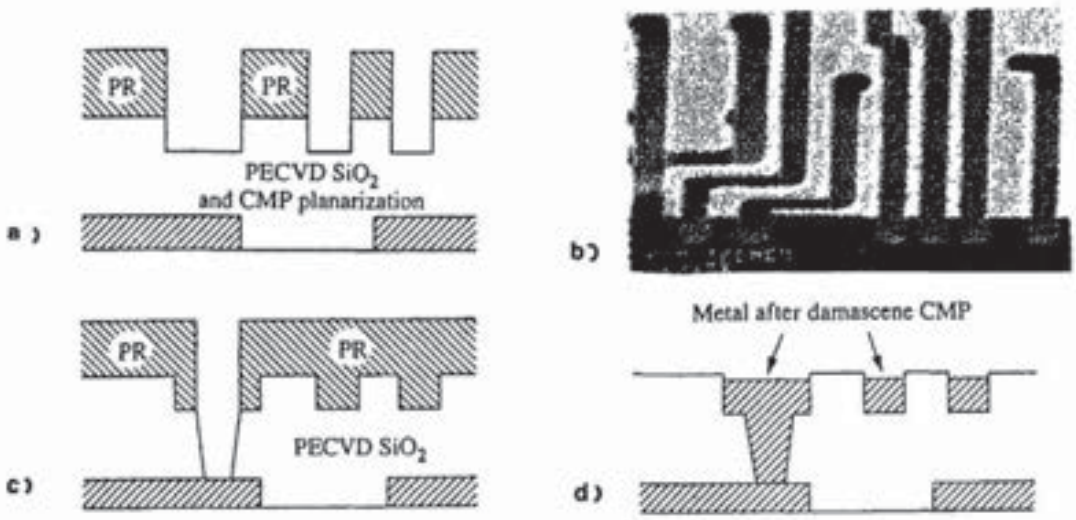


Fig. 15-3 *Dual-Damascene* process: (a) IMD is deposited by CVD and planarized by CMP. Trench is defined by PR #1 and then etched; (b) SEM of the etched trench after PR #1 has been stripped; (c) Vias are defined by PR #2 and then etched, using PR #2 to protect other regions of the IMD from etching; (d) Metal is deposited to simultaneously fill trenches and vias. CMP is used to remove excess metal. The sequence is repeated for the next level of metal.

7000Å-deep. After the via and trench recesses are etched, the via is filled in the same metal-deposition step that fills the trench. 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. As in the single-damascene process, once a planarized surface is achieved, it is no longer necessary to perform CMP on the dielectric layers. Thus, one of the CMP steps needed for each metal level in the subtractive interconnect process is also eliminated.

It should also be mentioned that in early dual-damascene processes, the trench dimension was somewhat wider than that of the vias. This made the alignment of the via to the trench somewhat easier. However, with time, the via and trench widths have been made equal. (At the 0.18-μm node, the via and trench widths are 0.22-0.24-μm wide, while at the 0.13-μm node they will probably be ~0.18-μm wide.) Thus, modern dual-damascene processes are more sensitive to misalignment between the trench and via. However, this overlay sensitivity problem is mitigated to some degree by the nature of the damascene process, making it possible to maintain high interconnect packing density. That is, the alignment of the trenches to the vias in the dual-damascene sequences is an easier task than in the subtractive interconnect process. In the latter, aligning metal lines

to the underlying contact holes (or vias) must be done after the opaque, shiny, blanket-metal film has been deposited. In damascene approaches, aligning the trenches to the vias is done through the transparent dielectric film, which allows tighter design-rule tolerances to be used when performing this alignment.

When comparing the single-damascene process to the simple dual-damascene process described above, one observes one metal-deposition step and one CMP step are eliminated in the latter (as well as one dielectric-deposition step). The reduction in the number of processing steps is another of the benefits that have driven the development of dual-damascene processes. Note that in single-damascene structures, plugs are typically filled with W, but in the dual-damascene-interconnects the metal that fills the holes is Al or Cu. The technology of filling of vias and trenches with Cu is described in Chap. 16. A dual-damascene process using Al was described by workers at IBM in 1998.[3]

15.5 THE THREE DUAL-DAMASCENE PROCESS SEQUENCES

Three different fabrication sequences have been developed to produce dual-damascene structures:

1. Trench-First Dual-Damascene
2. Via-First Dual-Damascene
3. Self-Aligned (or Buried-Via) Dual-Damascene

It should be noted that the *via-first* and the *trench-first* sequences are the processes currently being used in mainstream production. The trench-first sequence was probably the first one to be adopted, but the via-first sequence is currently the approach that has gained the greatest acceptance (and will likely remain so going forward). As a result, we will focus on these two sequences, and will point out the advantages and drawbacks of each one. While the third approach (the self-aligned process), has not been implemented in production (for reasons we will also explain), we will provide a brief description of it as well. However, whichever sequence is used, the final damascene structure is the same (see Fig. 15-4).

Before outlining the three dual-damascene processes, it should be mentioned that the dielectric layer in which the trenches and vias are etched actually consists of at least three layers of dielectric (see Fig. 15-8b). That is, SiO_2 (or a low-k dielectric is deposited to a depth of 5000-7000Å (and this will be the layer in which the *vias* will be etched). Next, a thin silicon nitride layer - or other material (which will serve as an *embedded etch-stop layer*), is deposited (about 30-nm thick). Finally, another SiO_2 layer (or a low-k dielectric layer) is deposited

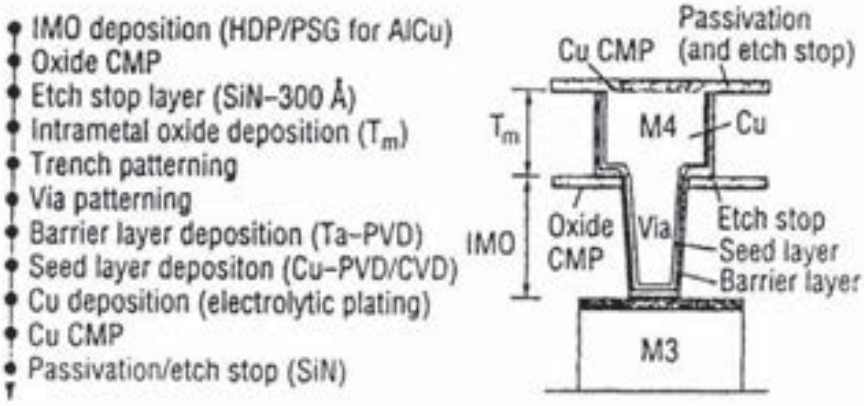


Fig. 15-4 A suggested dual-damascene-based Cu metallization flow and resulting cross-sectional structure.[8] Reprinted with permission of Solid State Technology. Published by PennWell.

(~4000-5000-Å thick). This will serve as the layer in which the *trenches* will be etched.

In the *trench-first* sequence (Fig. 15-5), the trench patterns are defined in the ILD first. That is, after spinning on the resist, the trench-pattern-mask is used to expose the resist. The trench is then produced by etching the dielectric down to the embedded etch-stop layer. After the trench-etching process, the first resist layer is stripped. A second resist layer is then spun on, and the via-pattern-mask is used to create openings in this resist layer - aligned to the trench that was etched previously. The resist protects the other parts of the wafer surface (including the etched regions of the trenches) so that the vias can be etched without further etching the dielectric in the trenches.

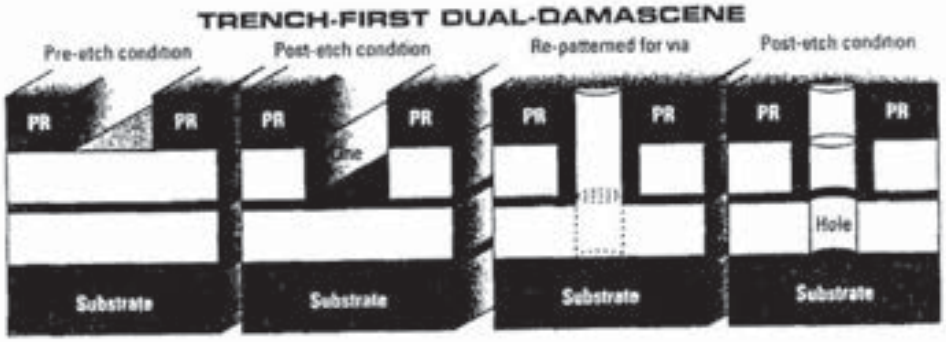


Fig. 15-5 In the trench-first approach, the vias are patterned & etched after trenches are etched.

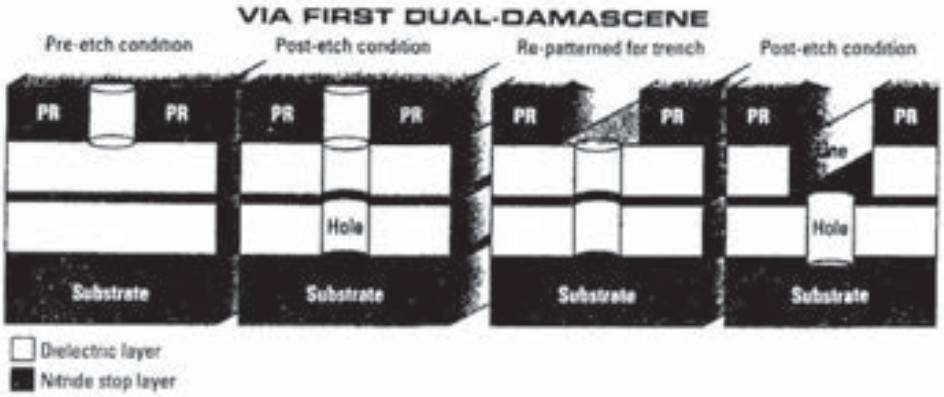


Fig. 15-6 In the via-first approach, the trench is patterned and etched after the via is formed.

The chief disadvantage of the trench-first sequence involves the fact that the vias must be patterned after the trench etch. That is, since resist is applied as a liquid onto the wafer surface, it fills recessed regions as water fills a lake. Hence, the top surface of the resist is planar. This means that the regions of resist over the damascene trenches is quite thick. Resolving fine features in thick resist is harder than in thinner resists. For features smaller than 0.25-μm the process latitude becomes too small for this to be a practical manufacturing process. Thus, while the trench-first sequence could be used for early dual-damascene processes (0.35-μm and some 0.25-μm processes), it is difficult to extend it to smaller technology nodes (i.e., in order to able to use it at such small dimensions, it may be necessary to implement a thin imaging-resist-layer process, as discussed later).

In the *via-first approach* the vias are defined first in the ILD, followed by patterning the trenches. As noted above, this is currently the most common method of fabricating dual-damascene interconnects (and is schematically shown in Fig. 15-6). The sequence of forming the damascene recesses in this approach begins by exposing the *via patterns* with the first mask. After etching the vias completely through the entire dielectric stack (except not through the barrier layer at the bottom of the dielectric stack) and stripping the resist, a second mask is used to pattern the trenches. The trenches are then created by etching the dielectric down to the embedded etch-stop layer. The barrier layer at the bottom of the vias is protected from further etching during the trench-etch either by resist or a BARC layer (applied for the trench-etch step) that floods the vias, or by using an etch process that is highly selective to the bottom etch-stop material (i.e., so that the etch-stop layer at the bottom of the via is not removed during the trench-etch step). After the resist is stripped and the etch-stop layer at the bottom of the via is removed by dry-etching, the metal that fills both the vias and the

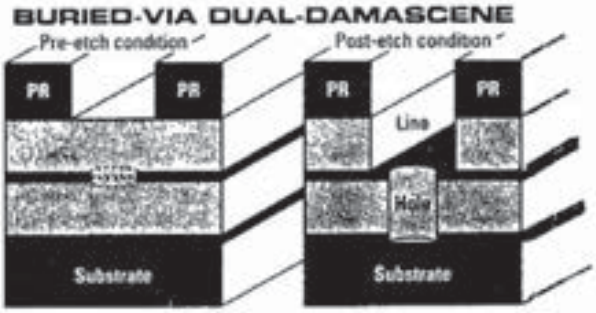


Fig. 15-7 In the self-aligned, or buried-via approach, a hard mask is patterned and etched. The trench and via can be etched in one step.

trenches can be deposited. After deposition, it is polished back to create the dual-damascene structure.

In the *self-aligned* dual-damascene process (Fig. 15-7), the via pattern is created in the embedded etch-stop layer with a lithography-and-etch sequence that is performed before the top dielectric layer of the stack is deposited. A special etch process is required for this step. After the via pattern is thus etched, the top dielectric layer is deposited. Finally, the trench mask is aligned to the via openings in the embedded nitride layer, and both the trench and via are opened with a single etch step. This sequence requires nearly perfect trench-to-via alignment, otherwise the via may no longer be round (but instead only half-moon shaped), resulting in high via resistance. In addition, an etch-process with high nitride-to-oxide selectivity is required. For these reasons this scheme has not found use in production.

15.6 THE EVOLUTION OF DUAL-DAMASCENE FROM SiO_2 TO LOW-k DIELECTRIC STACKS

In the first generation of dual-damascene processes (i.e., at the 0.25-μm node), Cu was used together with a conventional undoped PECVD-SiO_2 dielectric (with a PECVD-SiN layer beneath the SiO_2, which acts as both a Cu diffusion barrier and an etch-stop when the vias are etched, Fig. 15-8a).* The nitride and oxide layers may be deposited sequentially in the same tool (e.g., a PECVD reactor).

* This SiN layer is deposited first, and as noted, serves as a barrier against the diffusion of Cu between metal levels and also as an etch-stop in the dielectric etch that defines the vias. Because SiN has a high value of k, this layer should also be as thin as possible to avoid adding significantly to the interlevel capacitance. At the same time, the SiN layer must be low-stress, dense, and pinhole–free, so that its diffusion-barrier properties are not compromised. High-density is also critical to the nitride layer's role as an etch-stop.

This Cu/SiO_2 structure gave improved interconnect performance over the Al/W-plug subtractive process because Cu interconnects (i.e., with Cu in both the lines and vias) have a lower resistance than Al lines combined with W-plugs. Thus, the RC delay term was reduced through a decrease in R. A significant problem that occurs with using this dielectric structure is that the trench formation has no etch stop. Since the trenches are etched to an intermediate point part way through the dielectric layer, within-wafer etch-non-uniformity translates directly into within-wafer non-uniformity of the trench depths.

In second generation dual-damascene processes (down to about the 0.18-μm node), the CVD-SiO_2 material was replaced by a first generation low-k material, namely a PECVD-SiO_2 layer doped with fluorine (*fluorinated silicate glass, FSG*) whose $k = 3.5$. (FSG has many properties similar to SiO_2 and thus requires the smallest number of process integration changes.) A SiN layer, embedded halfway through the thick dielectric layer, was also added to the dielectric stack to serve as an etch-stop layer when the trenches are etched. However, since the k value of SiN is $k = 7.0$, the embedded-etch-stop SiN layer increases the net-k value of the resulting dielectric stack. In fact, the overall capacitance of the FSG dielectric stack with an embedded-SiN etch-stop layer was about the same as that of first generation SiO_2 layers - but it had the benefit of being able to ensure the trench is only etched halfway through the dielectric layer. The corresponding drawback is that fabricating a multi-layer dielectric stack adds cost and complexity to the manufacturing sequence.

The next step in the evolution of the dual-damascene dielectric stack was to incorporate a second generation low-k material ($k = 2.7$) into both the upper and lower segments of the dielectric stack, as this now allows the wiring capacitance to be reduced. (Note that the wiring capacitance is reduced by about 11% when a SiLK/SiN dielectric stack is implemented, compared to the wiring capacitance obtained when using a FSG/SiN stack.*) The first dual-damascene structure that incorporated a 2nd generation low-k dielectric in the stack (together with Cu), was announced by IBM in April 2000. It was intended for use in the IBM's Cu-15

* In a paper published in late 2001, an alternative dielectric stack to the SiLK/SiN stack was proposed. That is, it used FSG and a new embedded layer material (SiCN), whose k value of 5 is lower than that of SiN. The capacitance of this stack was about 9% lower than that of the FSG/SiN stack, making its wiring capacitance comparable to the SiLK/SiN stack. However, the authors claim that the SiLK/SiN stack is mechanically weak, and tends to fail during bonding and packaging processes, whereas the FSG/SiCN stack is more robust.[5]

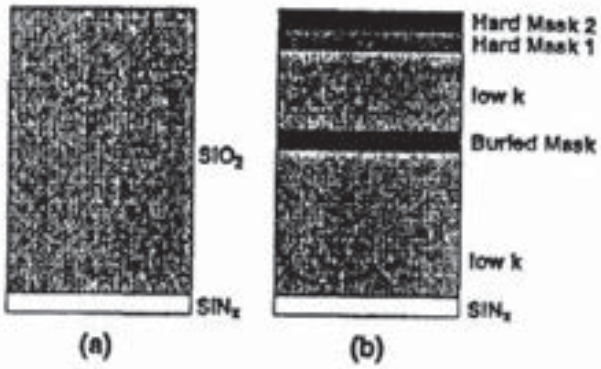


Fig. 15-8 a) First generation dual-damascene ILD stack, consisting of SiO_2 and a bottom SiN layer. b) Later generation dual-damascene ILD stack with low-k dielectric, embedded etch-stop layer, and hard-mask/capping layer (SiO_2).

0.13-μm CMOS technology.[4] The low-k dielectric material selected for this process was SiLK™ from Dow Chemical (see Chap. 14). The dual-damascene process is used to form the #2 through #5 interconnect levels of the 5-level-metal interconnect structure used for this technology node.

It should also be noted that when a 2nd generation low-k dielectric replaces the FSG (or SiO_2) material in the stack, a hard mask must also be added to the top of the stack. The material chosen by IBM for this hard mask was SiO_2, giving the dielectric stack a structure as shown in Fig. 15-8b. SiO_2 was selected for this role because it has a significantly lower k-value than SiN. In addition, this SiO_2 hard mask remains in place during metal deposition, so it can later serve as a polish-stop during metal CMP. Since use of an SiO_2 hard mask maintains the same materials system encountered during the CMP of metal and a PECVD-SiO_2 dielectric stack (Cu/TaN/SiO_2), the developed consumables, equipment, and processes for Cu-CMP on plain SiO_2 should be extendable to Cu-on-SiO_2-capped low-k. However, since this oxide cap becomes a primary supporting structure against material loss by dielectric erosion during CMP, the allowable dielectric erosion (see Sect. 8.11.1) in a capped dielectric stack is dramatically decreased.

It should also be noted that in some cases different materials may be used for the dielectric at the via level than for the trench level. Since most of the capacitive coupling issues that a low-k material helps solve are encountered between adjacent metal lines, the *trench dielectric* will be where the lowest-k

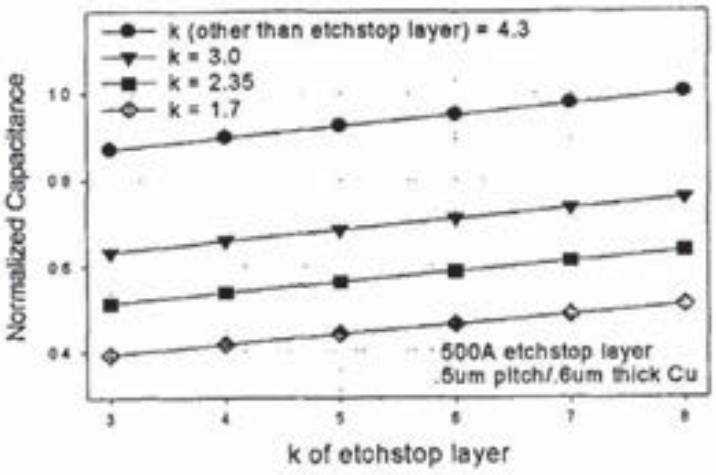


Fig. 15-9 Capacitance impact of the k-value of the etch-stop value in a dual-damascene structure for various dielectric constants. [32]

materials are first used. Since the via dielectric is less critical from a capacitance perspective, implementing a low-k material there may be delayed.

The use of SiN as an embedded etch-stop layer increases the capacitance. Figure 15-9 shows the effect of a 500-Å SiN on dielectric stacks with different low-k dielectrics.[32] To allow the capacitance of the dielectric stack to be reduced even further, a new material has been introduced to replace SiN for the embedded-etch-stop and hard-mask roles. This material is SiC (with k = 4.5). Figure 15-10 shows how much the use of Si-C instead of SiN reduces the effective k-value of the damascene dielectric stack.[33] Applied Materials introduced this material in 1999 under the name BLOk (barrier low-k).[6,7] In

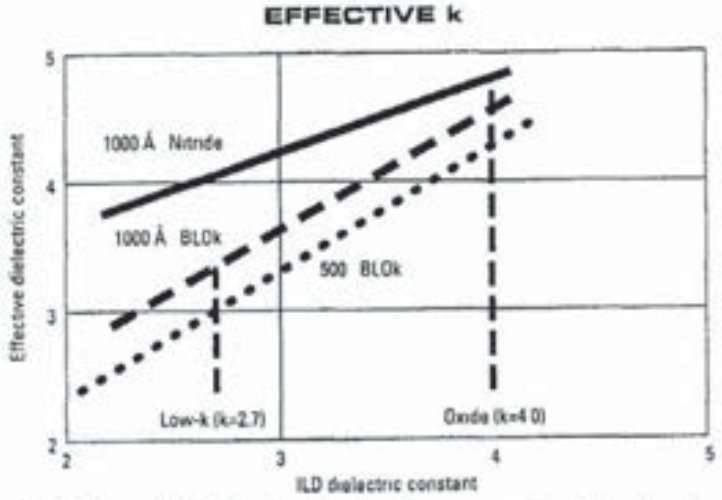


Fig. 15-10 The SiC-based dielectric barrier impacts the effective k value much less dramatically than the traditional nitride film. Courtesy Applied Materials.

addition, this material offers twice the etch selectivity of nitride to oxide-like films, has better Cu diffusion barriers than nitride, and shows good adhesion ot SiO_2, FSG, and organic-based low-k dielectrics. Furthermore, it exhibits the optical properties needed to make it behave like an anti-reflective coating in DUV imaging. The film is deposited in an HDP-CVD chamber using trimethylsilane at 350-400°C. The resulting Si-C film can also be etched using a "soft etch," which means it can also be used as the capping layer for Cu (i.e., at the bottom of the via of the via-first dual-damascene structure). Some companies have reportedly incorporated the Si-C films into damascene dielectric stacks which employ undoped SiO_2 or FSG, while others are using it with stacks containing lower-k films. Another vendor that supplies a tool and a process to deposit such an Si-C film is Novellus.

It is thought that the Si-C material will replace SiN as the etch-stop layer in low-k dielectric stacks for the 0.13-μm node. However, as devices are scaled even further, it is predicted that even the k value of 4.5 of Si-C will prove to be too large. Thus, it has been projected that at the 0.1-μm node, the embedded-etch-stop layer will be removed entirely. The ramifications of removing the embedded etch-stop layer will be discussed in a later section.

15.7 TRENCH-FIRST DUAL-DAMASCENE PROCESS FLOW

Here we present an example of a trench-first dual-damascene process flow, using undoped SiO_2 as the main dielectric material and Cu as the main interconnect metal. We assume that an STI isolation structure is used, so that the substrate is flat, up to the time of gate-stack formation. We will also see that both single-damascene and dual-damascene processing are used. Many variations of this example damascene-interconnect structure have been published.[8,9,10,11,12] Some advanced interconnect structures may elect to use both conventional W-plugs and etched Al-wire interconnects for some metal levels, and Cu-damascene for others. One such example is given in Ref. 13.

The fabrication sequence of this interconnect structure is begun by carrying out two single-damascene process sequences to form both the Metal-1 lines (Cu) and CVD-W plugs in the contact holes. The first process sequence is started by depositing a dielectric layer over the gate stack. This film serves as the *pre-metal dielectric* (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 process is then used to fill the contact holes. The W is then planarized, so that W-plugs are formed such that their top surface is flush with the top surface of the PMD.

The second single-damascene process is then performed. That is, an additional

dielectric film is deposited onto the PMD layer, and trenches (aligned to the W-plugs) are etched into it, deep enough so that the tops of the W-plugs are cleared of dielectric. A liner film (e.g., TaN) and a Cu seed-layer are sequentially blanket-deposited deposited by PVD. The main Cu film is then blanket deposited by electroplating. The Cu film (and TaN liner layer) are polished off of the top surface of the dielectric film with CMP, leaving the Metal-1 Cu lines inlaid in the grooves. The main challenge here is to obtain the same polishing rate on the tantalum nitride (which is a hard refractory-metal-containing nitride), and the copper (which is comparatively soft). This completes the formation of the Metal-1 and contact plugs by two single-damascene processes. The surface of the wafer is flat for the next dielectric film deposition.

The subsequent layers of metal are formed by a dual-damascene process sequence. The sequence begins by depositing a thin, low-stress silicon nitride layer. This layer forms a top encapsulation layer on the level-1 inlaid Cu lines, and also acts as an etch-stop layer when the vias are etched in the dual-damascene sequence. Next, a dielectric layer (here, undoped SiO_2) is deposited by a CVD process to a thickness required for the via-hole depth. Since the surface is flat, gap-filling by this layer is not required. Then, the embedded etch-stop layer is deposited by CVD (typically PECVD-silicon-nitride with a thickness of 30 nm). Finally, the inter-metal-dielectric stack (IMD) is completed by depositing a second dielectric film, whose thickness is determined by the next-metal-level thickness. (Note that the multi-layer dielectric film, containing two nitride etch-stop layers adds cost and complexity, as well as increasing the interlevel capacitance of the interconnect stack.) Photoresist is applied to the wafer surface. The trench pattern in the resist is formed by exposing the resist through a mask containing this pattern (Fig. 15-5). An etch step is performed next, with the etch creating the trenches in the upper part of the IMD stack. The trench etch is stopped by the silicon-nitride etch-stop layer embedded in the dielectric stack. The resist is then stripped, and a second layer of resist is applied. Another mask (this one containing the via pattern), is used to expose the second resist layer, with the via aligned to the previously etched trench-pattern. The via pattern is then etched in the lower part of the IMD stack. This may be a two-step etch process, as etching the nitride etch-stop layer at the bottom of the via to expose the Cu of Metal-1 may require a separate plasma chemistry than the one used to etch the underlying oxide material. Note that in practice, the trench width may be only slightly larger than the via width.

After the second resist layer is stripped, the formation of the metal line is begun. Again, a TaN (or Ta or TiN) liner and a Cu seed layer are sequentially

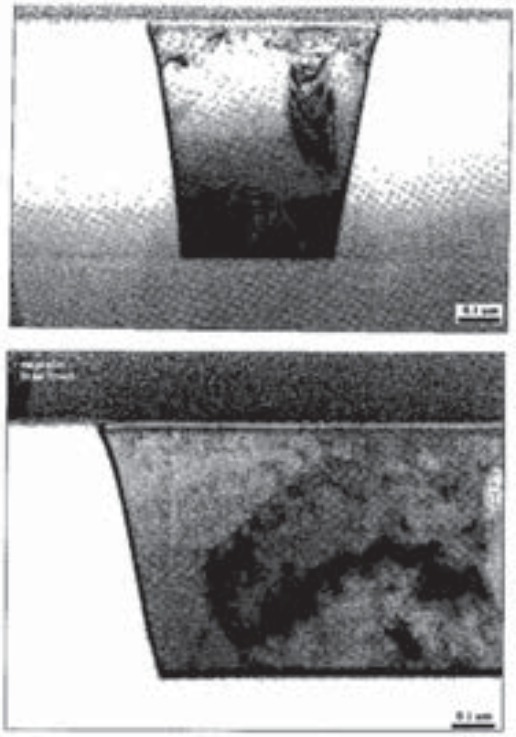


Fig. 15-11 SEM of an in-laid copper line in a dual-damascene interconnect-structure.[34] Reprinted with permission of Semiconductor International.

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 then used to remove the excess metal. A passivation/etch stop layer (silicon nitride) is deposited onto the top of the layer containing the grooves (trenches + vias) in the dielectric. This completely encapsulates the Cu with diffusion barriers. The entire interconnect structure is finally covered with an SiO_2/SiN passivation layer. Figure 15-11 shows an in-laid Cu-line in a dual-damascene interconnect structure.[34]

15.8 VIA-FIRST DUAL-DAMASCENE PROCESS WITH SiO_2 DIELECTRIC STACK

Many of the issues of forming the vias and trenches in the via-first process are common with the trench-first process and were described above. These include use of a single-damascene process to form the first layer of metal, and technologies for filling the damascene recesses with metal and polishing it back with CMP. We also assume that the via-first dual-damascene structure is being formed on the first level of metal, which has been covered with a SiN layer (that acts both as an etch-stop to the via etch and a barrier layer covering the top surface of the Metal-1 copper layer, to prevent its diffusion into the Si substrate).

However, there are also issues particular to the via-first sequence, and we will discuss them here.

For the purposes of opening this discussion, we assume that the dielectric stack consists of CVD-SiO_2/SiN/CVD-SiO_2. (The implications of using low-k material instead of SiO_2 in the stack were described in Sect 15-6.) The vias in this sequence are, of course, patterned first, using the via mask to align the vias to the Metal-1 lines produced previously (Fig. 15-6). The etch process used to open the via needs to be a multiple-step process because the first etch mechanisms must be *non-selective* (so that they can etch both SiO_2 and the embedded SiN etch-stop layer at about the same rates). However, the nitride capping layer at the bottom of the via should not be removed, and thus the etch mechanisms operating when the vias are close to being opened must be *selective* (to nitride).

After the via s are opened (but with the barrier-SiN layer at the bottom of the vias remains intact), the first layer of resist is stripped. Note that since SiO_2 is used in the dielectric stack, it can withstand the resist-stripping process (i.e., no extra hard mask is needed, as will be the case when most low-k materials are used in the ILD stack). The second layer of resist is then applied, and the trench mask is aligned to the opened vias. Note that the nitride at the bottom of the via must also remain intact throughout second (i.e., trench-etch) process.

Three methods are used to prevent removal of the nitride during trench etch:

1. The resist applied for the purpose of patterning the trenches will flood the open vias. If this resist is not exposed during the trench-patterning exposure step (i.e., because it is so deep beneath the top surface of the resist layer, that some may not develop after the exposure step. Thus, it remains to protect the via bottoms). However, due to pattern-density issues discussed later, such resist remaining in the vias may vary in thickness across the circuit. Thus, this approach will not always work.

2. A trench-etch process is used that is highly selective to the SiN that remains at the bottom of the open vias.

3. A bottom-anti-reflective coating (or BARC) is spun-on to partially (or completely) fill the open vias. This BARC remains after the resist is developed to protect the nitride at the via bottoms against the trench etch.[15] While this approach adds process control difficulties and process complexity (as the BARC must be applied, and separately removed later), this has been the method most widely adopted for preventing nitride-barrier attack during trench etching. We will continue the discussion of this approach in Sect 15.9.

In any case, the trench etch is carried out to create the trenches, stopping on the embedded SiN etch-stop layer. At that point the second resist layer is stripped.

After the second resist layer is stripped, the nitride at the bottom of the via must finally be opened. However, this procedure itself can cause some difficulties. First, the SiN-removal-step should be a "soft" etch. That is, the ions that bombard the surface being etched must not have enough energy to sputter Cu during the inevitable overetch time. If such sputtered Cu ends up on the via sidewalls it could be trapped on the wrong side of the subsequently deposited barrier-layer (exactly where you don't want it!). Another potential problem involves the possibility of slightly misaligning the via and the edge of the Cu film below. In this case a deep slit can be etched in the dielectric under the nitride. This slit would be difficult to fill with the barrier material, again leading to copper-diffusion problems. The way to avoid such slit-formation in the oxide, is to use a nitride-etch-process that is highly selective to oxide.

Another desired goal of the trench etching process is to reduce the handling of wafers between the various steps of the trench etching process (and the variability in waiting time between these steps). That is, the process actually contains four etch steps: **1**) organic ARC open (typically required for DUV resists); **2**) trench etch; **3**) resist strip; and **4**) nitride removal from the bottom of the via. Etching tools are now being made available that allow all four of these steps to be carried out in a single chamber, and use of such tools may be critical to achieving consistent device yield and lower overall cost-of-ownership.

15.9 PROCESS INTEGRATION ISSUES OF THE VIA-FIRST DUAL-DAMASCENE PROCESS WITH LOW-*k*-DIELECTRIC STACK

The process flow of forming each level of metal in a via-first dual-damascene interconnect system consists of the following steps:

- DEPOSIT Dielectric Stack (including the SiN barrier-layer over inlaid Cu);
- LITHO Step I (Via Patterning);
- Via ETCH;
- RESIST-STRIP I;
- Post-Via -Etch-Residue CLEAN;
- LITHO Step II (Trench Patterning);
- Trench ETCH;
- RESIST STRIP II;
- Post-Trench-Etch-Residue CLEAN;
- ETCH Barrier Film at Bottom of Via;

- Pre-Metal-Deposition CLEAN;
- Deposit Barrier Metal and "Seed" Cu Layer by PVD;
- Deposit Cu film by ELECTROPLATING;
- Post-Cu-Deposition ANNEAL;
- Remove Excess Cu/Barrier Metal (and do an SiO_2 Buff) by CMP;
- Post-CMP Residue CLEAN;
- CLEAN Top Surface of Cu Before Deposition of SiN Barrier/Etch-Stop Layer.

It can be seen that this process sequence contains: **1**) One Dielectric-Film-Stack Deposition sequence; **2**) Two Lithography processes; **3**) Two Resist-Strip processes; **4**) Three Etch processes; **5**) Five Cleaning processes; **6**) One CMP process; **7**) One PVD Deposition sequence; and **8**) One Electroplating (and Anneal) process.

Although many of these processes were mentioned earlier, there are some additional important process integration issues that should be discussed when considering the overall "via-first Cu/low-*k* dual-damascene" process flow. Thus, we will discuss these relevant process integration details here (Fig. 15-12). However, we will cover only those involved with Processes **2** through **5** above (since Process **1** is discussed in both Chap. 14 and in Sect. 15.6, and Process **6** is covered in Chaps. 8 and 16, and Processes **7** and **8** are described in Chap. 16).

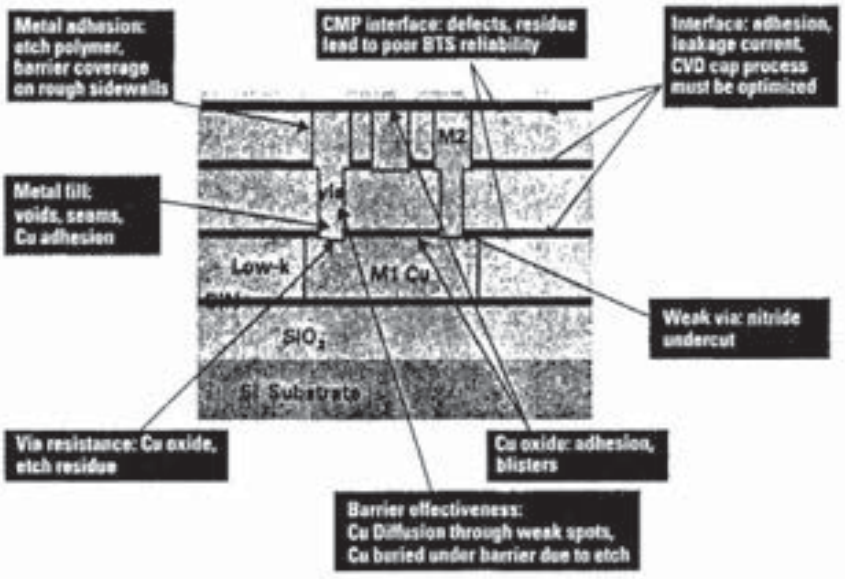


Fig. 15-12 Possible failure types associated with damascene processing pertain to film adhesion, barrier and metal fill capability, copper oxidation and CMP defects.

Besides the technical issues, there are those involving economics. In the face of increasingly difficult integration, the rise in the number of process steps per layer, and the 300-mm conversion, there will be the daunting economic challenges of containing interconnect processing cost, which in 2001 accounts for about 50% of the total wafer processing cost.[16]

15.9.1 Via-First-Damascene Integration Issues: Lithography

With the application of Cu/low-*k* dual-damascene integration, lithography and etch become more difficult because of the complex dielectric stack and intermediate topography. Here we consider the issues involved in carrying out lithography steps for the leading dual-damascene scheme, namely the *via-first approach*. There are two lithography processes needed in the via first approach, and each as its own challenges. They are: **1**) the *via-patterning lithography step*; and **2**) the *trench-patterning lithography step*.

Note that the issue of misalignment of the trench pattern to the via pattern is less severe for the via-first approach than for the trench-first approach. That is, as shown in Fig. 15-13b, even if the trench pattern and the via pattern are somewhat misaligned, since the via is already completely formed during the via etch process, it makes complete contact to the metal layer beneath it (provided it has been properly aligned to the underlying metal line). In the trench-first process, however, if the via is misaligned to the trench pattern, during the second etching step (via-etch) the entire via pattern will not be transferred to the bottom of the dielectric stack. This will cause the via resistance between the 2nd and first levels of metal to increase. Nevertheless, alignment of the via pattern to the underlying metal level below remain critical for both the via-first and trench-first processes, for reasons described in Sect. 15.9.2.3 (see also Fig. 15-13a).

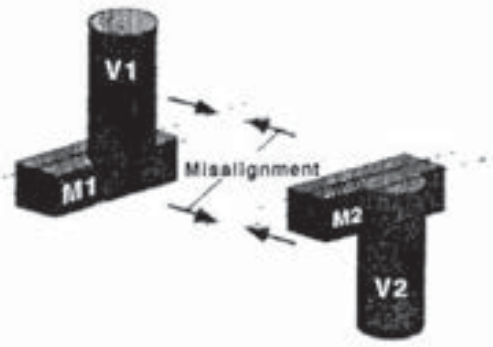


Fig. 15-13 Via-first damascene: a) Misalignment of the via to the underlying metal line is a problem; b) Partial misalignment of the trench to the previously formed via does not cause a problem because the via is completely etched through the ILD stack film.

15.9.1.1 Via-Patterning Lithography Issues: The major concern for the via-patterning lithography step is the ability of the resist to withstand the via etch process. This problem becomes worse as the vias are scaled because thinner resists have to be used to image smaller features, and the etch resistance of 193-nm resists is known to be inferior to that of 248-nm resists. The use of a bilayer resist process (as discussed in Sect. 15.9.1.3 below) may be required, since the imaging resist-layer contains silicon, making it more resistant to dry-etch attack than conventional DUV resists.

15.9.1.2 Trench-Patterning Lithography Issues - DUV Resist Poisoning: As is described in Chap. 6, DUV resists are sensitive to contamination by trace quantities of amines, a problem referred to as *DUV-resist poisoning*. While the most common source of such contamination is amines in the processing ambient, another source is amines in materials that are in direct contact with the resist. Some low-*k* dielectrics contain such amines (nitrogen compounds). For example, the SiOCH films (formed by PECVD using trimethylsilance and N_2O) entrap NH_3 products in the film as they are deposited. Such amines, if near the surface of the SiCOH films, can diffuse into and poison the DUV resist which is in direct contact with the dielectric surface. If an oxide hard mask is present on the top of the dielectric stack, the resist poisoning problem will not occur during the via-first etch. However, during the trench etch, the resist will come into contact with the exposed sidewalls of the SiOCH film under the hard mask, and resist poisoning can then occur during the trench-etch process. Performing an oxygen-plasma pre-treatment of the via sidewalls (perhaps as a part of the via-first resist stripping process) has been proposed as a way around this problem. Such a pre-treatment can leach out the amine compounds from the SiOCH surface on the sidewalls of the via. However, there is some concern that such a plasma pre-treatment may also take carbon out of SiOCH films and reduce their *k*-value.

As noted in Sect. 15.9.3, a plasma-treatment of the sidewalls of nanoporous low-*k* films with an NH_3 plasma has been proposed as a way to strengthen their surfaces against subsequent resist–stripping process of the via-patterning step. However, such a plasma-step can also trap amines in the film. These amines can later diffuse out and poison the resist during the trench-patterning step. (Figure 15-14 shows such a poisoned resist in the via after trench patterning.) Amines can readily diffuse in most low-density and nanoporous films, with diffusion lengths of hundreds of microns. The deposition processes, as well as etch and strip, must therefore be designed to avoid exposing the low-*k* film to any source of nitrogen. As mentioned above, if an SiO_2-capped low-*k* film is used, resist

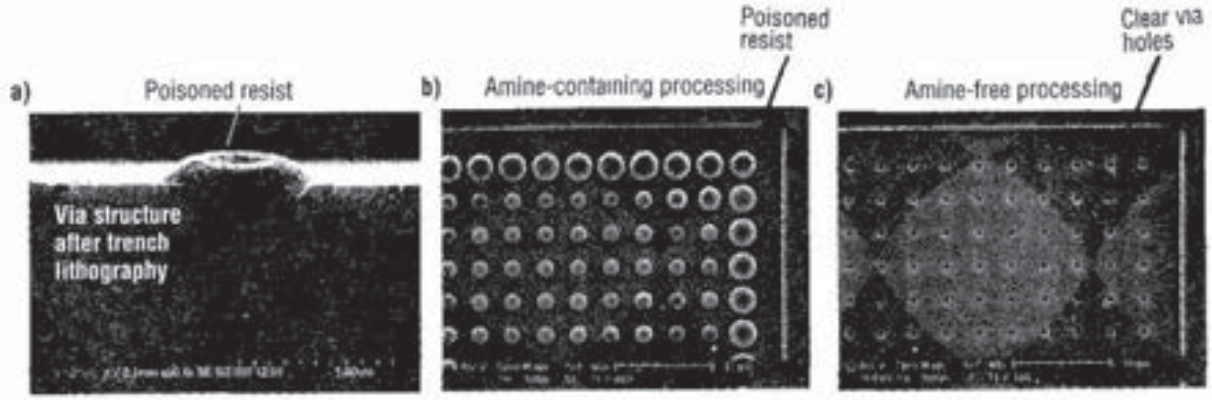


Fig. 15-14 DUV photoresist poisoning is indicated after trench patterning in a) cross-section, and in b) plan view. Poisoning was prevented in c) when the dielectric film was not exposed to nitrogen-containing plasma.[29]

poisoning may not be observed at the via lithography step, but poisoning can occur at the trench lithography step as a result of the trench-patterning resist coming in contact with exposed via sidewalls.[17]

15.9.1.3 Trench-Patterning Lithography Issues - Trench CD Control Issues: To be able to meet the stringent CD-control requirements (i.e., for trench width) for 0.18-μm technologies (and smaller), the resist film spun on a wafer should have the same thickness everywhere (to within ±100Å – see Vol. 1, 2nd Ed. Chap. 12). However, in the via-first damascene approach, since the resist floods the vias that were previously etched, the local thickness of the resist over the top surface of the oxide can be impacted by the presence of these vias. Namely, as shown in Figs. 15-15, and 15-16, the resist is considerably thinner in regions where a dense array of vias exists, compared to regions where only isolated vias exist. In addition, near the edge of such an array of vias, the resist thickness varies significantly, as also shown in these figures. Since higher resolution is generally

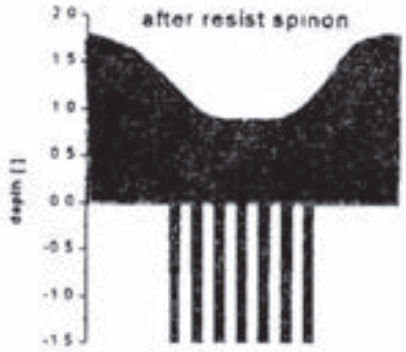


Fig. 15-15 Non-uniformity of resist layer within a chip due to pattern density variations.

