J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*

SEMICONDUCTOR LAYERING PROCESSES

The pervasion of silicon-based integrated circuits (*IC*s) in electronic systems continues as we step into the gigachip age with gigascale integration (*GSI*). Progress in scaling of semiconductor devices has enabled the exponential growth in complexity and functionality of ICs over the past decades. Limits to this development, originating from device physics, circuit architecture, and from materials and processes. The work on material/process technology may well lead to new ways of further increased IC functionality, if appropriate architectures and solutions to fabrication problems can be found. With continual demands for improved circuit density, performance, and reliability, the cost per function has resulted in challenging demands on IC fabrication processes. We have already reached the boundary of many conventional processes of the large-scale integration (*LSI*) era, and we have to leapfrog to a new paradigm in process technology to attain the above goals. Layering processes in semiconductor, dielectrics, and metal interconnects in complementary-metal-oxide–semiconductor (*CMOS*) technology is an outgrowth of one such divergent thinking process.

In thin-film deposition processes, the final structure and properties of the deposited material depend strongly on the deposition conditions, substrate surface structure and chemistry, and subsequent processing treatment. In most IC processes, the substrate material is generally fixed and cannot be altered. Moreover, processing budgets (thermal, chemical, structural, etc.) may not permit subsequent treatments of the deposited material. Therefore, artificial modulation of the structure of these films to attain various desired electrical, mechanical, chemical, microstructural and substructural properties by periodic variation with in situ processing is necessary. This article summarizes the concept of multilayering and multilayered structures in (1) silicon (Si) semiconductors, (2) oxide dielectrics, and (3) metal interconnects in Si ICs.

Semiconductor: Poly-Silicon

Key to the fabrication of dense CMOS chips is the use of polycrystalline Si (*poly-Si*) as gate-electrode material. The use of poly-Si allows realization of a self-aligned structure, greatly improving the device characteristics by reducing parasitic capacitance. It also permits more complex structures to be fabricated because of its compatibility with high-temperature Si IC processing. Similarly, circuit performance and increased density in bipolar (*Bi*) devices and BiCMOS devices are realized with the use of poly-Si in active regions. Today, advanced silicon IC technologies often use more than one layer of poly-Si, such as submicron FLASH electrically erasable programmable read-only memories (E^2*PROMs*), linear CMOS and BiCMOS technologies.

Multilayered Poly-Si for Dielectric Isolation. In most ICs, lateral separation is accomplished by interposing between the single-crystal Si regions, a region of electrical insulating material having a thickness approximately equal to the depth of the active regions of the single-crystal materials being separated. However, for some high-voltage ICs, the use of lateral isolation is not sufficient because of interactions between separate active regions. To prevent such undesirable electrical interaction between two active regions, vertical isolation, in addition to lateral isolation, is employed.

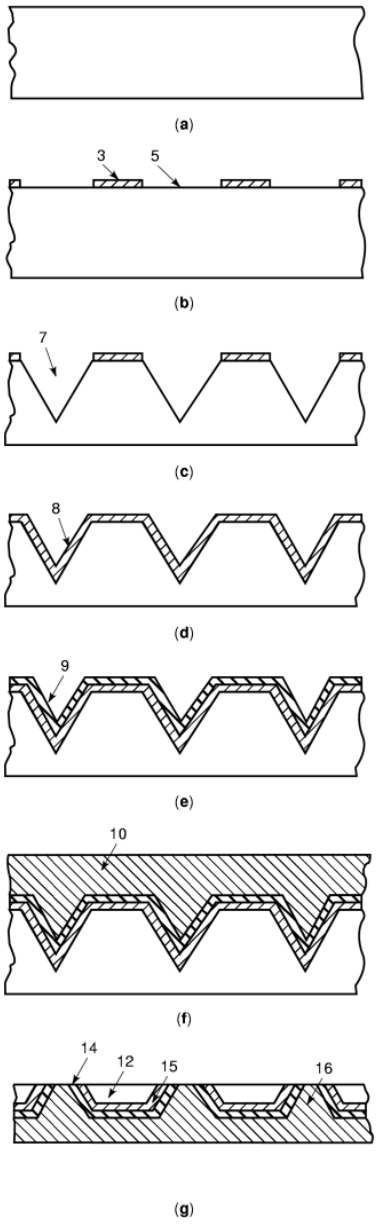

Fig. 1. Dielectric isolation process sequence.

Various processes have been employed to produce a component having both lateral and vertical isolation. Gated-dielectric-crosspoint (*GDX*) employed in advanced telecommunications switches uses the following dielectric isolation (*DI*) process [see (1)]. Here, a low-defect-density Si substrate is coated with an insulating material such as SiO_2 (3 in Fig. 1), and holes (5 in Fig. 1) are formed in the oxide using conventional techniques—for example, photolithography followed by chemical etching. Grooves (7 in Fig. 1) are then etched in the exposed portions of the Si underlying the holes in the dielectric material. Optionally, a region of n^+ Si (8

in Fig. 1) is produced on grooves (7 in Fig. 1). The n^+ Si is, in turn, coated with an insulator SiO_2, (9 in Fig. 1). The insulator is once again, in turn, coated with a layer of poly-Si (10 in Fig. 1). The structure produced is dominated in Fig. 1(f). The entire structure is then inverted; the Si substrate is then ground off until the structure shown in Fig. 1(g) is obtained. In this structure, the remaining high-quality Si is denoted by 12, SiO_2 layers are indicated by 14 and 15, and poly-Si is indicated by 16. Thus, the final structure has isolated single-crystal Si, 12, islands on SiO_2.

During DI wafer processing (Fig. 1), the substrate undergoes a variety of deformations due to the large stress generated during thick poly-Si deposition by a conventional chemical vapor deposition (*CVD*) process using hydrogen reduction of trichlorosilane, $SiHCl_3$. The severity of internal stresses increases significantly when the following factors are considered:

(1) $\langle 110 \rangle$ is the preferred texture in poly-Si films.
(2) Relatively high temperatures are associated with poly-Si deposition.
(3) A large thickness ($\sim$500 μm) of poly-Si support structure.

Therefore, important considerations in stress accommodation [see (2,3)] by substructural layering are needed, which is achieved by deliberate periodic oscillations in process parameters during poly-Si deposition, which creates virtual interfaces in the support structure.

One typical process, the thermal multilayering (*TML*) poly-Si process (2,3), first involves the deposition of a nucleation layer. Once the nucleation layer is formed, the remaining poly-Si material is deposited at elevated temperatures which are deliberately varied in a preprogrammed cyclic manner over a fixed period of time. The strain field produced by the multilayering thermal sequence generates structural instabilities in the poly-Si film, which often induces recrystallization in the nucleation layer. An alternative method, referred to as chemical multilayering (*CML*), achieves substructural layering by periodic fluctuations in H_2:$SiHCl_3$:SiH_4 relative concentrations under isothermal conditions.