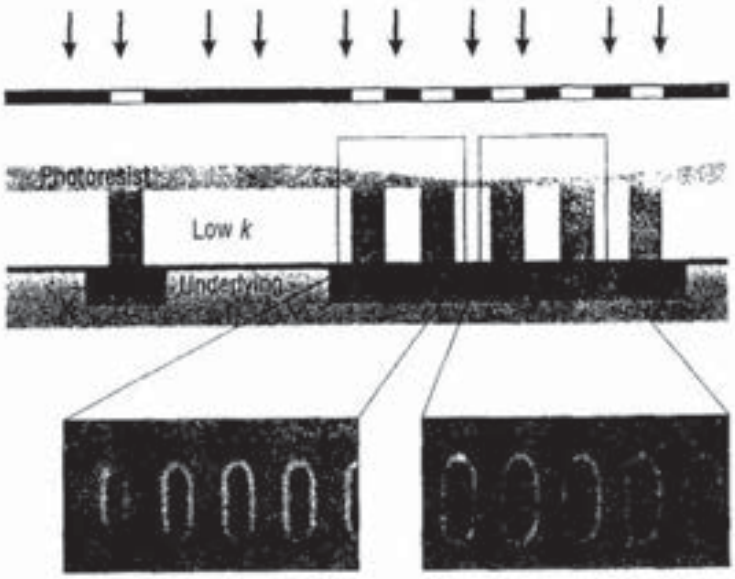


Fig. 15-16 Via-pattern density variations can lead to significant resist thickness non-uniformity during trench patterning in the via-first approach, and subsequent differences in trench widths. a) Trench widths vary at the edge of the via array. b) Trench widths do not vary at the center of the array of visas, where the local resist thickness is uniform.[20]

possible for thinner resist thicknesses, as the resist gets thicker the resolution degrades (e.g., the trench width in this case gets narrower). Figure 15-16 shows that the trench width in trenches patterned at the edge of an array of vias varies, compared to the trench widths patterned over vias at the center of the array. We will see in Sect. 15.9.1.5 that using bi-layer resist process (in which a planarizing underlayer produces more uniform resist thickness over the entire wafer) is a technique being explored to overcome this problem.

15.9.1.4 Trench-Patterning Lithography Issues - BARC Veils: Using a standard DUV lithography process for the trench patterning lithography step, a bottom antireflection coating (BARC) is applied. As noted earlier, the BARC serves two purposes. First, it improves pattern resolution by reducing substrate reflections. Second, the resulting partial fill of the vias with a BARC (Fig. 15-17b)[18] helps protect the SiN barrier layer from being attacked during the trench-etch step. This allows minimization of the barrier layer thickness. The BARC that partially fills the vias remains in place during trench etching and is removed later.

However, it is important to be able to control the depth to which the BARC fills the vias. (The BARC material that fills the via in this procedure is referred to a *via plug*.) That is if the via plug is not recessed below the buried SiN layer in



Fig. 15-17 Using a BARC to partially fill the vias prior to trench-PR deposition.[18]

the dielectric stack, "veils" (or "fences') may be created around the via opening at the bottom of the trench.

How such veils may arise is shown in Fig. 15-18. Here the BARC material (via plug) fills the via to a height that is above the top of the buried embedded-etch-stop SiN layer. As the trench is etched, the sidewalls of the via plug above the SiN layer are exposed to the etch ambient that removes the dielectric material to form the trench (Fig. 15-18a). The upwardly-protruding sidewalls of BARC can be chemically converted by an exposure to the plasma etching processes. This altered BARC-sidewall material may not be removable by the same resist stripping step that also removes unexposed BARC material. Thus, it may remain behind in the form of a "veil" surrounding the via opening. Figure 15-18b shows

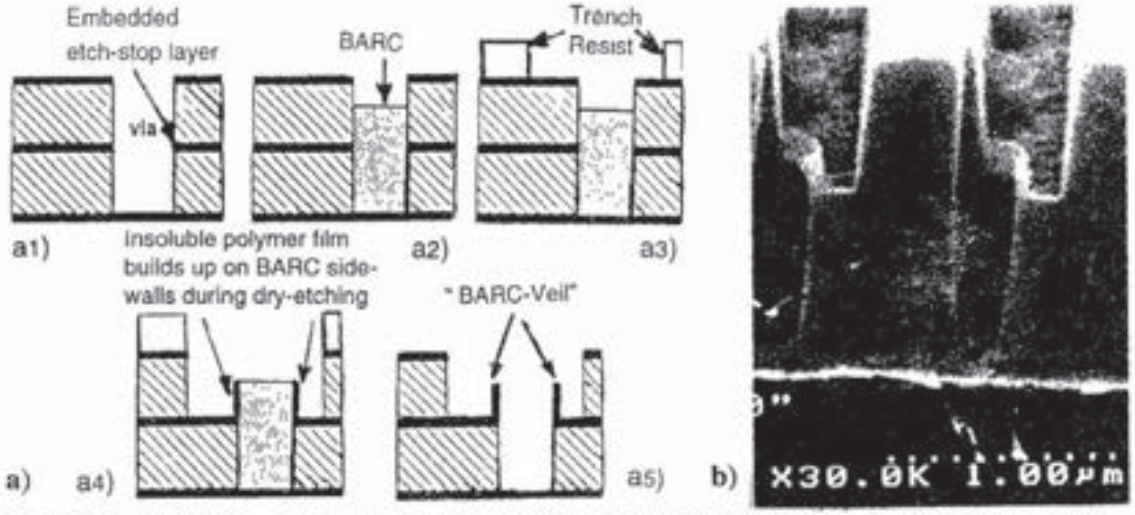


Fig. 15-18 a) BARC-veil formation mechanism in the Via-First Dual-Damascene process: a1) Etch Via, strip resist; a2) Via is filled with BARC to a level greater than that of the embedded etch-stop layer; a3) Deposit and pattern Trench photoresist; a4) Etch Trench. (*Note that during the dielectric etch step, neither the resist nor the BARC is attacked, BUT, an insoluble polymer builds up on the sidewalls of the BARC pillar that sticks up above the etch-stop layer.*); a5) After Trench resist and BARC are stripped, the insoluble film that was created during the dry-etch on the BARC-pillar sidewalls, remains behind as "BARC-Veils." b) SEM of BARC via-veils after trench-patterning and resist-strip.[18]

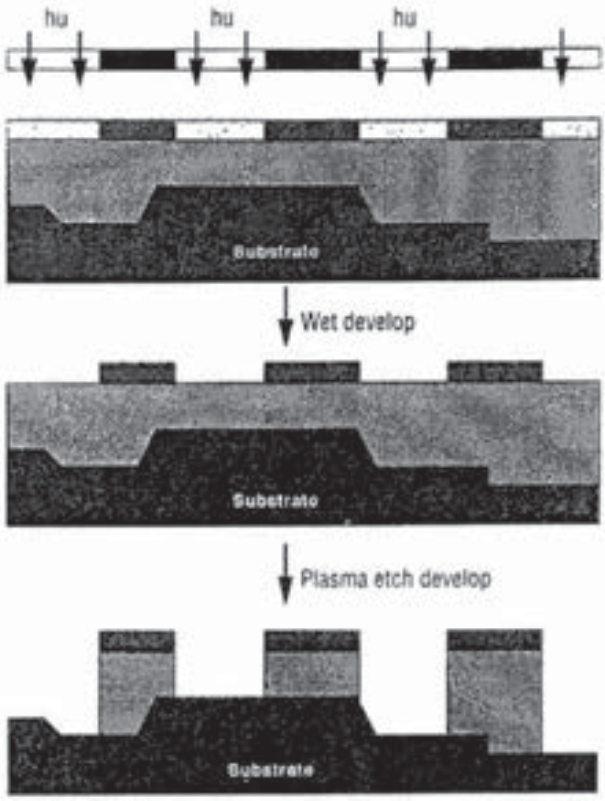


Fig. 15-19 TIS-2000 process flow with a thick planarizing underlayer and a silicon-containing imaging layer.

a SEM of such veils that remain in the trench around the via opening. Filling the trench/via openings with metal in the presence of such veils is problematic, and thus it is important that they be prevented from being allowed to form. To ensure that veils are not created by the trench etch step, it is necessary to control the degree to which the BARC fills the vias.

However, optimizing the BARC-fill of damascene vias is difficult, and often requires considerable process development and materials testing. It is possible to fill vias of different pitch equally by careful selection of the right BARC and coating conditions, but this only works if all the vias have the same size. A new BARC optimization also has to be done for each via diameter.

15.9.1.5 Bi-Layer Resist Process: To overcome the issues of resolution, critical-dimension control over arrays of vias, and the problem of via-plug depth-control that occur when the trenches are pattered in the via-first dual-damascene method, an alternative lithography approach has been proposed, namely a *bilayer, thin surface-imaging scheme* (Fig. 15-19) This method is based on a bilayer resist system developed by Arch Chemicals, called TIS-2000. The technique uses a

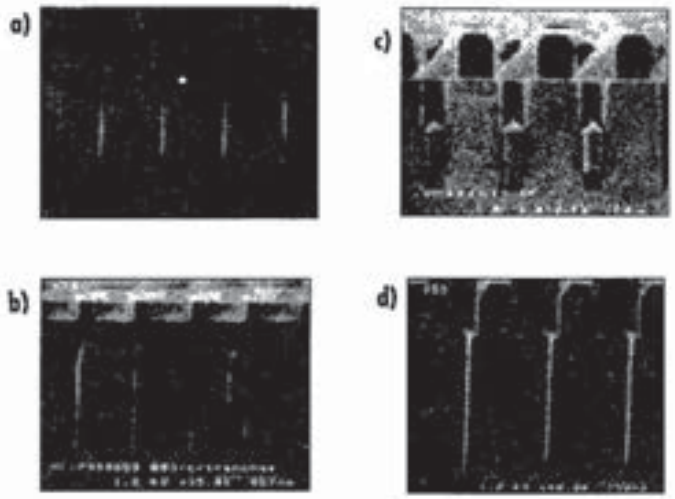


Fig. 15-20 SEMs of cleaved dual-damascene structures patterned with the TIS-2000 process: a) before exposure with vias filled; b) after wet development; c) after partial etch of underlayer and trench; d) after full underlayer etch (top resist still in place).[19]

thin silicon-containing top-imaging layer on top of a highly planarizing fully absorbing underlayer (chemically similar to a novolac resist). The processing sequence is similar to a process with an organic BARC. While this technique was developed for 248-nm lithography, it appears that it can be extended to 193-nm lithography (by using an alternate imaging layer and underlayer).

The thin imaging-layer on a planarized layer allows for very high resolution. The silicon present in the imaging-resist-layer also gives it a very-high etch resistance when the underlayer is dry-etched (developed). The underlayer effectively planarizes the topography of the via-containing dielectric stack (Fig. 15-20a).[19] In this figure (in which the underlayer fills 0.25-μm vias), the vias are completely filled and the top of the underlayer is flat without any topography caused by the vias underneath. The thickness of the underlayer over the substrate in the via-array region is about 350-nm, which is enough to insulate the imaging layer from any optical effects by substrate reflections (i.e., the underlayer also acts as an effective BARC layer). Figure 15-20b shows an array of vias after the *imaging layer* has been exposed and wet developed.

The next step is a dry-develop process done with an SO_2/O_2 plasma. This transfers the pattern created in the imaging-resist-layer to the underlayer. Dry-developing the underlayer to the dielectric surface below defines the trench pattern. At this point the vias are still completely filled with the underlayer material. If we stopped the develop step at this point, and proceeded with the etch step to remove the dielectric material which creates the trench, veils would again arise (in the same way they occur if the via plugs made from BARC material

filled the vias too full). Thus, the dry-develop step is extended so via-plugs (now consisting of underlayer material instead of BARC) are recessed. (Since trench open-areas are large compared to via areas, an optical endpoint signal can be used to determine when the recess etching begins, providing good control.)

Figure 15-21 shows the patterned trench-resist and the recessed via-plug (whose level is below the top of the buried-nitride etch-stop layer). When the trench dielectric is now etched, no veils remain as residue within the trenches. Figure 15-22 shows the trench and via after trench-etch and resist-stripping. The trench and via profiles are vertical and veils are not present. Since the via-plug covers the bottom SiN barrier-layer during the trench etch, the trench-etch process does not have to be made selective to the nitride. This also allows a trench etch process with higher etch rates to be used.

Comparison of trench lithography at the edge of large via arrays shows another advantage of the top-surface bilayer-imaging technique. As is shown in Fig. 15-23, while trenches over vias at the edge of dense array vary considerably in width for standard DUV lithography (Fig. 15-23a), those patterned with TIS-2000 show no variation in width.[20] This is because the underlayer of the TIS completely fills the vias. Thus, the resist thickness is constant and there is little (or no) trench-width variation.

15.9.2 Via-First-Damascene Integration Issues: Dielectric Etching

The plasma etching and resist/polymer strip steps are among the most critical for the successful integration of a *low*-k dielectric. The etch process must be able to

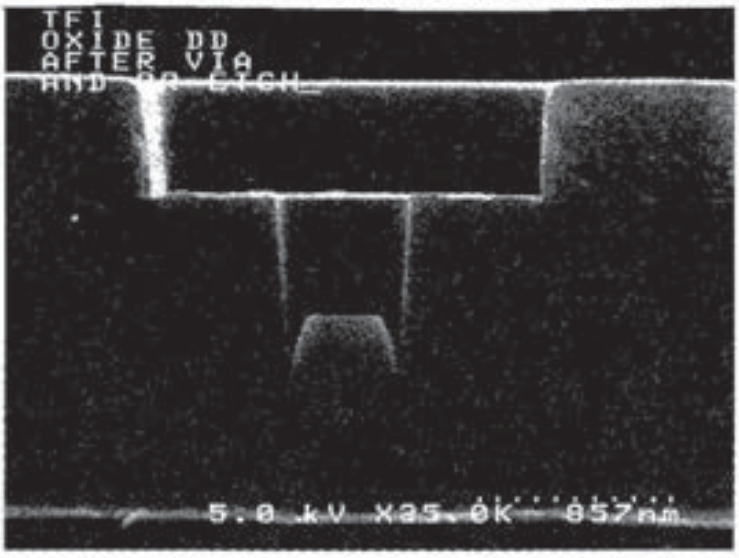


Fig. 15-21 Trench patterning using TIS-2000 bilayer resist. The dry-developed underlayer with recess etch brings the level of the via plug along the sidewalls below the nitride etch stop, which is important for the subsequent trench etch.[18]

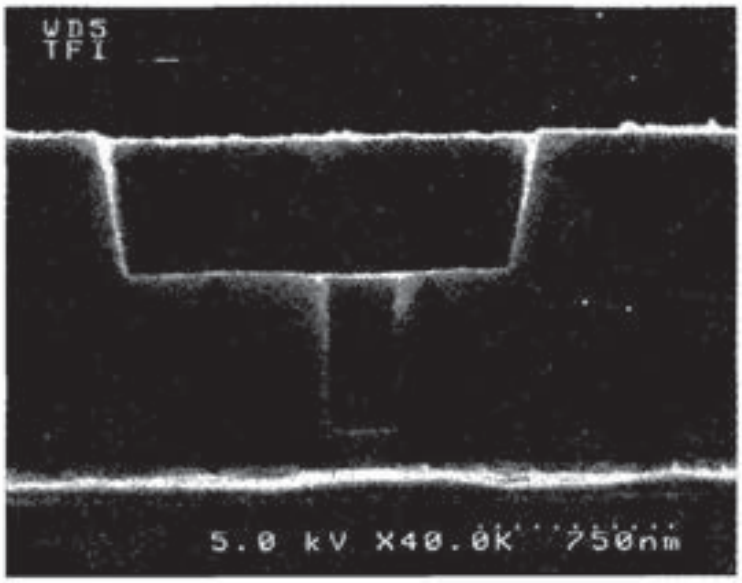


Fig. 15-22 Fully etched and stripped trenches over a via using the TIS-2000 bilayer resist process. The dual-damascene structure is nearly ideal in that there are no veils or faceting of the via due to trench etch.[18]

effectively perform the following tasks: **1**) Remove multi-layers of dielectric; **2**) Maintain a notch-free profile for optimum subsequent metal deposition; **3**) Achieve a high selectivity to resists (which is especially important as resists are getting thinner and 193-nm resists are less resistant to plasma etching than those used for 248-nm lithography); **4**) Retain the low-k value during the etch and strip processes so that ultimate device speed is not compromised; **5**) Achieve high selectivity to etch-stop layers, (especially since we want to make them as thin as possible to minimize the stack's effective k-value); **6**) Minimize microtrenching at the etch front; and **7**) Minimize aspect-ratio-dependent etch (ARDE) effects.

15.9.2.1 Via Etching for Low-k Processes: In the "via-first" dual-damascene process sequence, the vias are patterned and etched before the trenches. The via-

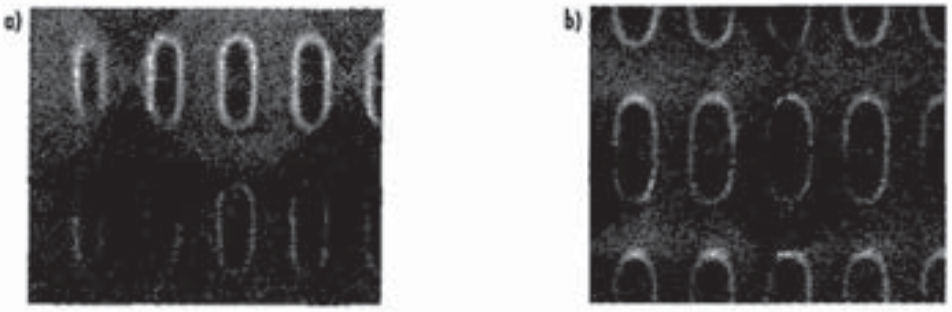


Fig. 15-23 Trenches over vias near the edge of a dense array: a) trench widths vary considerably for standard DUV lithography; b) the same area with TIS-2000 resist.[18]

etch process has some challenges, one of which was described in Sect. 15.9.1 (the resist must be able to withstand the deep via-etch process). To open the via, the following materials must be etched by the via-etch process in a low-k dielectric stack: **1**) BARC; **2**) SiO_2-hard-mask; **3**) upper low-k dielectric layer; **4**) thin SiN (or SiC) embedded etch-stop layer; and **5**) lower low-k dielectric layer. The etch step itself will therefore have to employ a multiple sub-steps (Fig. 15-17a).[18] The first such sub-step should be a non-selective etch, so that the BARC, SiO_2 hard-mask layer, low-k dielectric layers, and the embedded-etch-stop layers are all etched at about the same rate. This would ideally produce a via profile that is free of notching and bowing. This first sub-step would be terminated prior to removing all of the second low-k dielectric layer. The etching sequence would then switch to the second sub-step. The etch characteristic of this second-sub-step would be that it is highly selective to etching oxide and not nitride. This would allow it to remove the remaining lower low-k dielectric material from all of the vias without attacking the SiN etch-barrier layer at the bottom of the vias (i.e., a sufficiently long overetch time would be permitted to ensure that the lower low-k material was entirely removed from all of the vias).

While the above process is the desired goal, an even more complex etch process might be needed to produce a via-profile free of notching and bowing. That is, the etching characteristics of various low-k dielectrics are different from SiO_2, regardless of whether a spin-on organic-based material (such as SiLK® or FLARE®), or an SiCOH-CVD dielectric (such as Back Diamond,® Coral,® or Flowfill®) is used.

For example, a three-sub-step etch process might be needed for a SiO_2-hard-mask/SiLK stack. That is, fluorocarbon chemistries with high bias power might be used for the first-sub-step to etch the top SiO_2-hard-mask material and the embedded SiN layer (because this process can disrupt the strong Si-O and Si-N bonds), but when the SiLK layers are etched, a second-sub-step using oxygen as the active etchant is used. However, in an HDP tool, the high degree of gas-phase dissociation leads to high concentrations of atomic oxygen, which may produce some bowing of the via profile. Also, hard-mask undercut after via (and trench) etching can occur (Fig. 15-24), which would make it more difficult to achieve void-free metallization.[21]

When the low-k dielectric is a SiCOH-CVD film (having a higher porosity than normal CVD-SiO_2 films), the gas mixture for etching these low-k materials might be mixtures of CF_4 (the principal source of F atoms) and CH_2F_2 (a polymerizing gas for improving selectivity and providing sidewall passivation). Using such an etch procedure has reportedly allowed 0.3-μm vias to be etched in



Fig. 15-24 Etched via profiles in SiLK® low-k dielectric showing hard-mask undercut.[21]

such stacks with the desired via profiles (Fig. 15-25).[22] The choice of gas mixture depends on the precise application being considered. That is, since these films contain some C and H, if their composition is changed (i.e., the more C, the lower the k-value), it will be necessary to re-tune the etch process in order to meet the selectivity and etch requirements.

15.9.2.2 Trench Etching: Following the via-etch-step in via-first, dual-damascene

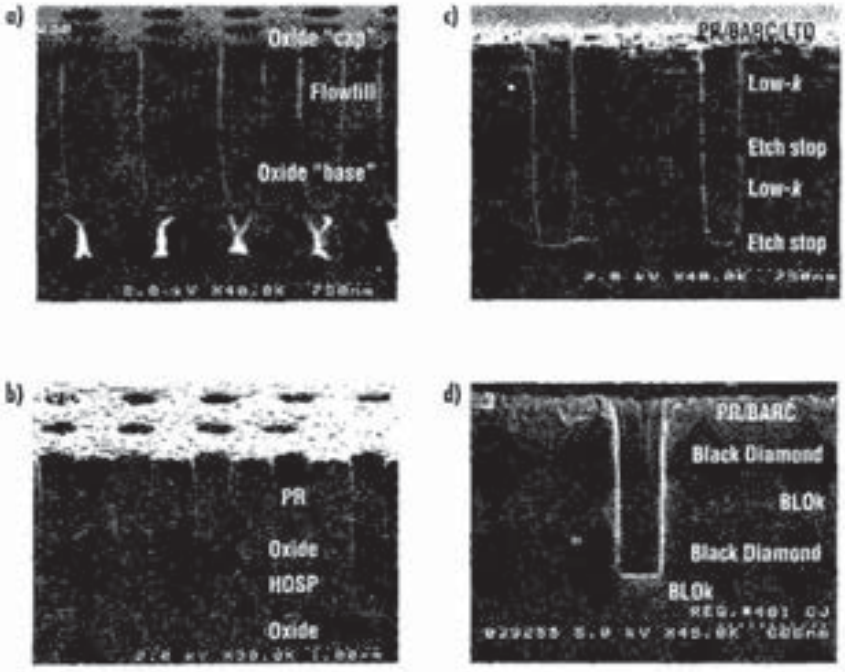


Fig. 15-25 Etching of: a) 0.3-μm vias through Trikon's low-k Flowfill (k = 2.9); b) 0.4-μm vias through Honeywells HOSP (k ~2.5); c) 0.28-μm vias through a Trikon dual-damascene stack (k ~2.8); and d) 0.29-μm vias through a dual-damascene stack containing Applied Materials Black Diamond film (k ~2.6) and BLOk (k ~4.5) etch stop.[22]

applications, trenches are etched down to the buried etch-stop layer. The issues involving protecting the barrier nitride at the bottom of the via (exposed during the previous via-etch step), from being etched during the trench-etch-step were previously considered in Sect. 15.9.1.

When a 2nd generation low-*k* dielectric is used, the top surface of the stack generally employs an SiO_2 hard-mask and a BARC to improve lithography resolution. Thus, the trench-etch process is also a multi-step procedure, involving: BARC-open; SiO_2-hard-mask etch; and low-*k* etch (down to the embedded etch-stop layer - Fig. 15-17b). Following the trench etch, the resist must be stripped, the BARC removed from the vias, and the barrier layer at the bottom of the vias etched away.

In early via-first, dual-damascene process flows, these steps were performed in several discrete tools. However, in the interests of improving cycle time (less handling and waiting), tool vendors are seeking to provide equipment that will allow an integrated-trench-etch process. That is, BARC opening, SiO_2-hard-mask etch, low-*k* dielectric etch, resist/BARC strip, and SiN barrier-layer open would all be carried out *in situ* in a single tool and a single process run.

15.9.2.3 Etching the SiN Barrier Layer at the Bottom of Vias: After the trench is completed and the second resist layer stripped, the nitride at the bottom of the via must be opened. As noted earlier, this etch process can cause some problems. First, the SiN etch step should be a "soft" etch. That is, the ions that bombard the SiN being etched should not have enough energy to sputter Cu during the inevitable overetch time. If such sputtering causes Cu to end up on the via sidewalls, it could be trapped on the wrong side of the subsequently deposited Cu-barrier-layer. Thus, any such sputtered Cu would have to be removed before the barrier metal was deposited. While a conventional dip in dilute HF is efficient for sidewall Cu contamination cleaning, this inevitably also cause some CD loss.

Another problem arises if the via is misaligned with the edge of the Cu line beneath it. In that case, during any overetch time of the nitride, a narrow slit would be etched in the dielectric alongside the Cu line (see Fig. 15-26).[31] This slit is hard to fill with barrier material, again leading to unwanted Cu diffusion. This issue can be addressed by employing a nitride etch process that is highly selective between nitride and oxide, and between nitride and Cu.

15.9.3 Via-First-Damascene Integration Issues: Resist Stripping

With the introduction of low-*k* and ultra-low-*k* porous dielectrics, existing resist-strip technology cannot be used without at least some modifications. Dry-etching in an oxygen plasma has become the most widely used stripping technology

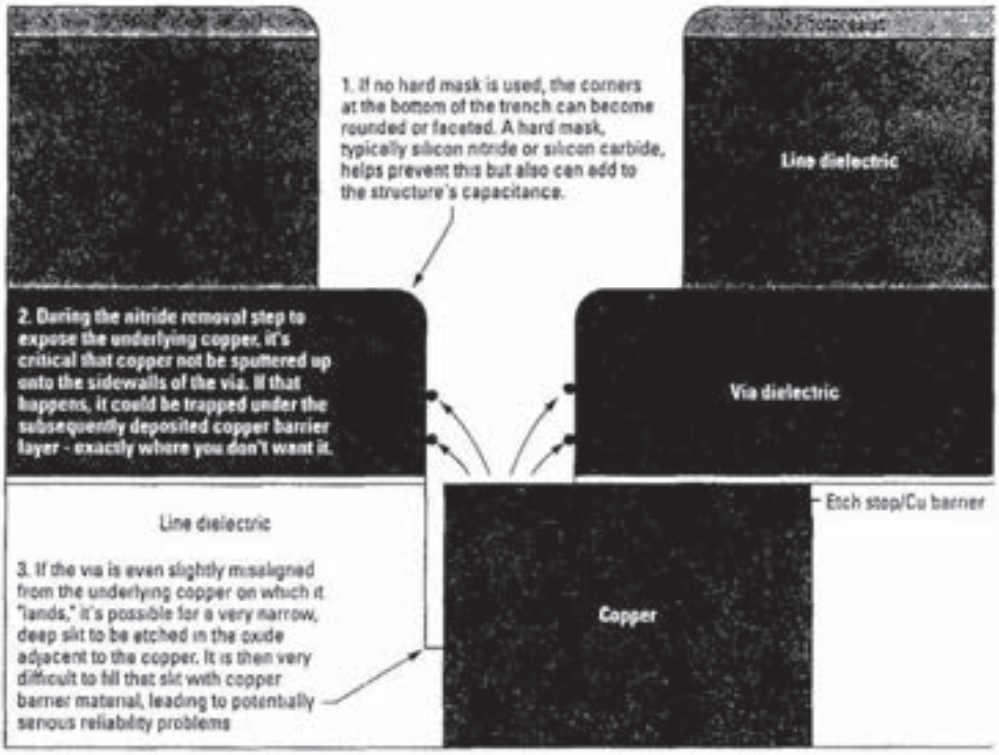


Fig. 15-26 Common problems in the via-first approach include rounding of the corners at the bottom of the trench; sputtering of the Cu onto the via sidewalls during nitride removal; and hard-to-fill slits etched next to the underlying Cu in the case of misalignment.[31]

(especially for back-end-of-line [BEOL] processes). However, many of the low-k dielectric materials are also etched by oxygen plasmas, and in some cases their k value may be degraded by reactive oxygen atoms as well. Here we discuss the issues facing resist stripping in such applications.

To begin, as noted earlier, SiO_2-hard-masks will typically be used on the top of the low-k dielectric stacks. These layers will actually serve multiple functions, including: **a**) as a hard-mask during via- and trench-etching; **b**) as a polish-stop layer for the Cu-CMP process; and **c**) as a dielectric anti-reflective coating (DARC). The use of a hard mask is mandatory if organic low-k films are used (since they are attacked by the same chemistries used when stripping the resists after an etch step). Even with a hard mask, however, difficulties arise when etching organic low-k materials. That is, if an etch-process is used that is based on an O_2 plasma, very reactive oxygen radicals are present in the plasma. These isotropically attack organic low-k dielectrics, causing bowing of the via and/or trench profiles (as shown in Fig. 15-27).[23] To stop such attack of the dielectric layer, sidewall-passivating gases must added to the O_2 etch gas. In one paper,

ethylene was added to the O_2 etch gas when an organic low-k material was being etched. Undercutting of the hardmask was prevented, and good CD control was maintained, as the hardmask itself was not eroded.[24]

Even if an HSQ or MSQ-based low-k material is used instead of an organic film (see Chap. 14), the resist stripping process is still a challenge. That is, while the resist is being stripped, the oxygen radicals will react with the exposed low-k material of the trench sidewalls. These reactions convert the Si-CH_3 bonds in the MSQ, and the Si-H bonds in HSQ films, back to Si-O bonds. This causes the k-value of the dielectric to rise back toward that of SiO_2 (~4.0). As a result, an alternative resist stripping process is needed. That is, an alternative anisotropic ashing process must be available.

One method of making low-k materials more robust against attack by oxygen radicals is *plasma treatment of the exposed low-k material surfaces.* An example of such a plasma treatment was reported to have carried out on an organic low-k material. After exposing this dielectric to an NH_3 plasma for 10 minutes, a nitrided layer is formed on its surface. The nitrided low-k layer exhibited lower leakage current after an ashing plasma exposure than did the untreated material. The film also acted as a better barrier against Cu diffusion.[25] In another paper, three such NH_3-plasma treatments in were performed on a dual-damascene via-first etched dielectric stack consisting of HSQ/PECVD-SiO_2/HSQ. The first treatment was carried out prior to depositing the embedded etch-stop PECVD-SiO_2 layer on the first HSQ segment of the stack. This improved adhesion of the embedded etch-stop layer to the HSQ. The second treatment was carried out after the second HSQ layer was deposited and the trench etch was completed. This purpose of this treatment was to strengthen the exposed HSQ sidewall surfaces against erosion during the resist stripping step.[26]

Carbon-doped low-k materials such as Coral® and Black Diamond® are ozone



Fig. 15-27 Cross-section SEM picture of single damascene structure made in 3M's low-k dielectric after strip and etch in a low-pressure, low-temperature strip.[23]

resistant. Resist stripping in processes that use these dielectric materials might potentially be carried out with a DI ozone strip process.

15.9.4 Via-First-Damascene Integration Issues: Cleaning

Cleaning issues are critical to producing dual-damascene interconnects with high yield and reliability. The cleaning challenges are especially severe for dual-damascene structures compared to front-end-of-line (FEOL) processes, because more aggressive chemicals can be used on silicon and thermal oxides. That is, the strong acids and bases in the FEOL flow effectively remove contamination and resist residues after critical FEOL steps. However, metals in back-end-of-line (BEOL) cannot be subjected to such chemicals since they are susceptible to chemical corrosion. Fortunately, the CMP process used in dual-damascene interconnects tends to remove particles on the dielectric surface during polish, another inherent advantage over the subtractive interconnect approach. However, other residual films must still be removed, as will be discussed here.

15.9.4.1 Post-Dielectric-Etch-and-Resist-Strip Residue Removal: After photoresist is stripped with a dry-stripping step (ashing), residue from the etch and the ashing step often remains on the via and trench sidewalls, and must be removed. Typically this is done with wet-chemical residue cleaners, in a separate cleaning step after the dry-strip step. One goal (as a way to improve cycle time), is to integrate these two steps into a single tool.

As long as SiO_2 was the material used in the dielectric stack, the post-etch residue removal step was easier, and proprietary organic solvents sold by such vendors as EKC Technologies, ASCI, J.T. Baker, and Olin could effectively remove the residues (Fig 15-28). However, with the introduction of low-k dielec-

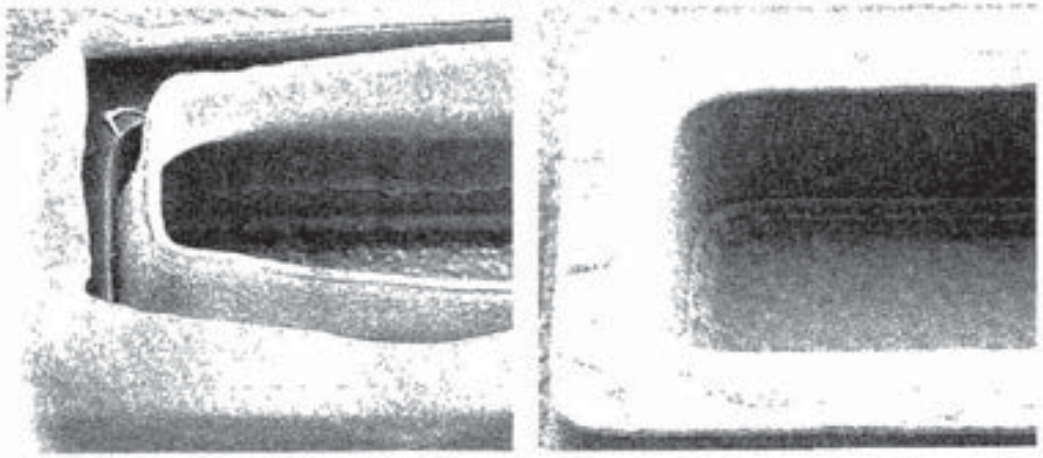


Fig. 15-28 Aqueous residue removers specifically target organometallic and inorganic residues left after ashing. Courtesy of J.T. Baker.

trics, the situation has changed. New cleaning chemicals must be used, and suppliers are working on chemistries for residue cleaning in Cu/low-k applications. However, a major issue in pursuing this goal is that it has not yet been decided exactly which low-k materials will be used for this application, and what are all of the material properties of the candidate materials. In addition, it appears that there will be a continuous transition from one material to another over the next few years, as ever lower-k materials are integrated into damascene structures. However, some general trends can be foretold.[27]

First, it must be ensured that residue-removal chemistries are compatible with specific the low-k dielectric and the metallurgy on the wafer. This is complicated by the lowering of the porosity of the new low-k materials and the thinning of the dielectric layers. Furthermore, since each low-k choice requires different plasma etching processes and chemistries to etch vias and trenches into the low-k materials, the nature of the residues can be different. Thus, no one product will likely to be able to meet all Cu/low-k applications. Nevertheless, it appears that some form of wet cleaning after etch and ashing will still be needed until below 0.13 μm. Fundamental barriers to wet cleaning approaches do not appear to be reached until well below 0.1 μm. At some point, however, there are concerns about liquids being able to go into and out of small, high-aspect-ratio features. At that time a dry-approach (or the use of supercritical fluids – see Sect. 15.9.4.5) may need to be used.

15.9.4.2 Pre-Metal-Deposition Clean: The pre-metal-deposition clean is carried out just before the barrier layer metal and Cu seed layers are deposited. The Cu film at the bottom of the vias must be free of organic residues and any native Cu oxide in order to ensure that low-resistance contact between the deposited barrier metal and the underlying Cu is fabricated with high yield. This clean step normally involves sputter etching of the wafer in the same tool in which the barrier layer metal and the seed Cu layer are sputter-deposited. Argon ion gas is used, and a dual-frequency rf sputtering module with a high-density plasma (HDP) source is required (so that a high degree of directionality can be obtained with the sputter-clean process).

During this pre-metal-deposition cleaning process, copper ejected from the lower layer can be re-deposited onto the via sidewalls, increasing the risk of line-to-line leakage. By 1999, however, the magnitude of this problem had not yet been quantified. It is hoped that the quantity of copper re-deposited will not be sufficient to degrade device performance. Nevertheless, if the problem becomes worse than expected, alternate via-clean methods will need to be developed that

selectively remove the etch residue and copper oxides without removing any copper.

Another special problem occurs when trying to integrate Cu with a fluorine-containing dielectric material, such as FSG. That is, the F bonds in many FSG films are reactive (particularly with water), and they can be broken during other process steps to produce free fluorine. At the interface between the FSG film and the barrier metal, Ta can react with such free fluorine to form volatile by-products at elevated temperatures. The resulting weakened interface can fail when subjected to the stress of an overlayer (such as that produced by the SiN diffusion barrier layer). A number of solutions to this problem have been proposed, including: use of a TaN barrier material (which is less reactive than Ta); using a more stable FSG film; or using a hydrogen-plasma treatment before depositing the Ta to remove mobile fluorine from the surface.

15.9.4.3 Post-CMP Clean: The post-CMP clean step is a critical one for Cu CMP. The bevel and backside of the wafer will most likely be exposed copper-containing solutions during the CMP step. Any copper remaining on the bevel or backside of the wafer risks contaminating other tools within the fab. As in the electroplating process, simple cleaning methods are available to remove residual copper, but the risks associated with copper cross-contamination demand careful monitoring of the post-CMP clean (see also Chap. 8).[28]

15.9.4.4 Clean of Top-Surface of Cu Prior to Diffusion-Barrier-SiN-Deposition Process: After the CMP step of the metal, and post-CMP clean, a SiN layer is deposited on top of the damascene structure. As noted earlier, this SiN film serves as barrier to keep the Cu from diffusing into the dielectric layer above it, and also as an etch-stop layer for the via etching step. However, the SiN may fail to adhere to the Cu surface if that surface is not properly cleaned. That is, the SiN may not adhere to the Cu interface if even a small amount of residue is left on the Cu surface following CMP. It turns out that there are two classes of such contaminants that must be removed from the Cu-surface: 1) residual organics that are present in the Cu-CMP slurry; and 2) native oxide that has formed on the Cu surface.

The residual organics can be removed by a pre-treatment with an *in situ* exposure to an aggressive hydrogen plasma just before depositing the SiN layer. Another technique is to remove such organics (corrosion inhibitors, usually BTA) from the CMP process. In addition, Cu forms a native oxide that can interfere with SiN adhesion. That is, if this oxide is not completely removed from the Cu surface, the Cu/SiN interface will be weak. Thus, a cleaning step

(consisting of a copper oxide reduction step using the same hydrogen plasma used to remove the residual organics) is recommended.[29]

15.9.4.5 Supercritical Fluids as Wafer Cleaners:* The temperature at which the vapor pressure above a pure liquid reaches one atmosphere is known as the normal boiling point – for water it is 100°C. In an open container, liquid water's temperature cannot be raised above 100°C because it would cause the vapor pressure to rise above one atmosphere, exceeding ambient pressure conditions. However, in a sealed container, liquid water can be heated to higher temperatures. As the water is heated, its density decreases (through thermal expansion), and simultaneously, the vapor density in the container increases.

This heating can be continued until the liquid's density has been so reduced – and that of the vapor so increased – that both densities are equal. This point is known as the *critical temperature*, and the pressure of the vapor at that point is called its *critical pressure*. Since the temperature and density of the water inside the sealed container are equal everywhere, thermodynamics dictates that the critical pressure inside the container also be equal everywhere. A fluid brought to conditions above its critical temperature and pressure (the *critical point*) is called a *super-critical fluid*.

Carbon dioxide (CO_2) is a material that is particularly attractive as a practically useful super-critical fluid. This is because it is non-toxic, non-flammable, and inexpensive. In addition, its critical conditions are easily achievable with existing process equipment. This unique set of characteristics has led workers to examine the possibility of using supercritical CO_2 instead of other solvents in such semiconductor processes as cleaning and photoresist stripping.

The attractiveness of using supercritical CO_2 as a replacement for traditional solvents in cleaning and stripping applications stems from the combination of gas- and liquid-like properties that it possesses. Table 15-1 compares the physico-chemical properties of a typical organic fluid in it's liquid, supercritical, and gaseous states. It can be seen that the supercritical fluid state possesses *gas-*

Table 15-1 PHYSICO-CHEMICAL PROPERTIES of TYPICAL ORGANIC FLUIDS

	Diffusivity (cm^2/sec)	Viscosity (mPa-sec)	Density (kg/m^3)
LIQUID	10^{-5}	1	1000
SUPERCRITICAL FLUID	10^{-3}	10^{-2}	700
GAS	10^{-1}	10^{-2}	1

* Written by James Rubin, Los Alamos National Laboratories. The material originally appeared in Semiconductor International. Reprinted with permission.

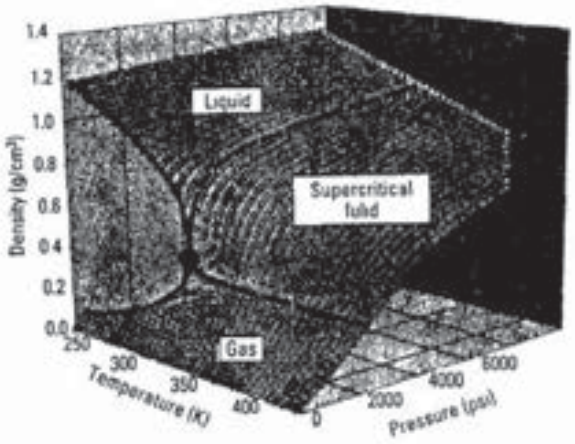


Fig. 15-29 Pressure-temperature-density surface for pure CO_2. The critical point is the small circle at the intersection of the contour lines.

like values of viscosity and diffusivity, and *liquid-like* density.

Figure 15-29 shows the density-pressure-temperature surface for pure CO_2. The critical point for pure CO_2 is 31°C (304K) and 1072 psi. To a first approximation, a fluid's solvent power is proportional to its density. Thus, high liquid-like densities achievable in supercritical fluids allow for liquid-equivalent solubilities. Figure 15-29 also indicates that small changes in pressure or temperature near the critical point result in large density changes. This implies that the density – and hence the solvent power – of a supercritical fluid is "tunable."

Figure 15-30 is a schematic pressure-temperature phase diagram for CO_2, incorporating the process flow for a closed-loop supercritical fluid treatment system. No uncontrolled waste streams can exit such a closed-loop system. All materials solubilized in the supercritical fluid are retained in the separation vessel

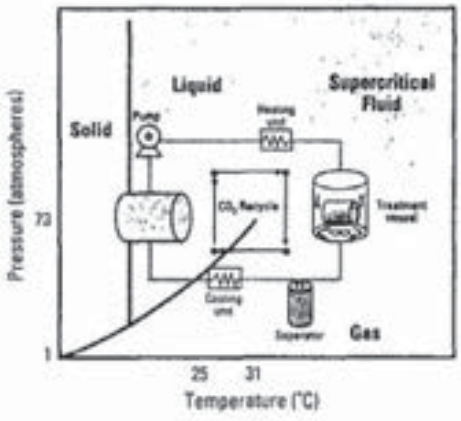


Fig. 15-30 Pressure-temperature phase diagram of pure CO_2 with a superimposed flow diagram for a closed-loop supercritical fluid treatment process.

for analysis, recycle, treatment, and/or disposal. In addition, these materials are concentrated in the separator, greatly reducing waste volume.

Several companies are working on commercializing tools that employ supercritical fluids in cleaning processes, including Supercritical Systems (a TEL-America company), and GT Equipment Technologies.

15.10 VIA-FIRST DUAL-DAMASCENE DIELECTRIC STACK WITH NO EMBEDDED ETCH-STOP LAYER

Eventually, the increase in the *k*-value of the damascene dielectric stack due to the embedded etch-stop layer (whether it is composed of SiN or SiC material) becomes too great. At that point, it may be necessary to use a damascene dielectric stack without an embedded etch-stop layer. Some predict that this approach will be needed for the 0.1-μm technology node.

The elimination of the etch-stop layer, however, produces some significant process integration issues. First, since there is no etch-stop layer, etch systems will have to have extremely good control of uniformity to ensure all trenches are of the same depth across the wafer, and from wafer-to-wafer. This could be difficult as there is no endpoint indicator, and the depth of the trench can only determined by the time of the etch step. Making the problem even more difficult is that low-*k* materials may not lend themselves to that kind of control. Second, if a hard etch-stop layer is not present, the top corner of the via can become rounded during the trench etch process (due to faceting, see Vol. 1, Chap. 11 - Physical Vapor Deposition, and also Fig. 15-26). A second issue involves the creation of microtrenches at the corners of the trenches themselves. This phenomenon is likely to be caused by enhanced ion-bombardment-flux at the corners of the trench (due to ions reflecting at shallow angles off of the sidewalls). Such microtrenches would provide locations that are poorly covered by the Cu diffusion barrier films. This could give rise to Cu diffusion through the barrier layer at such points where the barrier layer is thinned (or is even discontinuous).

15.11 PROCESS INTEGRATION ISSUES OF DUAL-DAMASCENE DIELECTRIC STACK WITH ULTRALOW-*k* DIELECTRICS

If damascene dielectric stacks are fabricated with ultra-low-*k* materials, additional process integration problems are encountered. That is, special techniques must be developed to protect these porous, less-robust materials during ashing and CMP. Two novel damascene processing steps have been proposed to handle such fragile films. The first is termed *half-via etch*. In this approach, the via is again etched first, but only half of the via is opened with this

process, down to the depth of the trench. Then the rest of the via is opened during the trench-etch step. This provides a simpler etch and alignment process and makes the lithography problem of resolving small feature sizes easier.

Another change may involve returning to the trench-first etching sequence. With a porous material, the via cannot be etched first because, after the via is etched and you begin etching the trench, problems will arise because the resist that flooded the vias will begin to be absorbed by the porous low-*k* dielectric.

REFERENCES

1. P. Singer, "Making the Move to Dual Damascene Processing," *Semiconductor International,* August 1997, p. 79.

2. B. Luther *et al.,* "Planar Copper-Polyimide Back End of the Line Interconnections for ULSI Devices," *Proceedings of the 1993 VMIC Conf.,* p. 15.

3. R. Iggulden *et al.,* "Dual-Damascene Aluminum for 1-Gbit DRAMs," *Solid State Tech.*, November 1998, p. 37.

4. P. Singer, "IBM to Use Dow's SiLK in Low-*k* Process," *Semiconductor International,* May 2000, p. 44.

5. H. Aoki *et al.,* "Robust 130-nm-Node Cu Dual Damascene Technology with Low-*k* Barrier SiCN," *Tech. Dig. IEDM* 2001, p. 76.

6. "New Barrier/Etch-Stop for Copper Damascene," *Semiconductor International,* July 1999, p. 28.

7. S.-M. Cho *et al.,* "Low Dielectric Constant Material Deposited by HDPCVD for Barrier and Etch Stop Applications in Cu Damascene Structure," *Abs. Mtg. Electrochem. Soc.,* Spring 2000, Abs. 231.

8. X.W. Lin and D. Pramanik, "Future Interconnect Technologies and Copper Metallization," *Solid State Technology,* October 1998, p. 63.

9. 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.

10. Y. Morand *et al.,* "Copper Integration in Self Aligned Dual Damascene Architecture," *Dig. of Tech. Papers 1997 Symp. on VLSI Technology,* p. 31.

11. S. Venkatesan *et al.,* "A High Performance 1.8V, 0.2 μm CMOS Technology with Copper Metallization," *Tech. Dig. IEDM,* 1998, p. 769.

12. D. Edelstein *et al.,* "Full Copper Wiring in a Sub-0.25 μm CMOS ULSI Technology," *Tech. Dig. IEDM,* 1998, p. 773.

13. 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.

14. P. Singer, "Tantalum, Copper, and Damascene: The Future of Interconnects," *Semiconductor International,* June 1998, p. 90.

Ramon CA, p. 183, 1999.
16. L. Peters, "Leaders Describe "Surmountable" Copper and Low-*k* Hurdles," *Semiconductor International,* September 2000, p. 25.
17. T. Mountsier *et al.,* to be published.
18. B. Gadsen *et al.,* "Top Surface Imaging Improves Copper Process Resolution," *Solid-State Technology,* March 2001, p. 77.
19. M. Neisser *et al.,* "Applying a Thin Imaging Resist System to Substrates with Topography," *Solid State Technology,* August 2000, p. 127.
20. M. Maenhoudt *et al.,* "248 nm and 193 nm Lithography for Damascene Patterning," *Solid State Technology,* April 2001, p. S15.
21. M. Moussavi "Advanced Interconnect Schemes Toward 0.1 μm," *Tech. Dig. IEDM* 1999, p. 611.
22. D.J. Thomas *et al.,* "Meeting the ITRS Roadmap Challenges with Low-*k* Dielectric Etching," *Solid State Technology,* March 2001 p. 108.
23. S. Vanhaelemeersch *et al.,* "Dry Etch & Clean Aspects for Advanced Integration of Low-*k* Dielectrics," *Mtg. Electrochem. Soc.,* Spring 2000, Abs. 224.
24. G. Rajagopalan *et al.,* "Development of Damascene Etch Processes for Organic Dielectric Materials," *Abs. Mtg. Electrochem. Soc.,* Fall 1999, Abs. 702.
25. I.-C. Deng *et al.,* "Using NH_3 Plasma Treated Spin-On Low-*k* Film as Barrier Metal-Free Dielectric to Suppress Copper Diffusion and Improve Its Ashing Resistance," *Abs. Mtg. Electrochem. Soc.,* Spring, 2000, Abs. 225.
26. T. Fukuda, *et al.,* "0.5-μm Pitch Copper Dual-Damascene Metallization Using Organic SOG (k = 2.9) for 0.18 μm CMOS Applications," *Tech. Dig. IEDM* 1999, p. 619.
27. A. Braun, "Photostrip Faces 300 mm, Copper and Low-*k* Convergence," *Semiconductor International,* September 2000, p. 88.
28. D. Hymes *et al.,* "The Challenges of the Copper CMP Clean," *Semiconductor International,* June 1998, p. 117.
29. V. Shannon, "Techniques to Improve Cu/Low-*k* Integration," *Solid State Technology,* September 2001, p. S22.
30. J. Rubin, "Supercritical Fluids as Wafer Cleaners," *Semiconductor International,* September 2000, p. 84.
31. P. Singer, "Dual-Damascene Challenges Dielectric Etch," *Semiconductor International,* August 1999, p. 69.
32. S. Russell *et al.,* "Integration Challenges of Ultra-Low-*k* Dielectrics," *Adv. Metallization and Integrated Systems for ULSI Applications,* 1997, p. 289.
33. L. Peters, "Low-*k* dielectrics: Will Spin-On or CVD Prevail?" *Semiconductor International,* June 2000, p. 108.
34. A. Braun, "CMP Battles Low-*k* Fundamental Barriers," *Semiconductor International,* October 2000, p. 66.

Chapter 16

COPPER INTERCONNECT PROCESS TECHNOLOGY

As described previously in Chap. 12, interconnects in deep-submicron ICs comprise a multilevel structure of fine wiring located on the top of the transistor circuitry of logic and memory chips. To avoid significant degradation of circuit speed, on-chip interconnections should permit rapid signal transmission among the various parts of the circuitry. Ever since the development of the integrated-circuit in the early 1960's, the materials most widely used for the wiring structures have been aluminum as the conductor (or eventually an aluminum-copper alloy for better reliability), and silicon dioxide as the insulator. In 1997 the transition to copper as the conductor was initiated with the announcement by IBM that they had developed a technology for forming copper interconnect lines and vias for IC applications. This was one of the most important changes in materials that the semiconductor industry has experienced since its inception. Since that breakthrough, AMD reportedly applied Cu to its microprocessor circuits at the 0.18-μm technology node, enabling it to introduce the first GHz-rated microprocessor in March 2000, ahead of competitors. In addition, circuit designers have also started to explore innovative layouts to take full advantage of the multiple benefits of Cu (as opposed to simply replacing Al with Cu in an existing design). Thus, it appears that the Age of Copper is dawning!

16.1 WHY COPPER FOR DEEP-SUBMICRON IC INTERCONNECTS?

As noted above, copper is now poised to take over as the main on-chip conductor for all types of ICs.[1,2] Early adopters of Cu technology put copper in production at the 0.25-μm node. Recent market-research studies indicate that for the 0.18-μm generation about 5% of the interconnect structures will be implemented with Cu. It also appears that most Cu in production (or pilot-line production) in 2001 was at the 0.18-μm and 0.13-μm technology nodes. For some manufacturers, 0.13-μm is a second-generation of Cu, for others 0.13-μm is the entry point for placing Cu into production (Fig. 16-1). It is predicted that Cu will represent about 60% of the interconnects for the 0.1-μm generation. The reasons why copper is displacing Al as the dominant conductive material for deep-submicron IC interconnects are explained next.

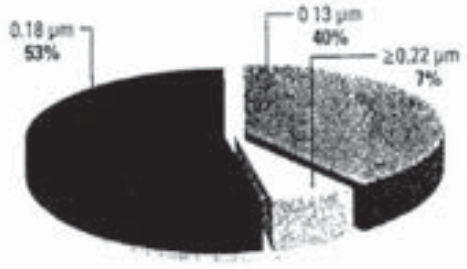


Fig. 16-1 Market analysis of device manufacturers using pilot production in 2001.[39]

16.1.1 The Advantages of Copper Interconnects

There are several critical advantages that copper interconnects provide over the traditional aluminum interconnects that have been used ever since ICs were invented in 1961:

1. Copper has a resistivity that is approximately 40% lower than Al (ρ_{Cu} = 1.7-$\mu\Omega$-cm versus ρ_{Al} = 2.7-$\mu\Omega$-cm), and 50% lower than the Al:Cu-alloys used in IC interconnects ($\rho_{Al:Cu(0.5\text{-at}\%)}$ = 2.9-3.3-$\mu\Omega$-cm). By referring to Eq. 12-7 for the interconnect *RC*-time-delay, it can be seen that *RC* scales directly with ρ. Specifically, if copper can be used instead of aluminum-alloys, the resistance of the wiring structures can also be proportionately reduced. That is, a reduction in the wiring resistance (and thus also the *RC*-delay) of as much as 45% can be obtained when copper replaces Al as the core-interconnect-metallurgy material.

It is important to note that a more complete analysis of the wire resistance must take into account not only the core-material of the interconnect (Al:Cu or Cu), but also any other (and usually higher-resistivity) materials used to surround the core-conductor. A comparison between the effective resistivity ρ_{eff} of Al:Cu wires and Cu wires is illustrated in Fig. 16-2. That is, the Al:Cu lines in Fig. 16-2 are fabricated by etching an Al:Cu film that has both bottom and top Ti layers. During sintering the Ti-layers react with the Al to form $TiAl_3$ (whose resistivity is ~35-$\mu\Omega$-cm, and the thickness of the resulting $TiAl_3$-layers is about 3.5 times the original Ti-layer thickness). The Cu wires are formed by a damascene process (see Chap. 15). This requires that the bottom and sides of the Cu wires be clad with a thin layer of a diffusion-barrier material. While this barrier-layer film also has high resistivity, it does not react with the Cu (unlike Ti, which reacts with Al), and this material also serves as an adhesion-layer. In any case, it is clear from Fig. 16-2 that copper offers a very significant reduction in the effective resistivity for both present and future technologies compared to aluminum.

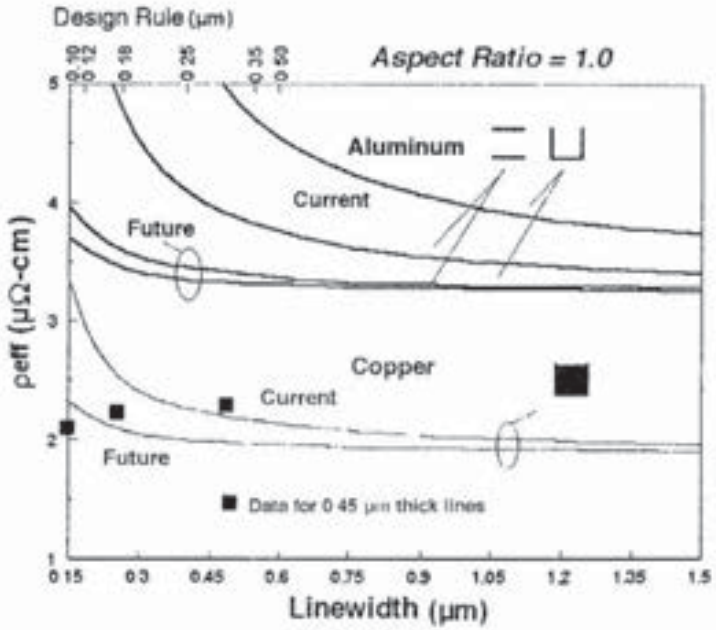


Fig. 16-2 Effective resistivity for aluminum and copper interconnects as functions of line-width and design-rule generation. Computational results (*solid lines*) have been obtained for various combinations of conductor material and barrier location shown in the right side of the figure. Experimental points correspond to published results from real structures.[46]

This lower effective-resistivity of Cu-wiring can be exploited in one of two ways to reduce the *RC*-time-constant of Cu interconnect-systems:

i) The total resistance R_{int} of an interconnect line can be reduced (by retaining the same metal thickness [and cross-sectional area]), or;

ii) The side-by-side capacitance C_{LL} of interconnect lines can be reduced by decreasing the metal-film thickness. This allows R_{int} to remain the same as it was for a thicker Al line, but it reduces C_{LL}. Meanwhile, the same current is carried in a smaller-cross-sectional feature, without fearing EM-failure (see following paragraphs).

The primary clock-speed benefit derived from Cu is gained by using it in the upper-interconnect-levels of a multilevel-metal structure (where the conductor lengths can be of the same order as the chip size). As was shown in Chap. 12, Fig. 12-1, by utilizing Cu together with a low-k dielectric to replace Al and SiO_2, the overall delay (and power consumption) in technologies below 0.25-μm is dramatically reduced.[3]

As also discussed in Chap. 12, the interconnect RC-delay is predicted to become larger than the *RC*-delay of MOSFET transistors for technology nodes below about 0.5-μm. As an example, Fig. 16-3 shows that the *RC*-

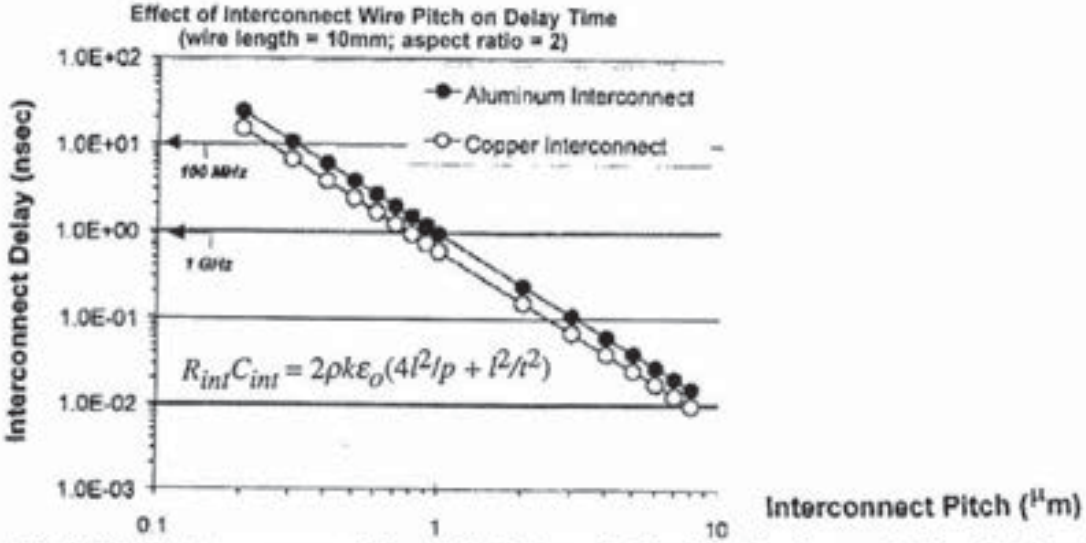


Fig. 16-3 The interconnect delay of both Cu and Al wiring has been calculated using Eq. 16-4 for *RC*. The delay-time in nanoseconds (1 nsec = 10^{-9} sec) is plotted versus the pitch of the wiring in microns (1 μm = 10^{-4} cm) for a line with an aspect-ratio = 2 and length l = 10-mm. The dielectric constant of the ILD has been set to k = 4.0, which is approximately equal to that of CVD deposited silicon-oxide. The graph shows that as the wire-pitch is decreased below 1-μm, the delay-time can become greater than the intrinsic switching-time of the transistors. The speed limit is increased by using Cu over Al.

delay of a dense array of 10-mm-long Al-alloy interconnect-lines with an aspect-ratio of 2 is longer than 1-nsec for pitches smaller than 0.8-μm. Since 1-nsec is equivalent to a frequency of 1-GHz, this implies that signal delays in IC wiring will limit circuit speed to about 1-GHz, (even though individual transistors may have faster intrinsic clock-rates). The interconnect thus represents a real limiter to high-performance ULSI chips – analogous to owning a sports car but only being allowed to drive it on a highway at rush hour. Switching to Cu is one way of improving this situation.

As was done in Chap. 12, Sect. 12-6, equations can be derived to calculate the interconnect delay-time $R_{int}C_{int}$. That derivation is repeated here, but for a simplified special case that exemplifies the impact of switching to copper. The following symbols apply to those aspects of the interconnect system depicted by the test-structure shown in Fig. 12-19.

That is, the metal-line *width* is w, the metal-line *thickness* is t, the *spacing* between the lines is s, the metal-line *pitch* is p ($p = w + s$), and the *thickness of the ILD between the top of the metal lines and the grounded planes* is d. Also the *length* of the metal lines is l, and the *resistivity* of the metal is ρ. Some equations can now be derived that allow the contributions made to R_{int} and C_{int} from: **1**) the materials properties; and **2**) the

interconnect architecture, to be expressed. Using this approach, the resistance of a metal-line is calculated from:

$$R_{int} = \rho l/wt \qquad \textbf{(16.1)}$$

Next, the capacitance between the metal lines and the grounded planes C_{LG} (separated by an ILD whose relative constant is k), is calculated from:

$$C_{LG} = k\varepsilon_o wl/d \qquad \textbf{(16.2)}$$

and the intralevel capacitance C_{LL} (assuming the same ILD) is found from:

$$C_{LL} = k\varepsilon_o tl/s \qquad \textbf{(16.3)}$$

where ε_o is the permittivity of free space. For the $R_{int}C_{int}$ time-delay-constant, it follows that:

$$R_{int}C_{int} = 2\,R_{int}(C_{LG} + C_{LL}) = 2\rho(l/wt)k\varepsilon_o(wl/d + tl/s) \qquad \textbf{(16.4)}$$

To simplify the discussion, we assume that the pitch is equal to twice the line width ($p = 2w$), which also implies that $s = w$, and the metal-thickness is equal to the insulator thickness ($t = d$). In this case, Eq. 16.4 becomes:

$$R_{int}C_{int} = 2\rho k\varepsilon_o(4l^2/p^2 + l^2/t^2) \qquad \textbf{(16.5)}$$

Equation 16-5 is the relationship that is used here to calculate the value of $R_{int}C_{int}$. As also pointed out in Chap. 12, the dependence on ρ, k and l in Eq. 16-5 shows that $R_{int}C_{int}$ can be reduced by choosing metals with low-resistivity, insulators with low-k-values, and/or making line-lengths as short as possible.

Using Eq. 16-5, it is also possible to calculate the dependence of $R_{int}C_{int}$ on metal-pitch and metal-thickness. The results of such a calculation is illustrated in Fig. 16-4, which shows that for a given metal-

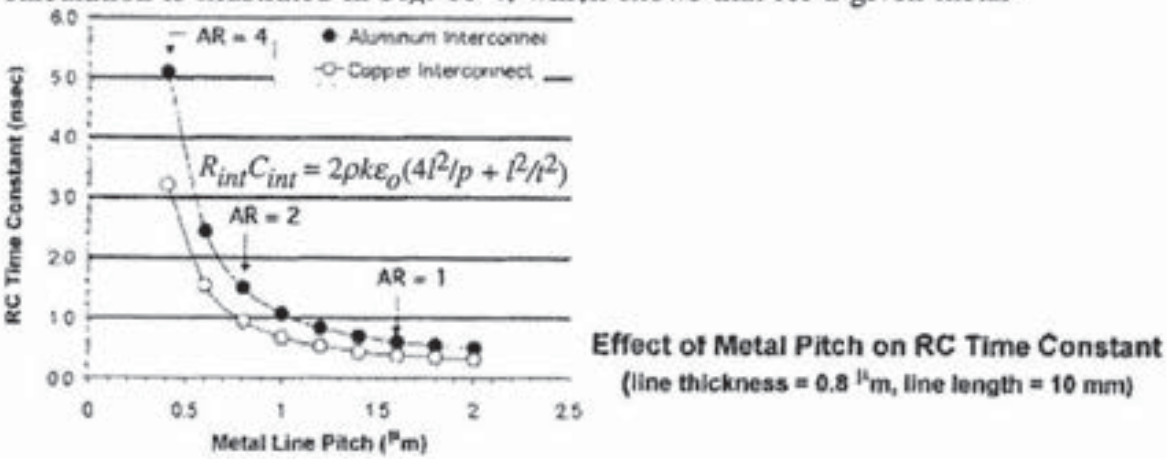


Fig. 16-4 The effect of Cu and Al metal-line-pitch on RC time-constant for a line thickness of $t = 0.8$-mm, a line-length of $l = 10$ mm for the case of both Cu and Al. Eq. 16-5 is shown on the graph and was used to calculate $R_{int}C_{int}$

pitch and line-width that $R_{int}C_{int}$ can be reduced by increasing the aspect–ratio of the line (which of course reduces R_{int} due to the cross-sectional-area increase). However, for aspect-ratios greater than 2, this reduction in the $R_{int}C_{int}$-product is offset by a corresponding increase in the intra-line capacitance C_{LL}. This fact (and the practical difficulty of fabricating metal-lines with very-high-aspect-ratios), is consistent with industry roadmaps predicting metal-line aspect-ratios of less than about 2.7 through the 70-nm node.

After IBM debuted Cu interconnects in 1997, several other companies in 1998 announced the development of Cu-multilevel interconnect systems that they would implement in commercially-available ICs.[4,5] The resistivity of composite copper interconnects in these reports (with the diffusion-barrier in place) did not exceed 2.2-$\mu\Omega$-cm, a 40% improvement over that of comparable Ti/Al:Cu lines.

2. Copper films exhibit much better electromigration (EM) characteristics and stress-void resistance than do Al films. Electromigration is an atomic-scale phenomenon in which the electrons that constitute electric-current collide with the metal-atoms of an interconnect line and push them in the direction of the current-flow. If this "*electron wind*" is strong enough, a significant migration of the metal atoms will occur. Over time this transport of material can lead to the electrical failure of the interconnect in one of two ways: **1**) by a break in the line (causing an open-circuit) due to local depletion of material; or **2**) by the formation of a short-circuit to an adjacent line due to local accumulation of material. The strength of the electron wind increases as the current-density is increased, and the upper-limit of current-density to prevent electromigration in Al:Cu interconnects is $5x10^5$ A/cm^2 (for a 20-year lifetime). This limit puts restrictions on circuit-design and chip-performance. (For example, a current of 5-mA in an interconnect-line of cross-section 0.5-μm x 1-μm (i.e., 0.5-μm^2 = $5x10^{-9}$ cm^2), represents a current-density $1x10^6$ A/cm^2, which is twice as great as the current-density-limit for Al:Cu-lines. Scaling of interconnect wires to increase functional-density and chip-performance inevitably leads to increased current-densities (see also Chap. 12), making Al:Cu interconnects less attractive for future technology structures.

Since Cu-atoms are more massive and more tightly bound together than Al atoms, interconnects made of Cu are generally more resistant to electromigration than those made of Al or Al:Cu (i.e., they exhibit 10-100x

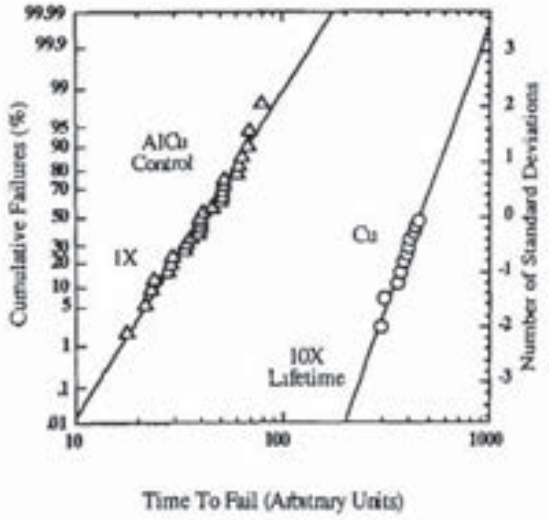


Fig. 16-5 Electromigration performance improvement using copper metallization.[5]

more EM-resistance). About an order of magnitude longer EM lifetimes are obtained for Cu lines of the same cross-section as Al lines, allowing Cu to deliver more reliability for a given interconnect design.[4] In addition to electromigration, stress-voiding (i.e., due to mass transport resulting from thermally-induced stress gradients), is another reliability concern.

Electromigration tests of 0.3-μm lines conducted at an elevated temperature (295°C) and a stress-current of 2.5×10^6 A/cm^2 showed a mean-time-to-failure (MTF) that was more than 10 times longer for Cu than for Ti/Al(Cu)/Ti lines (see Fig. 16-5). Excellent results were also obtained for stress migration testing.[6] Both sets of results argue strongly in favor of using copper interconnections for higher usable current densities and better overall performance.

Another benefit of Cu interconnects with respect to electromigration failures involves the current-flow in vias. As shown in Fig. 16-6, a small diameter via serves as the electrical bridge for current flow from metal-level-n to metal-level-n +1. A problem known as "current-crowding" arises because the current from the rectangular interconnect is compelled to flow through the much smaller circular-cross-section of the via. Due to the change of geometry and reduction in cross-sectional area, the current-flow in the via can be highly non-uniform, with local current-density much higher than in the interconnect line. If the via current-density exceeds the design specifications, excessive heating and/or electromigration can result. In turn, this can lead to circuit-reliability problems. Replacing W-plugs in

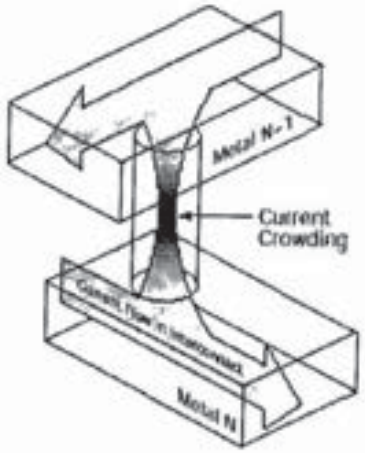


Fig. 16-6 Copper-filled vias can provide additional protection against structural failures associated with current crowding, thereby maintaining via reliability within designed specifications. In particular, the uncertainty in calculating the exact 3-D current-flow in the via is mitigated by the increased tolerance of Cu to electromigration at high current density.

the vias with Cu-plugs significantly reduces the electromigration problems observed in interconnects consisting of Al:Cu wires and W-plugs.

3. The third major benefit of employing copper (and some say it's the most important one) is that lower manufacturing costs can be obtained compared to using aluminum. The cost is lowered for two reasons.

First, because copper is difficult to etch, a new strategy called "damascene" or "in-laid" patterning was developed to form the Cu interconnect lines (see Chap. 15). IBM (which pioneered the process), claimed that the damascene approach requires 20-30% fewer steps than traditional subtractive patterning. Not only are there fewer steps, but dual-damascene eliminates or reduces the need for some of the most costly ones - including aluminum etch, and many of the tungsten-CMP and dielectric-CMP steps.

Second, using copper reduces costs because smaller lines can be employed to carry the same amount of current. Thus tighter packing-density can be achieved per level. This also implies that fewer levels of metal may be needed (see Fig. 16-7),[7] which significantly reduces manufacturing costs. This data indicates that for certain circuit applications only about half the number of levels are required if Cu is used instead of Al-lines and W-plugs.

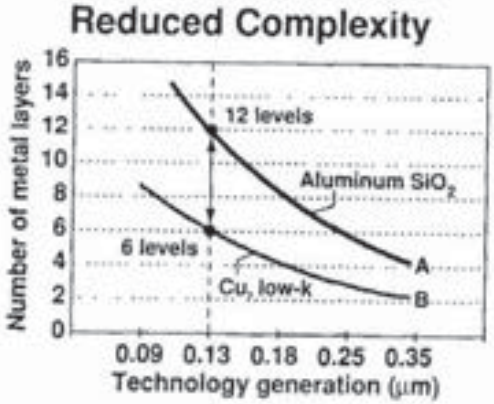


Fig. 16-7 One of the main benefits of copper, beyond the ability to increase chip speed and power consumption, is that the number of metal levels can potentially be reduced by as much as half.[1] Reprinted with permission of Semiconductor International.

4. The currently favored method for Cu deposition is electrolytic-plating, and this allows for much lower deposition temperatures that does Al sputtering or W-CVD. This enables the use of low-k dielectric materials whose thermal stability is lower than that of SiO_2.

16.1.2 Technological Challenges of Using Copper Interconnects

Although many powerful benefits can be gained if Cu is substituted for Al in the IC interconnect structure, a new set of technical challenges arise, including the following:

1. Cu does not form volatile copper fluorides and chlorides, making plasma etching of Cu impractical at temperatures below 200°C. Thus, another technique for forming Cu wires has been developed. Instead of patterning and etching Cu films, the Cu-films are first blanket-deposited over the entire wafer surface and into recesses in the dielectric layers. Then the Cu film is polished away by CMP in the flat areas, but is allowed to remain behind, in-laid in the recesses. This is the basis of *damascene* interconnect structure formation (as discussed in Chap. 15).

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.

3. Copper readily diffuses through SiO_2. If it is present in sufficient concentrations in a dielectric it can lead to shorts or leaky paths between conductors.

4. If the Cu manages to diffuse through the SiO_2 and then enters the silicon lattice, it will create deep impurity levels in the silicon bandgap. These levels act as *recombination/generation (R/G)* centers for free charge-

carriers. The presence of such R/G centers degrades transistor performance and causes *p-n* junction leakage. Copper can also precipitate at (or near) the silicon wafer-surface, and such metallic precipitates are known to cause defects in the gate-oxide as it is being grown. Cu is also a fast-diffuser in single-crystal silicon (i.e., it exhibits the highest diffusivity in silicon of all metal elements), and thus will quickly distribute itself everywhere within the bulk of the wafer (including in the device-regions near the wafer surface, where it can do the most harm). For these reasons, Cu is viewed as "poison" if it is present in the silicon lattice. As a result, special Cu processing areas are also set up in wafer fabs to keep trace amounts of Cu from contaminating any other areas of the cleanroom.*

5. Unlike Al (which reacts with SiO_2 to form a strong interface with excellent adhesion), Cu adheres very poorly to SiO_2.

6. The deposited copper films are polycrystalline in nature, and the preferred orientation of the crystal structure is a high (111)-orientation. Control of this orientation is important because resistivity, stress, adhesion strength, and electromigration resistance of copper layers are largely affected by the films crystal orientation.[17,53]

To prevent Cu from diffusing to the device areas of an IC, the interconnect structure must have the built-in reliability of several protective measures against this possibility. The primary line of defense involves totally encapsulating the Cu wires and plugs at all locations and at all levels to prevent copper from escaping. In dual-damascene interconnect processes, this is done in two-steps. In the first one, a *conductive* diffusion barrier-layer (e.g., Ta, TaN, or TiN) is deposited over the wafer after the trenches and vias of each level of metal have been etched in the dielectric stack, (but before the copper layer is put down). Thus, the insides of the trenches and vias of the damascene dielectric-stack are lined with a diffusion-barrier layer prior to deposition of the Cu into them. Such conductive barrier-layers must be effective against Cu out-diffusion and must simultaneously be thin enough to maximize the cross-section of the Cu-wires and plugs in the recesses. The second diffusion-barrier-layer is formed after the Cu-layer is deposited and the Cu structures have been polished with CMP. That is, a *non-conductive* diffusion-barrier-layer is deposited onto the wafer to seal the top-side of in-laid Cu structures (typically a CVD-silicon-nitride layer is selected to serve in this

* The use of Cu as an alloy in Al:Cu interconnects generates a completely different set of requirements for containing Cu atoms. That is, Cu-atoms alloyed with Al are immobilized by the aluminum matrix, which acts as a very efficient getter.

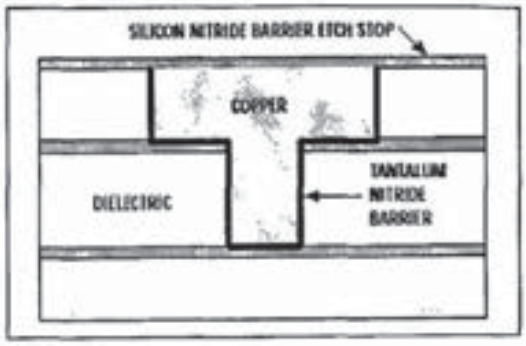


Fig. 16-8 Dielectric (silicon nitride) and conducting (tantalum nitride) barriers encapsulate copper in a damascene structure and protect against copper migration from interconnect structures to neighboring dielectrics and silicon.

role). Thus, Cu structures are ultimately enclosed on all sides by either a Ta (or TaN) or a silicon-nitride diffusion-barrier (Fig. 16-8).

The secondary on-chip line of defense is provided by W-plugs and phosphosilicate glass (PSG). The tungsten-plugs are used in the contact holes as a secondary barrier that keeps the first level of Cu-lines physically separated from the Si substrate (and this also means that a *single-damascene* process is used to form these W-plugs as well as the first level of Cu wires). PSG is known to getter metal atoms. The last level of protection on the chip itself is provided by a bottom silicon-nitride-layer, which further blocks the underlying silicon substrate from migrating Cu atoms.

While manufacturing protocols must also be modified to tightly control cross-contamination, the Cu-contamination problem does not seem to be as difficult as it first appeared to be. For example, fabrication tools cannot be shared between front-end and back-end processes, but copper manufacturing lines can be built in shared cleanroom facilities with aluminum interconnect manufacturing lines. Cleaning procedures must also be implemented to ensure that no detectable traces of copper are present on the bevels and backsides of wafers after processing.

Additional procedures have also been developed to heighten awareness of production workers and to contain copper contamination. For example, copper personnel in some fabs (especially maintenance technicians) wear cleanroom gowns that are used only in the copper bays, and they may not enter non-copper stations when wearing them. Specially colored wafer-carriers and conspicuous labels (indicating the presence of copper-products) are used to clearly identify wafers that have received copper processing and to segregate them from wafers that have not been subjected to copper processes.

Protocols also require that once a wafer has entered a copper bay it cannot return to a noncopper bay. In practice, however, wafers that leave the *front-end-of-line* (FEOL) after the contact holes to the silicon substrate have been opened, move to the *back-end-of-line* (BEOL), where the copper fabrication occurs. There is really no reason why the wafers should have to re-enter the FEOL part of the fab after that part of the process sequence is completed. Additional information on preventing cross-contamination from Cu is given in Sect. 16.7.

To counteract the problem of poor adhesion of Cu to oxide, the conductive diffusion-diffusion films under the Cu layers must also serve as adhesion promoters. The adhesion/diffusion layers used under the inlaid Cu features in damascene processes must therefore meet several criteria:

1. They must act as good adhesion-promoters for Cu onto SiO_2. That is, the adhesion/diffusion-layer materials should adhere well to oxide, and then the Cu films should adhere well to them.

2. They must be effective barriers against Cu diffusing through them under 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-$\mu\Omega$-cm to meet the needs of advanced interconnect processes. This will keep the total resistivity of the composite Cu/barrier interconnect-structures below the resistivity of comparably-thick Al-alloy/barrier interconnect lines.

4. They should retain their diffusion-barrier properties when they are very thin, because they do not have the low-resistivity of Cu. That is, thinner diffusion barriers for Cu will be needed than for Al because in-laid-Cu interconnects require the barrier-layer to be formed on at least three sides of the Cu conductor (Fig. 16-9). To keep the lower-conductivity barrier-layers from impacting the resistance of the interconnect line, they must thus be relatively thinner for Cu lines.

5. They should protect the Cu from oxidation and corrosion.

6. They should be able to make a low-resistance contact to the copper layer of the metal-level under it, and this barrier-layer/Cu interface should exhibit good electromigration reliability characteristics.

7. They should provide a surface which will facilitate the nucleation of the Cu-seed layer deposited on it. Furthermore, the surface-condition and micro-

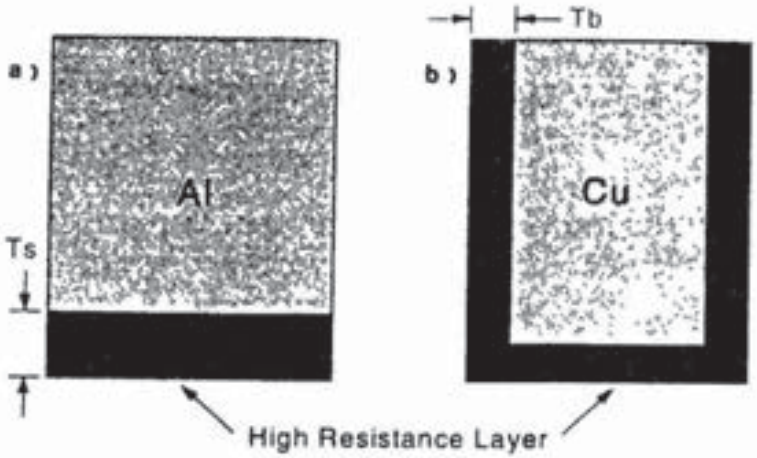


Fig. 16-9 Interconnect cross-sections showing the impact of higher resistance shunt or barrier layers: (a) Al interconnect; (b) Cu interconnect.[7] (© 1995 IEEE)

structure of the barrier-film can strongly affect the texture and grain-size of the deposited Cu (and these in turn, are critical in determining electromigration reliability).[50] Thus, the barrier-layer is also vital to seeding the desired Cu microstructure. Since the growth of Cu layers with a (111) preferred orientation provides Cu films with the best overall characteristics, the barrier-layer should enhance such Cu-structure.

In addition to the above materials-science issues associated with the use of copper, there are a number of difficult manufacturing challenges that will also have to be met when integrating Cu into an IC fabrication sequence. That is, getting a copper process (or any process) online and productive quickly is the key to profitability. The integration issues involved in bringing a Cu process to full-production are summarized in Fig. 16-10,[15] and they will be discussed in more detail in later sections of this chapter.

16.2 OVERVIEW OF COPPER PROCESS TECHNOLOGY

Before the details of depositing, polishing and cleaning copper IC interconnect structures are described, it is useful to give brief overview of the entire sequence of processes that are required to form a single-level of a copper dual-damascene structure. The sequence of process steps is as follows (and is also depicted in Fig. 16-11):

1. Non-Conductive Diffusion-Barrier-Layer Deposition
2. Dielectric Stack Deposition

3. Dual-Damascene Dielectric-Stack-Etch (Via, Trench, and Diffusion-Barrier-Layer Etch)
4. Conductive Barrier-Layer Deposition
5. Copper Seed-Layer Deposition
6. Copper Electrofill (and Annealing after Deposition))
7. Copper and Barrier-Layer CMP
8. Copper Clean
9. Non-Conductive Diffusion-Barrier-Layer Deposition.

The sequence starts by depositing a non-conductive diffusion-barrier-layer (e.g., a thin film of silicon nitride by PECVD) over the entire wafer (which has been planarized by a previous CMP step and may already contain one or more levels of metal). Next the remainder of the ILD stack is deposited (as described in detail in Chap. 15, Sect. 15.6). Using a lithography process, the vias and trenches of the dual-damascene structures are aligned to the previous layers of the circuit and are then etched with a dry-etch process. (This requires a sequence of two lithography and etch procedures, including stripping the resist after each etch step.) The SiN diffusion-barrier-layer at the bottom of the vias is also opened during the second of these etch steps, exposing the metal-layer of the previously formed level of the interconnect structure. Next, a conductive diffu-

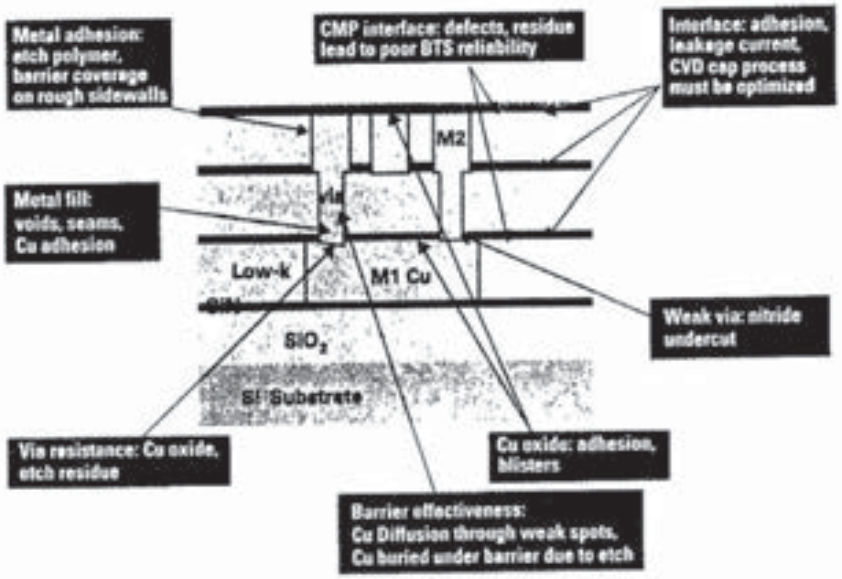


Fig. 16-10 Possible failure types associated with damascene processing pertain to film adhesion, barrier and metal fill capability, copper oxidation, and CMP defects.[9]

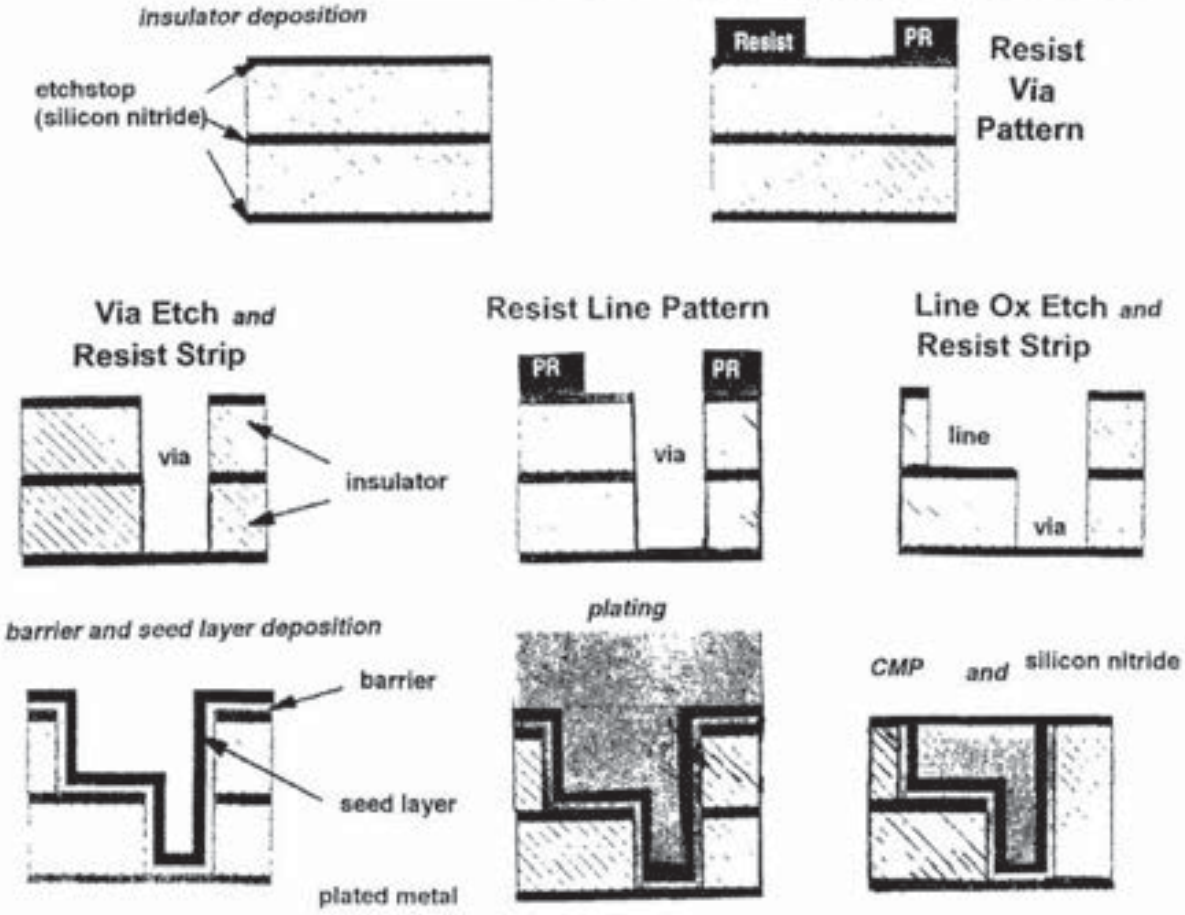


Fig. 16-11 Process sequence for forming a single-level of Cu metal in a dual-damascene interconnect structure (using a *via-first* approach).

sion-barrier/adhesion-layer is blanket-deposited onto the wafer, typically by PVD (with either a collimated-, a long-throw-, or [most typically] an ionized-metal-PVD process). Various materials have been selected to perform the role of this diffusion-barrier/adhesion-layer including Ta, TaN, TaSiN, TiN, TiSiN, WN, and WSiN (see Sect 16.3).

A copper seed-layer is deposited onto the diffusion-barrier/adhesion-layer to provide a low-resistance path for the copper electroplating process that comes next. The Cu-seed layer is also deposited *in situ* (i.e., without an air break) by one of same type of PVD-processes used to form the diffusion-barrier/adhesion layer film. Completing the process of filling the trenches and vias with Cu material is done with an electrolytic-plating technique. The key attributes of such Cu electrolytic-plating-processes include: **a**) the ability to fill the recessed structures without defects (i.e., no voids or seams should exist in the plated films after filling); **b**) the ability to achieve high-deposition rates; and **c**) the ability to

obtain uniform deposition-thickness across-the-wafer, and from wafer-to-wafer (e.g., ±2% across a 200-mm wafer with a 5-mm edge exclusion).