Thermal Multilayering (TML) Process. The deposition cycle begins by introducing the stabilized $SiHCl_3$:carrier H_2 gas mixture into a reactor chamber after the thermal stabilization of the susceptor at 1100°C [see Fig. 2(a)]. The nucleation phase takes place during the first 15 min of the poly-Si deposition. The H_2:$SiHCl_3$ reactant gas mixture with a 0.13 mole fraction of $SiHCl_3$ is obtained by bubbling 33.5 L/min of H_2 through a liquid $SiHCl_3$ source. The source gas mixture is further diluted by the mainstream hydrogen (75 L/min) to a 0.047 mole fraction of $SiHCl_3$ in the reactor chamber.

After completion of nucleation, the first cycle of the oscillatory deposition sequence starts with a 5 min temperature ramp to 1180°C for 10 min followed by a thermal quench to 1150°C, by programmed reduction of radio frequency (*RF*) power, and continued deposition for 10 min at 1150°C. The second and other subsequent multilayering cycles are the same as the first. Typical average deposition rates for the temperatures of interest and for mole fraction $SiHCl_3 = 0.047$ are 2.4 μm, 3.5 μm, and 4.0 μm/min for 1100°, 1150°, and 1180°C, respectively.

After the end of the seventh complete cycle, the deposition is stopped by turning off the $SiHCl_3$:carrier H_2 gas mixture and the RF heat source and allowing the system to cool below 650°C in the main hydrogen stream (75 L/min) for 10 min. Subsequent cooling to room temperature is done under N_2 purge.

Chemical Multilayering (CML) Process. The deposition cycle of the nucleation layer is identical to that of the TML process [see Fig. 2(b)]. After completion of nucleation, the temperature is ramped up to 1175°C. The onset of the first cycle of the oscillatory deposition sequence (*ODR*) starts with the introduction of SiH_4 (1.5 L/min) in the reaction zone for 10 min followed immediately by shutting off the SiH_4 for the next 10 min to complete one cycle. The second and other subsequent chemical multilayering cycles are identical to the first. All reactant gas flow and thermal sequencing is accomplished by a programmable process controller. After the end of the seventh complete cycle, the deposition is stopped and cooled similar to the TML process. Typical average

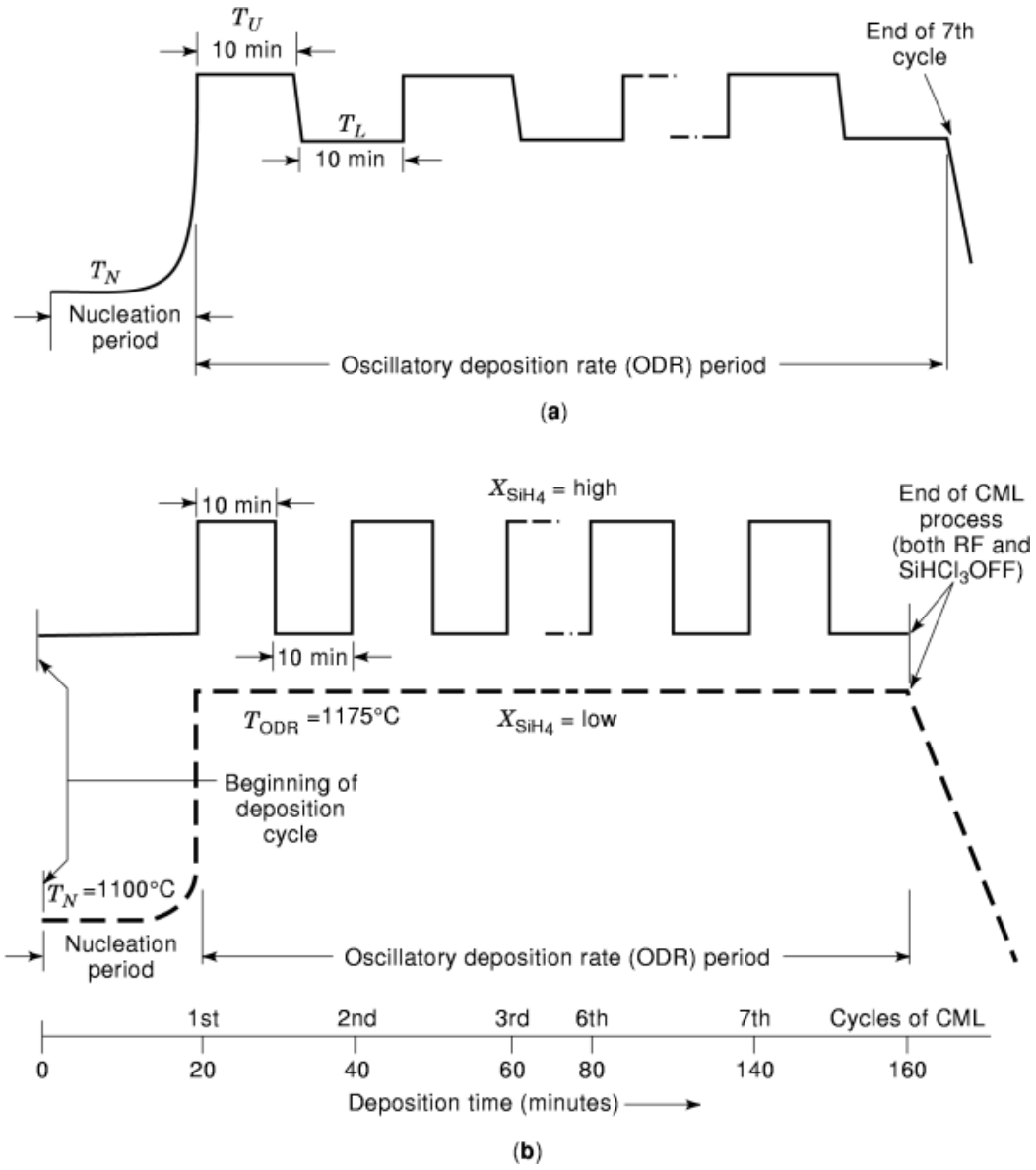

Fig. 2. Thermal (a) and chemical (b) multilayering poly-Si deposition.

deposition rates at 1175° and 1100°C are 3.75 μm/min and 2.6 μm/min, respectively. With SiH_4 injection (mole fraction 0.0136) the deposition rate increases to 4.35 μm/min at 1175°C.

Both the thermal and chemical multilayering processes are derived from the same concept of substructural layering via periodic oscillations in the poly-Si deposition rate. Therefore, both processes generate DI substrates with similar warpage and structural characteristics.

For the thermal multilayering process, the deposition rate variation was achieved by a temperature change of 30°C, whereas for chemical layering, a third reactant (SiH_4) was periodically introduced to the H_2:$SiHCl_3$

Table 1. Characteristics of Ultrathin Gate Oxides

	Defect Density, D_0 (number/cm^2)	Charge Fluence, log N_{bd} (C/cm^2)	Breakdown Voltage, V_{bd} (V)	Tunneling Voltage, V_{tun} (V)
		35 A		
O_2 grown	0.55	−0.57	6.51	4.58
N_2O annealed	0.47	−0.37	6.73	4.78
N_2O stack	0.25	−0.28	7.73	5.52
		50 A		
O_2 grown	0.35	−0.09	8.85	6.15
N_2O annealed	0.18	0.14	9.35	6.52
N_2O stack	0.10	0.10	9.85	6.55
		65 A		
O_2 grown	0.23	0.28	12.1	7.68
N_2O annealed	0.12	0.3	12.3	7.76
N_2O stack	0.05	0.31	12.53	7.70

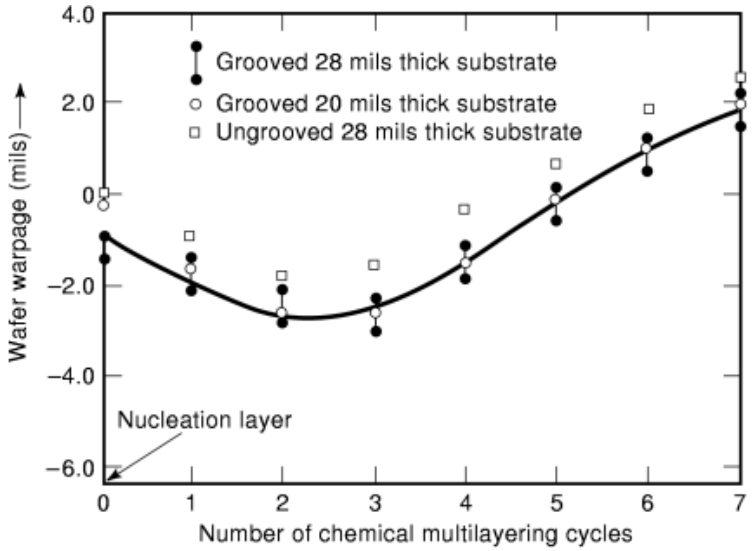

Fig. 3. Wafer warpage after various stages of multilayering. Bow indicates concave warpage of the poly-Si film.

mixture. This SiH_4 injection increases the deposition rate from ~3.75 μm/min to 4.35 μm/min, generating the desired deposition rate fluctuation.

The mechanical state of DI substrates at various stages in the ODR period of the chemical multilayering process is characterized by the wafer deformation. This deformation is measured with a conventional Heidenhain gauge and X-ray diffraction (*XRD*) line profiling [see (2)] of the (440), (422), and (531) peaks obtained from the top 8 μm to 10 μm layer of poly-Si films. Figure 3 shows the typical wafer warpage that results when the ODR period is terminated after different cycles of chemical multilayering for various DI substrates. The trend in wafer warpage variation after the various cycles of multilayering remain unaltered irrespective of thickness and groove design in DI substrates. The mechanical state of DI wafers by multilayering in accordance with

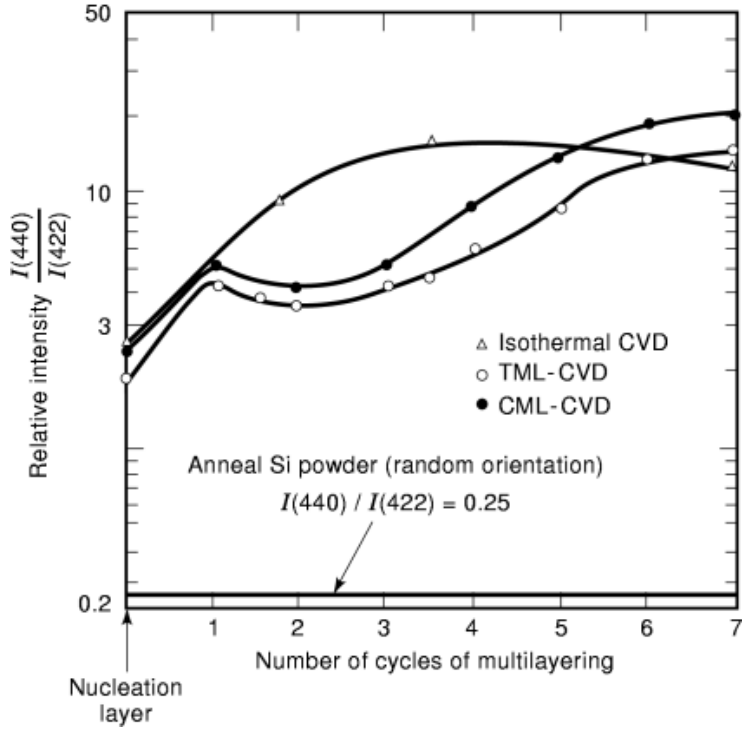

Fig. 4. Variation in [110] Si texture during various stages of multilayering.

the present technique is the net effect of (a) internal stress incorporation as the poly-Si films grow thicker and (b) the stress accommodation triggered by the deposition rate fluctuation. The interaction of these two opposing effects results in an inflection in the observed wafer warpage between the first and third cycles of multilayering.

Figure 4 illustrates the variation in the relative intensity of (440) with respect to (422), $I_{(440)}/I_{(422)}$, at the end of various stages of the chemical multilayering process. Also shown for comparison are similar measurements made on poly-Si layers that were grown under 1175°C isothermal conditions and by the thermal multilayering process. The conventional isothermally deposited poly-Si layers show that $I_{(440)}/I_{(422)}$ increases with increasing poly-Si thickness and saturates at a thickness of ∼250 μm to 300 μm. There is no further increase in the value of $I_{(440)}/I_{(422)}$ or the degree of ⟨110⟩ texture for thicker layers. $I_{(440)}/I_{(422)}$ obtained from nucleation layers (50 μm to 60 μm thick) always exhibit values equal or greater than 2, indicating that ⟨110⟩ growth and hence (110) islands are kinetically favored during the initial nucleation and subsequent growth of these poly-Si films. The increase in ⟨110⟩ texture is quite similar for both the conventional isothermal and multilayering (TML and CML) depositions. After this initial 20 min period, the ⟨110⟩ texture begins to differ, depending on the deposition conditions. For both the TML and CML processes, however, a change in the ratio begins after the first ML cycle, and the large ⟨110⟩ signals are not recovered until after five complete cycles. This indicates that there are structural rearrangements within the poly-Si layer that are taking place, which in effect reduce the degree of ⟨110⟩ texture of the poly-Si. This structural anamoly is also reflected in Bragg broadening data of the Si (440) reflection. The mechanical instability indicated by the inflection in wafer warpage, and the structural anomaly reflected in XRD peak profile data around the second cycle of multilayering, occurs due to the recrystallization in the nucleation layer.

The presence of the (400) reflection from the multilayered films indicates that these large recrystallized grains have the ⟨100⟩ texture. The energy required to rearrange the atoms from the ⟨110⟩ texture to this ⟨100⟩ texture is provided by the strain field produced by the multilayering technique. Under the deposition conditions used, at least two complete cycles are required to initiate recrystallization.

Layered Polysilicon for Submicron CMOS.