A CMP process is carried out next to remove any excess copper above the vias and trenches (called the *overburden*), but allowing metal to remain in the damascene recesses. During the CMP process, the diffusion-barrier-layer must also be removed. A key technological challenge involves achieving the same polishing rate on the diffusion-barrier layer, because such materials as Ta or TaN are hard, refractory metals (or metallic compounds), while copper is a comparatively much softer metal. After the CMP process is finished, it must be ensured that the wafer is cleaned, such that no copper residue remains on the oxide on the frontside of the wafer (nor on the wafer backside or beveled edges). Finally, another non-conductive diffusion-barrier film (e.g., CVD silicon-nitride) is deposited onto the wafer surface, which serves to cap-off the copper (i.e., this completes the encapsulation of the recessed-copper structures, ensuring they are surrounded by a diffusion-barrier-material on all sides).

The details of the first three steps of this sequence are described in Chap. 15, Sects. 15.6 and 15.9 (and thus will not be repeated here). The discussion that follows therefore begins with the Step **4** (Diffusion-Barrier-Layer Deposition, and the properties of such materials) and then proceeds to cover the details of the remaining steps of the process sequence.

16.2.1 The Transition from Aluminum to Copper Interconnects

The process of incorporating copper into the interconnect process of an IC technology may be done in one of two ways. Some manufacturers have decided to use Al:Cu (and W-plugs) for several of the lower levels of a multi-level metal interconnect structure, and limit copper to the upper levels. In one such example, Al:Cu was used in the four lower levels of the 6-level metal interconnect structure and Cu for the upper two levels (Fig. 16-12).[10] Another process described by TSMC also uses four aluminum levels of metal and two of copper.[11] The interface between the top aluminum interconnect level and the lower Cu interconnect level can be implemented either with a W-plug or with a Cu-damascene via structure.[12]

On the other hand, other manufacturers decided to use copper for the entire interconnect structure (with the exception of W-plugs being used in the contact holes). An example of such a 6-level Cu-interconnect-structure (with W-plugs used to make contact to the Si-substrate) was described by IBM (see Fig. 16-13).[13] Figure 16-14 shows a SEM of a perspective view of a six-level metal Cu interconnect structure (see also Fig. 12-16 in Chap. 12).

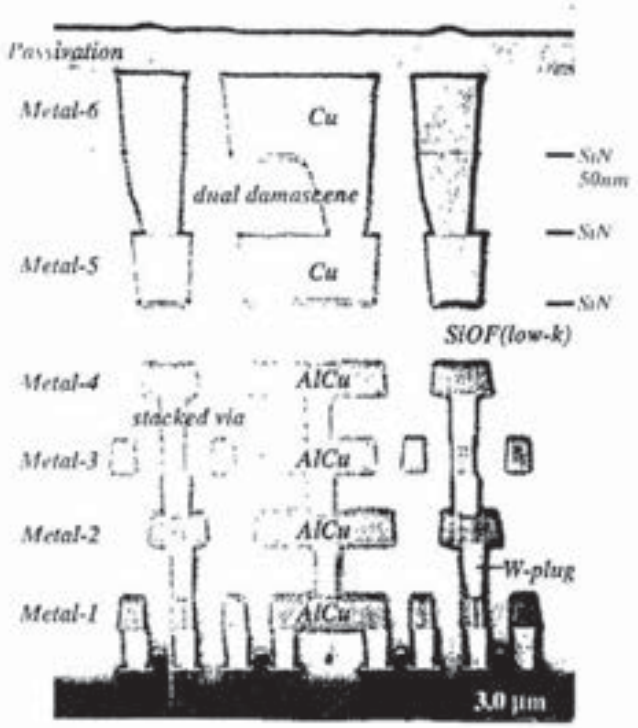


Fig. 16-12 Cross-section of six-level metal structure with the first four-levels being Al:Cu/W-plug interconnects and the top two levels of Cu metal.[10]

16.2.2 The Key Goal of Depositing Copper into Damascene Recesses

One of the most important aspects of depositing copper successfully into the recesses of a dual-damascene ILD stack is ensure it done without any voids or defects. The key to being able to do this is to fill the recesses with copper from

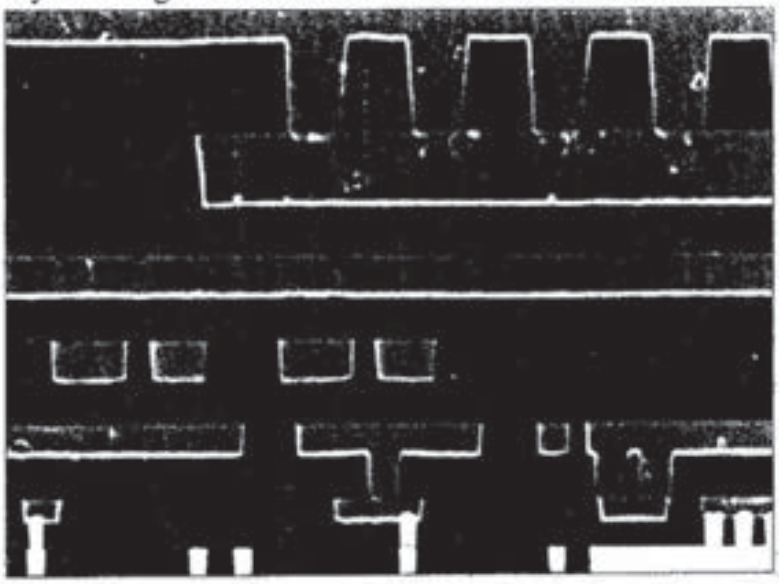


Fig. 16-13 Cross-section of six-level metal structure with all levels of Cu metal.[13]

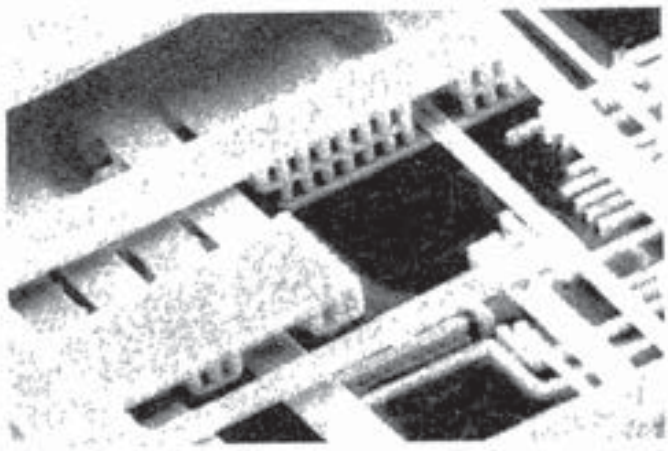


Fig. 16-14 An SEM of a perspective view of a six-level metal Cu interconnect structure.[32]

the bottom up, a technique called "bottom-up" filling, or "superfilling." However, successful superfilling is affected not only by the electrofill process but also by the dielectric etch process, the barrier-layer deposition-process and the seed-layer deposition-process. Thus, this topic has to be addressed in a piecemeal fashion, as will be done in the three upcoming sections.

16.3 BARRIER LAYERS FOR COPPER INTERCONNECTS

As discussed previously, to stop Cu from diffusing into the dielectric materials in which it is embedded (or into the Si substrate), it must be surrounded by a barrier-layer on all sides. The underlying layer is deposited prior to depositing the copper (as noted in Sect 16.2). In this section, the technology for forming these conductive barrier-layers is described. The subjects that will be covered are:

1. The impact of the barrier-layer on the Cu-electroplating process;
2. The technology of depositing barrier-layers into damascene recesses by PVD;
3. The material properties of tantalum (and TaN & TaSiN) barrier-layers;
4. The other candidate materials for Cu-technology barrier-layers (including CVD processes for depositing such materials).

16.3.1 The Impact of the Barrier-Layer Deposition Process on the Filling of Damascene Recesses with Copper

To recap, a barrier-layer for Cu interconnects must prevent Cu diffusion into the ILD and must assure good adhesion between the Cu and the ILD. Furthermore, the barrier must be thin, conformal, and have a low electrical resistivity. Deposition should be at a low temperature with few particles being produced.

The barrier-layer material should also be compatible with CMP of Cu and with the Cu electroplating process. Finally, there a number of other aspects involved with the deposition of the barrier-layer that impact the ability to completely fill damascene recesses with copper without voids. These include depositing a barrier-layer that is continuous and adequately thick everywhere along the recess surfaces (typically 25-30-nm-thick), and with little (or better yet, no) overhang. The PVD-process used to deposit the barrier-layer impacts these last two factors, and the PVD techniques for depositing seed and barrier layer is discussed next.

16.3.2 Introduction to the Technology of Depositing Barrier-Layers for Cu Interconnects

The conductive barrier-layer deposited into recessed damascene features before the Cu is deposited can be done by PVD or CVD. While PVD has been the most widely used method down to at least the 0.13-μm technology node, there may be a change to CVD at some point. One reason for such a technology replacement involves the formation of a film within a trench or contact hole by a *physical vapor deposition process*. Assume a contact-hole exists with the top face being open with the dimensions of a cube of side **a**. This represents a contact hole with an aspect ratio of 1:1. In such a case there are five inner surfaces of the cube that must be covered by material which would otherwise deposit on area $\mathbf{a}^2$ on the top surface of the cube (i.e., if the top of the cube was not open, see Fig. 16-15). However, if the top of the contact hole is open, the total area inside the hole that must be covered now is $5\mathbf{a}^2$. Thus, if the material that would have covered $\mathbf{a}^2$

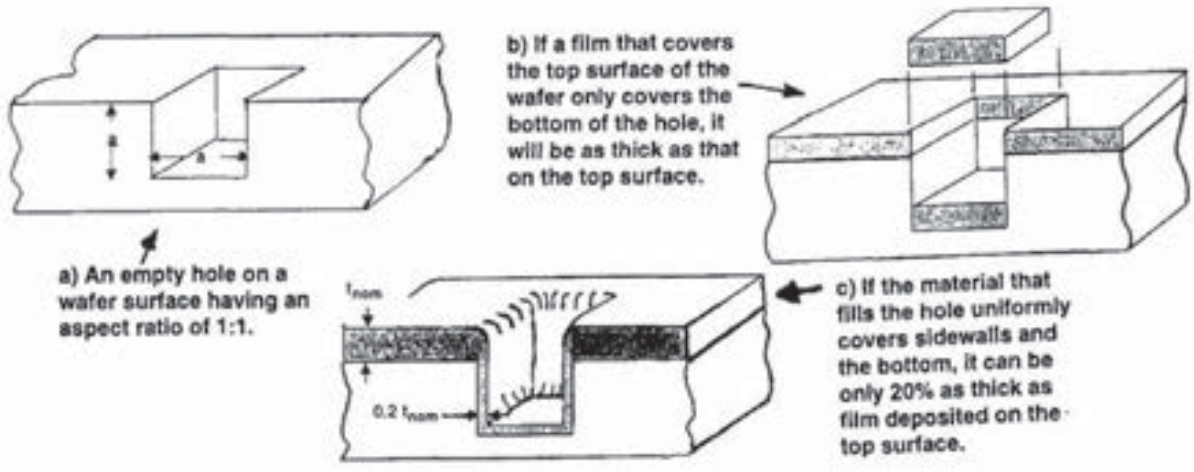


Fig. 16-15 One dilemma faced when forming films in high-aspect-ratio holes with PVD.

is evenly distributed onto all five insides of the contact hole, the film thickness will only be 20% as thick as that on the flat areas outside of the contact hole. Obviously, if the area of the hole-opening remains constant while its depth is increased, the aspect-ratio will also increase (e.g., if the square contact-hole opening-size remained at $\mathbf{a}^2$, but the depth became $2\mathbf{a}$, the aspect ratio will increase to 2:1, and the area that will have be covered by the material of $\mathbf{a}^2$ will now be $9\mathbf{a}^2$). Thus, for the case of a physical-vapor deposition process, as the aspect-ratio of a recessed feature is increased, the material entering the feature must be spread over a larger area, making the thickness per unit area decrease. Furthermore, if the film on the upper flat regions must be etched or polished away after the hole has been filled, a much thicker film now has to be removed (which makes the etch or CMP process task significantly more time consuming and difficult). Deposition by CVD may be able to avoid this problem because CVD processes can be reaction-rate limited (allowing deposition to be almost 100% conformal at a very high-aspect-ratio, see Vol.1, Chap. 6). Thus, CVD may be employed for depositing barrier-layer films for Cu in the lower levels of deep-submicron interconnects (where the aspect-ratios continue to increase).

In the previous discussion it was noted that it becomes progressively harder to fill contact holes by PVD as the aspect-ratio increases. A related problem further increases the difficulty of filling high aspect-ratio-recesses. Recall that the sputtered-atoms undergo collisions as they are transported in the gas-phase in the space between the target and the substrate. In addition, sputtered atoms leave a planar-target surface from each point at which sputtering occurs in directions that roughly exhibit a cosine distribution. As a result of these two mechanisms, in a conventional sputtering apparatus the sputtered-atoms will arrive at the upper corners of the trench or contact hole on the wafer from a larger arrival-angle than they will at surfaces inside the recess. Thus, the film deposition-rate at the top corners of the recess will be higher than along its sidewalls or bottom. As a result, the top opening will begin to close off (i.e., an *overhang* will start to form), as shown in Fig. 16-16a. Thus, for this case, as time goes on (Fig. 16-16b) the deposition-rate onto the sidewalls and bottom will continually decrease. If the overhang gets thick enough, the hole will close entirely, leaving a void or keyhole.

Even if deposition does not continue long enough to completely close-off the opening, the overhang will shadow the bottom corners of the contact hole, where the arrival angle for sputtered atoms is already the smallest (see Fig. 16-16b). Thus, the deposited film will be thinnest at the bottom corners of the hole. Since a barrier-film must have a minimum adequate thickness along the entire bottom

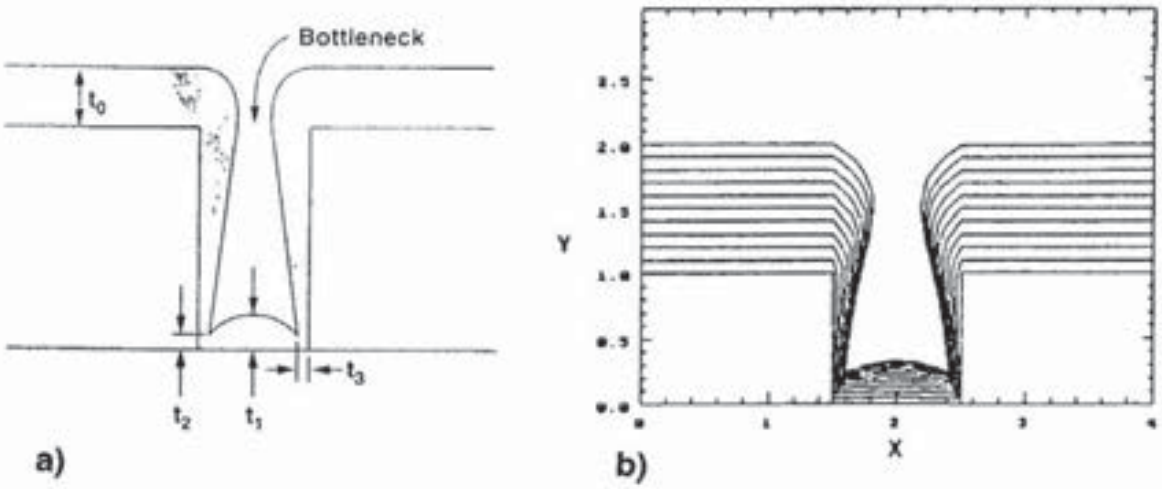


Fig. 16-16 a) Deposition of a sputtered film into a contact hole or via having a high aspect ratio. b) Development of a bottleneck due to the buildup of an overhang on the top corner.

of the hole in order to function effectively, such thin-spots at the corners can jeopardize the effectiveness of the barrier-layer.

Various methods have been explored to overcome this thinning of the film at the bottom corners of contact holes (including the use of CVD). However, modifications to the sputtering process have also been pursued. Two such early methods were *collimated sputtering* (in which a honeycomb baffle - i.e., a *collimator* - is inserted into the space between the target and the wafer to trap obliquely-directed vapor) and *long-throw-sputtering* (in which the target and wafer are separated by a distance 2–4 times longer than in a conventional magnetron source, and the plasma is also operated at a lower pressure). However, these two methods have proven to be inadequate for deep-submicron, dual-damascene applications. It is unlikely that they will even be commercially developed for 300-mm wafer processing. Thus, they will not be discussed further here (but more information on them is available in Vol.1, 2nd Ed, Chap. 11).

16.3.3 Ionized Magnetron Sputter Deposition

While the two sputtering techniques described above produce adequate barrier-layer films for contact-holes and vias for 0.35-μm (and some 0.25-μm) CMOS technologies, they appear to fall short for other 0.25-μm (and smaller) contact holes and vias when the aspect-ratio exceeds about 3:1. Thus, an alternative to collimated (or long-throw) sputtering, namely *ionized magnetron sputtering*, or *I-PVD* (also termed *ion metal plasma* [*IMP*] by Applied Materials), has been developed to deposit the next generation of ULSI barrier-layers with PVD. In this method, a large fraction of the neutral sputtered-metal-atoms are ionized while in flight between the target and the wafer. A bias on the wafer then causes an

electrical acceleration of these ionic species to the wafer-surface with a controlled energy and direction (i.e., perpendicular to the wafer surface).[14,15,16] There have been several sputtering-source configurations developed to implement the ionized magnetron sputtering technique.

Figure 16-17a shows a schematic drawing of an ionized magnetron sputtering chamber used in the first commercially available I-PVD tools. A conventional magnetron source (powered with a dc power-supply) remains in place to sputter neutral metal-atoms from the target. (The magnetron produces a plasma and dark space that is confined near the target surface.) However, an rf-powered (and water-cooled) coil is now also located within the chamber (Fig. 16-17b). This coil inductively couples energy from a second 13.56 MHz rf-power supply into a secondary plasma downstream of the magnetron plasma. By increasing the pressure in the chamber to 20-40-mtorr (up from the 5-mtorr of a conventional magnetron sputtering process), and by increasing the depth of the secondary plasma to 8–10 cm (by increasing the distance between the cathode and the wafer chuck), the number of collisions experienced by the sputtered atoms before they

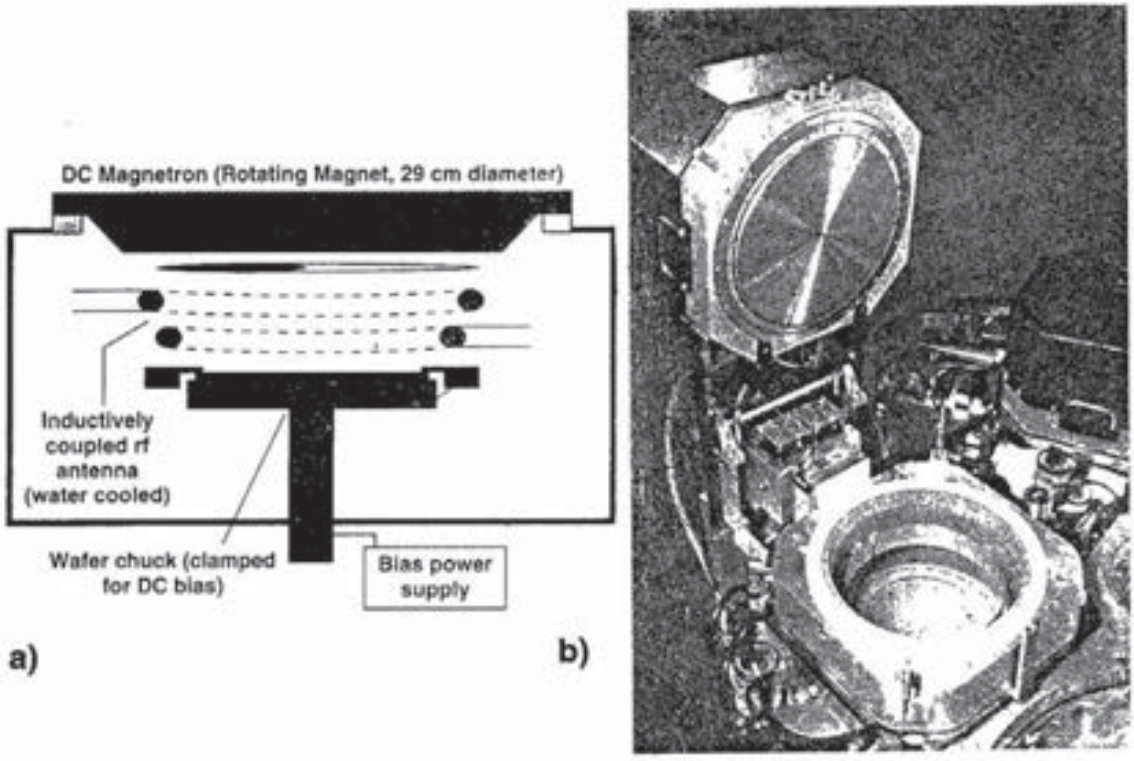


Fig. 16-17 a) Schematic of a magnetron sputter chamber fitted with an rf-coil which allows an ion magnetron sputtering process to be implemented. b) Photograph of an Applied Materials IMP system showing the internal coil. (Courtesy of Applied Materials)

arrive at the wafer (as well as their transport time through the gas phase), is increased. This longer transport time, and the fact that the ionization potential of metal atoms is much lower than that of Ar atoms, allows a much higher fraction of the metal atoms to be ionized by this secondary plasma (e.g., for a Ti-sputtering process, typically 50-70% of the sputtered Ti-atoms are ionized). A medium-density secondary-plasma is thus created in the region between the cathode and the wafer ($5x10^{11}$-10^{12} ions/cm^3).

The chief advantage of the I-PVD approach can now be exploited. That is, by applying a negative-bias to the wafers, the ionized metal-atoms can be accelerated toward the wafer surface (at voltages ranging from a few to several-hundred volts, depending on the specific application). Thus, the metal ions will strike the surface exactly at normal incidence to the wafer surface, regardless of their original trajectory. In addition, the third (substrate-bias) power-supply allows the kinetic energy of the metal ions striking the wafer to be controlled. Using an rf-power-supply to produce this bias overcomes the problems with insulating wafer surfaces (or backsides). Note that the ion-flux onto the wafer results in power dissipations ranging from about 25 watts to several hundred watts, which calls for some degree of wafer cooling.[16]

The bottom-coverage of high-aspect-ratio holes with I-PVD is better than that observed with collimated sputtering. That is, 40–80% bottom coverage has been achieved on 0.18-μm holes with aspect-ratios of up to 8:1.[17] Note that while the bottom-coverage is high, the sidewall-coverage is much thinner (although the film on the sidewall is continuous, see Fig. 16-18).

Several other advantages of I-PVD processes should be noted. First, much better target-utilization is obtained than with collimated sputtering (as almost all of the sputtered atoms arrive at the wafer surface, instead of some being intercepted by the collimator). Sputtering tool throughputs are therefore also increased. Second, the high flux of metal ions (as well as some Ar ions) directed perpendicularly onto the wafer surface appears to re-sputter the atoms that would otherwise form an overhang at the top-corners of a recess. This keeps the top opening of contact holes and vias from closing up. Figure 16-18 shows the coverage of the sidewall and bottom of a high-aspect ratio hole with I-PVD. Third, the negative-bias applied to the wafer not only causes the ionized metal atoms to preferentially strike the bottom of the via, but it allows the kinetic energy of these metal (and Ar) ions bombarding the wafer to be adjusted. For appropriate bias levels, re-sputtering from the feature base can be produced, which redistributes sputter-deposited material onto the lower sidewalls (Fig. 16-19). This is exactly where the lowest PVD coverage occurs in collimated and

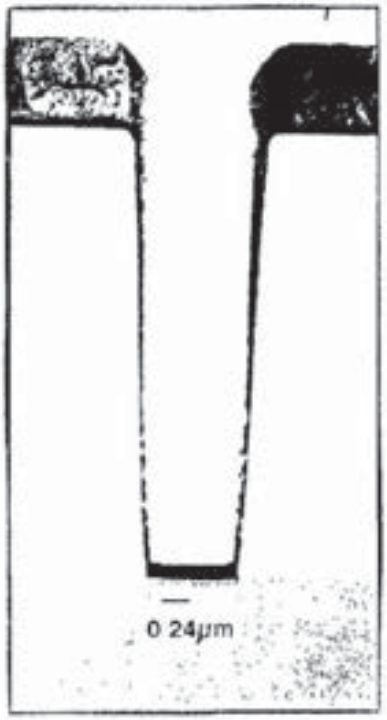


Fig. 16-18 SEM showing the coverage of the sidewall and bottom of a high-aspect ratio hole deposited by I-PVD. (Courtesy of Applied Materials)

long-throw PVD-processes. The process of local re-sputtering can also be used to tailor the conformality of the deposit to a greater degree, and this has extended the use of PVD for depositing barrier-layer and seed-layer films.

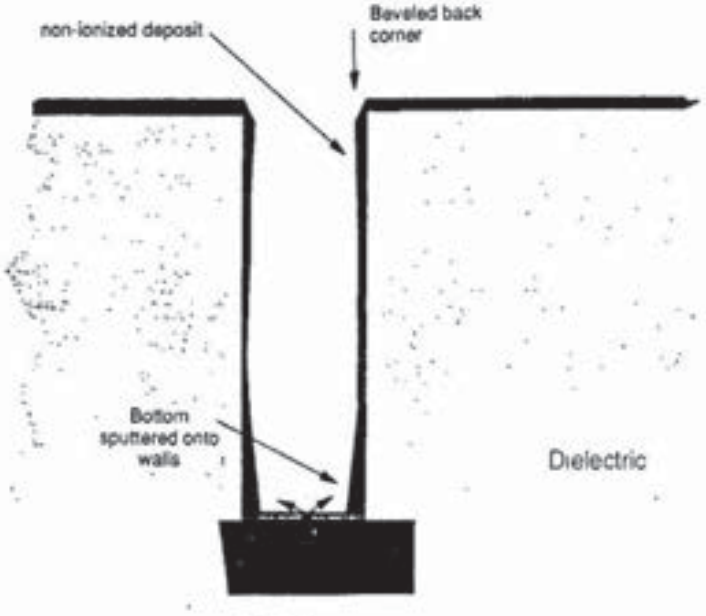


Fig. 16-19 Schematic drawing of I-PVD deposition of a high-aspect-ratio recess showing the effects of re-sputtering of the deposited film.

While the concept of I-PVD has the many advantages described above, the internal coil-configuration has some practical difficulties. First, the coil can get coated with the sputtered metal. If this gets too thick, the film may flake off and produce particles. Second, material from the coil itself can be sputtered, with some material ending up being re-deposited on the wafer. While a copper coil can be used to circumvent this problem when depositing Cu-seed-layers, this does not provide a fix when depositing Ta or Ti barrier-layers. Instead, a coil made of Ta (or Ti) would be needed, but forming such coils (which must also accommodate the flow of cooling water) is not a trivial task. As a result, two other I-PVD source configurations have recently been commercially introduced.

16.3.4 Hollow-Cathode Magnetron Sputtering Source

The *hollow-cathode magnetron* (HCM™) used in the Novellus INOVA™ sputtering system (Fig. 16-20) employs a similar technique to ionize sputtered atoms as they are transported to the substrate.[42,55] A uniform plasma throughout the volume of the cup-shaped hollow cathode target provides sputtering from a large fraction of the area of the target surface. Due to the geometry of the cup-shaped target a significant fraction of the sputtered neutrals are re-deposited on the target surface, where they can be sputtered again (instead of depositing on the chamber walls). This provides the benefit of high target utilization. In addition, no internal rf-coil is needed to generate the high-density plasma that ionizes the sputtered metal atoms. Instead, the orientation of the permanent magnets allows

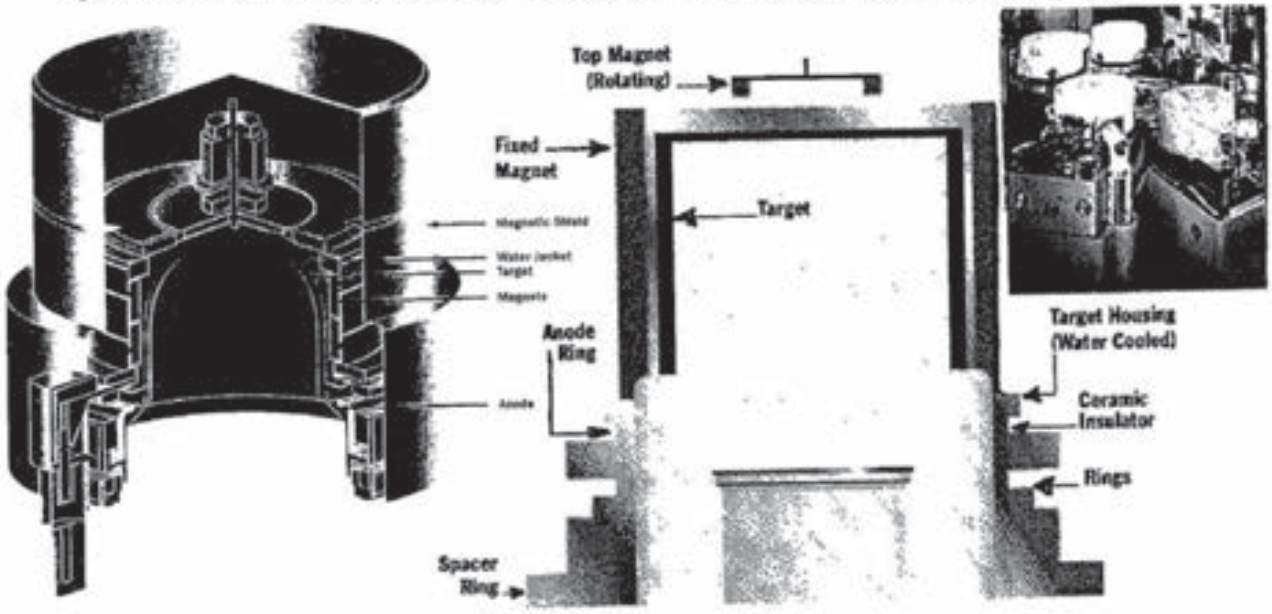


Fig. 16-20 Cutaway drawings of the hollow-cathode magnetron source used in the Novellus Systems's INOVA® sputtering system. Courtesy of Novellus Systems, Inc.

the electrons emitted by the target to efficiently ionize the sputtered neutrals (with the degree of ionization exceeding 90%). A bias on the wafer-chuck accelerates the ionized sputtered atoms toward the wafer in a direction perpendicular to its surface, allowing good coverage of high-aspect-ratio holes. The HCM can be used to sputter Ta, TaN, TiN, and Cu.

16.3.5 Self-Ionized-Plasma Sputtering Source

Another new sputtering source for implementing the I-PVD technique was recently introduced by Applied Materials. They refer to this source as *self-ionized-plasma* (SIP), and the internal rf-coil used in their IMP™ source to ionize the sputtered metal atoms is not needed. While few details about this source have been published, some of its aspects have been described.[18,19] The source operates at relatively low pressures (< 2 mTorr) with a target-to-substrate distance that is increased above the distance used in conventional magnetron sputtering sources. These measures increase the mean free path of the metal-ions and provide additional self-collimation to the metal-atom trajectories. The term *self-ionized-plasma* probably refers to the phenomenon in which some of the sputtered metal atoms that are ionized by the source, return to the target with enough energy to sputter other metal-atoms (and also produce additional electrons). These additional electrons, in turn, ionize more of the sputtered metal atoms, which keeps the plasma self-sustaining and produces more metal-ions. The phenomenon is apparently an efficient way of ionizing metal-atoms that have a small ionization energy (such as Cu). However, questions have been raised about whether it is effective for metals that have higher ionization energies (such as Ta).

16.3.6 Material Properties of Tantalum and Refining Tantalum Metal For Semiconductor Applications

The metal *tantalum* (or such tantalum-based compounds as TaN or TaSiN), have been the most commonly used barrier-layer materials for Cu-damascene interconnects. As such, it is useful to give some background about this element.

16.3.6.1 Material Properties of Tantalum: Tantalum is an uncommon (in nature) refractory metal, with an atomic-number of 73, an atomic-mass of ~181, a body-centered-cubic (BCC) crystal structure, and a steel gray color. It is an attractive Cu-barrier-material because of its high melting point (2996°C), and because it is immiscible with Cu. It is also a highly reactive metal that forms strong metal-metal and metal-oxygen bonds (much like titanium). Thus, it provides a low-resistance ohmic contact and good adhesion to the copper above (and to the

dielectric material below) it in the Cu-damascene metal scheme. 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. Such TaN barrier-layers will be discussed further later.

16.3.6.2 Refining and Forming Tantalum Metal For Sputtering Targets: Tantalum exists in nature as an oxide Ta_2O_5. The principle ore of tantalum is *tantalite* (Fe, Mn)Ta_2O_5. Mines in western Australia are the largest producers of high-grade tantalite, and such concentrates are shipped to the tantalum metal manufacturers for refining.[20] The main market segments for refined tantalum are capacitor-grade powder (about 50%), mill-products and wires (16%), chemical compounds (such as tantalum carbides, 14%), and use as a metal-alloy additive (6%). High-performance tantalum products such as *sputtering targets*, represent less than 10% of the tantalum mill-product and wire business.

Tantalite ore is not the only source of tantalum. Nearly 25% of the total Ta production is recycled. In fact, if the market for high-quality sputtering targets outpaces the rate of increase in tantalite production, it might become necessary for wafer manufacturers to return spent targets to the Ta manufacturers in order to be assured of future availability.

The extraction and refining processes for tantalum conclude with a chemical precipitation reaction that yields Ta metal in powder form. The high-density, metallurgical-grade powders (and recycled Ta scrap) are used as feed for *vacuum melting*. (Because tantalum exhibits a great affinity for oxygen, nitrogen, and carbon, thermal treatments must be conducted in a vacuum or inert gas atmosphere.) For demanding applications (such as sputter deposition), the refined tantalum must possess a number of material properties that can be obtained by subjecting the tantalum powders to such vacuum melting processes.

Typically, tantalum ingots are melted several times to purge most of the impurities from the metal. Because the melting point of Ta is much higher than most materials, this is effective in volatilizing the interstitial gases (such as O, N, C) and many metal impurities. However, the other refractory metal elements (i.e., niobium, molybdenum, and tungsten) are not volatilized during vacuum melting, but remain as soluble impurities in tantalum. However, these refractory metal contaminants can be minimized through specialized front-end chemical extraction processes, resulting in extremely pure-Ta ingots. For example, the Performance Materials Division of Cabot Corp. (Boyartown, PA) offers Ta-ingots that are 99.999% pure.

The pure ingots consist of poly-crystalline Ta with small-sized grains. These ingots can be cold-worked (i.e., at ambient temperature) to cause a re-crystallization of the grain structure in the finished annealed product. The desired microstructure for Ta sputtering targets is a fine, homogeneous polycrystalline structure. Tantalum crystal grains predominantly have a (111) or (001) type structure. It is preferable to have a Ta ingot with a preponderance of (111) grains. This is because Ta sputtering-targets which contain large (001) clusters exhibit lower sputtering yields. The reason is that in such body-centered cubic (BCC) metals as Ta, the erosion-rate of (111) incident planes is predicted to be far greater than for (001) crystals.[21] The proportion of (111) to (001) grains is governed primarily by deformation history.

In summary, attaining microstructural and textural homogeneity in tantalum sputtering targets through the use of proper thermochemical processing is critical to assuring the performance of such targets. It should also be noted, that all of the refining and thermochemical processes are expensive. In turn, this also makes Ta sputtering targets quite costly (reaching $40,000 or more for some target configurations).[22]

16.3.6.3 The Properties of Tantalum (and TaN & TaSiN) as Cu-Barrier Layers: To prevent the drift or diffusion of Cu atoms, a barrier-layer should not react with Cu. Such a barrier-layer should remain thermally stable up to the maximum temperature that the wafer will be subjected to once the Cu has been deposited. Since this temperature is not likely to exceed 500°C, the barrier-layers should be stable to this temperature. Ta and TaN films are observed to remain thermally stable up to about 550°C, and thus they are capable of meeting this requirement for back-end processes.[23]

However, as noted in Sect. 16.1, the barrier-layer materials must also serve as adhesion layers for Cu. There is typically a tradeoff between barrier integrity and adhesion performance with Cu. If a thin-film of the barrier-layer material does not react with Cu at all, it may exhibit poor adhesion. If the material reacts too easily, it may not function as a good diffusion-barrier. For good barrier properties and good adhesion with Cu, an ideal barrier would react with Cu to some self-limiting extent.

Tantalum films, sputter-deposited using a Ta target in an Ar ambient, have been investigated for such Cu barrier-layer applications. For the case of Ta, an as-deposited Cu-layer on a Ta film displays an abrupt interface with Ta. However, a 3-nm thick amorphous layer is observed at the Ta/Cu interface after annealing for 1 hour at 400°C.[24] The mechanism for the amorphous layer

formation is not clearly understood because Ta has no equilibrium solubility with Cu at 400°C, and does not form compounds with Cu. Nevertheless, it is believed that the thin amorphous-layer improves the adhesion between Ta and Cu. In addition, the Ta film has the advantage of serving as a nucleation layer for Cu, enabling Cu films with (111)-orientation to be deposited.[25]

On the other hand, Ta barrier-layer films deposited by PVD are polycrystalline films. Thus, even though the reaction between Ta and Cu at 400°C occurs over only a short distance, the Cu may diffuse along the grain boundaries of the Ta film and escape into the neighboring dielectric material. Thus, by itself, Ta may not prove to be an adequate barrier-layer.

TaN is a related material to Ta, having a high melting point of 3087°C and a more dense microstructure. It is also typically deposited by PVD using a Ta target, but with nitrogen added to the argon gas in the sputtering gas ambient. Thus, as the reactive Ta is sputtered onto the wafer, it reacts with the nitrogen and forms a deposited film of TaN. The TaN material is more inert than Ta and does not react with Cu. In addition, TaN can be deposited as a polycrystalline film or an amorphous film by adjusting the nitrogen flow rate.

TaN films have been found to be better barrier diffusion-barrier films than Ta films. However, since TaN does not react with Cu, it exhibits worse adhesion behavior than Cu. For some Cu-CMP processes its adhesion properties are inadequate. To overcome the limitations that each film (i.e., Ta and TaN) exhibits when used alone, the implementation of bilayer films (TaN/Ta)[17] or trilayer films (Ta/TaN/Ta) have been investigated. These multilayer barrier-films reportedly offer better overall characteristics for Cu damascene interconnects.

TaSiN is another candidate for a Cu-barrier-layer material. It is formed by PVD of a TaSi target (with a low Si concentration) in an Ar/N_2 ambient. TaSiN is an amorphous refractory ternary material, and as such does not have any grain boundaries (which can serve as fast diffusion paths in polycrystalline films). It has been reported by Hara *et al.* that TaSiN exhibits better adhesion and barrier-layer characteristics than even TaN/Ta bilayers.[26,85] Furthermore, the same group reported that Cu-films deposited on TaSiN barrier-layers have a stronger (111) preferred orientation than do Cu films deposited on Ta or TaN layers.[27] A survey of over 75 papers from 1999-2000, indicated that TaSiN can maintain its diffusion-barrier properties for Cu to higher temperatures than Ta or TaN. However, since the Si containing ternary compounds (TaSiN and TiSiN) have higher resistivity, the Si industry tends to use TaN or Ta for Cu barrier-layers.[28]

16.3.7 Other Candidate Materials for Cu-Barrier Layers

Various other materials have been studied as Cu barrier-layers, and the most important of these (i.e., TiN, TiSiN, and WN) are discussed in this section. TiN is a good barrier-layer material for Al interconnects, and has been widely used for this application. While it also performs well as a barrier-layer for Cu, its behavior as an adhesion-layer and seed-layer for Cu is not as good as the Ta-based materials. CVD-TiN films have also been considered as Cu barrier-layer materials but they are reportedly less effective as diffusion barriers than PVD-deposited Ta and TaN films.[18,41] In addition, Cu films deposited on TiN have a weaker (111) structure than those deposited on Ta.[23] Hence, TiN deposited by either PVD or CVD has not been widely adopted for use in Cu interconnect metallurgy.[24] However, TiN deposited by ALD has been mentioned as a possible barrier-layer material for Cu. Work to deposit TiN by ALD has recently been reported.[29]

On the other hand, TiSiN-films deposited by CVD have been reported to exhibit significantly improved diffusion-barrier properties for Cu compared to TiN. The Si introduced into the TiN matrix bonds with the nitrogen (and this plays a role in producing improved barrier-layer behavior).[18] The TiSiN also has the additional benefit of good wetting to Cu, which has been correlated to improved Cu electromigration performance.[30] The CVD-TiSiN is deposited using TDMAT (tetrakis [dimethyl-amido] titanium) and silane as the precursors.[31,56] (Additional information CVD-TiN with TDMAT can also be found in Vol. 1, Chap. 6, Sect. 6.7.3.) The deposition of barrier-layers by CVD may become necessary for sub-0.1-μm technologies.

Tungsten nitride (WN) films deposited by CVD have also been considered as materials for Cu barrier-layers. Limited development of CVD-WN was described by Genus in 1999,[32] but they apparently decided not to commercialize this approach. Their decision may have been based on results indicating that such films are less effective than PVD-films of Ta or TaN as Cu diffusion-barriers.[18]

16.4 SEED-LAYER TECHNOLOGY FOR COPPER INTERCONNECTS

A *thin Cu seed-layer*, deposited by a method other than electroplating (e.g., PVD) must be deposited on top of the diffusion-barrier-layer.[9] This layer needs to be present to ensure that the subsequent Cu-electroplating operation can be performed successfully. Specifically, there are a number of requirements that the seed-layer must satisfy in order to enable the electroplating fill-process to be effectively carried out.[33] That is, if the seed-layer exhibits any one of a number of flaws, including poor morphology, inadequate sidewall- and bottom-coverage,

excessive overhang, or surface contamination, the electroplating results will be unsatisfactory (even if the best electroplating process is available). The details of the characteristics that must be possessed by the Cu-seed-layer are described as follows:

1. *Low-Resistivity* - At the start of the electroplating process, the seed-layer is the primary-conducting layer that carries current from the edge of the wafer to the center (allowing the plating current-source to contact the wafer only near the edge). Thus, the seed-layer should provide a low-resistance path from the edge to the center of the wafer, to make the current-density during the Cu-electroplating process uniform across the entire wafer (especially when the plating process is being initiated). Uniform-current density across the wafer-surface helps maintain a uniform plating-rate across the wafer.

2. *Film Continuity Along the Sidewalls and Bottom of the Trenches and Vias* - The electroplating process must also occur within the recessed features on the wafer surface. Since, the seed layer carries current from the top surface into the bottom of vias and trenches, the electrical path to the bottom of such features must be continuous and adequately thick. If there is insufficient seed-layer thickness at the bottom or sidewalls of the vias, this will lead to a slowdown in the plating-rate, which, in turn can cause a void during the Cu-fill process (Fig. 16-21a). To get adequate Cu-seed-

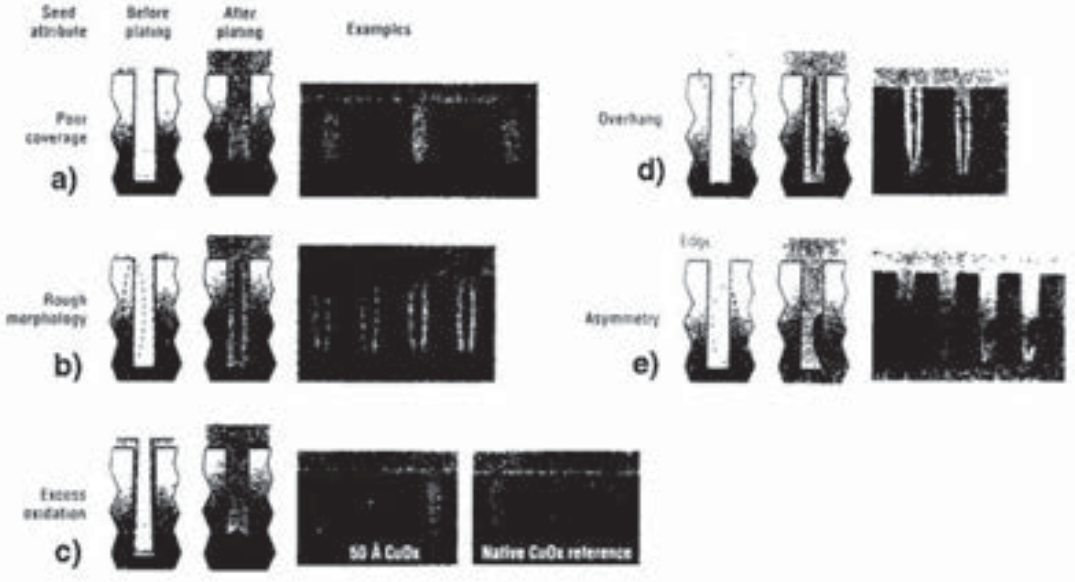


Fig. 16-21 Example set of illustrations and cross-sectional images showing key seed attributes and their impact on electroplating fill results.[33] Courtesy of Novellus Systems.