Applicaton to Silicide Technology. In recent years, refractory metal silicides [see (4,5)] have found widespread applications in silicon GSI processes as interconnects and contacts. With the transition to submicron channel lengths, the importance of silicide integrity, silicide/poly-Si interfacial characteristics, vertical topography, and local dopant profile are becoming more and more critical in polycide and salicide structures. Conventionally, low-pressure CVD (*LPCVD*, 0.25 torr to 0.5 torr) poly-Si deposited from SiH_4 pyrolysis (570° to 650°C) [see (6,7)] generates films with ⟨110⟩ preferred texture. The high degree of texture coupled with the variation of stress field within the poly-Si layer plus dopant segregation at the grain boundaries results in uncontrolled conductivity within the poly-Si layer. More importantly, the lack of stress accommodation in highly textured poly-Si films (>200 nm) results in unpredictable silicide formation by conventional metal deposition and subsequent rapid thermal annealing (*RTA*) at high temperature. Furthermore, due to stress relaxation during higher temperature processing (T >800°C), the dopant usually out-diffuses from the poly-Si to the silicide layer, thus decreasing the dopant concentration at the silicide/poly-Si interface and resulting in an increase in silicide-to-poly-Si contact resistance. Therefore, for any reliable polycide interconnect technology (8), one must have a controlled poly-Si substructure with a relatively stable dopant profile. This chapter discusses a novel in situ doped layered poly-Si deposition process where substructural layering was achieved through a periodic oscillation in deposition rate while simultaneously modulating the dopant profile by adjusting the mole fraction of dopant gas during the LPCVD pyrolysis of SiH_4.

Refractory metal silicides ($TiSi_2$, in particular) are relatively unreactive, possess low electrical resistivity, and are more amenable to process integration. The resistivity weighs very heavily in favor of $TiSi_2$ (13 $\mu\Omega$·cm to 25 $\mu\Omega$·cm). In addition, Ti is readily etched in a standard cleaning etch, leaving the silicides unreacted.

Poly-Si was deposited by the LPCVD SiH_4 pyrolysis technique onto an SiO_2 gate dielectric layer on a Si substrate (7) at a temperature of 620°C. The deposition rate was varied from a low rate of 1 nm/min to a high rate of 10 nm/min for a total of four half-cycles, as shown in Fig. 5(a). The deposition rate differences were obtained by varying the pressure from 0.4 torr to 1.3 torr in the CVD reactor. A total poly-Si thickness of ∼450 nm was achieved. During deposition, the poly-Si was doped in situ by adding PH_3 gas diluted by nitrogen carrier gas to the SiH_4 atmosphere. The dopant concentration was decreased as shown in Fig. 5(b), being in the range of 0.8 wt. % to 0.9 wt % for the first (bottom) sublayer and decreasing to the range of 0.3 wt. % to 0 wt. % for the last (top) sublayer. Next, a layer of titanium 70 nm to 100 nm thick was deposited by sputtering. The silicide was then formed by rapid thermal annealing (RTA) in a nitrogen atmosphere in two heating steps. In the first heating step, the structure was heated to 600°C to 640°C for a period of 1 min to 2 min. Then the unreacted Ti was removed by etching. In the second heating step, the structure was heated to 800° to 900°C for 10 s to 60 s. This achieved a low resistivity, disilicide phase (C54) ∼80 nm to 120 nm thick.

The above process generated a low sheet resistance of the overall polycide structure (∼0 Ω/□ to 2 Ω/□), due mostly to the silicide layer. In addition, the threshold voltage variability of field effect transistors having this polycide gate structure was significantly lower than those made by conventional techniques. This improved consistency is apparently due to the reduced spiking of the metal silicide into the poly-Si layer. To compare the stress characteristics of poly-Si layers, X-ray diffraction studies were conducted. The peak profiles of the Si (220) and Si (311) lines were determined. It was found that the layered poly-Si structures had peak positions that were intermediate between those for conventional poly-Si and single-crystal Si, indicating lower stress.

The titanium silicide/layered poly-Si interfacial characteristics were compared to conventional titanium silicide/poly-Si structures by the spreading resistance profile and secondary ion-mass spectroscopy (*SIMS*) concentration depth profile. The comparisons showed that a more shallow and more uniform titanium silicide layer, having a much sharper silicide/poly-Si interface. In addition, the planarity of the structures was sig-

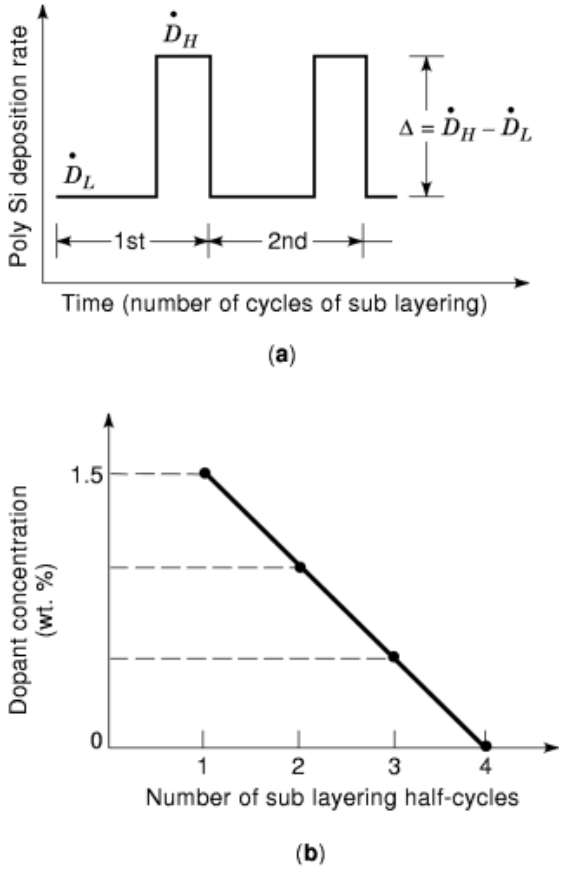

Fig. 5. Substructural layering and dopant concentration modulation by layered poly-Si deposition.

nificantly improved. This appears to be due to the reduction and accommodation in stress resulting from the multiple sublayers.

Uniform silicide formation with sharp interfaces on layered poly-Si is a direct consequence of stress accommodation and dopant concentration modulation. Its technological implication is rather substantial, for example, a relatively planar topography on poly-Si gate stack allows a reduction in stack height ($\lesssim$100 nm) without losing silicide integrity in silicided CMOS technology. Another important aspect of layering is the creation of virtual interfaces among sublayers as a result of fluctuation in deposition rates. These sublayer virtual interfaces act as scattering surfaces to high-energy source/drain implant species, thus reducing ion-channeling and defect generation in gate oxide during CMOS processing.

Application in E^2PROM. Layering provides stress accommodation within the poly-Si layer, allowing for improved dielectric formation. Other beneficial effects may accrue, including reduced channeling of an ion-implanted species through the poly-Si layer. For attaining the maximum benefits in terms of stress accommodation, and hence oxide quality formed on the deposited poly-Si, the number of sublayers is desirably maximized.

LPCVD poly-Si was deposited by the pyrolysis of silane (SiH_4). During portions of the deposition, the poly-Si was doped in situ by adding phosphine (PH_3) gas diluted by N_2 carrier gas to the SiH_4 atmosphere (6,7). The temperature of deposition, reactor pressure, and gas flow rate were adjusted to modulate the deposition

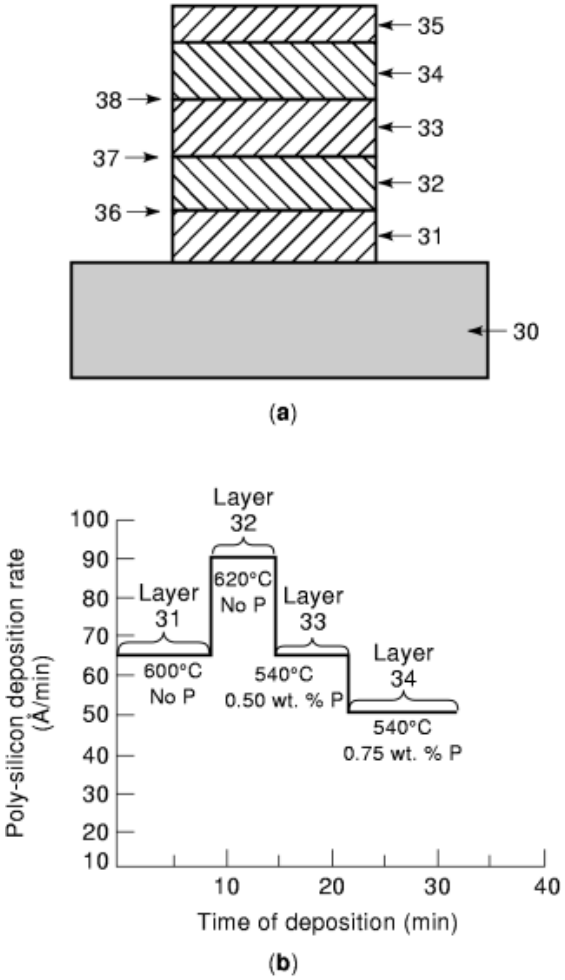

Fig. 6. Schematic in situ doped layered poly-Si depositon process.

process, as shown in Fig. 6(a). The first two layers (layers 31 and 32) were deposited at the temperatures indicated, at a pressure of 280 mtorr. However, the major change in the deposition rate was produced by a reduction in the SiH_4 flow rate from 165 standard cubic centimeters per minute (*sccm*) for layer 31 sccm to 65 sccm for layer 32. The temperature increased for the third layer (layer 33) concurrent with an increase of the pressure to 350 mtorr. However, the inclusion of the dopant gas (phosphine) at a rate of 15 sccm reduced the deposition rate for layer 33 as shown in Fig. 6(b), even though the flow rate of the silane was increased to 240 sccm for that layer. The reduced deposition rate of the final layer (34) was obtained by an increase in the flow rate of the phosphine to 26 sccm, even though the silane flow rate was increased to 260 sccm. The deposition rate was varied from a low rate of 5 nm/min to a high rate of 9 nm/min, for a total of four deposition sequences. A total poly-Si thickness of ∼200 nm was achieved. As a control, a poly-Si layer was deposited at a constant rate of 5 nm/min with 0.75 wt % phosphorus to obtain an equal thickness (200 nm) on both layered and control samples. An oxide layer (layer 35) of 42 nm and 58.5 nm was, respectively, formed by 900°C oxidation.

The ratio of oxide growth rates of the top-doped to control samples was 0.72. The breakdown fields of the two samples were also determined by depositing an additional doped poly-Si layer onto the top oxide layer

and applying an electric field thereacross. On a test vehicle consisting of two overlapping poly-Si layers, the breakdown field of the control sample was 4.1 MV/cm, whereas for the top-doped sample it was 5.0 MV/cm.

Although the highest quality dielectric is often the goal, there are applications that require a thin oxide with a well-controlled conductivity. Certain E^2PROM designs rely on a properly textured poly-Si surface to promote geometrically enhanced Fowler–Nordheim tunneling to transfer charge from a dielectrically isolated floating gate. Here a grain boundary or defined edge of the poly-Si floating gate can be accentuated during oxidation forming an emitting cathode. This effect arises from a differential oxidation rate of oxide due to surface curvature (concave and convex geometry) [see (9)] due to grain size and distribution, crystalline orientation of the individual grains, and dopant concentration and distribution. In order to meet specific device requirements, the shape of and distribution of these emitting surfaces and their oxidation must be controlled. The goal of the poly-Si deposition and subsequent oxidation and heat treatments is to achieve a density of appropriately shaped cathodes and dielectrics capable of achieving the necessary device performance. Measured electrical parameters include tunneling voltages (V_{tun}), trap-up rates (R_{t}), and charge to breakdown (Q_{BD}) in addition to breakdown voltages. These measurements are performed by stressing the dielectric with a constant injected current. The voltage developed across the dielectric after 10 s defines the tunneling voltage while the trap-up rate (9) is defined as the change in voltage across the dielectric per decade of time over the same 10 s interval. The injected current density is then increased by a factor of 100, and the total charge injected through the dielectric to initiate the breakdown is calculated.

Layered top-doped structures generated a V_{tun} of 12.2 V and an R_{t} of 1.06 V/decade. In comparison, oxide on conventional poly-Si generated a V_{tun} of 15.7 V and an R_{t} of 1.71 V/decade. For these tests a lower value is desirable, and thus a clear advantage is obtained with a top-doped layered structure. Charge-to-breakdown measurements for the control were 0.14 C/cm^2 and 0.79 C/cm^2 for the top-doped structure. A lower R_{t} and higher Q_{BD} are indicative of greater endurance for E^2PROMs.

Oxides: Gate Dielectrics

The pervasiveness of silicon-based semiconductors in electronic systems is a result of the unique ability of a single-crystal silicon to grow thermally an amorphous stoichiometric oxide with an interface (transition zone) only a few atomic layers wide. Metal–oxide–silicon–field-effect transistors (*MOSFET*s) are used in circuits ranging from memories to microprocessors to custom logic circuits for functions such as echo cancellation, voice recognition, data encryption, and high-definition television. Indeed, the power of silicon circuits is limited only by their size or the level of integration.

A major hurdle to achieving gigascale integration (GSI—>100 million transistors on a chip) has been the inability of process technologists to grow ultrathin oxides with low defect density and atomically sharp interface. This chapter describes the synthesis of a thin multilayer stacked oxide structure that should allow us to build circuits with physical features measuring $<0.5\ \mu$m.