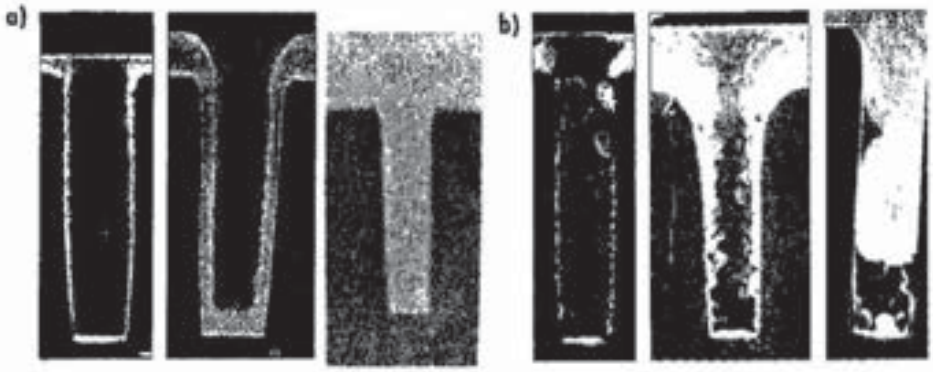


Fig. 16-22 Cross-sectional via images showing (left-to-right: a) Smooth seed coverage only, metal profile following partial fill, and complete fill by electroplating; b) Agglomerated seed coverage, metal profile following some electroplating, and resulting void at completion of electroplating.[33] Courtesy of Novellus Systems, Inc.

thickness along the sidewalls and bottom of deep-submicron, high aspect-ratio vias, either an ionized-PVD process or a CVD process is needed.

3. *Smooth Surface* - The seed layer must possess a smooth and continuous morphology. If the seed layer on the recess sidewalls is rough (or agglomerated) voids are also likely to form during plating. Rough sidewalls can lead to premature feature closure, and the discontinuity of an agglomerated film causes nucleation to fail at some locations. Either effect can give rise to creating a void (Fig. 16-21b). However, with a continuous, smooth seed-layer, plating readily occurs on all surfaces, thereby facilitating the subsequent as-desired, bottom-up filling (Fig. 16-22a).

It should be noted that the temperature of the wafer during the PVD process must be kept low to ensure that a smooth seed-layer surface is obtained (i.e., at room-temperature, or lower). Elevated wafer temperature during the seed-layer sputter-deposition process leads to a rough Cu surface (or even agglomeration, Fig. 16-23).

4. *Minimal Oxidation and Corrosion* – Oxidation and corrosion of the Cu seed-layer must be minimized both prior to plating and when the wafer is first immersed in the plating bath. Either effect can lead to seed-discontinuity and fill voiding. That is, the seed-surface must be largely free of oxide so that efficient charge-transfer can start to take place when the plating process is initiated. If the oxide on the PVD-Cu-surface is acid-soluble and sufficiently thin (i.e., in the thickness range of 10-20-Å), such initiation of the plating process is routine, and filling is complete.

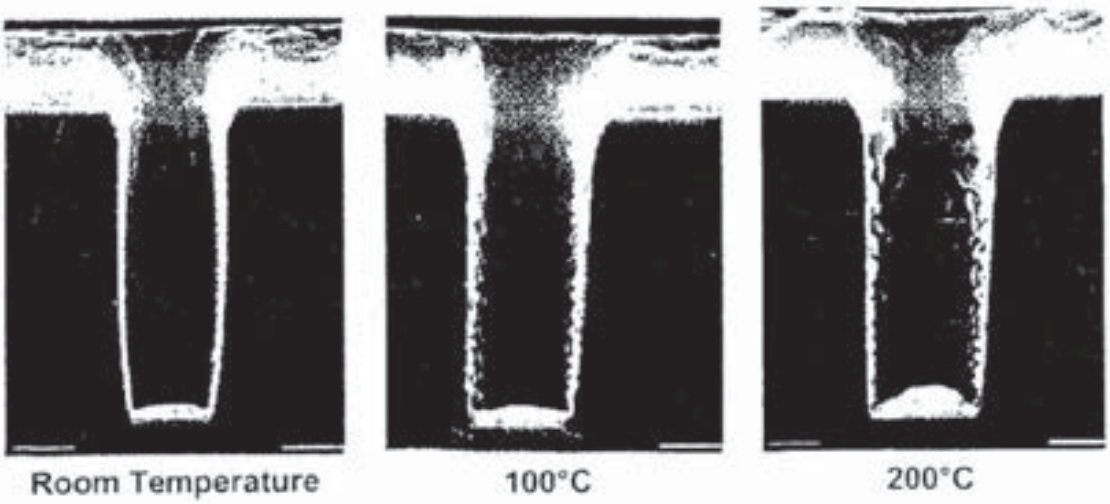


Fig. 16-23 Effect of wafer temperature on surface morphology of Cu deposited into vias by I-PVD. Courtesy of Applied Materials

However, when the cuprous oxides are thicker (i.e., 50-100-Å-thick), a greater degree of bottom voiding is observed (Fig. 16-21c).

5. *No Overhang at the Top Corners of Trenches and Vias* - Excessive overhang can lead to premature closure of the top of the recess during the plating process (Fig. 16-21d).

6. *No Asymmetry of Coverage Between Radially-Inward- and Radially-Outward-Facing Sidewalls* - Too thin a seed-layer on one side can impair nucleation and result in void-formation (Fig. 16-21e). That is, a much thinner seed-layer will form on that side than would be expected, and a void can thus result. This means that control of directional deposition during the PVD process must be maintained across the wafer plane from center-to-edge.

7. *Preferred Crystallographic Structure* – The crystallographic texture of the Cu seed-layer is influenced strongly by that of the underlying barrier-layer. This texture, in turn, determines the microstructure of the electroplated Cu deposited onto the seed layer. As noted previously, a (111)-orientation of the Cu is generally preferred because such films have the best overall film characteristics. A smooth Cu seed-layer-surface is also required to obtain this crystal orientation in the Cu layer that is subsequently deposited by electroplating.

In order to obtain a strong (111) Cu texture, it is also important that an *in situ* Cu-seed-layer deposition be performed. That is, the seed-layer

deposition process should be carried-out in the same cluster PVD-tool in which the barrier-layer PVD process is performed. This ensures that the surface of the barrier-layer (typically consisting of a reactive metal, such as Ta), does not form an oxide. If the barrier layer is exposed to air prior to Cu deposition, the resulting (111) Cu texture is much weaker than without an air break. This is probably attributable to oxidation of the barrier surface, and the resulting elimination of the pseudohexagonal surface that would otherwise allow for heteroepitaxial deposition.

Controlling the via shape itself is also important to ensure repeatable feature-filling. That is, the via-shape influences the filling by impacting the continuity of the seed-layer. Problems with the shapes of the vias often occur when trying to integrate an organic low-k ILD and Cu. That is, the via profiles of such materials frequently exhibit undercutting and notching as a result of the via-etch process (see Chap. 14). Such notches in the via sidewall profile can cause a discontinuity in the seed layer, and adversely impact the ability to completely fill the feature.

To get adequate seed-layer thickness into along the sidewalls and bottoms of high aspect-ratio recesses, either an ionized-PVD-Cu or a CVD-Cu process may be needed. (Ionized-PVD technology for such applications is discussed in Sect. 16.3, and additional information about seed-layer deposition with CVD is given in later paragraphs of this section.) As also noted in Sect. 16.3.2, the ionized-PVD process offers an additional advantage provided by the fact that a negative-bias can be applied to the wafer. This not only causes the ionized metal-atoms to preferentially strike the bottom of the via, but it allows the kinetic energy of these ions bombarding the wafer to be adjusted. For appropriate bias levels, this can cause re-sputtering from the feature base, which redistributes sputter-deposited material onto the lower sidewalls. The improvement in the conformality of the seed-layer thickness helps enhance filling of the recess without voids during plating. The impingement of the energetic ions on the top corners also helps keep the top opening from forming an overhang. Since the aspect-ratios of mid- and upper-level lines and vias is not likely to exceed 5:1, it is probable that ionized PVD techniques will continue to find use for these applications in future technology generations.

Unfortunately, as the feature sizes get smaller without a corresponding reduction in the ILD thickness, the aspect-ratio of the recesses gets larger. This is a situation that is likely to be encountered for the lower metal-levels of sub-0.1-μm interconnect structures. As noted in Sect. 16.3.1, for a PVD process the amount of material entering the recess must shared among all of the recess

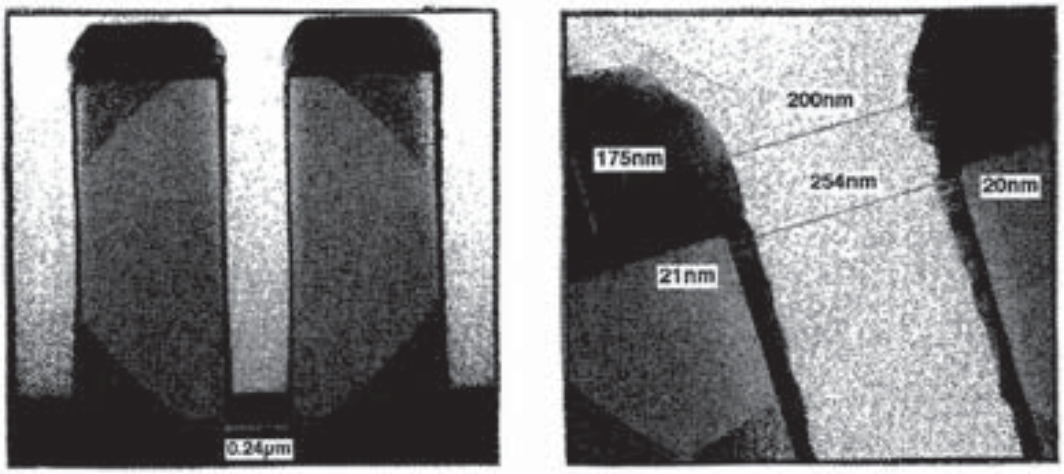


Fig. 16-24 SEM showing how thin the sidewall coverage is in high-aspect-ratio-features with I-PVD compared to the deposited layer thickness on the top surfaces of a wafer.

sidewall and bottom surface areas. For a via with an aspect-ratio of 5:1 the surface area of the sidewalls and bottom is 21x the surface area of the top opening. Thus, the material-thickness on all surfaces can only be 1/21 times as thick as the film deposited on the flat areas adjacent to the recess. Even at the 0.18-μm technology node, where the aspect ratio may be 5:1, a film-thickness of 2000-Å must be deposited in the flat regions to get a 200-Å-thick film on the via sidewalls (Fig. 16-24). Thus, at some time soon, it is predicted that PVD will no longer be capable of depositing an adequate seed-layer. Some have called the PVD-seed-layer the "weak-link" in the Cu-electroplating process.

At the point when the aspect-ratios for the lower levels of the interconnect system get to be too high to be adequately covered with any of the available PVD-processes (or if such a thick PVD-layer is needed that it causes too great an overhang), it may be necessary to switch to a CVD-Cu seed-layer. Some predict that this will occur for either the 100-nm or 70-nm CMOS nodes. Figure 16-25a shows a CVD Cu seed-layer on a CVD barrier layer deposited in a high-aspect-ratio recess that is only 0.14-μm wide. The CVD-Cu seed-layer is totally conformal and 500-Å-thick. The remaining 300-Å-wide slot was filled by electroplating without voids.[32] Figure 16-25b shows another example of a CVD-Cu seed-layer that is conformally deposited over high-aspect-ratio features.[34] One concern of using CVD-Cu seed-layers is the strength of their adhesion to the barrier-layer materials. Many freshly deposited barrier-layers (especially Ta-based), have highly reactive surfaces. These surfaces can degrade organic compounds resulting in interfacial contamination that prevents good adhesion.

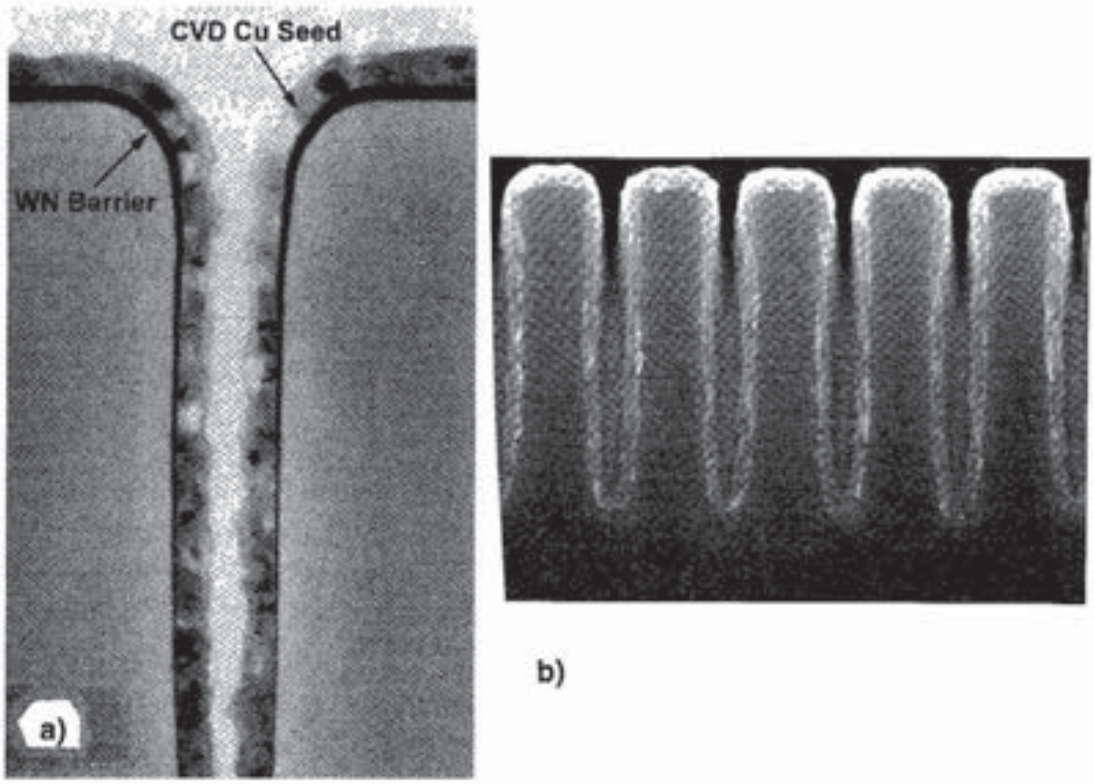


Fig. 16-25 a) SEM showing a CVD-Cu seed-layer integrated with clean and PECVD tungsten nitride (WN) barrier. (Courtesy Genus) b) SEM showing CVD-Cu seed-layer. (Courtesy Applied Materials)

This problem will need to be resolved by optimizing the barrier-layer choice, and improving the CVD deposition-process and precursor chemistry.

Two other alternative replacements for the PVD-Cu seed-layer have also been reported. The first is an electrochemical deposition process developed by Semitool. In their method, an ultra-thin Cu-layer (e.g., 200-Å-thick) is first deposited by PVD for adhesion purposes. This ultra-thin Cu-layer is then conformally thickened to provide a full-thickness seed-layer with an electrochemical process. Finally, the electrochemical process is switched over to a conventional electroplating chemistry to complete the feature fill.[35]

The other alternative replacement technology for PVD-Cu seed-layers is atomic-layer deposition (ALD - see Chap. 4, Sect. 4.6.1 for more background information on ALD).[36] ALD is a monolayer, stepwise growth-process that exploits the binding-energy-difference between chemisorption and physisorption. The growth proceeds by exposing the substrate-surface alternately to the precursors, and following each reaction, the excess species and by-products are removed with a nitrogen-pulse. The reactant-source pulses are thus separated by

a pulse of an inert gas. As a result, ALD is capable of depositing conformal-films with thicknesses that can be determined accurately by the number of deposition-cycles. The Cu precursor chosen for the work in Ref. 36 was $Cu(hfac)_2 \cdot XH_2O$ (see Sect. 16.5.3), and methanol, ethanol, and formalin were examined as the reducing agents. Smooth, conformal, near-bulk-resistivity Cu seed-layer films of 20-nm thickness with good adhesion were deposited by this ALD method.

As noted in Sect.16.3, another reported problem involving the PVD-seed-layer is its selective adhesion to barrier-layer materials. That is, the PVD-Cu seed-layers adhere well to Ta, but less well to TaN. Cu-films used with TaN barrier-layers can peel away during the later Cu-CMP process. Unfortunately Ta is not as good a diffusion barrier as TaN. To overcome this problem, use of a Ta/TaN bilayer has been reported.[37] Even better adhesion by the PVD Cu seed-layer was obtained by using a TaSiN barrier-layer.[26]

16.5 COPPER INTERCONNECT FILM DEPOSITION

Copper films for IC interconnect applications can be deposited in a number of ways, including: electrolytic-plating; electroless-plating; chemical-vapor-deposition (CVD); and physical-vapor deposition (PVD). Electrolytic-plating of Cu is the dominant technique used to fill recessed damascene features in deep-submicron ICs. Therefore, it is discussed in the greatest detail here (i.e., in Sect. 16.5.1). Electroless-plating of Cu and CVD-Cu are described more briefly in Sects. 16.8.2 and 16.5.3, respectively. PVD technology of Cu-films (used primarily to deposit Cu-seed-layers prior to the electrolytic-plating filling process), was covered in Sect. 16.4.

16.5.1 Electrolytic-Plating of Copper

As described in Chap. 15, dual-damascene processes have been developed for the multi-level Cu-interconnect structures of deep-submicron ICs. The damascene approach avoids the problems associated with having to etch Cu-films, but a challenge of a different sort arises from the need to fill the high-aspect-ratio damascene recesses without voids in the metal.

Line-of-sight PVD techniques are inadequate for filling such features because accumulation of deposits at the upper-corners of trenches closes-off the opening of the recess and this leads to void-formation. Chemical-vapor-deposition (CVD) has the potential for conformal coverage, but numerous chemical and hardware issues have hampered development of a manufacturable fill process for Cu. Thus, to the surprise of many in both the semiconductor and plating industries, electrolytic-plating has emerged as the deposition method of choice for

damascene-Cu interconnects (although advanced sputtering techniques are still needed for depositing thin barrier- and seed-layers, as described in Sects. 16.3 and 16.4) before the filling of the recesses with a Cu-plating process.

Electrolytic-plating (also often called *electroplating*) and *electroless-plating* of metals are older technologies still used in a variety of industries including hardware, auto, and electronics. In the electronics industry these processes have been (and still are) widely used in the printed-circuit-board (PCB) industry to plate copper layers on both sides of plastic boards for PCB interconnects. By the late 1990's, work was begun to develop electrolytic-plating processes for IC fabrication technologies (especially for depositing copper films for IC interconnect structures).

The key reasons why development of electroplating processes for on-chip metallization was pursued stemmed from the future requirements of IC interconnects. That is, it became clear that in deep-submicron ULSI technologies, low-temperature, low-cost, metal-deposition-processes compatible with thermally fragile low-k dielectrics would be needed. Electroplating of metals (especially Cu) represented a very attractive process for such applications. This is because deposition of Cu by metal-plating offers a high deposition-rate process (0.5-1.0-μm/min), which in turn permits high-throughput (50 wafers/hr), and thereby also a potentially much lower cost-of-ownership than PVD or CVD. In addition, such processes can be carried out at temperatures $\leq 40°C$, and a well-tuned electroplating-process can fill high-aspect-ratio structures without defects (i.e., trenches with aspect ratios of 5:1 can be completely filled). The details of electroplating copper-films for deep-submicron ICs are described here. As noted earlier, since electrolytic-electroplating requires a closed-conducting-path for continuous electric current, it is not possible to plate directly onto an insulator such as SiO_2. In this case, an electrically-conducting contact-layer must first be deposited (i.e., in Cu electroplating, a low-resistivity Cu seed-layer is deposited on the insulator).

Electroplating processes utilize conductive solutions containing ions of the metal to be deposited.[13,38] In the case of copper, the solution typically contains copper sulfate ($CuSO_4$ in the concentration range of 17.5-60g/L), sulfuric acid (H_2SO_4 in the concentration range of 45-325 g/L), and water. The $CuSO_4$ in this acidified solution breaks up into Cu^{2+} (cupric-ions) and SO_4^{2-} (sulfate-ions). The sulfuric acid added to the solution plays several roles: **a**) it dissolves any native copper-oxide on the copper seed-layer, so that the copper underneath is exposed, allowing the plating process to be reliably initiated across the entire wafer;* **b**) it

* Both CuO and Cu_2O (cuprous oxide) are unstable in a strongly acidic medium (pH < 3).

increases the conductivity of the electrolytic solution; and c) it improves the wetting capability of this solution. In commercial processes, proprietary organic additives are also added to the plating solution in order to improve the filling of recessed features by the deposited Cu film (as is described in more detail in Sect. 16.5.2.3). As with other types of deposition, the success of the process is measured by factors such as: deposition-rate, film-morphology, trench- and via-filling-capability, deposited-film-thickness-uniformity across-the-wafer and from wafer-to-wafer.** Plating chemistries were evolved from the plating bath formulations used in the PCB industry, and these were made to work for semiconductor applications.

In the Cu-electroplating-process for on-chip interconnects, wafers with a conductive top surface (which in this case is a PVD-Cu seed layer, as mentioned earlier) are immersed face-down in the electrolytic solution. The wafer-surface is also electrically connected to the negative-side of an external dc power-supply (by a conducting sealing-ring at the outer edge of the wafer, see Fig. 16-26a[50]), making it the *cathode* in this circuit. Electric-current is passed so that a chemical reaction occurs at the wafer surface which causes Cu plating to occur (with the cupric-ions that carry electric charge constituting the current in the electrolytic solution, see Fig. 16-26b). That is, as the positive cupric-ions arrive at the negative-biased-cathode they adsorb on its surface. They also acquire two electrons and are reduced to Cu metal (which plates out), with the reaction for Cu being:

$$Cu^{2+} + 2e^- \rightarrow Cu^0 \qquad (16.6)$$

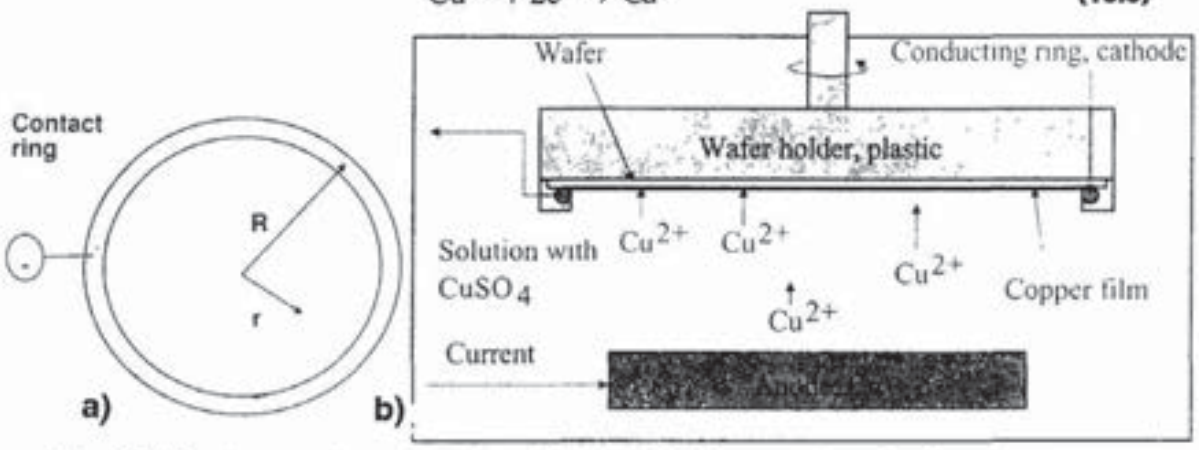


Fig. 16-26 a) Contact ring of a copper plating cell. b) Copper electroplating process.

** While the semiconductor industry has predominantly deposited Cu films electrochemically using acid baths, alkaline baths have also been considered. This is because alkaline baths can plate directly onto TiN diffusion barriers without a copper seed-layer and without the need for organic additives. On such bath studied by Semitool used an Cu-EDTA-complexed bath and the deposited Cu exhibited superior adhesion to TiN.[49]

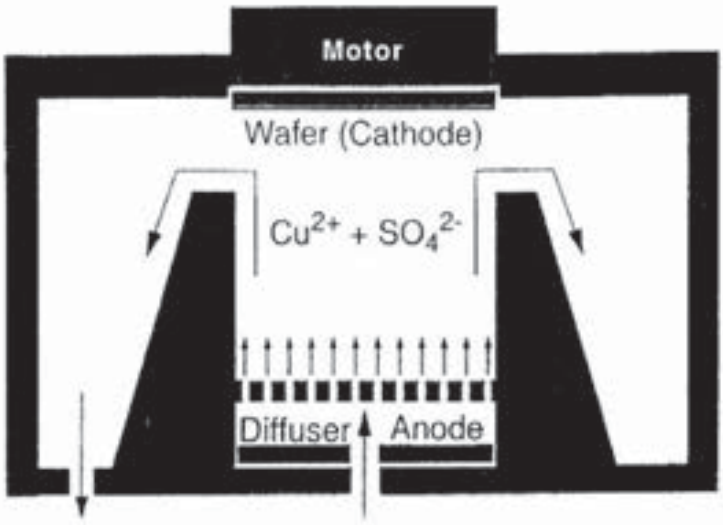


Fig. 16-27 a) Schematic drawing of a cross-section of a Cu electroplating cell.

Another electrically active surface (a solid-copper electrode which serves as the *anode*), is present in the conductive solution to complete the electric circuit (i.e., it is connected to the positive side of the power-supply). At the anode, the oxidation reaction occurs that balances the current flow at the cathode, thus maintaining the electrical neutrality in the solution. That is, all of the cupric-ions removed from solution by the plating of Cu onto the cathode are replaced by dissolution from a solid-copper anode immersed in the solution. (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.

If the applied potential is reversed, the cathode and anode are also reversed, and so are the locations where Cu deposition and dissolution occur. Alternating between deposition and dissolution (a process known as *pulse-plating*) may be useful in order to improve the filling capability of a plated film (see Sect. 16.5.2.5 for more details on pulse-plating). To enhance the uniformity of deposition across the wafer, the electrolyte is usually pumped against the wafer during the plating process, and the wafer may also be rotated at the same time (see Fig. 16-27).

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 the plating cell be designed so that no copper is deposited on the backside of the wafer.) The current delivered to a conductive surface is directly proportional to the quantity of metal deposited (according to

Faraday's Law of Electrolysis). By using this relationship, the mass deposited can be readily controlled through variations of plating current and time.

If no potential is applied across the interface between a metal and the electrolytic solution, an equilibrium potential exists between the two (see Fig. 16-28). Once the potential is shifted by an external power supply away from the equilibrium potential, a current will be driven across the interface. Under conditions typical of most plating processes, this current-flow is approximated by an exponential relationship known as the *Tafel equation*.

Figure 16-28 shows a current-vs.-potential curve of a copper electroplating process (also referred to as a *Tafel plot*). As the potential is increased from equilibrium to more negative values, the current increases in an exponential manner (Tafel region). In this region of operation the overall deposition-rate is determined largely by reaction-rate kinetics at the cathode. Ordinarily, conversion of Cu^{2+} ions to plated Cu-atoms occurs with nearly 100% efficiency, so that the electric current-density is a direct measure of the plating-rate. This also implies that ohmic effects (primarily through the uniformity of current distribution) are one of the important parameters that impact plating-thickness uniformity across a wafer. The strong dependence of current upon potential implies that the plating cells must be designed to enhance the uniformity of the potential across the wafer surface. The resistance of the barrier/seed-layer also impacts such ohmic effects, as will be described in later paragraphs.

As the potential continues to increase, mass-transfer effects become predominant and a limiting current-plateau is reached. This occurs because the Cu^{2+} species reacting at the cathode no longer reach the interface at a rate

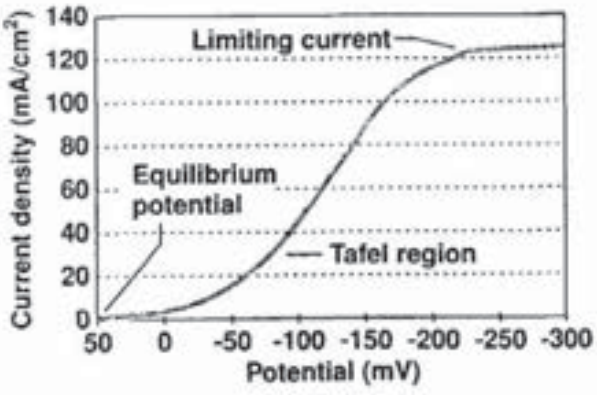


Fig. 16-28 A current-potential curve (Tafel plot) typical of a copper electroplating process.

sufficient to sustain the high reaction-rate. As a general guideline, Cu plating processes are operated at currents no greater than 30-50% of the limiting current to avoid undesirable film-deposition-characteristics.

To ensure that the rate of mass-transfer of Cu species to the interface is large compared to the reaction-rate (and that the mass-transfer rate is uniform across the wafer), the rates of migration, diffusion, and convection must be well understood and controlled. Convection is the most important mode of mass transfer in a liquid, and can vary from stagnant to laminar (or turbulent) flow. (See Vol. 1, p. 72 [Chap. 3] and p. 156 [Chap. 6] for a discussion of turbulent and laminar flow, and p. 157-8 [Chap. 6] for a discussion of mass-transfer in regions where movement by convection is zero [i.e., in stagnant regions].) In some electroplating systems, the wafer is also rotated while it is immersed in the electrolyte to get a uniform boundary-layer and to ensure that the fluid mechanics and material balance for the copper-diffusion match well everywhere on the wafer.

Generally, electroplating-processes have high deposition-rates (e.g., 400-1000 nm/min) and form desirable metallurgy. However, as noted in Sect. 16.4, 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 barrier/adhesion-layer. As noted in Sect. 16.4, the characteristics of the seed-layer are another of the critical factors that impact the ability to completely-fill recesses by electroplating without voids or other defects. 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 Ta, TaN, Ta/TaN or TaSiN is first deposited, with a Cu seed-layer applied in a second step.

Note that to obtain uniform deposition across the entire wafer, the current-density should be equal at all locations. Since the conducting layer has some resistance, there will be a voltage-drop from the edge-to-center of the wafer. This means that the edges of the wafer will plate faster than the wafer center (and the deposition-rate non-uniformity is greatest at the initiation of the electroplating process). The problem also gets more severe as the wafer diameter is increased, and thus is worse for 300-mm- than for 200-mm-wafers. To minimize such deposition-rate non-uniformity-effects, the voltage-drop from edge-to-center must be kept tolerably small. Since the barrier-layer is very thin and typically has a relatively high resistivity, a Cu seed-layer must be formed on top of the barrier-layer to increase the conductivity to a sufficiently high degree. (It has also been reported that reverse pulse-plating [as discussed later], will help reduce the non-uniformity of deposition, as the copper also dissolves more rapidly during the part of reverse-plating cycle in places where it is most heavily plated.[25])

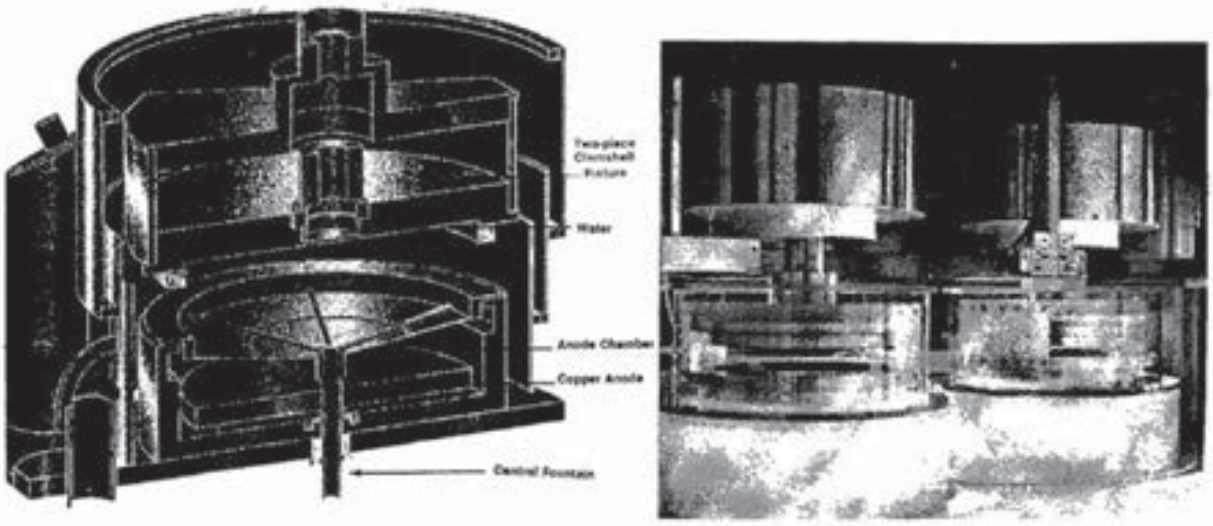


Fig. 16-29 Schematic drawings of plating cell of the Novellus SABRE™ electroplating tool. Courtesy of Novellus Systems.

Copper electroplating tools are commercially available from a number of companies. In 2000 the three companies with the largest market share were Novellus Systems, Semitool, and Applied Materials. The market for copper tools is expected to continue to grow significantly in the decade starting at 2001. Figure 16-29 shows illustrations of the Cu electroplating cell in the Novellus Systems SABRE™ elecroplating tool. Figure 16-30 is a drawing of the overall SABRE tool layout. This electrofill system enables void-free copper in extremely narrow high-aspect-ratio structures (up to 10:1) without backside contamination. Throughputs of up to 75 wafers are possible, with manufacturing availability above 90%.

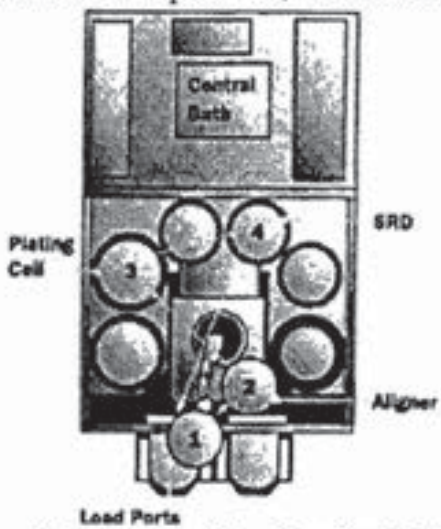


Fig. 16-30 Schematic drawing of the layout of the Novellus SABRE™ electroplating tool. Courtesy of Novellus Systems.

Figure 16-31 depicts the Applied Materials Electra™ Cu electroplating system. It has dual process cells and dual wafer handling blades for high throughput and reliability. Wafers are loaded into the twin plating cells. After plating they are processed in an edge-cleaning station followed by a spin-rinse-dry. The wafers are then moved to the integrated anneal chambers before being delivered back to the cassettes.

16.5.2 Filling High-Aspect-Ratio Recesses Without Voids Using a Cu Electrolytic Electroplating Process

The deposition process of Cu films for dual-damascene IC interconnects must not only have a high and uniform deposition-rate across the wafer, but it must also be capable of producing *void-free fill* of high-aspect-ratio damascene features. There a number of factors that impact the ability of the Cu electroplating process to fill such high-aspect-ratio recessed structures without the formation of voids or seams, including: **1**) the incoming feature profile; **2**) barrier-layer and seed-layer attributes; and **3**) key electroplating process parameters that are optimized to encourage acceleration of deposition near the base of damascene features ("bottom-up" fill). That is, for deep-submicron IC applications an accelerated plating rate is desired within submicron recessed features to achieve void-free fill. The attributes of the barrier- and seed-layers needed for permitting void-free fill were discussed in Sect. 16.3 and 16.4, respectively. The void-free filling factors involving the electroplating process are considered in the following subsections. They include: **1**) the electroplating solution must provide *rapid uniform wetting of the wafer surface*, especially within the area of high-aspect-ratio features; **2**) the deposition process must be *properly initiated*; **3**) appropriate

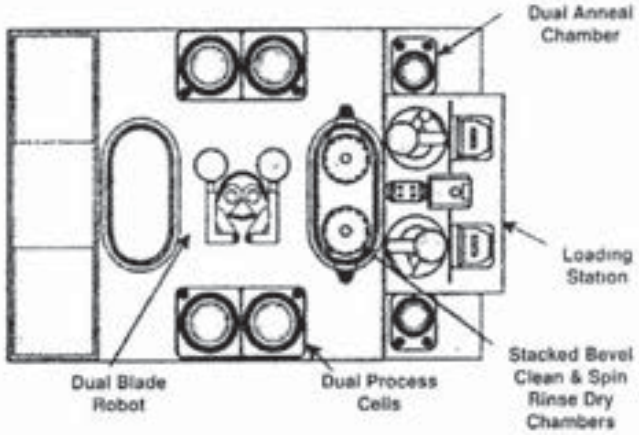


Fig. 16-31 Schematic drawing of the layout of the Applied Materials Electra™ Integrated ECP system. Courtesy of Applied Materials.

organic additives are added to the electroplating baths to facilitate void-free filling; and **4**) the *electric-current* waveform used for the deposition process can impact the filling characteristics. Topics **1**) and **2**) involving the electroplating process itself are described in Sect 16.5.2.1, while the effect of additives in the plating bath is covered in Sects. 16.5.2.2 and 16.5.2.3. The topic of plating-bath chemistry-management is discussed in Sect. 16.5.2.4 and the subject of pulsed plating in Sect. 16.5.2.5. The issue of topography in plated Cu-layers is addressed in Sect. 16.5.2.6.

16.5.2.1 The Electroplating Process Sequence: The electroplating process sequence typically consists of the following events. The wafers are immersed face down in the plating cell containing the electrolytic bath. Adequate surface wetting prior to initial deposition of the copper must occur. That is, rapid uniform wetting of the wafer surface and the inside of the high-aspect-ratio features is another critical requirement for good filling performance. For a solution with good wetting properties, capillary action should force solution into submicron features. Wetting is highly dependent on plating chemistry, seed-layer structure, composition, surface oxidation, and contaminants, as well as how the wafer is immersed into the plating bath and how the plating is initiated.[39]

The time the wafer is being wetted is referred to as the "zero-current-induction- time," or "induction-time" (see Fig. 16-32). During this time no external potential is applied, and thus no cathodic current flows. It represents the time period over which the wafer is exposed to the plating electrolyte at an uncontrolled (floating) potential. If the wafer is not polarized during this time, copper corrosion and cuprous oxide dissolution can occur.[40] To avoid the possibility of corrosion during the initial immersion period, the wafer may instead be introduced into the bath with an immediate small current flowing (or with a very-small cathodic potential being applied). In this way these corrosion reactions can be reduced or reversed. Immersing the wafers while applying a small cathodic potential is applied is thus referred to as *cathodic protection.*[33]

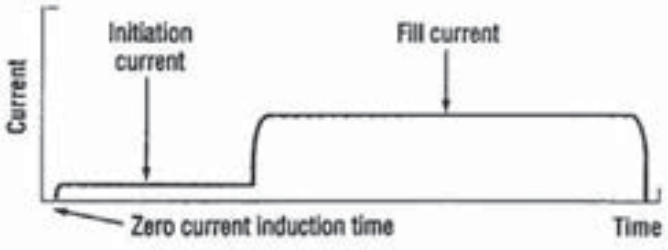


Fig. 16-32 Example of a two-step dc plating-waveform with three key features identified.[33]

After the induction time has passed, an "initiation current" is applied to the wafer (see again Fig. 16-32). During this time the current level is small, the deposition rate is slow, and the deposition occurs conformally. Thus, the seed-layer thickens uniformly during this period (allowing it to become continuous if a discontinuity in its coverage had occurred). In some cases, a thicker Cu seed-layer is needed to support the nucleation and high current density required in the subsequent fill step.

At the conclusion of the initiation-current stage, the "fill-step" stage is performed (Fig. 16-32). Since the damascene recesses are typically about 0.5-μm deep, and the Cu deposition-rate during the electrofilling stage of the plating process is ~1.0-μm/min, the fill-step lasts 20-40-sec, depending on bath conditions and feature size. Smaller features tend to fill earlier because they develop the conditions necessary for bottom-up fill (as will be described in the next sections). Since the larger features (i.e., with aspect-ratios much smaller than unity) fill more slowly, the copper layer must typically be deposited to a thickness somewhat greater than that of the damascene dielectric film stack. This issue will be considered in more detail later.

16.5.2.2 Additives in Plating Baths for Damascene Cu Interconnects: As noted earlier, the most important challenge for depositing Cu-films by plating for damascene applications is to completely fill vias, holes, trenches, and their combinations without voids or seams. For such defect-free filling, a higher deposition-rate along the sides and the bottom of the feature is needed than on the flat, top-wafer surface. As early as 1990 at IBM, it was discovered that plating from certain electrolytic solutions that contain additives leads to this kind of "bottom-up" filling. Since that time, work has continued to understand the role of such additives in plating solutions that will give the best filling results.

Additives to electroplating baths were originally developed to enhance the brightness and hardness of metallic deposits. Both of these attributes can be related to decreased grain-size in the deposits and reduction (or leveling) of the topography over rough surfaces. When the Cu-plating for submicron IC interconnects was being developed it was also observed that without additives the grain-size of the Cu was large (roughly 1-μm) and the film surface was rough. Additives led to a much finer grain-size (<1000-Å) and a smooth, highly reflective surface that had significantly less topography than the underlying substrate. From that work it was surmised that perhaps additional additives could be found that would also permit the damascene recesses to be filled without voids. It was discovered that a number of different additives (i.e., four or more)

need to be added to the plating solution to give the best filling results. (It also appears that the effects of these multiple additives are interactive.) Here such additives and their characteristics are described. In the next section a model is given for how these additives allow recesses to be filled from the "bottom-up."

The plating baths, as formulated by a number of specialty suppliers, include low concentrations (in the millimolar or micromolar range) of a variety of organic agents (generically referred to as *additives*) which serve to modify the properties of the deposited films as compared to results otherwise obtained from additive-free copper-sulfate/acid solutions. The manufacturing processes generally employ Cu plating-solutions purchased commercially, and the identity of the additives is proprietary. The examples of the various additives listed in this section come from the patent literature, as well as from various research articles that describe the additives being investigated. At the outset of the discussion the various classes of plating-bath additives are described. These classes include: suppressors, accelerators, and levelers (which, as will be noted, are also referred to by other names).

Accelerators are additives that speed or enhance deposition of the Cu during electroplating. Accelerators adsorb strongly on seed surfaces, where they also compete with suppressors for surface adsorption sites. Accelerators are usually unsaturated small organic molecules containing a polar sulfur, oxygen, or nitrogen functional group. In addition to increasing the Cu deposition-rate, they promote dense nucleation (and consequently the growth of films with fine grains). As a result, accelerators are also referred to as *brighteners* (i.e., because they cause the Cu-film to have a smooth, well-textured [i.e., bright] finish). Accelerators are added to plating baths in low concentration-levels (1-25 ppm). Examples of accelerators reported in the literature include: bis(3-sulfopropyl)disulfide (SPS);[41] sulfonated acetylthiourea;[42] 3-mercapto-1-propanesulfonate (MPSA);[33] and dibenzyl-dithio-carbammate.[43]

Suppressors (also called *inhibitors, carriers, surfactants or wetting-agents*) are plating-bath additives that reduce the Cu plating rate. Suppressors are usually present in the plating bath at higher concentrations (200-2000-ppm), so that their concentration at the interface is not strongly dependent on their rate of mass-transfer or diffusion to the wafer surface.