Driving Force behind Stacked Oxide Structure. As metal–oxide–semiconductor (*MOS*) technology continues to advance and feature sizes shrink, a scaling down in the vertical dimension also occurs. Critical to the success of these advanced devices is a reliable and a high-quality gate dielectric with a low defect density (D_0) and a high breakdown field strength (F_{bd}) that retains its quality during advanced processing. Previous reviews (10) indicate that with careful processing, such as the use of RCA-type pregate oxidation cleanup, and a 2% to 5% chlorine-bearing species [hydrochloric acid (HCl), trichloroethane (*TCA*), or trichlorothylene (*TCE*)] in an oxidant during the oxide growth, D_0 and the breakdown strength of the gate dielectric can be greatly improved.

With a multilayered oxide-nitride (SiO_2–Si_3N_4) dielectric, Watanabe et al. (11) achieved a genuine lowering of D_0 in the range of 0.1 cm^{-2} to 0.5 cm^{-2}. However, the Si_3N_4–SiO_2 interface is invariably associated with a high density of interface states (Q_{it}) that cannot be annealed out easily because the Si_3N_4 is impervious

to diffusion of oxidizing species. These multilayered dielectrics are unsuitable as gate dielectrics in advanced CMOS ICs, because the interface states can cause a charge-induced shift in threshold voltage and can reduce channel conductance during operation.

The concept of stacking thermally grown and chemical vapor deposited (CVD) SiO_2 is borne out by the mutlilayered SiO_2–Si_3N_4 structure. The interface between grown and deposited SiO_2 layers is virtual [see (12)], but it serves the same purpose as the real interface in SiO_2–Si_3N_4 structures; that is, it reduces the D_0 by misaligning defects across the interface. More important, the interface traps in stacked oxide structure can be removed easily by an oxidizing anneal, since the top deposited SiO_2 layer, unlike the Si_3N_4 film, is transparent to oxidizing species, that is, transports them by diffusion. The stacking concept can be applied to any dual dielectric structure (SiO_2) plus another dielectric, namely, a high-k dielectric, with similar results as long as the top dielectric is transparent to the oxidizing species. This allows us to grow a structurally superior third oxide layer during a subsequent oxidizing anneal.

Background on Stacked Oxide Process Design. Major factors contributing to defects in thin-oxide gate dielectrics are growth-induced micropores and the intrinsic stress within the oxide layer, as shown by Irene (13). The micropores are 1 nm to 2.5 nm in diameter, with an average separation of ~10 nm. The pores form at energetically favored sites (such as heterogeneities created by localized contaminants, ion-damaged areas, dislocation pileups, and other defect areas) on the silicon surface as a result of retarded oxidation in these sites. The pores grow outward as oxidation continues to consume silicon around the pore. Thus, a network of micropores exists in SiO_2 (shown schematically in Fig. 7). The micropore network forms potential short-circuit paths for diffusional mass transport and for current leakage.

In addition, the stress within SiO_2 layer, often accentuated by complex device geometry and processing, usually increases both the size and the density of the micropores. Therefore, in developing thin dielectrics with ultralow D_0, one must reduce local stress-gradients near the Si–SiO_2 interface by providing a stress-accommodating layer, such as an interface within the dielectric that acts as a stress cushion and a defect sink.

During densification/oxidation annealing (at 850° to 1150°C) in Si–SiO_2 structures, the oxide grows with the inward advance of the Si–SiO_2 interface (δ) from the reaction of the silicon with the incoming oxidizing species to which the SiO_2 layer is transparent. The newly formed Si–SiO_2 interface, in this case, virtually mimics the interface before the densification anneal, as shown schematically in Fig. 7(a). The Si–SiO_2 interface in an Si_3N_4–SiO_2 dual electric structure, on the other hand, remains virtually intact during the oxidizing anneal [Fig. 7(b)] because the Si_3N_4 is opaque to the diffusional transport of the oxidant, and only the top layer of the Si_3N_4 film gets oxidized to form a silicon oxynitride layer. The top deposited SiO_2 layer in a stacked oxide structure, shown by Roy and Sinha (12), unlike Si_3N_4, is transparent to the diffusional transport of the oxidizing species. During a densification anneal, the SiO_2 growth occurs by the interfacial reaction of the oxidizing species and silicon with a simultaneous densification of the top deposited layer.

The newly grown SiO_2 is structurally superior because the growth occurs in near-equilibrium conditions in the presence of a stress accommodating virtual interface layer between the deposited and the thermally grown SiO_2 layers (12). The newly formed Si–SiO_2 interface [Fig. 7(c)] is structurally smoother with very little local stress variation and interfacial asperities (14). Furthermore, the interface states get annealed during densification, and the stacked oxide structures therefore have superior charge trapping characteristics.

Process Synthesis. The synthesis is a three-step process that involves growing, depositing, and growing SiO_2 layers on silicon by thermal oxidation, chemical vapor deposition, and densification/oxidation, respectively.

The First Layer: Grown SiO_2. Thermally grown SiO_2 passivates the semiconductor silicon surface more than any other kind of deposited oxide film. Passivation is the reduction in the number of surface states (10^{15} cm^{-2}) arising from unsatisfied chemical bonds at the free surface of Si [see (13)]. Thermal oxidation of Si lowers the number of dangling bonds to $\sim 10^{10}$ states/cm^2 with the formation of a very stable SiO_2, and the first layer of the stacked oxide is therefore grown thermally. The silicon surface oxidizes by reaction with the oxidant as

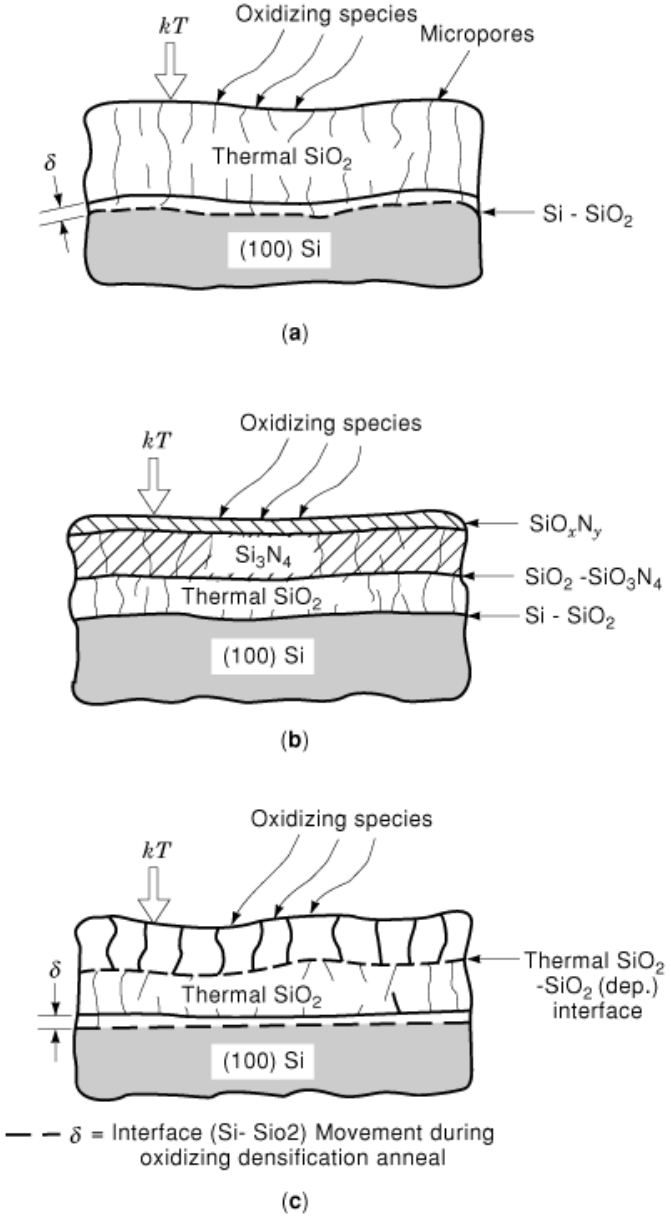

Fig. 7. Schematic showing effects of oxidizing densification anneal: (a) SiO_2 films, (b) SiO_2–Si_3N_4 dual dielectric, (c) Thermal SiO_2-deposited SiO_2 multilayered stacked films.