The suppressors are generally polymeric surfactants (with 1500 molecular-weight polyethylene glycol [PEG] reportedly being the most widely-used material for this purpose).[33] The suppressor molecules slow down the deposition-rate of Cu by adsorbing on the wafer surface, where they form a monolayer which offers a diffusion barrier to the cupric ions. In addition, they enhance the

cathodic polarization needed for fine-grain structure. In order for the polyethylene glycol to act as an effective suppressor, however, chlorine ions must also be added to the bath (in the form of HCl), and thus the chloride ion Cl^- can be considered a co-suppressor. In fact, HCl is the only non-organic additive in plating solutions.[33]

The role of chloride ions should be discussed in more detail. Chloride ions are believed to adsorb on the wafer surface, where they attract PEG molecules, and cause them to be more strongly adsorbed there as well. In fact, it has been observed that low chloride concentration leads to weak or negligible PEG adsorption.[44] This effect results in poor suppression of current at the wafer surface, and therefore the deposition on the wafer surface is not suppressed (which impairs the desirable condition of a differential deposition-rate between the top surface and inside the recesses that assists the complete filling of recesses). On the other hand, as the Cl^- concentration is allowed to increase to the appropriate level, good suppression of Cu deposition on the wafer surface is obtained. The levels of Cl^- within the recesses, however, might not be sufficiently high to maintain a high level of PEG adsorption, allowing accelerator molecules to be adsorbed on those surface instead. Thus, the conditions for good "bottom-up" filling are not disrupted. But too high a concentration of Cl^- concentration can interfere with "bottom-up" filling, because Cl^- can compete successfully with the accelerator molecules for surface sites in the recesses, and subsequently attract suppressing PEG molecules within the recess.

Levelers are plating-bath additives used to reduce surface roughness and to smooth the topography of the plated-film. They perform this role by locally reducing the Cu deposition-rate (i.e., by adsorbing on surface sites, blocking them from serving as locations where cupric ions can plate). Hence, levelers can also be considered as kin to suppressor additives. However, unlike suppressors, levelers are added to plating baths only in low concentrations (i.e., 1-25 ppm, comparable to the accelerator concentration). Hence, unlike suppressors, the concentration of levelers at the interface is mass-transport dependent and the large size of the leveler molecules also impedes their migration into trenches. In these ways, isolated locations (such as the insides of recesses, where mass-transport is limited) have fewer leveler species adsorbed on their surfaces. Thus, these regions experience less deposition-reduction effects caused by leveler adsorption than do the top, flat-surface regions of the wafer.

On the other hand, protruding-surfaces on wafer flat areas, or trench corners (to which mass-transport by diffusion or surface-migration is more efficient), experience a greater suppression of the deposition-rate. That is, the number of

leveler molecules adsorbed on wafer surfaces is continually being depleted during the plating process – by incorporation into the film or by electrolytic decomposition. The suppression of plating at protrusions on the wafer flat-areas and the upper corners of the trenches is maintained by constant replenishment of the leveler species that block surface sites (provided the leveler is present at the correct concentration). The suppression of the growth of protrusions on the wafer surface results in a more *level* film surface on the top-surface of the wafer (hence the name *leveler*). The suppression of deposition on the corners of the trenches allows the tops of the recesses to remain open while the "bottom-up" filling proceeds to completion. Examples of leveler additives listed in the literature are Janus Green B (JGB) a kind of dye,[45] and quatenary ammonium salt (QAS).[43]

For Cu-plating in dual-damascene applications, the composition and concentration of accelerators and suppressors are selected such that the accelerator surface-concentration dominates that of suppressors on the interior surfaces of trenches and vias. Local deposition-rates are thus suppressed on the flat surfaces on top of the wafer relative to the insides of recessed features, leading to the desired "bottom-up" deposition and void-free metal filling.

As a result, when these various additives are correctly balanced in the plating bath solution, it is possible to get a correct fill without voids and seeds. However, high precision is required to balance the roles of these materials. This topic is addressed in Sect. 16.5.2.4, where the issues of managing plating-bath chemistries are covered.

It should also be noted that in addition to void-free filling of recesses by Cu layers, it is also desirable to end up with a smooth (not rough) and topography-free (i.e., level) film after plating. This will allow a CMP process without an extensive overpolish time (i.e., that would be needed to get rid of the high spots that accompany rough and non-planar films) to be used. If an extensive overpolish time is required, this will lead to more severe dishing & erosion problems. Adding levelers to the plating bath and employing reverse-pulse plating waveforms are measures used to try and planarize Cu-layer topography.

16.5.2.3 Models of How Additives in Plating Baths Impact "Bottom-Up" Filling:

A model that describes the effects of the plating-additives on the "bottom-up" filling of damascene recesses was published by Reid in 2000.[33] Reid's model is based on a large volume of experimental work, and it proposes that the mechanism for "bottom-up" fill is primarily due to the accumulation of *accelerator species* within damascene features. It supercedes a model by

Andricacos in 1998 that held that the "bottom-up" filling was driven primarily by the establishment of a gradient of *suppressor polymers* in the recesses.[46]

Before presenting the mechanisms that Reid proposes are responsible for "bottom-up" recess-filling, a few preliminary remarks are in order. First it has been observed by others that if the co-suppressors PEG and Cl^- are the only additives present in a plating bath, feature filling is poor. If PEG and Cl^- and an accelerator (e.g., SPS) are the additives present, better filling is seen to occur. If all three additive types (i.e., suppressor, accelerator and leveler) are present, the best filling results are obtained.[47] Furthermore, the concentrations in the bath must be correctly balanced to obtain the best filling effects. This implies that the effects of the three types of additive are interactive, as is also proposed in Reid's model.

Figure 16-33 shows the possible ways for a profile of plated copper to evolve during plating. In conformal plating, a deposit of equal thickness at all points of a

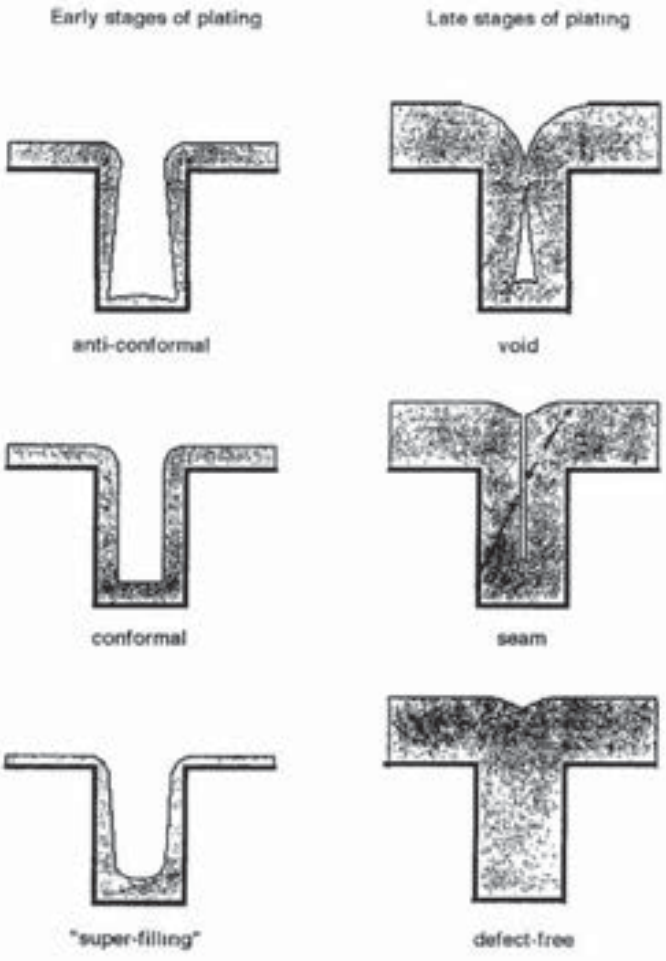


Fig. 16-33 Types of profile evolution in damascene plating.

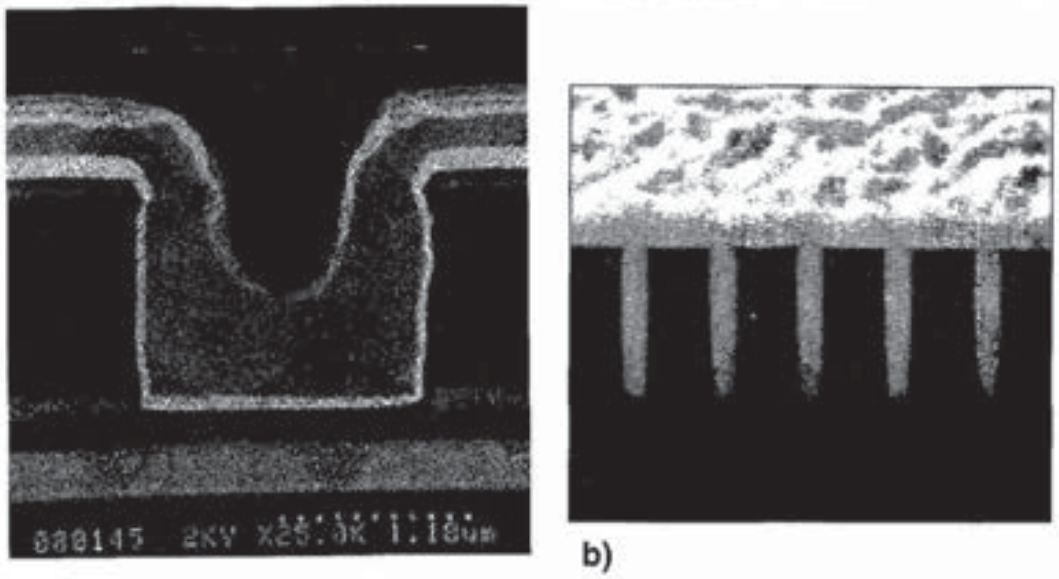


Fig. 16-34 a) Cross-section of a partially-filled line showing the profile of electroplated copper that exhibits the "bottom-up" filling phenomena.[48] b) Vias completely filled with Cu.

feature leads to the creation of a seam. If the shape of the feature is re-entrant, a void is formed. Subconformal plating also leads to void formation, even in straight-walled features. Subconformal plating occurs when substantial depletion of the cupric ions in the plating solution leads to significant concentration overpotentials, which in turn, cause the current to flow preferentially to more accessible locations outside of the feature. Thus, for defect-free filling, an increasing deposition-rate along the sides and bottom of the feature is desired ("bottom-up" filling). Fig. 16-34 shows a SEM of partially-filled line showing the profile of electroplated copper that exhibits such a "bottom-up" filling phenomenon.[48] Figure 16-35 shows SEMs of the time-evolution of filling a trench with a Cu-layer in a "bottom-up" manner.[49]



Fig. 16-35 SEMs of how "bottom-up" filling of Cu damascene recesses occurs versus time.[49]

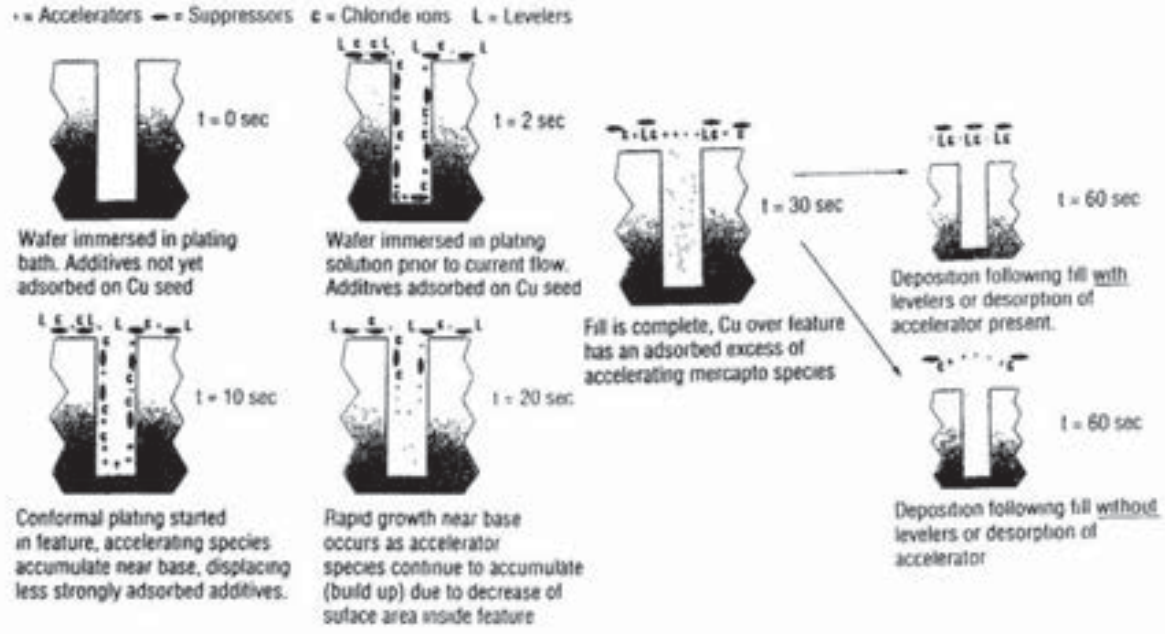


Fig. 16-36 Illustration of additive adsorption behavior during plating. A mechanism for the establishment and propagation of "bottom-up" fill is suggested. By permission of Solid State Technology. Published by PennWell.[33]

Reid's model of the sequence of events that accompanies "bottom-up" filling during Cu-plating is shown using Fig. 16-36. It begins by assuming that when a wafer is initially immersed in the plating bath (t = 0 in Fig. 16-36), that no additives are adsorbed on the Cu seed. However, within a few seconds (i.e., t = 2 sec in Fig. 16-36) both the suppressor and accelerator additives have been adsorbed on all areas of the wafer (including the insides of recesses) at their equilibrium levels. Thus, when the plating process is initiated by applying an external voltage to the cathode, the initial currents on all surfaces should be approximately equivalent on all surfaces. This leads to a deposition-rate that is constant everywhere. This perception is supported by the observed conformal deposition seen in the first 5-10 sec of a filling process (see also Fig. 16-35b - 10 sec). As was shown in Fig. 16-32, the plating current is kept low during this time (initiation current), to keep this conformal deposition thin.

After some time (which depends on current density, chemical concentrations, and feature size - e.g., t = 10 sec in Fig. 16-36), two phenomena are thought to begin to contribute to turn the fill process from a *conformal* filling-mode to a "*bottom-up*" filling-mode. First, the accelerator molecules start to accumulate at higher concentrations at locations near the bottom of narrow recesses. The increase in accelerator concentration near a recess bottom occurs because the surface area within the recess decreases as it fills (even if the fill is conformal),

and the accelerator molecules are neither incorporated in the depositing Cu nor are they desorbed into solution. (Recall from Sect. 16.5.2.2 that accelerator species adsorb strongly on seed-layer surfaces.) Meanwhile the PEG and Cl^- species are held less tightly on the seed-layer surface, and so they are displaced by the accelerator molecules and desorb into the solution. Thus, with a higher concentration of accelerator molecules in the areas of geometric concentration, the current-density (and the Cu-deposition-rate) is increased (t = 20 sec in Fig. 16-36).

The concentration of accelerator species in the solution must remain within an appropriate range. Too much accelerator in a bath disrupts fill because the accumulation of accelerating species also begins to take place on the top wafer surface between the recesses, and the differential deposition-rate with respect to the bottom of recesses is lost.

16.5.2.4 Electroplating Bath Control: Copper plating-baths are complex mixtures of chemicals, whose compositions contain various components that must be maintained at correct concentration levels. This is a challenging requirement because the composition of the bath changes as the plating process proceeds. The most dynamic ingredients of electroplating solutions are the organic additives. During normal electroplating the organic chemicals may be consumed in stoichiometric reactions (or they may break down in a variety of oxidative side reactions, some of which are non-stoichiometric). In any event, for consistent performance of the plating-bath, the original concentrations of these organics must be maintained, and their breakdown products must be removed. Even a small imbalance between components of the additives can cause defects in the filling of the trenches and/or degradation of the properties of the electroplated copper. As higher aspect-ratio-recesses are encountered, the ability to fill them properly will thus increase the importance of electrolyte management systems. (It should be noted that the bath solutions themselves are supplied by specialty-chemical vendors, such as Shipley Ronal.)

To enable the plating tools to be used in a high-volume production environment, the complex plating-bath chemistries have to be continuously monitored and controlled. Rigorous analysis of the plating bath chemistry has to be performed to determine how the bath composition changes with time as a consequence of plating, and then appropriate measures to counteract these changes must be taken. Such real-time analysis and feedback control of electrolyte are key to operating plating processes in a manufacturing mode. A self-adjusting, self-monitoring electrolyte is necessary to ensure precise control

of the plating chemistry to guarantee repeatable lot-to-lot fill. This has to be done on a large number of plating tools in a wafer fab that may be producing 20,000 wafers per month (with six or more levels of copper interconnect).

In some cases a separate, standalone, closed-loop analysis-and-dosing-system for multiple plating tools is used. In other cases the tool is designed with an on-board closed-loop electrolyte management system (e.g., as is done in the Applied Materials Electra ECP Deposition System™).[50] In the Electra tool, the bath management system measures (and then doses) the bath on-board at the system level, enabling precise control of suppressor, accelerator, leveler, copper, and acid concentrations. The system has a 200-liter tank with filtration on-board and six dosing channels. If the monitoring unit detects that, for example, suppressor x is under-concentrated, it immediately calculates how much is needed to dose the tank, and then does it. Feedback is on the order of minutes, rather than hours.

One method for monitoring the electroplating bath is a technique called *cyclic voltametric stripping* (CVS). The CVS principle has been widely used in the PCB industry, and was subsequently implemented for monitoring the electrolyte baths for IC Cu-plating. In the CVS method, a Cu film is deposited on an inert electrode and then this film is stripped by anodic dissolution. During stripping, the current is integrated over time to quantify the electric charge required for complete film dissolution. This charge is directly related to the molar quantity of the metal stripped (and thus to the amount initially deposited). From this information, calibration curves can be generated that provide reasonably accurate quantification of the total additive concentration in a plating bath. The stripping charge is monitored instead of the deposition charge because it is less sensitive to changing electrode surface states and impurity currents. Monitoring a process that proceeds to a well-defined endpoint also has inherent advantages.

The small plating cells used for plating Cu onto silicon wafers create very dynamic changes in chemical component activity, and this has generated a demand for a continuous, on-line, closed-loop analytical system. ECI Technology (Rutherford, N.J.) has developed such a CVS system for monitoring Cu-plating processes for silicon fabrication,[51] the Quali-Line QLC-5000.™ The CVS bath controller samples multiple plating tanks as well as a standard solution and then outputs the results.

The main limitation of CVS is its sensitivity to "matrix effects." Many bath components and their breakdown products display convoluted electrochemical interactions, so that the stripping-charge response is ambiguous if several components simultaneously change concentration. Thus, it is difficult to precisely correlate observed changes of electrochemical activity to variations in

individual bath components over extended times. These attributes make CVS less than ideal if the user's goal is individual monitoring of each bath component. However, under conditions where the principal requirement is to detect and correct changes in one or two electrochemically active components that have a rapid rate of consumption compared to others, the CVS method may be adequate for a closed-loop control system. In addition, CVS can provide rapid measurements with low-hardware investment, and little added waste-disposal burden.[52]

Another more elaborate plating-bath monitoring technology involves metal-free and ion liquid chromatography (also termed *high performance liquid chromatography* [HPLC] and *ion chromatography* [IC]). This method works on the principle of separating each of the individual components in the bath prior to detecting and quantifying them. It can provide information on the concentration and proportional balances of components in the plating solutions, as well as information on contaminants and degradation products.[53] A bath monitoring HPLC system for IC Cu-plating applications is being developed by Dionex Corp.[63] A short-coming of the chromatography methods include waste generation and the fact that the time required to perform the analyses can be long. It appears that a workable approach may be one in which both CVS and chromatography are integrated into a hierarchical method that exploits their relative strengths and minimizes their shortcomings.

When it has been determined that a plating bath needs to be adjusted (or replaced), several strategies can be employed. The simplest (and crudest) approach is to dump the bath when the plating results no longer meet the required specifications (and this is associated with a condition where the organic contaminant level exceeds a specified value). This technique is referred to as *batch dump,* or *bleed-and-dump-to-drain.* The level of contaminants in the bath is determined by correlating the Cu deposition tool throughput (and its related current-density) to the concentration of organic contaminants produced by breakdown of the organic additives. The increase in contaminant concentration is correlated to the amount of charge passed through the plating bath, and this relationship is expressed in units of *mg/amp-hr* (see Fig. 16-37). Thus, if this quantity is known (and the maximum acceptable contaminant-level is specified), the length time that a bath can be used before dumping is known. However, this method generates large quantities of waste that must be treated due to the copper and acid content. There is also no control (in either the general, or statistical sense), over the bath chemistry between dumps.

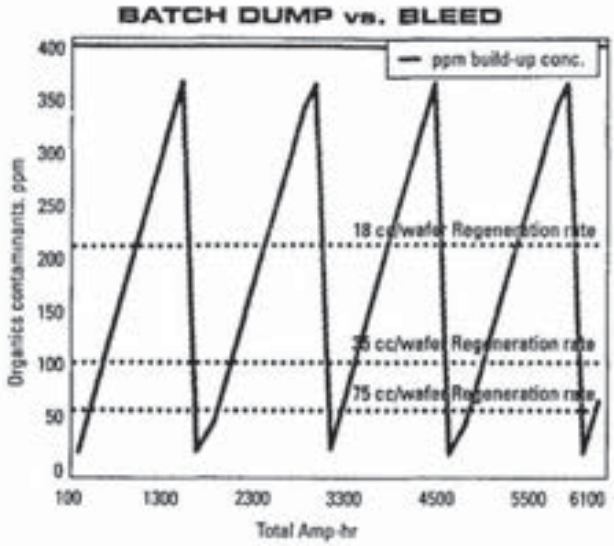


Fig. 16-37 "Batch dump" control versus "Bleed" control (0.3 amp-hr/wafer used to plate 1-μm thickness).[54] By permission of Solid State Technology. Published by PennWell.[33]

In a more refined technique, some fraction (typically about 10%) of the bath is removed daily, and it is replaced with new electrolyte solution. This is called the *bleed-and-feed* method. By analyzing a sample of the removed bath solution (e.g., with CVS), the concentration of additives in the new solution added to the bath can be adjusted to keep the total additive concentration within specified limits. However, as shown in Fig. 16-37, only a higher bleed-and-feed flow-rate can actively control the bath contaminant build-up that adversely impacts the bottom-up filling mechanisms. In addition, the bleed-and-feed method generates about the same amount of waste over time as the batch-dump method. Finally, while it does remove some of the contaminants that arise from breakdown of the organic additives, it does not completely remove them. It only dilutes them somewhat to a steady-state concentration. These drawbacks have led to a new approach to managing bath chemistry, referred to as a *remove-and-reconstitute* technique.

In this closed-loop method, the bath chemistry is continuously regenerated by stripping all organics (active and contaminant) from the recirculating bath. Then new organic additives (with the original concentration), are added back in. The bath solution is removed with flow-rates that are high enough to stabilize the concentration of contaminants at low levels. This flow rate is adjusted automatically, based on such inputs from the process tool as: amp-hours; wafer throughput; tool status, etc. The organic components are removed from the electrolyte by filtration, followed by their destruction using oxidizing agents

combined with heat and UV-light exposure. The organics are photochemically decomposed to carbon dioxide by exposure to the UV light - a clean process that involves no direct contact with the process fluid.

Process stabilization of both by-product and active organics are the benefits of this remove-and-reconstitute approach. The cost of consumables is also greatly reduced as only the organic additives are being removed (so only the organic additives need to be replaced). By using a closed-loop system, costly treatment of waste copper and acid is eliminated. A system in which this process is carried out is built by BOC Edwards. They call it the CuBIS™ (Copper Bath Integrated System), and it is shown in Fig. 16-38.[54] Its centralized monitoring system can simultaneously track five plating tools (from different equipment manufacturers).

16.5.2.5 Pulsed Plating: Utilizing pulsed-current waveforms (see Fig. 16-39b) during plating (instead of a dc-current) is a technique that complements the use of additives to obtain complete Cu filling of recessed features without voids. By periodically ceasing the current-flow, the plating reaction at the surface is halted in that time interval. The Cu filling process is improved for the following reason:

During electroplating a diffusion boundary layer is established between the bulk of the electrolyte solution and the wafer surface. In this boundary layer the Cu concentration trails off towards the wafer surface (since cupric ions are plated

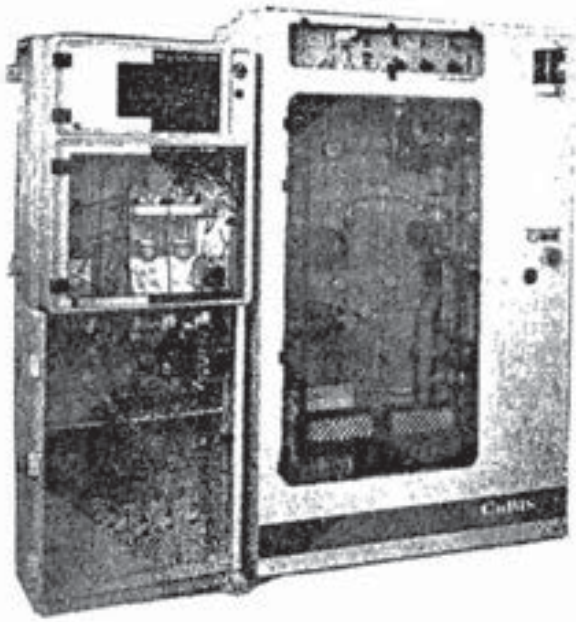


Fig. 16-38 Chemistry management is moving away from costly bleed-and-dump-to-drain approaches to bleed-and-feed. BOC's CuBIS Autorator uses a "remove-and-reconstitute" technique, which recycles the bleed stream, replenishing the bath at a rate automatically determined by plating conditions and by-product formantion. Courtesy of BOC Edwards.[54]

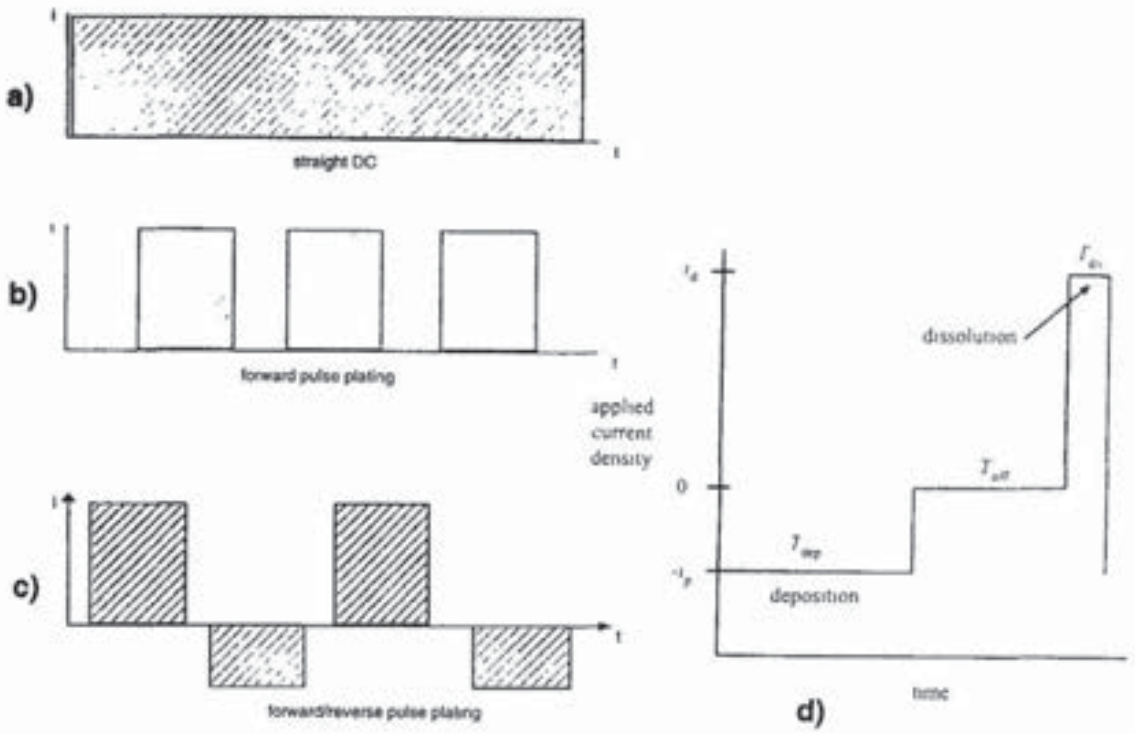


Fig. 16-39 a) dc current plating waveform; b) pulsed plating waveform; c) reversed-pulse-plating waveform; d) Example of reverse-pulse-plating waveform with optimized current amplitudes and pulse durations.[57]

out of the solution onto wafer surface). This boundary-layer is typically several microns thick. For a boundary-layer thickness that is larger than the pattern dimension, when the plating process reaches a steady-state condition, the copper-ion density is higher at the top-edges of trenches than at the bottom, Thus, for a dc-current-plating-process, a higher deposition rate is expected at the top of the trenches than at the bottom - and plating in this manner can lead to void formation.

However, by using pulsed-mode current, the concentration of cupric-ions depleted deep in the damascene recesses by the plating reaction can be restored to a higher level by Cu-ion diffusion to the bottom of the recess during such "rest-periods." This allows the deposition-rate at the bottom of the recess to equal the deposition-rate at the top when the current is restored during the next pulse of the waveform. This type of plating process is thus called *pulse-plating*. With an optimally shaped voltage pulse (both in amplitude and temporal shape), Cu filling of the recesses can be significantly improved.

Not only can the dc plating-current waveform be replaced with a pulse-waveform, but the polarity of the pulses can also be periodically reversed (Fig.

16-39c). The effect of reversing the polarity is to cause the Cu-film to undergo dissolution during the time polarity is reversed. Under the proper waveform conditions, the dissolution occurs more selectively at the top-corners of the trenches, thus allowing the trenches to remain open during the deposition-part of the plating waveform. Figure 16-39d shows an example of pulse-reverse waveform.[57] In one report, the application of periodic polarity reversal, together with adequately formulated and dosed surface-active additives allowed high aspect-ration via/trench filling capability. [58] One report observed that using small pulse periods gave better filling characteristics inside the trench. Using a small duty cycle and current density along with small pulse period appeared to give the best results.[59]

16.5.2.6 Copper Thickness Variation Within a Wafer After Electroplating: The copper topography across a wafer after the completion of a conventional electroplating process is not planar. Such within-wafer non-planarity is due to several effects. First, as discussed in Sect. 16.5.1, the higher current-density at the edge of the wafer next to the ring electrodes causes the Cu deposition-rate to be higher there than at the wafer-center. Thus, there is a buildup of copper thicker than at the center. One of the significant issues of achieving tight dishing and erosion specs after Cu-CMP is that these thicker Cu-regions require extra polishing time to get rid of them. This tends to increase dishing and erosion on other areas of the wafer.

Second, topography variations are caused by the feature-size-dependence of the filling process that is a characteristic of conventional electroplating. For example, the additives present in the electroplating baths used to enhance "bottom-up" filling of recesses with Cu can cause copper mounds to grow above regions containing dense arrays of narrow features (Fig. 16-40a). This occurs because the concentration of additives which accelerate the Cu-deposition rates in the recesses remain adsorbed on the top of the copper layer after the recesses

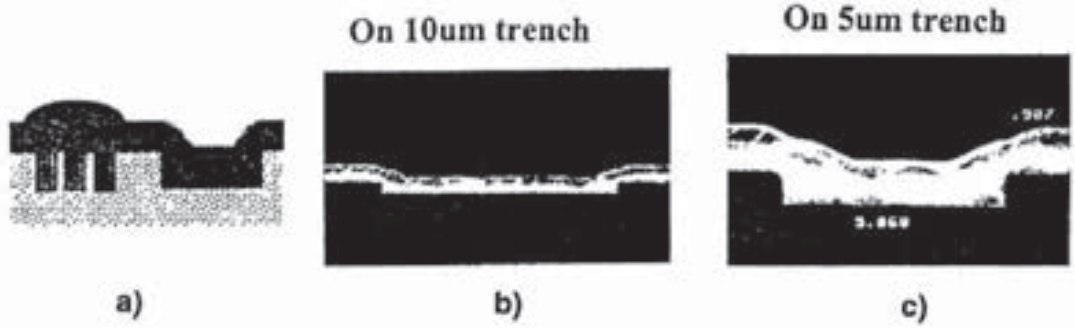


Fig. 16-40 a) Topography of a Cu-layer after plating into damascene recesses of varying widths; b) SEM of Cu plating into 10-μm-wide trench; c) into a 5-μm-wide trench.[61]

filling has been completed in those areas. This accelerates copper growth with respect to other areas of the wafer. If plating is done using electroltye solutions in which no additives are present to inhibit this effect, the copper layer over dense features can be twice as thick as the copper layer on wafer regions with no recesses. This reflects a continuation of copper growth beyond the time of filling completion, and the phenomenon is sometimes referred to as *overplating* (or *overburdening*, see also Fig. 16-36, t = 30 sec, and t = 60 sec).[60] The presence of mounds of excess copper in such regions again requires additional CMP over-polish. As a result, levelers are added to the plating bath (and reverse-pulsed waveforms are also used) to suppress the degree of such overburdening.

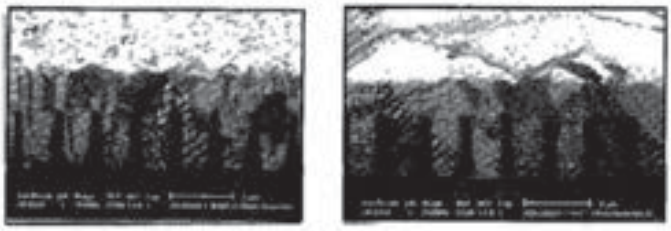
On the other hand, wide features do not fill as rapidly with Cu as narrow ones. Thus, there are steps in the copper layer at the edge of wide features. Figure 16-40a also shows such a step at the edge of a wide recess. Figures 16-40b and 16-40c show SEMs of Cu layers plated onto wide recesses, again showing the topography that results from a slower filling of wide recesses by plating relative that at which narrower recesses are filled.[61]

In summary, the topography that results from conventional electroplating processes exacerbates the problems of dishing and erosion that occur in a Cu-CMP process. Modifications to the plating process (such as those described in Sect. 16.5.4) have been proposed to alleviate this difficulty.

16.5.3 Post-Electroplating Cu Annealing and Recrystallization

When work began on copper-electroplating for IC interconnects, it was discovered that Cu-films are unstable after being deposited by electroplating, in that they self-anneal at room temperature (with the transient film-changes occurring over a period of hours to days). Furthermore, the transformation time is a strong function of film thickness. Thinner films take substantially longer to transform than do thicker films. In addition, the Cu seed-layer thickness also appears to play a role in the time and the degree to which annealing reduces the resistivity of the plated Cu-films. Greater resistivity reductions were observed as the plated films self-annealed on thinner Cu seed-layers than on thicker ones.[62]

In any case, if these variables are not controlled, each plated wafer headed for CMP may be different, depending on how much time has elapsed before it reaches the CMP step (a period that can vary depending on load size or tool selection). This can lead to wide distribution in electromigration behavior due to grain-size differences. To prevent such lack of control in plated-Cu-film characteristics, a post-ECP anneal at an elevated temperature was instituted. Such



a) No Anneal b) Integrated Anneal

Fig. 16-41 Comparison of grain size in damascene trenches: a) without; and b) with anneal.

post-plating anneals have since become a standard part of the Cu-deposition sequence.

The annealing process causes several changes in the properties of the as-plated Cu-films. First, grain-growth and strong (111)-texture is observed after 400°C annealing (Fig. 16-41). The as-plated grain-size of 1000 Å increases to several microns.[77] The increase in grain-size is significant because the film resistivity is also a strong function of grain-size. (Thin films with larger grains generally have a smaller resistivity than do those with smaller grains. This is because in films with large grains there are fewer grain boundaries – which cause carrier scattering, and thus degrade conductivity.) Thus, annealing of plated-Cu films also reduces their resistivity by up to 20%.[50]

Furthermore, annealing before CMP stabilizes both the Cu-grains and the Cu film-density (i.e., they no longer undergo changes with time at room-temperature). The result of such stabilization is not only a more controllable (and repeatable) Cu-CMP polishing-rate, but this rate also tends to be somewhat higher (by about 20-30%).[64] However, the temperature and time-duration of the annealing process can have an impact on the reliability of the Cu lines, and this must be considered when specifying the annealing conditions (as is discussed in more detail in Sect. 16.8.2).

To avoid having to raise the temperature of the Cu annealing process beyond the point where it makes no further improvements, studies of changes in the sheet resistance versus anneal-temperature and time were conducted. In one study it was found that a temperature of 200°C for 30 seconds would adequately reduce the resistivity and improve the Cu CMP removal-rate (Fig. 16-42). Thus, by designing a resistively-heated-pedestal-based annealing chamber within an electroplating tool enclosure, an annealing process can be integrated into the plating system. A chill-plate is also used in this arrangement to quickly cool the wafer to 50°C after annealing before exposing it to atmosphere (to keep the hot copper from oxidizing, as it would do if it were allowed to cool in air). The wafer

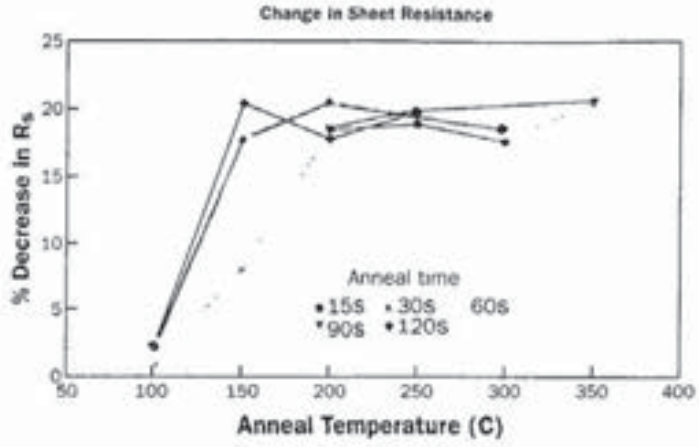


Fig. 16-42 Sheet resistance of Cu film as a function of anneal time and temperature after electroplating.[50]

is actually moved close to the chill plate (i.e., proximity cooling is used). To prevent the copper film from oxidizing during annealing, the ambient in the annealing chamber is maintained with < 5-ppm oxygen.

The Cu self-annealing effect can be explained as a recrystallization and purification process.[65] That is, an as-deposited plated-Cu-film is highly defective and the resistivity of the as-deposited film is high. The defects in the film are likely to be impurity inclusions caused by additives in the baths. (The large-grained deposits from additive-free baths do not undergo this recrystallization.) The inclusion of impurities at grain boundaries from the additives or their electrolyte by-products would be expected to cause pinning that would suppress grain growth. The as-deposited electroplated Cu film is not stable because of this high defect density. The film tends to reduce the defect level by recrystallization. During recrystallization, new grains are formed with fewer defects. These grains grow larger by consuming defective grains, leaving defects mainly in the grain boundaries and film interfaces. The stress in electroplated copper also changes from compressive-stress to tensile-stress after self-annealing. This is the macro evidence of film contraction caused by defect reduction.

16.5.4 ElectroChemical Mechanical Deposition (ECMD™)

As noted in Sect. 16.5.2.2, the Cu electroplating deposition-rate within recessed regions of a wafer exhibits a feature-size-dependence. That is, the Cu-deposition-rate is higher on areas of the wafer where there is a high-density of narrow, recessed damascene-features (and the plating process continues beyond the point

where the recess is filled). This results in mounding of the copper to a local thickness above those regions that is greater than in the flat regions devoid of recessed features (overburdening). In wide recessed-features, the deposition-rate is close to that on the flat-areas devoid of features, and hence a downward step can occur in the copper film when going from a flat area adjacent to a wide-recessed area. This has several detrimental consequences, especially with respect to the CMP process. That is, the thickness of the plated film must be increased to ensure that the wide recesses are completely filled. Thus, the process time of the CMP to polish the thicker Cu-film and the overburdened areas must be longer (which reduces the CMP throughput). There is also apt to be more Cu dishing in the wider-recessed areas. This problem would be overcome if the Cu could be deposited so that its surface after plating was globally planar. While techniques such as the use of leveler additives and pulsed plating can reportedly reduce the non-planarity to some degree, further improvement in the global planarity is still being pursued.

A novel process and deposition tool has been introduced by NuTool Inc. which seeks to address this issue directly. That is, in a process they call *electrochemical mechanical deposition* (ECMD™), as a copper layer is deposited in their tool (NuTool 2000™), a polishing pad is in physical contact with the wafer. This mechanical polishing coupled with the NuTool process chemistry simultaneously deposits copper and removes it. A differential growth rate between Cu deposition on the wafer flat areas and Cu deposition in the features is obtained. The speed, pressure, and mechanical path of the pad can all be controlled with software, eliminating the common problems of polishing copper, such as dishing and erosion.[66]

The ECMD process solves the problem of *excess-overburdening* (i.e., the buildup of excess copper over small densely arrayed features) and the downward steps that occur at the edges of wide recessed-features, by providing *aspect-ratio-independent filling* of features. Because the copper surface is not only globally planar but it is thinner in the field area than in a conventional Cu electroplating deposition process, the ECMD technique permits the CMP polish time to be reduced.

The tool and process were introduced in 2000, and thus information about its use in production is not available at the time of this book's publication. Some observers have commented that the ECMD process combines electroplating with some form of CMP on the wafer. Whether this combination of phenomena can be made to work simultaneously in a manufacturing environment will have to be demonstrated.

16.5.5 Electroless-Plating of Copper

Copper films can also be deposited by an electroless plating process, and this technology is already used in electronic packaging. Some work has also been done to develop Cu electroless plating processes for IC interconnect films. However, while this method has not gained the acceptance of electrolytic electroplating for forming Cu films, it is also briefly described here.

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-agents, and pH-adjusters on catalytic surfaces. The chemical process involves a CuO_4 and formaldehyde reduction:[67]

$$Cu^{2+} + 2HCHO + 4OH^- \rightarrow Cu^0(\text{solid}) + 2H_2O(\textit{liquid}) + H_2(\textit{gas}) \qquad \textbf{(16.7)}$$

An electroless Cu plating process for ICs was described in 1993 to give films with resistivities of 1.9 $\mu\Omega$-cm, close to that of bulk-Cu (1.7 $\mu\Omega$-cm).[68] Later work demonstrated filling of vias with aspect ratios of greater than 4:1 onto diffusion-barrier films of TiN.[69]

Although potentially a low-cost process because of its batch nature, electroless plating lost ground to electrolytic plating because of concerns due to hydrogen evolution, complexity of process control, low deposition rates, and potentially adverse environmental impact. However, it is possible that electroless plating will offer significant advantages for future wiring structures with improvements in sophistication and understanding. For example, some think that electroless plating may be used to deposit the Cu seed-layer at the 0.7-μm technology node.

16.5.6 CVD of Copper

While electroplating is the dominant method used to deposit Cu films for IC interconnects, work has been done on CVD-Cu deposition methods as well. One reason why interest remains in Cu-CVD processes is that it becomes increasingly difficult to deposit a uniform, conformal Cu seed-layer in the vias and trenches as their aspect-ratio continues to increase. At some point, it may not be possible to deposit a Cu-PVD-layer with sufficient thickness and/or continuity into such narrow features. A more conformal CVD-process could potentially produce a better Cu seed-layer in such high-aspect-ratio features. Thus, CVD-Cu technology approaches are briefly described here. (It should be noted however, that a copper PVD-process which deposits a very-thin nucleation layer will still be still needed before the Cu-CVD process.)

Organo-metallic sources are the most widely used precursors of CVD copper. These include Copper(II) {bis-hexa-fluoro-acetyl-acetonate copper} [$Cu^{II}(hfac)_2$] and Copper(I) {hexa-fluoro-acetyl-acetonatoL copper} or [$Cu^{I}(hfac)L$], 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*). $Cu^{II}(hfac)_2$ 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 $Cu^{II}(hfac)_2$. 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 $\mu\Omega$-cm.

The $Cu^{I}(hfac)L$ precursor (where the L = vtms) is commercially available as a product called Cupra Select,™ developed by Schumacher.[70] 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). This delivery method is needed for such precursors as $Cu^{I}(hfac)$-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:

$$2Cu^{I}(hfac)\text{-vtms}\ (gas) \rightarrow Cu^{0}(solid) + Cu^{II}(hfac)_2\ (gas) + 2\text{vtms}\ (gas) \qquad \textbf{(16.8)}$$

Both the $Cu^{II}(hfac)_2$ 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. Since the reaction is reversible, vtms can be used to dry clean the deposition chamber. By removing the Cu deposited on the parts inside the chamber, particle contamination by the film's cracking and flaking can be prevented. However, an issue that must also be addressed with CVD-Cu is wafer-backside Cu-deposition.