the oxidant migrates inward. Thus, the silicon surface is constantly renewed by oxidation, and many of the surface and bulk defects are removed.

The Second Layer: Deposited SiO_2. The passivation of a silicon surface is no longer critical after the growth of the first SiO_2 layer, and surface-state generation is minimized; hence the deposited second layer of SiO_2 layer is deposited on the first grown layer. More importantly, the second layer must be deposited

rather than grown if it is to form a virtual interface (12) that reduces the effective D_0 and acts as a stress accommodating layer. Thin SiO_2 films are usually deposited by LPCVD methods from oxidation of silane (SiH_4) with oxygen or nitrous oxide (N_2O) or from the pyrolysis of tetraethyl orthosilicate [*TEOS*, $Si(OC_2H_5)_4$]. These LPCVD methods generate conformal coverage with reproducibly uniform thickness.

The Third Layer: Oxidation/Densification. The final step of the synthesis is to grow a third SiO_2 layer underneath the first grown layer. This layer growth occurs during densification annealing in mild oxidizing conditions [see (12)]. The densification anneal is usually performed in an atmospheric hot-wall (750° to 900°C) furnace. During densification, the third oxide layer grows under near-equilibrium conditions in the presence of a stress accommodating virtual interface layer that serves as a defect sink. Two events occur: Traps get annealed out, and a planar and stress-free Si–SiO_2 interface is formed by the newly grown SiO_2.

Stress Measurements. The stress in the silicon layers near the Si–SiO_2 interface, which also reflects the state of stress within the oxide layer, was measured by Si(400) Bragg peak profiling (2,3) using X-ray microdiffraction (*XRMD*). Figure 8 schematically shows XRMD on an Si–SiO_2 structure under Bragg diffraction conditions. The Si(400) 2θ peak position is a direct measure of the interplanar spacing (d) of (400) planes. Any deviation from the unstressed value of $2\theta_0$ is a measure of lattice dilatation (d–d_0 or Δd), which can be converted to the stress in silicon (s_{si}) from the elastic stiffness values of silicon (15). Furthermore, the peak breadth in the Si(400) peak profile gives information about the silicon substructure (12) in terms of the defect state in silicon near the interface.

During stacked oxide synthesis, the first thermally grown 5 nm SiO_2 layer (B-1) has a peak position of 68.9260° and a peak broadening of 0.7520°, indicating a tensile stress of 7.83×10^9 dyn/cm^2 and a slight worsening in the silicon substructure near the Si–SiO_2 interface. The Bragg peak gets slightly broader (0.7700°) when the second 5 nm SiO_2 layer is deposited. The large peak shift of 69.2440° corresponds to a mild compression of -9.1×10^9 dyn/cm^2 as a result of stacking and stress accommodation by the virtual interface (B-2). After oxidation/densification annealing and the formation of the newly grown SiO_2 layer, the Si(400) peak position corresponds to a nearly zero stress value (0.20×10^9 dyn/cm^2). The simultaneous reduction in peak broadening to a value of 0.7250° indicates a superior substructure of silicon near the interface (B-3). Profile B-3 almost mimics profile C for unstressed (100) Si crystal, indicating that the Si–SiO_2 interface generated by this new oxidation process is stress-free.

Lattice Imaging of the Si–SiO_2 Interface. In a transmission electron microscope (*TEM*), forming an observable image depends on two interrelated processes: interaction of electrons with the specimen and transfer of the interaction to a photographic plate by the objective and image-forming lenses. In the high-resolution imaging mode, the objective lens of the TEM brings the various diffracted beams to interference in the image plane. Consequently, a lattice image, which is simply an electron interferogram, is formed. The contrast in the image is due either to electron diffraction (amplitude contrast) or to electron interference (phase contrast). With phase contrast, which was used in present investigation, one obtains the images of atomic planes of appropriate Miller indices by interference of diffracted beams with the undiffracted beams in the image plane.

Changes in the Si–SiO_2 interface roughness and asperities were observed directly from Si(111) lattice images using a TEM as shown by Roy and Kannan (16). Specimens were from Si–SiO_2 cross-sections cleaved parallel to the [100] direction by argon milling the interfacial area to a thickness of ~180 nm. Figure 9 shows lattice image pictures of the Si–SiO_2 interface for thermal and stacked SiO_2 films. There were a drastic reduction in the interfacial roughness (<0.8 nm) for the stacked interface. In comparison, the standard thermal SiO_2 interface has a roughness of ~1.5 nm to 3 nm. The silicon layer near the interface—indicated by parallel lines from Si(111) planes at 54.7° angle with respect to the [110] direction—showed different degrees of phase contrast in conventional and stacked oxide films. The relatively large contrast modulation in the lattice image of the silicon layer near the interface for thermal SiO_2 was due to localized strain fields. Strain fields were minimized in stacked SiO_2 films because the SiO_2 formed during densification generated very little stress in the underlying silicon. These results are confirmed by XRMD stress data. Figure 9 also shows the TEM