16.6 WET-ETCHING AND ELECTROPOLISHING OF CU FILMS

Work has also been carried out to develop processes that would replace (or shorten the process time) of Cu-CMP. Two basic approaches have been pursued: **1**) chemical *spin-etch planarization* (SEP) - in which the Cu film is etched with wet-chemicals using a special spin-processing tool; and **2**) electrochemically removing the Cu film with a process called *electropolishing*.

16.6.1 Spin-Etch Planarization of Cu Films

A SEP process was jointly developed by SEZ and Honeywell.[71] The SEP process removes layers of Cu material at a controlled rate by exposure to a reactive chemical solution while the substrate is spinning. Its goal is to remove the copper overfill material after a plating process with the same planarization results as CMP (but with no risk of damage to the low-k materials in the structure), and at much lower cost than CMP. Damage to the low-k is avoided and low-erosion of the dielectric is also achieved because there is no mechanical downforce as occurs in CMP.

During SEP the wafer is suspended horizontally on a nitrogen cushion above a rotating chuck (utilizing the Bernoulli principle). The substrate is held in place laterally with locating pins about the wafer perimeter. As the chuck and wafers are spun, wet-etchants are dispensed onto the wafer from a radially oscillating nozzle. An appropriate etching solution and spinning of the wafer result in a planar final surface (with the simultaneous radial and tangential etchant flow greatly enhancing planarization over simpler static etch processes).[72] DI water and nitrogen are then applied to the wafer to achieve rapid cleaning and drying. Etch rates of copper layers can range from 500-2000-nm/min, depending on the chemistry selected, and almost infinite selectivity exists with respect to etching of the underlying Ta and TaN barrier-layers.

By 2000, SEP was still in an early stage of development. However, it was demonstrated that the SEP process could provide a high degree of planarization for small features (0.35-0.5-μm-wide), and work was proceeding to improve the planarization of larger features in the 5-100-μm range. One way that SEP may become more practical is if a short CMP process (with a small downforce) could be used to knock off the large mounds of excess overburden present on parts of the wafer where high densities of narrow features are present. This would make the surface substantially more planar, making the SEP process more effective.

16.6.2 Electropolishing of Cu Films

Electropolishing of the overfill material after Cu electroplating is another process that has been pursued to eliminate or shorten the Cu CMP process. That is, after the electroplating step is completed, the current flow can be reversed, and the Cu film will then de-plate. The idea is to use such de-plating to remove the Cu overburden while leaving the Cu in the damascene recesses. Since reverse electroplating will remove Cu uniformly across the wafer, at first glance the concept appears feasible. However, it proves not to be practical as long as the

problem of aspect-ratio-dependent filling-effects occur during the initial electroplating process (see Sect. 16.5.2.6). That is, the non-uniform topography of conventional electroplating processes will result in overpolishing in the wide trench-regions and under-polishing in the circuit regions where there are dense-arrays of small-features. However, with the advent of a globally-planar Cu-deposition-process such as described in Sect. 16.5.4, electropolishing is once again being investigated.

In 2001 a Cu-electropolishing process was described in which the Cu-CMP step was either eliminated or significantly shortened.[61] A globally-planarized Cu-film was then deposited (such as was described in Sect. 16.5.4). With a globally planarized Cu film, the electropolishing can remove copper across the wafer without thinning the metal in wide trenches. In the Cu-CMP-free process that was described in Ref. 61, the Cu-overburden was electroplolished down to the barrier-layer. The barrier-layer was then removed by dry-etching with a fluorine-based plasma (that does not etch the exposed Cu remaining in the trenches). The wafer was then cleaned with brush scrubbing (Fig. 16-43). In the CMP-less process, the Cu-overburden was partially-removed by electropolishing and the remainder by conventional CMP. Less dishing could be achieved with the CMP-less process. A smooth Cu surface and highly (111) texture was also obtained by selecting the proper current density during the electropolish step.

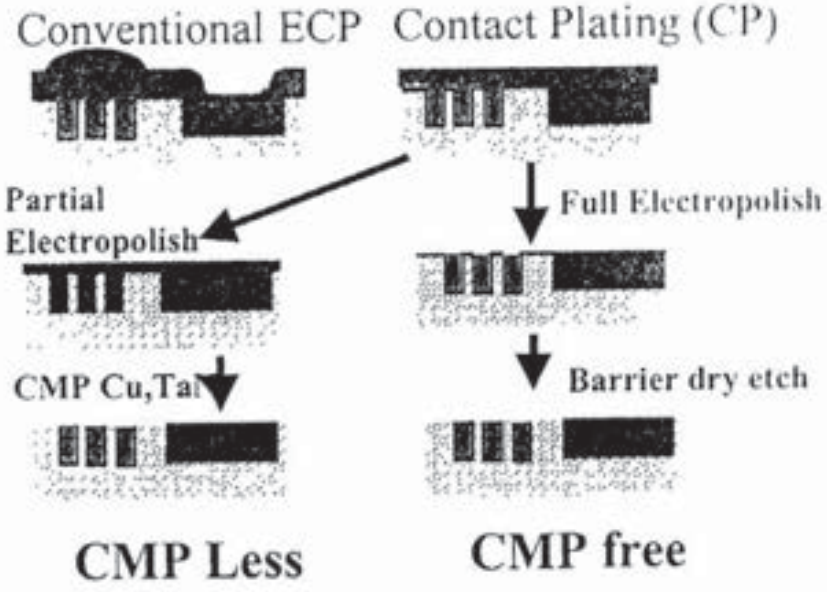


Fig. 16-43 Schematic diagram of CMP-less and CMP-free process flows by using a process that produced planar-Cu films after the plating process.[61]

A tool designed for precise electropolishing was introduced by ACM Research Inc. They refer to the process as *stress-free polishing* (SFP), and tool is called the ACM Ultra SFP.™ It is claimed that the SFP Ultra can polish Cu supported by ultra-low-*k* dielectrics (i.e., with a *k*-value less than 1.8).

16.7 CONTAMINATION CONTROL IN COPPER TECHNOLOGY

In Section 16.1 the issue of copper-contamination was introduced. Also briefly listed were the on-chip lines-of-defense and Cu-manufacturing protocols developed by wafer fabs to guard against this occurrence. Here additional information is provided about methods used to control copper-contamination in wafer manufacturing environments.

16.7.1 Sources of Copper Cross-Contamination in a Wafer Fab

In a completed device with a dual-damascene interconnect (and during many stages of the wafer manufacturing sequence), the embedded copper lines and vias are fully encapsulated with barrier-layers to prevent Cu out-diffusion into the surrounding dielectric and silicon areas. In those processes the likelihood of copper contaminating either other parts of the chip, or the wafer-fab environment, are small. However, in several process steps, Cu is exposed or is present outside of these sealed boundary areas (e.g., during those steps that occur from Cu-deposition to Cu-CMP). The most obvious risk of Cu-contamination occurs when large open areas of Cu are present on the wafer, or when Cu-containing chemicals or waste products - such as electroplating solutions or Cu-CMP slurry - are present. Another critical mode of Cu-contamination results from the high-energy bombardment of exposed Cu surfaces during plasma processing (e.g., via-open etch, and resist-strip). The following wafer processing steps (in roughly descending order of importance), can thus be considered to carry the highest risk of spreading Cu contamination - either into the deeper layers of an individual wafer or from wafer-to-wafer: **a**) Cu-CMP; **b**) Cu-deposition (including PVD seed-layer-deposition and electroplating); **c**) post-plating anneal: **d**) cap diffusion-barrier layer deposition; and **e**) via opening.[73]

During Cu-seed-layer deposition, electroplating, and post-plating anneal processes, the front side of the wafer surface is almost completely covered with a blanket Cu-film (and some Cu may find its way to the bevel and wafer backside during one or more of these steps). Therefore, cleaning of the wafer backside and bevel at certain critical steps during this stage may be necessary to prevent the spread of Cu particles from wafer-to-wafer by way of wafer-handling equipment (see Sect. 16.8.3). A barrier layer should also ideally extend over the whole wafer

surface, including around the bevel and on the backside to prevent Cu from diffusing into the wafer. This could be accomplished by using a wafer frontside barrier-layer (typically the conductive barrier-layer deposited before Cu), which overlaps a diffusion-barrier layer present on the wafer backside (e.g., CVD-SiN). Thus, the wafer bevel would also be completely covered by one or both of these barrier-layers. In many cases, the backside SiN diffusion-barrier-layer is not present. The backside and bevel cleaning is even more important in such cases. In some cases, if there is no backside diffusion-barrier layer, a silicon etch is performed after the backside and bevel cleaning step.[74] Its purpose is to remove a thin layer of silicon from the wafer backside that contains any copper that was deposited by the previous processes.

As noted above, contamination from copper can arise from tools and equipment involved in the deposition and handling of wafers. For example, a deposition tool that coats wafers with copper may cause it to be deposited on the bevel of the wafer. This wafer with copper on its bevel may be sent to a metrology tool equipped with a wafer handler that manipulates the edges of the wafers. (Such metrology tools are commonly shared among various process steps because manufacturers generally find that the acquisition of dedicated metrology equipment for copper processing is cost-prohibitive. These tools may include in-line SEMs for critical-dimension measurement, film-thickness monitors, and optical defect inspection devices.) The metrology-tool wafer-handler is now contaminated with copper, and can then cross-contaminate other wafers (for example, those that are on their way to an etch tool). The etcher thus eventually becomes contaminated with copper, and this contamination can spread throughout the fab (accumulating in plasma process chambers, wet benches, and lithography steppers). Contamination can also be spread by way of the wafer backside, as many wafer-handlers and robots hold and lift wafers by their backsides.

Wafer-handling by transport in cassettes is another likely source of copper cross-contamination. If copper is left on the wafer bevel as it exits a process tool, the carrier will invariably be contaminated with copper. Great care must thus be taken to either limit the use of wafer carriers to a specified area or to switch to a clean carrier once the copper is removed from the bevel. Wafer carriers may also need to be cleaned more frequently than otherwise required just for particle control. Furthermore, to prevent wafer-handlers and wafer-carriers from being contaminated by Cu particles on the wafer edge, it is preferable to maintain a sufficient edge exclusion of the Cu seed-layer or electroplated film. However, a full-faced seed-layer is typically required to allow electroplating with an edge-exclusion region of less than 4-mm. If dedicated cassettes are used to move the

wafers from the PVD-tool to the plating tool, and a cleaning step that removes the Cu from the bevel (and wafer backside) after electroplating is used, this problem can be circumvented.

Another source of possible copper cross-contamination is wafer breakage. Although the incidence of wafer breakage in processing tools has dramatically declined in recent years, it still occasionally happens. Thus, protocols for cleaning tools in which a wafer that contains copper has broken, are crucial to preventing factory-wide cross-contamination. A decontamination plan is required for each area and tool, which mandates appropriate cleaning methods and wipedowns in the immediate area in which the breakage occurred. However, there is no easy test for demonstrating the cleanliness of a tool after such breakage and cleanup, short of analyzing a test-wafer for copper-contamination.

Copper can reach the wafer bevel or backside during a deposition process. However, these surfaces can be protected by clamping and sealing the wafer hermetically during the plating operation so that only the face of the wafer is exposed during the process as required. The SABRE Electrofill™ tool from Novellus Systems uses a clamping method known as a *clamshell* to prevent the wafer backsides from being exposed to copper-containing solutions during the electroplating process. Surface analysis over a period of time during which 100,000 wafers were processed through a SABRE tool showed that a consistently low copper concentration of $< 10^{11}$ atoms/cm^2 were present (which was similar to the copper concentrations on control-wafers that did not undergo the plating process).[75]

The plating tool from Applied Materials (Electra™) uses a different approach. It contains a separate backside- and bevel-cleaning station integrated into the tool to treat each wafer after the electroplating step such that any copper that is deposited is removed (see Sect. 16.8.3 for more details).

16.7.2 Effects of Copper-Cross Contamination at the Back-End-of-Line (BEOL)

The concern for copper cross-contamination at the front-end-of-line (FEOL) stages of wafer processing is that the copper can diffuse into the active device regions and cause degradation of the gate-oxide-integrity or *pn*-junction leakage. The method of ensuring that copper-contamination is not occurring in the FEOL bays of a wafer fab is to monitor the MOS-C characteristics of wafers passing through the process line (much in the same way that is also done to safeguard against other forms of metal contamination - e.g., sodium).

However, the concerns at the BEOL phase of the wafer processing sequence are different from those feared at FEOL. That is, the solid solubility of Cu at

BEOL processing temperatures is relatively low, so not much Cu is expected to be incorporated into the silicon substrate from Cu on the wafer surface as long as the process temperature does not exceed 400°C. A test at SEMATECH took wafers with Cu plated on the backside, and put them into a furnace at 400°C.[76] Subsequent tests for leakage and threshold-voltage-shift showed no deviation from control wafers. Apparently the gettering mechanisms that protect devices from other contaminants can also handle levels of copper-contamination from such an event. As a result, worries about Cu entering the Si wafer at BEOL are probably somewhat exaggerated. On the other hand, the concern that copper will diffuse into the interlevel dielectric and cause leakage during BEOL processing is more real. The failure due to such a degradation of interconnect-leakage from copper contamination could also be one that develops slowly during the life of a part.

16.7.3 Cu Removal from the Wafer Bevel/Edge and Backside

Isolating Cu production from the rest of a wafer fab's manufacturing area is one approach to control copper contamination. But, many fabs share such critical and sophisticated tools as metrology and lithography equipment. Thus, such tools must accept Cu-processed wafers but it must be ensured that no contamination transfers to non-Cu wafers. While backside hermetic sealing and clamp rings are used to provide electrical contact and to mask the wafers exclusion zone from Cu deposition, they may not be 100% effective in preventing Cu deposition in the exclusion zone and on the wafer backside. This condition is further exacerbated by trends that minimize the exclusion zone to 3-mm (and even 2-mm), effectively putting the deposited Cu blanket onto the beveled edge of the wafer (Fig. 16-44a). Furthermore, since CMP cannot remove copper from the bevel and

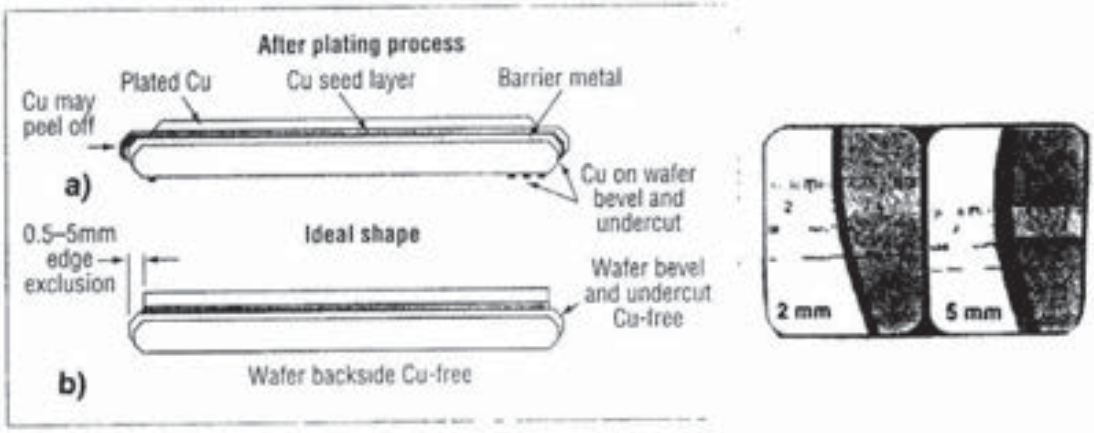


Fig. 16-44 a)Full coverage electrochemical Cu plating; b) The ideal Cu condition on the wafer after SPCE. By permission of Solid State Technology. Published by PennWell.[74]

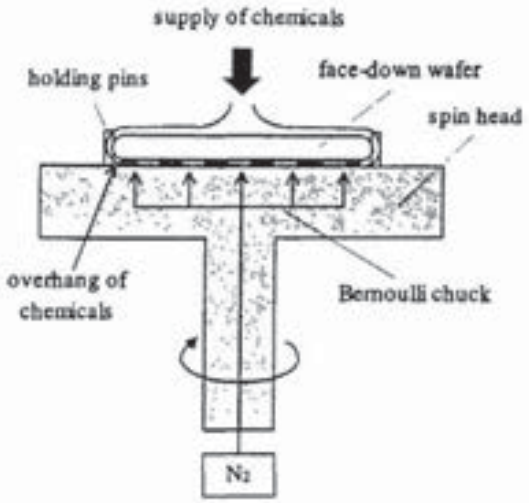


Fig. 16-45 Schematic diagram of the single-wafer cleaning system with a Bernoulli chuck.

backside, after a few layers, Cu-peeling (and resulting Cu cross-contamination) due to film buildup may occur.

The approach used to overcome this problem is to remove any copper that may have accumulated on the wafer-backside, bevel/edge, and exclusion-zone with a spin-etching process. Such a process is carried out immediately after the electroplating step is completed, and may be conducted either on a standalone spin-processor or on spin-processor that is integrated into the platform of the elctroplating tool.

A *standalone, single-wafer spin-cleaning-system* for backside Cu removal has been developed by SEZ Inc. They call their technology *spin-process contamination elimination* (SPCE™). During SPCE the wafer is suspended horizontally and frontside-down on a nitrogen cushion above a rotating chuck, utilizing the Bernoulli principle (Fig. 16-45). The substrate is held in place lateral with locating pins about the wafer perimeter. As the chuck and wafers are spun, the wafer is lifted about 1-mm. Wet-etchants are dispensed onto the wafer backside from a radially oscillating nozzle (e.g., modified HNO_3 chemicals are used for Cu removal), and the N_2 flow protects the entire front of the wafer (except the edge) from the chemicals. The degree of wraparound of chemicals onto the front of the wafer is a function of the N_2 flow-rate, the rotation speed of the chuck, and the viscosity and composition of the chemicals. This enables contamination and film removal from the back, edge/bevel, and front of the wafer without damaging the devices and without the use of masking materials on the frontside of the wafer. The width of the exclusion ring is tunable from 0.5-5-



Fig. 16-46 Edge-clean chambers in the Applied Materials Electra electroplating tool. (Courtesy of Applied Materials)

mm.[74] DI-water and nitrogen are applied to the wafer for rapid rinsing and drying. Fig. 16-44b shows ideal Cu coverage on a wafer after an SPCE process.

A backside and edge *Cu-removal spin processor* has also been *integrated* into the Electra™ electroplating tool of Applied Materials. After the electroplating process, the wafer is moved to the cleaning chamber, where it is loaded face down by a robot onto the self-centering wafer hoop. A wafer chuck holds each wafer, which is then lowered into a cavity with three chemical channels below and three above the wafer (Fig. 16-46). While the wafer is rotated, a DI-water rinse first rinses the electrolyte from the wafer surface. As the wafer is spun dry at high rotation speeds, the chemical dispensing arm moves into position to release its etching/cleaning chemical precisely at the edge of the wafer. Three sets of nozzles (120 degrees apart) enable simultaneous chemical dispensing for very short edge clean times. While the chemicals remove any edge and backside copper, DI-water is sprayed on the front side. By rotating the wafer and balancing the flows, it is possible to control how much material from the backside overwraps, either on the side or the bevel, or just above the bevel, or whether its kept from coming to the front side at all. Figure 16-47 shows the programmable edge exclusion with this edge clean process.[50]

An integrated bevel and backside cleaning process is also offered in the Novellus SABRE™ electroplating system. In this approach, the Cu is removed

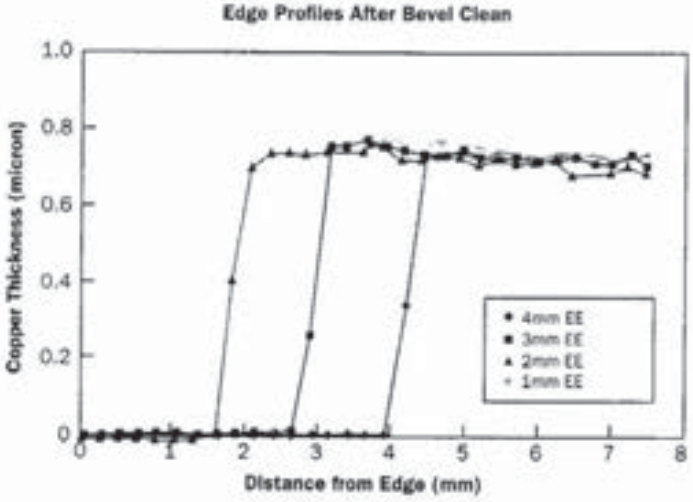


Fig. 16-47 Programmable edge exclusion with the applied Materials Electra™ edge-clean process. (Courtesy of Applied Materials)

from the bevel after the Cu seed-layer is deposited (i.e., prior to the electroplating process). The removal is carried out in the cleaning station of the tool.[78]

16.8 RELIABILITY OF COPPER INTERCONNECTS

As described in Sect. 16.1, copper lines offer the benefit of improved reliability over Al lines, primarily due to their superior electromigration lifetimes. However, the issue of copper interconnect reliability has other aspects as well, and some of these are described in this section.

16.8.1 Reliability of Copper Interconnects: "Missing-Metal" Defects

One Cu reliability issue involves post-CMP defects in narrow lines, which are referred to as "missing-metal" defects (Fig. 16-48). Such defects occur as a result

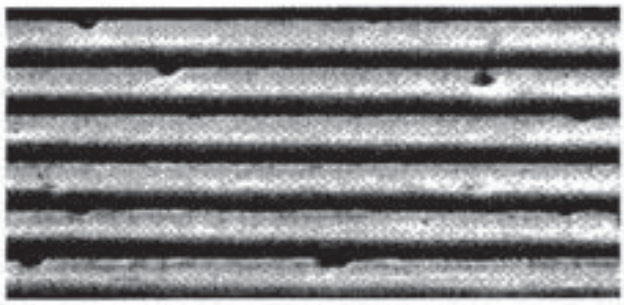


Fig. 16-48 Example of "missing-metal"-type defects. These defects arise from grain pull-out or corrosion and are exposed during the CMP process.[39] Reprinted with permission of Semiconductor International.

of corrosion or *grain-pullout* of lines exposed to a CMP-process. While the mechanism for these defects is poorly understood, it appears that their occurrence is influenced by a number of factors, including: plating-chemistry; plating-waveform; composition of the barrier-layer; and thermal-treatment. The latter effect appears to be most significant, and is thus discussed further here.[39]

Copper-lines are annealed after electrodeposition with the Cu overburden still in place. High thermal-excursions can cause large stress in the Cu films. That is, as the temperature is increased, the thermal expansion of the copper is greater than that of the surrounding dielectric material. This Cu expansion thus induces compressive-stress in the copper film that causes the film to re-crystallize and relax through dislocation motion. However, upon cooling the copper contracts and goes into tensile stress that is much higher under ambient temperature conditions than in the as-deposited film, and such stressed crystalline-films are prone to grain-pullout during CMP.

To reduce the incidence of defects related to thermal treatment it is necessary to minimize the temperature excursion of the post-deposition-anneal process (i.e., while the Cu overburden remains in place). Temperatures should be kept below the Cu plastic-flow limits (~200°C) so the copper can expand and contract elastically. Figure 16-49a shows the sensitivity of via-chain yield to the post-electroplating annealing temperature. Some have said that controlling the defects related to the copper process is perhaps the most important reliability issue facing the drive to achieve higher yields for copper in production.[39]

16.8.2 Reliability of Copper Interconnects: Plated vs. CVD Cu Films

Another reliability aspect observed in copper interconnects is that electroplated copper is more reliable than CVD-Cu with respect to electromigration failures.

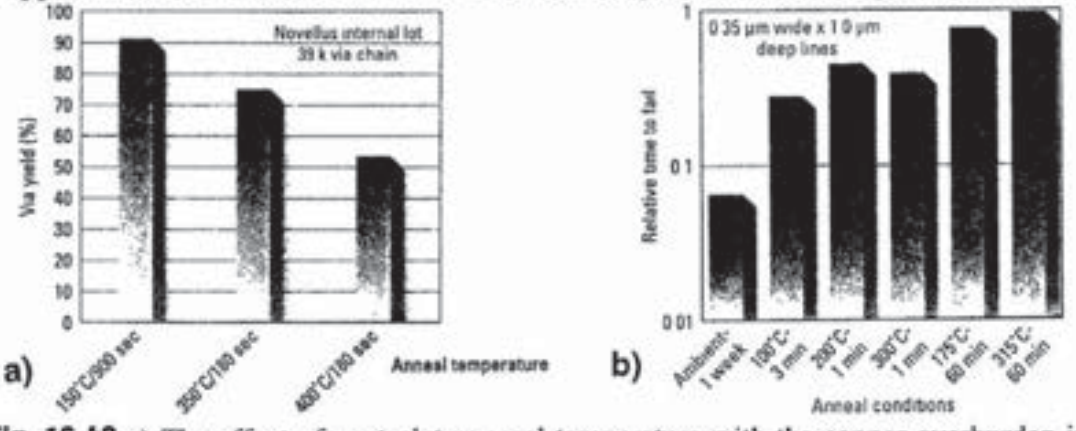


Fig. 16-49 a) The effect of post-plate anneal temperature with the copper overburden in place on via-chain yield.[39] b) Electromigration lifetimes for narrow lines (0.35-μm) with various pre-CMP anneals.[39] Reprinted with permission of Semiconductor International.

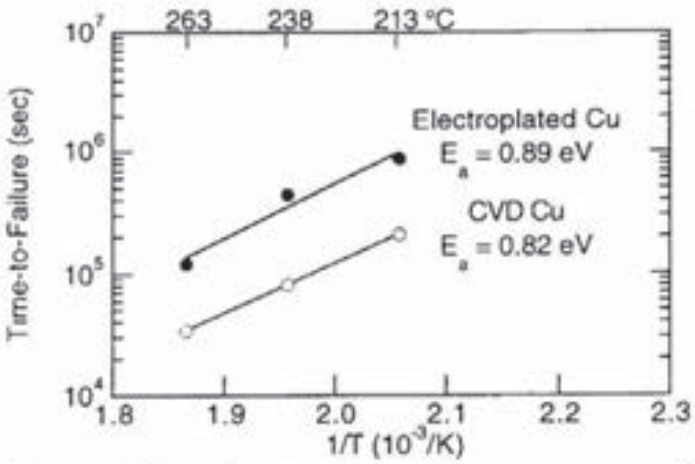


Fig. 16-50 Plot of time-to-failure of CVD-Cu and electroplated-Cu lines.[80] (© IEEE 1998)

As shown in Fig. 16-50, this is especially true for deep-submicron Cu linewidths. (The 0.4-μm-thick Cu-lines in Fig. 16-50 were tested for electromigration behavior after deposition, annealing [at 400°C for one hour], and CMP.[80]) Copper-lines deposited by electroplating exhibit longer lifetimes under electromigration-stressing than do CVD-Cu-lines because CVD-Cu has a smaller as-deposited grain size. In the narrow trenches of deep-submicron dual-damascene interconnects, electroplated-Cu has a bamboo grain-structure, which is conducive to superior electromigration behavior (whereas CVD-Cu has smaller grains after annealing than Cu deposited by electroplating, as can be seen in Fig. 16-51).[80,81]

As far as the electroplated-Cu-lines themselves, their electromigration reliability is impacted by the line-width and post-plating anneal-cycle. Wide lines anneal fully at the lowest anneal-temperatures, and when fully annealed they

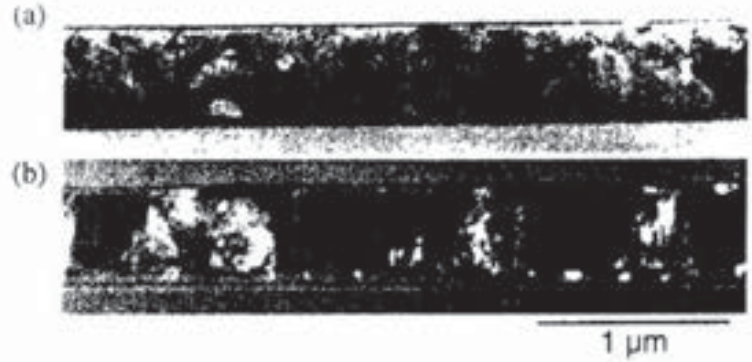


Fig. 16-51 Plan-view TEM micrographs of Cu lines in trenches after annealing: a) CVD-Cu; and b) electroplated Cu. (© IEEE 1998)

exhibit the longest electromigration lifetimes. Narrower linewidths require a higher-temperature anneal-cycle to achieve the same reliability behavior (see Fig. 16-49b, which shows the electromigration lifetimes of narrow copper-lines versus anneal-time and temperatures).[39] However, higher-temperature annealing cycles lead to greater "missing-metal" defect levels in narrow Cu-lines. Thus, a tradeoff must be made between deteriorating defect-performance and improved electromigration-performance as the anneal-temperature is increased.[39]

It was also reported that by incorporating a small amount of magnesium into the copper, various characteristics of the Cu interconnect were improved. That is, by forming a CuMg/Cu/CuMg stack, it was observed that the Mg diffuses through the copper and aggregates on the Cu surfaces. At the interface between the Cu seed-layer and the TiN barrier-layer this Mg improves the adhesion of copper to the TiN. At the top surface of the Cu-layers, the Mg forms a self-limiting oxide that hinders copper oxidation. Electromigration characteristics of such Cu/Mg stacked layers showed no difference relative to pure copper layers.[79]

16.9 ENVIRONMENT, HEALTH, AND SAFETY ISSUES OF COPPER TECHNOLOGY

The implementation of copper technology in IC fabrication introduces new chemistries, tools, and waste streams into the manufacturing arena. Before copper processes are put into production, it is important to address the *environment, health, and safety issues* (EHS) involved with them. Copper brings five new technologies to wafer fabrication, and their EHS impacts are discussed here: barrier-layer deposition; copper seed-layer deposition; Cu-fill by electroplating; copper CMP; and copper wafer reclaim.[82]

The first two of these processes, *barrier-layer deposition* and *Cu seed-layer deposition* are currently PVD processes which use solid sources, and hence do not involve any special EHS considerations. *Electrolytic plating for Cu fill*, however, is a new technology for the semiconductor industry, which also involves new chemistries, tools, and wastes. Nonetheless, commercially available electroplating chemistries used in the printed-circuit wiring board (PWB) industry have been evaluated and found to have no particular hazards or worker exposure issues. That is, all materials are TSCA listed (Toxic Substance Control Act). The primary concern is the two waste streams generated by the process: rinse water and a concentrated bath.

Copper CMP involves the use of a slurry that may or may not contain any chemicals that are environmentally hazardous. However, the process effluent is a potential environmental hazard, insofar as the it contains all the copper removed from the wafer by the CMP process. Development and manufacturing fabs

generate a large number of *test wafers that are reclaimed.* When these have a copper film, they cannot be reclaimed using existing methods. Nevertheless, relatively environmentally friendly techniques have been developed for this purpose (e.g., ammonium persulfate/sulfuric acid),[83] which is also compatible with electroplating waste. Nevertheless, the wafer reclaim process, like the electroplating process, generates the same two waste streams.

Thus, the two key steps in a copper dual-damascene structure – electroplating and CMP – are both wet-chemical processes. They use large amounts of water and create large amounts of acid and heavy-metal waste. They require large quantities of slurries, plating solution, and other liquid feedstocks. Bulk chemical distribution systems and waste-water treatment facilities for these materials can be built into new fabs. Accommodating them in an existing fab is more difficult. Typical fab delivery systems are designed around gases and DI-water, not slurries and plating chemicals.

16.9.1 Copper Waste Handling

The liquid chemical wastes from the various copper process steps described above must be evaluated for their toxic substance concentrations, and treated appropriately. It has been determined that the rinse water from the copper plating and wafer reclaim process ordinarily contains relatively low-levels of copper and contribute only minor increases in the total copper concentration of the industrial waste-water discharge. However, copper CMP contributes more than 80% of the discharged copper on a mass basis, and is likely to require treatment before being discharged to the industrial waste-water.

16.9.1.1 Handling Concentrated Waste from Plating and Wafer Reclaim: The liquid chemical wastes and slurries containing concentrated copper must be treated to remove the copper. A process similar to the electroplating process (called *electrowinning)* is one method for treating the concentrated plating baths and wafer reclaim wastes. That is, copper is reduced and plated out of solution at the cathode, and water is oxidized (split) to form oxygen at the anode. This process is capable of reducing the copper concentration in these waste liquids to levels below the current concentration based discharge limit for *publicly owned treatment works* (POTW). While, the time to reach such limits with the electrowinning process is 20-40 hours, it is still a cost-effective method and it does not generate other waste streams (e.g., sludge).

16.9.1.2 Handling Copper CMP Effluents: The treatment of the Cu-CMP effluent is done in another way. The primary goal of such treatment is to remove the

copper from the effluent. Recycling of the water is a secondary goal. In many CMP effluents the copper is entirely soluble and is not incorporated into the abrasive particles of the slurry residue. Thus, the treatment first involves microfiltering the effluent to remove the nonhazardous abrasive solids (which are then discharged into the fabs industrial wastewater and ultimately to the POTW). Next the water containing the slurry chemicals and copper is treated. The slurry oxidizer (e.g., H_2O_2) is removed by appropriate chemical addition. Then the copper can be removed by several different methods:

1. An ion exchange process can be employed, using cation-exchange canisters. These require periodic on-site or off-site regeneration (which again involves an electrowinning procedure). Such a treatment scheme can achieve very high levels of copper removal from the relatively dilute CMP effluent.

2. Coagulation and microfiltration is another method that can effectively reduce the copper content to less than 0.1 ppm.[83]

3. To insure copper removal to sub-0.1 ppm levels due to stringent POTW discharge requirements, it may be necessary to use water-soluble heavy-metal removal agents. These materials typically contain sulfide functionality, which bonds strongly to transition metals and forms insoluble precipitants.[84]

REFERENCES

1. P. Singer. "Tantalum, Copper, and Damascene: The Future of Interconnects," *Semiconductor International*, June 1998, p. 90.
2. R.L. Jackson *et al.*, "Processing and Integration of Copper Interconnects," *Solid State Technology*, March 1998, p. 49.
3. S.C. Sun, "Process Technologies for Advanced Metallization and Interconnect Systems," *Tech. Dig. .IEDM*, 1997, p. 765.
4. T. Takewaki *et al., Symp. on VLSI Technology, Dig. of Tech. Papers*, (1995), p. 31.
5. S. Venkatesan et al., "A High Performance 1.8 V, 0.20 μm CMOS Technology with Copper Metallization," *Tech. Dig. IEDM*, 1997, p. 769.
6. D. Edelstein *et al.*, "Full Cu Wiring in Sub-0.25-μm CMOS Technology," *Tech. Dig. IEDM*, 1997 p. 773.
7. M. Bohr, "Interconnect Scaling: The Real Limiter to High Performance ULSI," *Tech. Dig. IEDM*, 1995, p. 241.
8. C. Ryu *et al., International Reliability Physics Symp. Proc.*, 1997, p. 201.

9. L. Peters, "Leaders Describe 'Surmountable' Copper and Low-*k* Hurdles," *Semiconductor International*, September 2000, p. 25.
10. 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.
11. A.E. Braun, "Aluminum Persists in a Copper Age," *Semiconductor International*, August 1999, p. 58.
12. P.K.K. Ho *et al.*, "Direct Integration of Damascene Copper onto Aluminum Interconnect," *Abs. Mtg. Electrochem. Soc.*, Fall 2000, Abs. 525.
13. P.C. Andricacos, "Copper On-Chip Interconnections,," *Electrochemical Society Interface*, Vol. 8, No. 1, Spring 1999, p. 32.
14. S. Rossnagel and J. Hopwood, *Applied Physics Letters*, **63**, 3285 (1993).
15. S. Rossnagel and J. Hopwood, "Metal Ion Deposition from Ionized Magnetron Sputtering Discharges," *J. Vac. Sci Technol.*, **B12**, 449 (1994).
16. S.M. Rossnagel, "Ionized Magnetron Sputtering for Lining and Filling Trenches and Vias," *Semiconductor International*, February 1996, p. 99.
17. G.A. Dixit *et al.*, "Ion Metal Plasma Deposited Titanium Liners for 0.25 μm/0.18-μm Multi-level Interconnections," *Tech Dig. IEDM*, 1996, p. 357.
18. B.L. Chin *et al.*, "Barrier and Seed Technologies for Sub-0.1 μm Copper Chips," *Semiconductor International*, May 2001, p. 107.
19. P. Gopalraja and J. Forster, *Appl. Phys. Letts.*, Vol 77, 2000, p. 3526.
20. C. Michaluk *et al.*, "Tantalum 101: Economics and Technology of Ta Materials," *Semiconductor International*, July 2000, p. 271.
21. H.E. Roosendaal, "Sputtering by Particle Bombardment I," in *Physical Sputtering of Single-Element Solids*, Behrisch (Ed.), Springer-Verlag, Berlin 1981, pp. 219-56.
22. D. Tracy, "The Emerging Copper Revolution Will Impact ICs with Critical Dimensions Smaller than 100-nm," *Chip Scale Review*, March 2001, p. 83.
23. C. Ryu *et al., Proc. Mtg. Material Res. Soc.*, Spring 1998, p. 75.
24. C. Ryu *et al.*, "Barriers for Copper Interconnections," *Solid State Technology*, April 1999, p. 53.
25. S.S. Wong *et al,. Proc. International Interconnect Conf.* (IEEE) 1998, p. 107.
26. T. Hara and K. Sakata, "Stress in Copper Seed Layer Employed in Copper Interconnects," *Electrochem. Soc. Letts.*, March 2001, G77.
27. T. Hara *et al.*, "Control of the (111) Orientation in Copper Interconnection Layer," *Electrochem. Soc. Letts.*, March 2002, C41.
28. A.A. Istratov and E.R. Weber, "The Importance of Understanding the Physics of Copper in Silicon for Development and Characteristics of Copper Diffusion Barriers," *Abs. Mtg. Electrochem. Soc.*, Fall 2000, Abs. 522.

29. M. Juppo *et al.*, "Atomic Layer Deposition of Titanium Nitride Thin Films Using *tert*-Butylamine and Allylamine as Reductive Nitrogen Sources," *Electrochem. Soc. Letts.*, January 2002, C4.
30. S. Hirao *et al., Dig. Tech. Papers Symp. VLSI Technology,* 1997, p. 57.
31. J.-S. Min *et al.*, "The Mechanism of Si Incorporation with the Digital Control of Si Content during the Metallorganic Atomic Layer Deposition of Ti-Si-N Thin films," *J. Electrochem. Soc.* **147**, October 2000. P. 3868.
32. A.E. Braun, "Copper Electroplating Enters Mainstream Processing," *Semiconductor International*, April 1999, p. 58.
33. J.D. Reid *et al.*, "Factors Influencing Damascene Feature Filling Using Copper PVD and Electroplating," *Solid State Technology,* July 2000, p. 89.
34. A.E. Braun, "ECP Technology Faces Chemical, Dielectric Hurdles," *Semiconductor International*, May 2000, p. 60.
35. I. Chen and T. Ritzdorf, "Seed-Layer Enhancement in Ultra-Thin Seed for Inlaid Copper Metallization," *Abs. Mtg. Electrochem. Soc.* Spring 1999, Abs. 275.
36. R. Solankey and B. Pathangey, "Atomic Layer Deposition of Copper Seed Layers," *Electrochem. Soc. Letts.*, October 2000, p. 479.
37. T. Nogami *et al., Proceedings of the IITC*, p. 298 (1998).
38. J.D. Reid, "Fundamentals of Copper Electroplating," *Semiconductor International*, June 1998, p. 94.
39. V. Shannon and D.C. Smith, "Copper Interconnect for High-Volume Manufacturing," *Semiconductor International,* May 2001, p. 93.
40. S. Mayer *et al.*, "Integration of Copper PVD and Electroplating Processes for Damascene Feature Filling," *Abs. Mtg. Electrochem. Soc.*, Fall 1999, Abs. 732.
41. J. Horkans and J.O. Dukovic, "Monitoring of SPS-Based Additives in Cu Plating," *Abs. Mtg. Electrochem. Soc.*, Spring 2000, Abs. 360.
42. J.C. Hu *et al.* "Investigation of the Leveling Effect on Electrodeposited Cu Films for ULSI Applications," *Abs. Mtg. Electrochem. Soc.*, Spring 2000, Abs. 388.
43. P.L. Cavalotti *et al.* "Copper Electrodeposition for On-Chip Interconnections: Role of Different Additives," *Abs. Mtg. Electrochem. Soc.* Spring 2000 Abs. 359.
44. M. Yokio *et al.*, "Adsorption Behavior of Polyethylene Glycol on a Copper Surface in Acid Copper Sulfate Bath," *Denki Kagaku*, Vol. 52, p. 218, 1994.
45. J.L. Kelly *et al.*, "Leveling and Microstructural Effects of Additives for Copper Electrodeposition," *J. Electrochem. Soc.*, **146**, July 1999, p. 2540.
46. P.C. Andricacos *et al.*, "Damascene Copper Electroplating for Chip Interconnections," *IBM J. Res. & Develop.* **42**, 567 (1998).
47. J.J. Kelly and A.C. West, "Experimental and Numerical Study of Leveling of Submicron Features by Organic Additives," *Abs. Mtg. Electrochem. Soc.*, Spring 1999, Abs. 262.
48. P. Andricacos *et al.*, "Damascene Copper Electroplating for Chip Interconnects," *Abs. Mtg. Electrochem. Soc.*, Spring 1998, Abs. 254.

49. W.Y. Hsieh, "Interconnect Manufacturing Issues," *Advanced Interconnects: Design, Process, Integration, IEEE Short Course,* IEDM 2000.
50. S. Subrahmanyan ***et al.,*** "Integrated Electroplating – Key to Chip Production," ***Vacuum Technology & Coating***, March 2000, p. 50.
51. G. Chalyt and P. Bratin, "On-Line Analysis of Organic Additives in Damascene Copper Processes, *Semiconductor International,* April 1999, p. 66.
52. T. Taylor *et al.,* "Electroplating Bath Control for Copper Interconnects," *Solid State Technology* November 1998, p. 56.
53. "Analysis of Copper Plating Baths," Dionex Corp. Website: www.dionex.com.
54. O. Blachier and J. Clark, "Taking a New Approach to Copper ECD Bath Management," *Semiconductor International,* July 2000, p. 229.
55. M. Biberger *et al.,* "Low Pressure Sputtering of Copper and Related Barriers for Seed Layers and Complete Planarization," Novellus Systems Website.
56. C. Marcadal *et al.,* "Metallorganic Chemical Vapor Deposited TiN Barrier Enhancement with SiH_4 Treatment," *J. Electrochem. Soc.* January 2002, p. C52.
57. A.C. West *et al.,* "Pulse-Reverse Copper Electrodeposition in High-Aspect-Ratio Trenches and Vias," *J. Electrochem. Soc.*, **145**, September 1998, p. 3070.
58. S. Lopatin, "Extendability of Electrochemical Deposition for High-Aspect-Ratio Cu Interconnects," *Abs. Mtg. Electrochem. Soc.,* Spring 1999, Abs. 260.
59. D. Varadarajan *et al.,* "A Mass Transfer Model for the Pulse Plating of Copper Into High-Aspect-Ratio sub-0.25 Micron Trenches," *Abs. Mtg. Electrochem. Soc.,* Spring 1999, Abs. 268.
60. P. Singer, "Copper CMP: A Question of Tradeoffs," *Semiconductor International,* May 2000, p. 73.
61. M. Tsai *et al.,* "CMP-Free and CMP-Less Approach for Multilevel Cu/Low-*k* BEOL Integration," *Tech. Dig. IEDM,* 2001, p. 80.
62. B. Baker *et al.,* "The Interactions Between PVD Copper and ECD Copper as Observed in the Self-Annealing Process," *Abs. Mtg. Electrochem. Soc.* Spring 2000, Abs. 364.
63. E. Kaiser *et al.,* "Improved Determination of Additives in Acid Copper Plating Baths," *Abs. Mtg. Electrochem. Soc.,* Fall 2001, Abs. 1303.
64. T. Hara *et al.,* "Chemical Mechanical Planarization of Copper and Barrier Layers by Manganese (IV) Oxide Slurry," *Electrochem. Soc. Letts.,* December 2001, G109.
65. E. Shan *et al.,* "Effect of Deposition Temperature on ECD Copper Self-Annealing," *Abs. Mtg. Electrochem. Soc.* Spring 1999, Abs. 274.
66. J. Chappell, "NuTool Unveils ECMD Tool," *Electronic News,* 12/18/2000.
67. V.M. Dubin *et al.,* "Selective and Blanket Electroless Cu Platingby Contact Displacement for Deep Submicron Via and Contact Filling," *Proc. VLSI Multilevel Interconnect Conf.,* 1995, p. 315.

68. J. Van der Meerakker, "On the Mechanism of Electroless Plating - I - Oxidation of Formaldahyde at Different Electrode Surfaces," *J. Appl. Electrochem.*, **11**, p. 387 (1981).
69. A. Contolini *et al.*, "Copper Electroplating Process for Sub-Half Micron ULSI Structures," *Proc. VLSI Multilevel Interconnect Conf.*, 1995, p. 322.
70. J.A. Norman *et al.*, "New OMCVD Precursors for Selective Copper Metallization," *Proc. VLSI Multilevel Interconnect Conf.*, 1991, p. 123.
71. D. DeBear *et al., Solid State Technology*, March 2000, p. 53.
72. J.A. Levert *et al.*, "A Novel Spin-Etch Planarization Process for Copper Dual-Damascene," *Abs. Mtg. Electrochem. Soc.* Fall 1999, Abs. 731.
73. D. Dornisch *et al.*, "Cu Contamination Control for Advanced Interconnect Manufacturing," *Solid State Technology*, May 2000, p. 137.
74. P.S. Lysaght and M. West, "Addressing Cu Contamination Via Spin-Etch Processing," *Solid State Technology,* November 1999, p. 63.
75. T. Cacouris, "Preventing Cross-Contamination Caused by Copper Diffusion and Other Sources," *Micro*, July 1999.
76. L. Peters," SEMATECH Explores Copper Contamination," *Semiconductor International*, May 2000, p. 89.
77. M.E. Gross *et al,* "Implications of Damascene Topography for Electroplated Copper Interconnects," *Solid State Technology*, August 1999, p. 47.
78. For more information on this process visit www.novellus.com.
79. L. Peters, "Following a Steep Learning Curve for Copper," *Semiconductor International*, October 2000, p. 25.
80. C. Ryu *et al.*, "Electromigration of Damascene Copper Interconnects," *Dig. Tech. Papers VLISI Symp.* 1998, p. 156.
81. H.S. Rathore *et al., 16th International VLSI Multilevel Interconnect Conf.* (1999), p. 89.
82. L. Mendocino and P.T. Brown, "The Environment, Health, and Safety Side of Copper Metallization," *Semiconductor International*, June 1998, p. 105.
83. R.D. Letterman (Ed.) *Water Quality and Treatment*, McGraw-Hill, New York, 1999, 5th Edition.
84. J.H. Golden *et al.*, "Evaluating and Treating CMP Wastewater," *Semiconductor International*, October 2000, p. 85.
85. T. Hara *et al.*, "Improved Barrier and Adhesion Properties in Sputtered TaSiN Layer for Copper Interconnects," *Electrochem. Soc. Letts.*, March 2002, G36.

INDEX

Abruptness of doping profile in SDE, 508
Accelerators (Cu plating additive), defined, 757
Accumulation-layer resistance (R_{ac}), 198-200, 509
additives to Cu plating baths, 756-63, 764-9
Aluminum, CMP of, 418, 711-3, 715-7, 726
  Electromigration behavior of, 716-7
  Hybrid Al + Cu interconnects, 726
  Resistivity of. 712-3
Aluminum oxide (Al_2O_3), 146, 172-3
  ALD of, 172-3
Amine contamination of DUV resists, 230-3
  Filtering to prevent, 232-3
Annealing of Plated Cu films, 770-2
  Sheet resistance vs. annealing temp., 771
Anode-hole model of oxide breakdown, 94-7, 97-8
Anti-reflective coatings (ARCs), 237, 244-51
  Bottom ARCs, 245, 246-50
    Inorganic, 249-50
    Organic, 248-9
    PVD ARCs, 250
  Top ARCs (TARs), 245
  UARLs, 251
ARC-open step, 248
ArF excimer lasers, 265
Aspect-ratio, defined, 316
Atomic-layer deposition (ALD), 170-2, 172-3
Atomic-layer epitaxy (ALE), 170, 508

Backbone polymers in resists, 238, 242
  In 193-nm resists, 238
  In 157-nm resists, 242
Back-gate in FD-SOI MOSFETs, 524
BARC in dual-damascene process, 678, 686, 692-4, 698, 700
BARC veils, in dual damascene lithography, 692-4
Barrier layers for Cu interconnects, 720, 725, 728-40
$BaSrTiO_3$, 145, 146, 152, 162-9
  Deposition of BST films, 163-5
  Integration issues, 167-9
  Microstructure of, 166-7
Bath (Cu plating) control, 763-9
Beveled edge, wafer,
  Cu cleaning from, 403
  Polishing of, 59
BiCMOS, Si:Ge, 479, 491, 496, 498
Bilayer resist process, 237, 694-6. 697
  in dual-damascene process, 694-6. 697
Binary imaging mask (BIM), 285
Bird's beak, in LOCOS, 433
Bleed & dump-to-drain (Cu plating bath Management, 765
Bleed & feed (Cu plating bath management), 766
Body contact in PD-SOI MOSFETs, 520-22
Bond-and-Etch-Back (BESOI), 541-2
Boron penetration of gate oxide, 112, 138, 139
"Bottom-up" filling of recesses with Cu plating, 728, 756, 759-63, 769
  Models for, 759-63
BOX (see Buried oxide)
Bridging, in $TiSi_2$ formation, 615
Bright-field inspection systems, 410, 412
Brighteners (in Cu plating baths) def., 757
BST (see $BaSr\ TiO_3$)
Buff polish step, in CMP, 404, 416
Buried oxide (BOX)
  in STI, 435-6, 7
  in SOI, 502, 530, 532-3, 536-7

Calcium fluoride (CaF_2), 267, 267, 269, 270-2, 289-90
  Extrinsic birefringence, 271
  Intrinsic birefringence, 271-2
CARL resist process, 237
Catadioptric optical system, 269
Cathodic protection, 755

C-49 phase of $TiSi_2$, 615, 618, 619, 620, 621, 622
C-49 phase of $TiSi_2$, 615, 618, 619, 620, 621, 622
Chained implants, 186-7
"Champagne-glass" via, 579
Charge-to-breakdown (Q_{BD}), 98-9
Charge storage capacity in DRAM cells, 150
Chemically-amplified resists (CA), 228-9
Chemical mechanical polishing, 313-428
of Aluminum, 418
Cleaning issues, 392-406, 407, 408
Brush scrubbing, 399-404
Buff polish step, 404-5
Corrosion during, 418
Equipment, 406-7, 408
Megasonic cleaning, 404, 405
Metallic contaminants, 405
Steric Stabilization of particles, 398-9
VanderWaals force, 393-4
Vibrational scrubbing, 404
Zeta potential, 394-6
of Copper, 417-8, 726
Dishing, 421-26
Direct polishing, to prevent, 424-5
Dummy-active regions, to prevent, 423-4
in STI, 462-4
Nitride cap, to reduce, 423, 467
RIE/CMP, to prevent, 424, 464-5
Economics of CMP, 427
Edge effect in CMP, 364-5
Efficiency of planarization (EOP), 326
Endpoint detection in CMP, 385-9
Frontside laser interferometry, 388
In-line, defined, 385-7
In-situ, 387
Motor-current monitoring, 391
Optical monitoring, 390
Spectroscopy based, 389
History of, 333-5
of Interlevel dielectrics, 414
of Low-*k* Dielectrics, 420-1
Metrology of CMP,
Degree of planarity, 408
Thickness measurement, 407
Modeling of the Mechanisms of, 335-52
Metal removal, 349-52
Oxide removal, 341-8
Pads, 375-8, 382-4, 414
Slurry-free pads, 382-4
Pad conditioners, 378-82
Pattern dependence of planarization rate, 346-9
of Photoresist, 419
Polishing tools, 352-8
Fixed Table, 353, 356-7
Linear, 353, 355-6
Orbital, 353, 355
Reliability of, 411-2
Rotary, 353, 354
of Polysilicon, 418-9
Process integration issues of, 412-3
Removal rate, defined, 325
Slurries, 366-71
For metal CMP, 370-1
For oxide CMP, 367-9
Ceria, 369
Colloidal silica, 368-9
Fumed silica, 367, 368-9
Slurry distribution systems, 371-4
Slurry-free pads, 382-4
in STI, 438-9
of Tantalum (Ta), 417
of Tungsten (W), 415-17, 500-2
Wafer-to-wafer non-uniformity (WTWNU), 330-1
Within-die non-uniformity (WIDNU), 327-9
Within-wafer non-uniformity (WIWNU), 329-30, 339-40
Child's Law, 237
Cleaning Issues of CMP, 392-406
CMP (see chemical mechanical polishing)
Cobalt Silicide ($CoSi_2$), 203-4, 604, 610-11, 620, 624-34
Diode leakage in, 628-30, 633
Epitaxial $CoSi_2$, 630-1
Process sequence for forming, 625, 627-8

Silicon consumption during formation, 626-7
as Source/Drain diffusion source, 631-2
TiN capping layer, 629-30
Collimated sputtering, 731, 733
Contact region of drain, 10, 203, 516
Resistance of (R_{co}), 203-4, 208, 610, 612-14
Contact resistivity (ρ_c), 610
Salicide to silicon, 610
Copper (Cu), 711-89
Advantages of, 712-9
Barrier layers for, 720, 725, 728-40
CMP of, 417-8, 726,
CMP cleaning of, 403
Contamination control of Cu processes, 778-81
Effects of Cu on IC reliability, 780-82
Removal of Cu from wafer bevel and backside, 781-4
Sources of Cu contamination, 778-80
Corrosion during CMP, 418
CVD of Cu, 774-5
Difficulties of using Cu, 719-23
Diffusion barriers for, 728-40
Conductive, 720, 722-3, 728-40
Silicon nitride, 721
Tungsten-plugs, 721
Dishing during CMP of, 425
Effect of Cu on the reliability of ICs, 719-20, 723-4
Electroless plating of Cu, 774
Electrochemical mechanical deposition (ECMD), 772-3
Electromigration behavior of, 716-7
Electroplating of Cu, 747-69
Accelerators, def. 757
"Bottom-up" filling, 728, 756, 759-63, 769
Models for, 759-63
756, 759-63, 769
Brighteners, def. 757
Basics of, 747-51
Levelers, def. 758
Plating-bath control, 763-9
Cyclic voltametric stripping (CVS), 764-5
High-performance-liquid-chromatography (HPLC), 765
Plating tools, Cu, 753-4
Process sequence, Cu plating, 755-6
Pulse-plating, 752, 767-9
Reverse-pulse-plating, 768-9
Supprssors, def. 757-8
Electropolishing of Cu films, 775-8
Spin-etch planarization of Cu films, 776
Environmental, Health, and Safety issues of Cu processing, 787-9
Manufacturing protocols for, 721-2
Propagation delay in, 713-16
Resistivity of, 712, 713
Seed-layers of Cu, 725-6, 740-7
ALD of, 746-7
CVD of, 745-6
Desired characteristics of, 740-1
Preferred crystal orientation, 742-3
PVD of, 743-4
Self-annealing of plated Cu films, 770-2
Supprssors, def. 757-8
Topography of Cu films after plating, 769-70
Wet etching of Cu films, 776
COPs (see crystal-originated particles)
Corner effect in STI, 434, 451, 469-72
Corner parasitic MOSFETs in STI, 436
Corner Rounding in STI, 451, 458
Critical dose in high-dose SIMOX, 532
Crystal-originated particles (COPs) 21-32, 27-9, 47, 48, 64-5
Annihilation of, 27-9, 30
Defined, 23
Formation of, 23-5
In epitaxial wafers, 48
In fast-pulled wafers, 47
Current crowding in Cu vias, 717-8
Current-gain cutoff frequency (f_T), 481, 490, 494, 498
Cyclic order skipping, in film thickness measurements, 407

Cyclic voltametric stripping (CVS), 764-5

Damascene (see Dual Damascene)
DARC, 251
Dark-field inspection systems, 410, 412
Decel in low-energy implanters, 207
Deep-trench isolation, 490
Deep-UV resists, 228-44
Amine contamination of, 230-2
248-nm, 228-37
193-nm, 237-41
157-nm, 241-4
Poisoning, 690-1
T-top formation in, 230-2
Defects in CMP, 408
Types, 408
Degree of planarization, 313-5, 326
Measurement of, 408
Denuded zone, 39, 40-1, 45, 50-2
In "New" wafers, 48-50
Magic Denuded Zone, 50-2
Dielectric isolation, 517
Direct polish process to prevent dishing in STI, 424, 467-8
Direct tunneling leakage current in gate oxides, 101-4
Dishing, 421-26
in Cu CMP, 417
in CMP, 421-6
in STI, 462-4
Direct polishing, to prevent, 424-5
Dummy-active regions, to prevent, 423-4
Nitride cap, to reduce, 423, 467
RIE/CMP, to prevent, 424, 464-5
DNQ/novolac resists, 207-8
Dose-loss during shallow-junction formation, 214-7
Out-diffusion, 215-6
Screen oxide effect, 216-7
Self-sputtering, 214
Double-level metal interconnects, 576-7, 580
Dual damascene interconnects, 582, 583, 584, 671-709
BARC veils, 692-4
Buried via, 676, 679
Defined, 672, 674-5
Dielectric stack in, 679-80
Embedded etch stop layer, 676, 681-2
Silicon carbide, 682-3
Silicon nitride, 681-2
Hard mask in dielectric stack, 681
Process flow, basic, 584, 675-6, 687-8
Process integration issues, 687-706
Cleaning, 703-8
Post-CMP clean, 705
Pre-metal deposition, 704-5
Resist-strip residue, 703-4
Supercritical fluids, using, 706-8
Etching, 697-700
SiN layer, 700
Trench, 699-700
Via, 698-99
Lithography, 689-96
BARC veils, 692-4
Bi-layer resist process, 694-6, 697
DUV-resist poisoning, 690
Trench CD control, 691-2
Resist stripping, 700-3, 703-4
Ultra-low-*k* dielectrics, 708-9
Self-aligned process, 676, 679
Trench-first process, 676, 677-8, 683-5
Via-first process, 676, 678-9, 685-7, 687-8, 708-9
Dual-gate MOSFETs, 1, 524-5, 656-6
Dual-gate oxide thickness IC, 121
Dual-doped poly, in CMOS, 12, 13
Dummy active regions to prevent dishing in STI, 423-4, 464, 465-6

Edge effect in CMP, 364
Edge exclusion, in CMP, 356-6
Edge polishing of wafers, 59. 60
Edge-rounding of wafers, 59, 60
Electrochemical mechanical polishing (ECMD), Of Cu, 772-3
Electromigration behavior of Cu interconnects, 716-7
Electron beam projection lithography (EPL), 397-9

Electroplating of Cu (see Cu Interconnects)
Electroless plating of Cu, 747, 748, 774
Electropolishing of Cu films, 775-8
Elevated source/drains, 203, 205-6
ELTRAN, 547-50
Endpoint-detection in CMP, 385-9, 407
Frontside laser interferometry, 388
In-line, defined, 385-7
In-situ, 387
Motor-current monitoring, 391
Optical monitoring, 390
Spectroscopy based, 389
Epi-optimized wafers, 12, 35, 45
Epitaxial layers, Si:Ge, 485-6, 487, 492
Compatibility with ion-implantation, 493
Compatibility with thermal oxidation, 493-4
On Patterned wafers, 492
RTP-CVD of Si:Ge, 489-90
Selective Si:Ge, 490
Thermal stability of, 492
UHV-CVD of Si:Ge, 488-9
Epitaxial wafers, 11
EPL (see Electron-beam projection lithography)
Equivalent oxide thickness, 130, 146-7, 175
Erosion, in CMP, 425
Etalon, 262, 264
EUV (see Extreme UV lithography)
Excess overburden, in Cu plating, 770, 773
Excimer, defined, 260-1
Excimer lasers for DUV lithography, 259-65
248-nm (KrF), 262-5
193-nm (ArF), 265
157-nm (F_2), 265
Spectral bandwidth of, defined, 261
Exposure tools for DUV lithography, 266-73
For 248-nm lithograhy, 266-7
For 197-nm lithograhy, 267-8
For 157-nm lithograhy, 268-9
Mix-and-match lithography, 274, 299
300-mm Exposure tools, 272-3
Extinction coefficient in BARCs, 246
Extreme UV lithography, 301-7
Masks for EUV, 305-6
Multilayer reflector for, 305

Fast-pulled silicon ingots, 30-2, 45, 47
Fe-B (iron-boron) pairs, 12, 38-9
Ferroelectric dielectrics, 151, 163
Ferroelectric RAMS (FeRAMs), 154, 155, 163
Fixed oxide charge (Q_f), 82
Floating-body effects in Partially-depleted SOI MOSFETs, 519-20, 559-61, 562-3
Transient, 562
Fluorinated oxides, 420, 650, 654-5, 680
CMP of, 420
Fluorine (F_2) excimer lasers, 265
FOSB, 62-3
Fowler-Nordheim tunneling, 100-1
Front opening shipping box, 52-3
Fully-depleted SOI MOSFETs, 518, 522-6

Gallium-arsenide, 476-7, 478
Gas immersion laser diffusion, (GILD), 508
Gate current, 99-104
Gate oxides
Breakdown of, 90-9
Anode-hole Model, 94-5
Current-driven model, 94-5
E-field driven model, 93-4
Fluorinated oxides, 120-1
Growth models of thin oxides, 104-9
Interfaces charges, 82-90
Microstructure of, 80-2
Nitrided oxides, 112-20
Rapid thermal oxidation (RTO), 109-12
Scaling of, 76-8
Soft-Breakdown of SiO_2, 126-8
Gate oxide integrity, 12, 22, 31, 47, 48, 64-5
Gettering, 32-45
Basic methods, 34-6
Extrinsic, 35, 36-7
Impurity pairs, 38-9
Intrinsic, 35, 37-9, 41-4, 51, 66
Mechanisms of, 34-6
Oxygen precipitates, 37-8, 39
Phosphorus gettering, 35
Global planarization, defined, 313-4, 323
GOI (see gate oxide integrity)
Grain noise, in post-CMP processing, 409

Hafnium oxide (HfO_2), 146, 175-6

HALO implant, 7
HBT (see heterojunction bipolar transistors)
HDP-CVD of trench dielectrics, 459, 461-2
Hemispherical silicon grain (HSG), 151, 159
Heterojunction bipolar transistors (HBTs), 477, 478, 480-86, 498
  Large-bandgap emitters, 482-4
  Small bandgap emitters, 484-6
  Si:Ge, 478-80
HF-de-glaze steps in STI, 452
Hi-Wafers, 28
High-density plasmas for depositing trench dielectrics for STI, 460-2
High-*k* Dielectrics, 145-178
High-dose SIMOX, 531-5
High-performance-liquid-chromatography (HPLC), 765
Hollow-cathode magnetron sputtering, 735-6
Hot-carrier injection, 8
Hybrid-organo-siloxane-polymer (HOSP), 658
Hydrogen annealing of Si/SiO_2 interface, 85-6
Hydrogen annealing of Si wafers, 27-9
Hydrogen annealing to achieve corner rounding In STI, 458
Hydrogen silsesquioxanes (HSQ), 651-3

Implant shadowing effects, 188-90
Indium channel implants, 193-4
Induction time in Cu plating, 755
Indium implants for HALO applications, 221-2
Interface trapped charge (Q_{it}, D_{it}), 82, 83-9
  Effect of interface traps on ICs, 86
Interconnects for ICs (see Multilevel interconnects)
Interstitials, self, in Si, 22, 50-2
Initiation current in Cu plating, 755-6
Intrinsic point defects, in Si, 22
Ionized magnetron sputtering (I-PVD), 731-4, 744
Ionized metal plasma (IMP), 731
ITOX (internal thermal oxidation) 536-7, 538
ITRS Roadmap, 505, 506, 514, 642

"Kink effect" in STI, 437, 438, 469-72
"Kink effect" in SOI, 520, 559-60
Kooi effect in STI, 454
KrF excimer lasers, 262-5
KrF (see Krypton fluorine lasers)

Laser marking of wafers, 56
LATIPS, 8, 219-20
LDD (see lightly doped drain
Leakage paths in STI structures, 441-2
Levelers (Cu-plating-bath additive), 757
Lightly-doped drain, 9, 195-6
Line jogs, 282
LLDs (localized light scatterers), 24
LOCOS isolation, 13, 14, 433, 434, 436
  Mini-LOCOS in STI, 458
  Recessed LOCOS for fabricating thin-body FD SOI MOSFETs, 525-6
Long-throw sputtering, 731, 734
Low-angle tilt-implants to dope STI trench sidewalls, 444-5
Low-dose SIMOX, 530, 535-7
Low-energy implants, 206
  Decel in, 207
  Tools, 207
Low-*k* Dielectrics, 639-59
  BCB materials, 657, 658
  Carbon-based materials, 644, 649
  CMP of, 420-1
  CVD materials, 643, 654-5, 656, 659-62, 662-3
  Defined, 639
  Density impact on *k*-value, 640-1, 664-5
  Desired characteristics of, 646-8
  Embedded low-*k* films 650,
  Fluorinated oxide films (FSG), 650, 654-5, 680
  Hybrid-organo-siloxane-polymers (HOSP) materials, 658
  Hydrogen silsesquioxanes (MSQ), 651-3
  Inorganic materials, 649
  Ionization impact on *k*-value, 641
  Methyl silsesquioxanes (HSQ), 651-3
  Organic materials, 649
  Organosilicate glasses (OSG), 659-62, 698

PAE materials, 657, 658
Parylene materials, 662-3
Polarization impact on *k*-value, 641
Porous ultra-low-*k* films, 668-9
Process integration isues of, 648-9
Silicon based materials, 649
Sol-gel materials, 664
Spin-on materials, 643, 650, 651-3, 653-4, 656, 657-8
Ultra-low-*k* materials, 663-9
Xerogel materials, 665
Low-temperature epitaxy, 487-8, 508
of Si:Ge layers, 487-8
LSTDs, 23

Magic Denuded Zone, 50-2
Magnesium oxide (MnO_2), CMP slurry, 421
Mask-error factor (MEF), 292-5
Maximum frequency of oscillation (f_{max}), 481, 490, 494, 498
MEF (see Mask error factor)
MEEF (see Mask error factor)
Megasonic cleaning, 404, 405
Metal-insulator-metal capacitors (MIM) in DRAMS, 153, 160
Metal-insulator-silicon capacitors (MIS) in DRAMS, 152, 155
Methyl silsesquioxanes (MSQ), 653-4
Mini-LOCOS process for STI, 458
Mix-and-match lithography, 274, 299
Mobile ionic charge (Q_m), 82
Molecular airborne contamination, 137
Oxide growth problem, 137
Molybdenum silicide ($MoSi_2$), 604, 605
Multilevel interconnects for ULSI, 573-601
Contact resistance in, 598
Crosstalk in, 596-7
Double-level metal interconnects, 576-7
ESD protection in, 599
Geometrical aspects, 585-8
Orthogonal tier architecture, 585-7
Global interconnects, def., 585
Local interconnects, def., 585
Power dissipation in, 595-6
Propagation delay (RC) defined, 588-9
Repeaters in, 593-5
Single-level metal interconnects, 575-6
Trends in, 599-600

Nanocleave, 544-7
Narrow-line effect in $TiSi_2$, 618-22
Next-generation lithography (NGL), 301-9
EPL, 307-9
EUV, 301-7
NGL (see Next-generation lithography)
Nickel silicides (NiSi), 610, 611, 634-5
Nitride cap process to reduce dishing in STI, 423, 467
Nitrided oxides, 112-20
Nitrogen doping of Si ingots, 30-1
Nitrogen implantation of Si to inhibit growth of oxide, 121
Non-optical lithography, 301-9
Notches, in Si wafers, 53-4

Octohedral voids in Si wafers, 23
Off-axis illumination, 275-9
Annualr apperture, 278
Quarupole apperture, 278
Offset-spacer to form SDE overlap, 217
157-nm DUV resists, 241-4
193-nm resists, 237-42
Line edge roughness, 240
Outgassing, 241
Resist collapse, 240-1
Shrinkage from e-beam, 241
Onium salts, 228-9
Optical proximity correction (OPC), 279
Organo-silicate glasses (OSG), 659-62, 698
OSF ring (see Oxidation Stacking Fault ring)
Out-diffusion, dopant loss during shallow-junction formation, 216-7
Overburden, 726, 770, 773
Defined, 726
Excess, 770, 773
Overlap of SDE and gate edge, 216-7
Overplating, 770
Oxidation-induced stacking faults (OSF), 26
Oxide/nitride/oxide (ONO), 151
Oxide trapped charge (Q_{ot}), 89-92

Effects on ICs, 90
Oxygen precipitates, 37, 38-9, 41-4
Oxygen stacking faults (OSF) ring, 26-7, 30
Oxynitridation of Si in N_2O, 113-6
Oxynitridation of SiO2 in N_2O or NO, 116-7

Pads, CMP, 375-8, 382-3 (see also CMP)
Contribution to dishing, 422, 463-4
PAE low-*k* materials, 657, 658
PAG (see photoacid generator)
Paraelectric dielectrics, 163
Partially-depleted (PD) SOI MOSFETs, 518, 519-20, 559-61, 562-3
Body contacts in, 520-2
Floating-body effects in 519-20, 559-61, 562-3
Particle adhesion mechanisms, 393-5
Parylene low-*k* materials, 662-3
"Perfect" silicon, 29-30, 45, 47
Perovskites, 163, 165, 166, 168
Phase-shift masking (PSM), 283-92
Alternating PSM (ALT-PSM), 286-7
Attenuating-PSM, 288
Half-tone PSM, 288
Rim-shift PSM, 288
Sidewall-chrome alternating aperture, 287
Strong, 288, 291, 292
Weak, 288, 292
Photoacid generator, 228-9, 239, 241
Photo-islands, 252-6
Photoresist processing systems, 251-4
Track systems, 252-4
Physical-vapor deposition (PVD), 729-36
Collimated sputtering, 731, 733
Filling recesses with, 729-33
Hollow-cathode magnetron sputtering, 735-6
Ionized magnetron sputter deposition (I-PVD) , 731-4
Ionized metal plasma (IMP), 731
Long-throw sputtering, 731, 734
Self-ionized plasma sputtering (SPS), 736
Planarization, partial (in double-level metal), 578-9
Planarization factor, 315
Plasma immersion ion implantation, 531
Plasma-induced damage of oxides, 134-5
Platinum, 153, 167
Platinum group of metals (PGM), 153, 161
"Plus-One" approximation of TED, 209
Pocket implant, 8, 219-21
Polycides, 4, 603, 605-8
Polyfluorotetraethyl, low-*k* material, 666
Polysilicon depletion, 128-33
Polysilicon, CMP of, 41
Pores in low-*k* films, 668
Post-exposure bake (PEB), 230-2, 235
Pourbaix diagram, 351
Praseodymium oxide (Pr_2O_3), 176-7
Pre-amorphization implant (PAI) in $TiSi_2$, 620-2
Preston's constant, 337
Preston's Law, 336-8, 340-2
PREVAIL, 309
Propagation delay (RC) in IC interconnects
defined, 588-95, 639-40
in Cu interconnects, 713-6
PSM (see Phase-shift masking)
Pseudomorphic-epitaxial layers in Si:Ge, 405-6
Puller, Si ingot, 20
Pull rate, Si ingot, 20, Pull-speed, Si ingot, 20
Pulse plating of Cu interconnects, 750, 767-9
Reverse-pulse plating of Cu, 768-9
PZT, 145, 154

Quad Implants, 188-90, 221
Quadruple off-axis illumination, 278

RELACS, 235-6, 470-1
Reliability issues of SOI, 567-8
Replacement gate MOSFETs, 1
Remove-and-reconstitute (Cu-plating-bath management), 767
Resist stripping in dual-damascene process, 700-3
Resolution-enhancement technologies (RETs), 274-91
Off-axis illumination (OAI), 275-9
Optical proximity correction (OPC), 279-84
Optical proximity correction features, 280-84
Line jogs, 282
Scattering bars, 283
Serifs, 282

Phase-shift masking (PSM), 283-92
Alternating PSM (ALT-PSM), 286-7
Attenuating-PSM, 288
Half-tone PSM, 288
Rim-shift PSM, 288
Sidewall-chrome alternating aperture, 287
Strong, 288, 291, 292
Weak, 288, 292
Rapid-thermal oxidation (RTO) process, 109-14
Wet RTO processes, 111-2
Rapid-thermal annealing of shallow junctions, 210-11
Retrograde wells in CMOS, 12, 182-91
Reverse-narrow-channel effect, 471
RIE/CMP process to reduce dishing in STI CMP, 423, 464-5
Ruthenium, 153, 158, 167, 168, 169
Ruthenium oxide (RuO_2), 158, 160, 167, 168
Sacrificial gate oxide , 454,
Sacrificial layer etchback, 320-1
SADS (see silicide as diffusion source)
Salicide, 4, 203, 603, 604, 605, 608-35
$CoSi_2$ (see Cobalt silicide)
Defined, 605
Nickel silicide, 610, 611, 634-5
$TiSi_2$ (Titanium silicide)
Sawing wafers, 53-7
Scanning capacitance microscopy (SCM), 219
Scanning tunneling microscopy (SCM), 219
SCALPEL, 307
Scattering bars, 283
SDE (see source/drain extension)
Seed-layers of Cu (see Cu Interconnects)
Selective epitaxial growth (SEG) of Si:Ge films, 490, 508
Self-annealing of plated Cu films, 770-2
Self-heating effects in SOI MOSFETs, 566
Self-interstitials in Si, 22, 25
Self-ionized plasma sputtering (SPS), 736
Self-sputtering during low-energy implants, 313-4
Separation by Implanted Oxygen (SIMOX), 518, 529-37
High-dose, 530, 531-5
Low-dose, 530, 535-7
ITOX, 536-7, 538
Serifs, 281-2
Shallow-junction formation, 206-10, 213-7, 506-7
Dopant loss during, 313-4
Out-diffusion, 215-6
Self-sputtering, 213-4
Screen oxide effect, 216-7
Doping profile abruptness, 508
RTA of, 210
Shallow-trench isolation, 433-72
Baseline process flow, 446-55
Corner rounding in, 451, 458
in CMOS, 422
CMP, polish stop layer, 447
Dishing in, 422
"Kink effect" in STI, 437, 438, 450
Leakage paths in, 441-4
Si trench depth issues, 457
in SOI, 556, 557
Trench filling process, 458-62
Sheet resistance, 509, 611, 618, 624
$CoSi_2$, 611, 624
$TiSi_2$, 611, 618
Short-channel effects in SOI MOSFETs, 564-6
Silicide, 603
Silicon
COPs in, 21-32, 27-9, 47, 48, 64-5
Epitaxy optimized wafers, 12, 32, 45, 46, 47-8
Grown-in defects in, 21-32
"Perfect" Si, 29-30, 45, 47
300-mm Si wafer growth, 17-21
Silicon carbide (SiC), 682-3
Silicon as diffusion source (SADS), 631-4
SiLK, low-*k* dielectric, 657-8, 680-1
Silicon:Germanium (Si"Ge), 475-98
Base-after gate process, 495, 496
Base-during gate process, 495-6
Base layer, 486
HBTs, 478-80
Process integration issues, 491-4

Process technology for making Si:Ge HBTs, 486-90
Vertical scaling of, 494-5
Silicon-on-Insulator, 14, 501-68
Fabricating SOI wafers, 526-50
BESOI, 541-2
ELTRAN, 5467-50,
Nanaocleave, 544-7
SIMOX, 518, 529-37
Smart-Cut, 542-4
Silicon-on-Sapphire (SOS), 517, 527-8
Wafer Bonding, 518, 537-50
Fully-depleted SOI MOSFETs, 518, 522-6
History of SOI, 516-8
Isolation issues in, 502-5
in CMOS, 504-5
Lateral isolation in, 550-6
LOCOS isolation, 551-4, 555
Mesa isolation, 55-1, 552, 553
STI, 556
Partially-depleted (PD) SOI MOSFETs, 518, 519-20, 559-61, 562-3
Process technology of fabricating FD SOI MOSFETs, 525-6
Reliability Issues of SI, 501, 567-8
Self-heating effects in SOI, 566
Short-channel effects in SOI, 565-6
Thick SOI, 518
Thin SOI, 518
Silicon-on-Sapphire (SOS), 517, 527-9
Silicon dioxide (see Gate oxides)
SIMOX (see Separation by Implanted Oxygen)
Single-damascene process, 673-4, 676
Single-event upset (SEU), 517
Slurry, CMP, 366-70 (and see CMP)
Slurry distribution systems, 371-3
EHS issues, 373-4
Smart-Cut, 542-4
Soft-Breakdown of SiO_2, 126-8
Soft-error rate (SER), 150, 513
Sol-gel technique, 664
Solid-phase epitaxial (SPS) growth of SOS films, 529
Double SPS (DSPS) films, 529
SOI (see Silicon-on-Insulator)
Source/Drain Extension, 10, 208, 216-7, 219-20, 516
Abruptness of doping profile, 218
Offset spacer in, 217
Overlap with gate edge, 216-7
Resistance of (R_{DSE}), 196
Spacers, offset, to form SDE-overlap, 217
Spacers, sidewall, 9
Spreading resistance of drain (R_{sp}), 198-203, 508-9
SDE doping profile abruptness, 218-9
Spike anneals, 208, 211-2, 507
Spin-etch planarization (SEP) of Cu films, 776
Spin-process contamination elimination of Cu (SPCE), 782-3
SSR (see Super-steep retrograde channel doping)
Stacked gate dielectric films, 138-9
Steric stabilization of particles in a solution, 398
STI (see Shallow trench isolation), 13
Stress-induced leakage current, 126
Subtractive interconnect process, 671
Super-steep-retrograde channel doping (SSR), 6, 7, 191-95
Surface-channel MOSFETs, 13
Subtractive interconnect process, 575
Supercritical fluids, 664
"Superfilling" of recesses with Cu plating, 727-8
Suppressors (Cu plating bath additive), 770-2
Surface grinding, Si ingot, 53-4

TAETO, 155-6

Tafel equation, 751
Tafel plot, 751
Tantalum, 417, 720, 725-6, 728, 735-6, 738-9, 740, 747, 752, 776
CMP of, 417
Diffusion barrier properties of, 738-9, 740, 747
Material properties of, 736-7
Refinng of, 737-8
Tantalum nitride (TaN), 684, 720, 725-6, 737-40, 747, 776
Diffusion barrier for Cu, 739-40, 747, 752

Tantalum silicide ($TaSi_2$), 604, 605
Tantalum silicon nitride (TaSiN), 725, 728, 736, 739, 752
Tantalum pentoxide (Ta_2O_5), 145, 146, 148, 149, 151, 154-61
as DRAM capacitor dielectric, 154-60
Oxygen annealing of, 157
as MOSFET gate dielectric, 160-1
Thick SOI MOSFETs, 501, 518, 520
Thin SOI MOSFETs, 518
(311)-defects, 209
Time-dependent-dielectric breakdown (TDDB), 91, 97, 125
Time-to-breakdown (t_{BD}), 93, 94, 98-9
Titanium aluminide ($TiAl_3$), 712
Titanium arsenide (TiAs), 623, 632
Titanium boride (TiB), 623, 632
Titanium nitride, as Cu diffusion barrier, 720, 725, 736, 740
Titanium silicon nitride (TiSiN), 725, 739-40
Titanium silicide ($TiSi_2$), 203, 204, 604, 610, 611, 614-24, 625, 627, 632
Bridging during formation, 615-18
Nitrogen stuffing to prevent, 616-7
C-49 phase of $TiSi_2$, 615, 618, 619, 620, 621, 622
C-49 phase of $TiSi_2$, 615, 618, 619, 620, 621, 622
Dopant redistribution during formation, 622-4
Molybdenum implant into, 621
Narrow-line effect, 618-22
Pre-amorphization implant (PAI), 620-22
Process sequence for forming, 622-4
Silicon consumption during formation, 622-4
Track systems, photoresist, 252-6
Transient-enhanced diffusion, 209-10, 211-2
Trench filling processes for STI, 458-62
Trench-first dual-damascene process, 676, 677-8, 683-5
T-top formation in DUV resists, 228-9
248-nm DUV resists, 228-37
Post-exposure baking of, 235-6
RELACS, 235-6
Tungsten
CMP of, 415-17
W-Plugs, 580-2, 683
and Al lines, 583
Tungsten hexafluoride (WF_6), 606, 607
Tungsten nitride (WN) as Cu diffusion barrier, 725, 740
Tungsten silicide (WSi_2), 604, 605-6
Deposited with SiH_4 and WF_6, 606
Deposited with SiH_2Cl_2 and WF_6, 607
Tunneling currents in oxides, 99-104
Direct, 102-4
Fowler-Nordheim, 100-1

UARL, 251
UHV-CVD of Si:Ge layers, 488

Vacancies, in Si, 22, 25, 50-2
Vacancy-rich region in Si ingot, 26, 45
Van der Waals force, 393-4
Vertical gate MOSFET, 1
Via-first dual damascene process (see Dual damascene interconnects)
Void formation during dielectric film deposition, 322-3
Void formation during trench dielectric deposition in STI, 460-2
Void-free filling of recesses with Cu, 754-6

Wafer Bonding, 518, 537-50
BESOI, 541-2
ELTRAN, 5467-50
Generic, 538-41
Nanaocleave, 544-7
Smart-Cut (Hydrogen implantation), 542-4
Wafering, 52-66
Economics of, 66-8
Edge rounding of wafers, 59
Edge polishing of wafers, 59
Flat grinding, 53
Lapping and grinding of wafers, 57-9, 61
Laser marking, 56
Notch grinding, 53
Sawing of wafers, 55-7
Shipping wafers, 62-3

Specifications of 300-mm wafers, 63
Well dopants in CMOS, 6
"Well engineering," 184
Well formation in CMOS, 181
Retrograde wells, 182-191
Wire sawing of wafers, 56-7
Within-wafer non-uniformity from CMP, 408, 426-7

Xerogel low-*k* materials, 665

Zeta potential, 394-6
Zirconium oxide (ZrO_2), 146, 173-5
Yttrium stabilized (YZT), 174

ORDER FORM

LATTICE PRESS
POST OFFICE BOX 340
SUNSET BEACH, CA 90742, U.S.A.

www.latticepress.com
Phone: (714) 840-5010
FAX: (562) 592-1976

Please send the number of copies indicated below of **SILICON PROCESSING FOR THE VLSI ERA - Vols. 1, 2, 3,** and/or **4 (please specify!)** A check, money order, or Credit Card payment for the full amount for the number of copies ordered is enclosed. Make checks payable to: **LATTICE PRESS.** I understand that I may return any book within 30 days for a full refund if not completely satisfied!

_____
NAME (and COMPANY - if part of your address)

_____
ADDRESS

_____
CITY

_____
STATE ZIP CODE COUNTRY

_____ _____
Daytime Phone Number e-Mail Address

_____ _____
Credit Card Number Expiration Date

____ copies **Vol. 1** @ $199.95 each, plus $10.00 for each copy - shipping-handling fee.
(*California Residents*: Please add $15.00 per copy for Sales Tax.)

____ copies **Vol. 2** @ $139.95 each, plus $10.00 for each copy, shipping-handling fee.
(*California Residents*: Please add $11.00 per copy for Sales Tax.)

____ copies **Vol. 3** @ $149.95 each. plus $10.00 for each copy, shipping-handling fee.
(*California Residents*: Please add $12.00 per copy for Sales Tax.)

____ copies **Vol. 4** @ $199.95 each, plus $10.00 for each copy shipping-handling fee.
(*California Residents*. Please add $15.00 per copy for Sales Tax.)

International Ordering Information:

Enclose full payment in Travelers Checks (in $US), International Money Order, or by a Check payable to **Lattice Press** in **US dollars drawn on a U.S. bank.** Or send Credit Card payment.

Include a $12.00 US shipping/handling fee per copy for Overseas Orders shipped via SURFACE MAIL. (Allow **8-10 weeks** for Surface Mail delivery.)
Include a $30.00 US shipping/handling fee per copy for Overseas Orders shipped by AIR MAIL, with tracking capability. (Allow **2 weeks** for Air Mail delivery.)

VISIT OUR WEBSITE AT: ***www.latticepress.com***

ORDER FORM

LATTICE PRESS
POST OFFICE BOX 340
SUNSET BEACH, CA 90742, U.S.A.

www.latticepress.com
Phone: (714) 840-5010
FAX: (562) 592-1976

Please send the number of copies indicated below of **SILICON PROCESSING FOR THE VLSI ERA - Vols. 1, 2, 3,** and/or **4 (please specify!)** A check, money order, or Credit Card payment for the full amount for the number of copies ordered is enclosed. Make checks payable to: **LATTICE PRESS.** I understand that I may return any book within 30 days for a full refund if not completely satisfied!

_____
NAME (and COMPANY - if part of your address)

_____
ADDRESS

_____
CITY

_____
STATE ZIP CODE COUNTRY

_____ _____
Daytime Phone Number e-Mail Address

_____ _____
Credit Card Number Expiration Date

____ copies **Vol. 1** @ $199.95 each, plus $10.00 for each copy - shipping-handling fee.
(*California Residents*: Please add $15.00 per copy for Sales Tax.)

____ copies **Vol. 2** @ $139.95 each, plus $10.00 for each copy, shipping-handling fee.
(*California Residents*: Please add $11.00 per copy for Sales Tax.)

____ copies **Vol. 3** @ $149.95 each. plus $10.00 for each copy, shipping-handling fee.
(*California Residents*: Please add $12.00 per copy for Sales Tax.)

____ copies **Vol. 4** @ $199.95 each, plus $10.00 for each copy shipping-handling fee.
(*California Residents.* Please add $15.00 per copy for Sales Tax.)

International Ordering Information:

Enclose full payment in Travelers Checks (in $US), International Money Order, or by a Check payable to **Lattice Press** in **US dollars drawn on a U.S. bank.** Or send Credit Card payment.

Include a $12.00 US shipping/handling fee per copy for Overseas Orders shipped via SURFACE MAIL. (Allow **8-10 weeks** for Surface Mail delivery.)
Include a $30.00 US shipping/handling fee per copy for Overseas Orders shipped by AIR MAIL, with tracking capability. (Allow **2 weeks** for Air Mail delivery.)

VISIT OUR WEBSITE AT: ***www.latticepress.com***

THE COMPANION VOLUMES TO THIS BOOK

Silicon Processing for the VLSI Era

Volume 1: Process Technology By S. WOLF & R.N. TAUBER

TABLE OF CONTENTS - **Ch. 1** Silicon: Crystal Growth & Wafer Preparation; **Ch. 2** Crystalline Defects & Gettering; **Ch. 3** Vacuum Technology; **Ch. 4** Thin-Film Basics; **Ch. 5** Contamination Control & Cleaning; **Ch. 6** Chemical Vapor Deposition; **Ch. 7** Epitaxial Growth; **Ch. 8** Thermal Oxidation; **Ch. 9** Diffusion in Silicon; **Ch. 10** Ion Implantation; **Ch. 11** Physical Vapor Deposition; **Ch. 12** Photoresist Processing; **Ch. 13** Optical Aligners & Photomasks; **Ch. 14** Dry Etching; **Ch. 15** Multilevel Interconnects; **Ch. 16** CMOs Process Integration; **Ch. 17** Assembly & Packaging.

2000 LATTICE PRESS 910 pp. ISBN 0-9616721-6-1

Volume 2: Process Integration By STANLEY WOLF

TABLE OF CONTENTS - **Ch. 1** Process Integration for VLSI and ULSI; **Ch. 2** Isolation Technologies for Integrated Circuits; **Ch. 3** Contact Technology and Local Interconnects for VLSI; **Ch. 4** Multilevel Interconnect Technology for VLSI and ULSI; **Ch. 5** NMOS Devices and Process Integration; **Ch. 6** CMOS Process Integration; **Ch. 7** Bipolar and BiCMOS Process Integration; **Ch. 8** Semiconductor Memory Process Integration; **Ch. 9** Process Simulation.

1990 LATTICE PRESS 752 pp. ISBN 0-9616721-4-5

Volume 3:The Submicron MOSFET By STANLEY WOLF

TABLE OF CONTENTS - **Ch. 1** Role of Process & Device Models in Micro-Electronics Technology; **Ch. 2** Numerical Methods for Solving the Partial Differential Equations which Model Submicron Devices and Processes; **Ch. 3** BasicMOS Physics & MOS Capacitors; **Ch. 4** Long-Channel MOSFETs; **Ch. 5** The Submicron MOSFET; **Ch. 6** Isolation Structures in CMOS; **Ch. 7** Thin Gate Oxides: Growth & Reliability; **Ch. 8** Well Formation in CMOS; **Ch. 9** Hot-CarrierResistant Processing & Device Structures.

1995 LATTICE PRESS 722 pp. ISBN 0-9616721-5-3

Order by Using the Form on the Last Page of the Book

ISBN 0-9616721-7-X