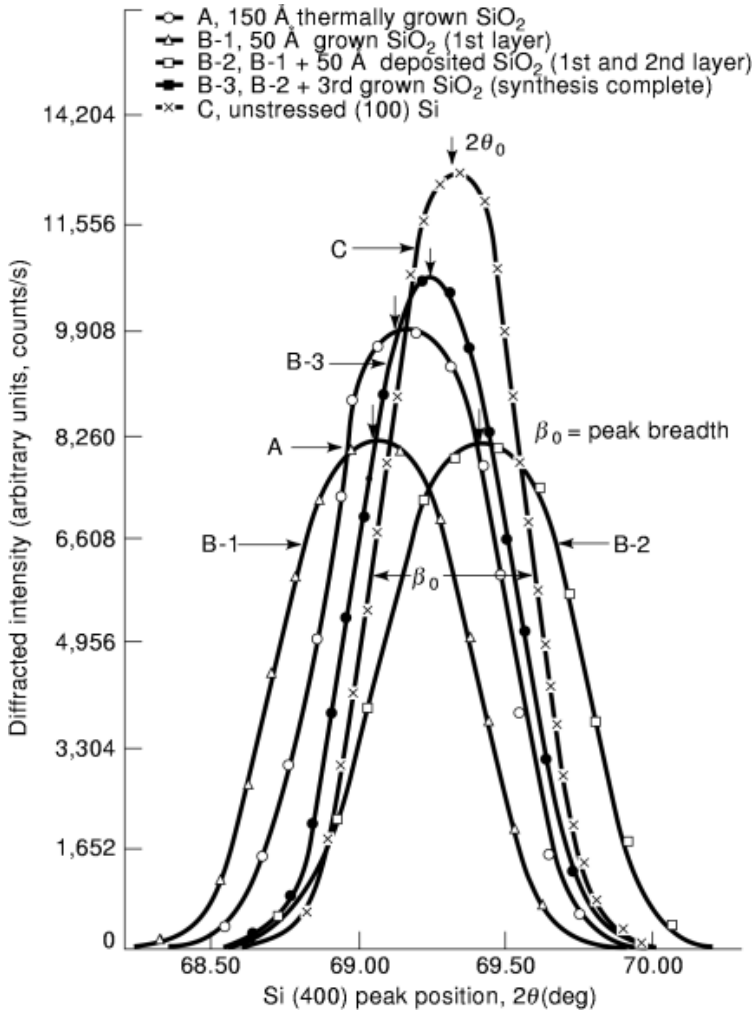

Fig. 8. XMRD Si (400) peak position profile in (100) Si–SiO_2 structure for stacked oxide films (15 nm) at various stages of processing.

bright-field picture of a virtual interface formed between the grown and the deposited SiO_2 layers before the densification anneal.

There are three major advantages for stacked oxidation:

(1) Mismatch of the micropores present in the thermal oxide and the CVD layer reduced the defect density dramatically.
(2) Si substrate consumption is less than conventional thermal gate oxide, and thus fewer substrate defects are incorporated into the bottom thermal oxide.
(3) Stress compensation between the bottom thermal oxide layer and the CVD layer reduces the stress at the thermal oxide–Si substrate interface.

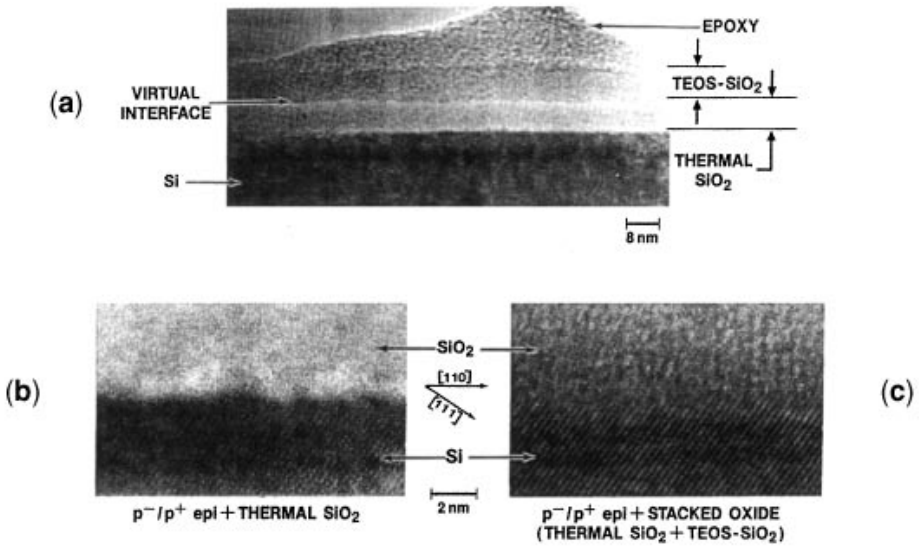

Fig. 9. Bright-field electron micrograph of virtual interface (a) and Si (111) lattice images at the Si–SiO_2 interface for two different gate oxides (b) and (c).

These advantages make the stacked gate oxide very attractive for submicron technology, as demonstrated by Tseng et al. (17). The benefits are clearly shown in the breakdown voltage histograms of submicron *SRAM* array capacitors comparing thermal oxide and stacked gate oxide. The four times lower incidence of low-voltage failures for the stacked dielectric is similar to the improvement observed for area capacitors. Micropores in the bottom thermal oxide are sealed by the CVD layer, thus minimizing their effect and thus providing lower defectivity.

Nitrides

In conventional semiconductor device processing, nitride layers are utilized in isolation structures and as part of an interlevel dielectric structure [ONO or ONON dielectrics, e.g., that use alternating layers of oxide (O) and nitride (N)]. Presently, an LPCVD process is used to deposit the nitride in a single step on the oxide surface. In particular, NH_3 and SiH_2Cl_2 are reacted at temperatures between 700° and 850°C to form the nitride film (18). This process results in generating significant stress between the deposited nitride and underlying oxide, forming defects in the structure and a large variability in the nitride thickness. These problems limit the usefulness of the nitride film in submicron applications where film thickness needs to be controlled. The stress also results in the formation of microcracks and pinholes in the structure and is problematic when used as part of an ONO or ONON dielectric.

An improved technique for depositing nitride layers in semiconductor device processing that addresses the concerns listed above is now discussed (19). In particular, a multilayered nitride deposition process has been developed where the deposition rate is varied in a controlled fashion (low–high–low–high···) to create a multilayer structure that includes stress accommodation at the interface between layers. The internal stress relief results in an overall nitride layer of improved uniformity in thickness with a significant reduction in microcracks and other stress-related characteristics.

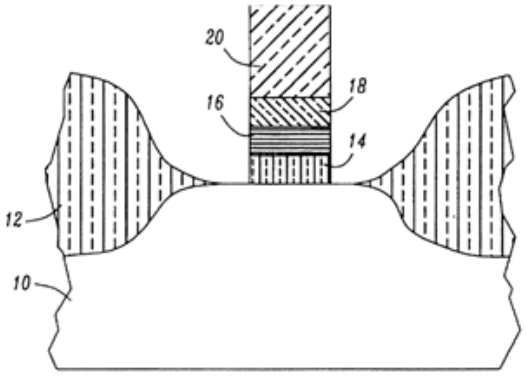

Fig. 10. Gate oxide structure with an oxide-nitride-oxide (ONO) dielectric region utilizing a multilayered nitride structure.

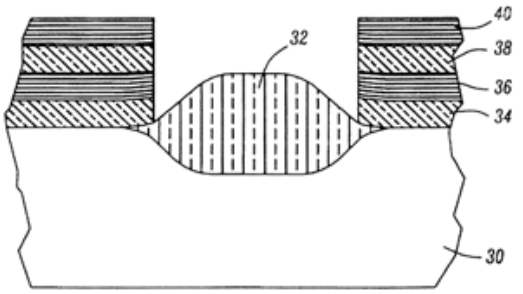

Fig. 11. Schematic diagram of an exemplary LOCOS isolation region with an ONON dielectric region utilizing a multilayered nitride structure.

There are many applications where nitride layers are used in isolation structures and as part of an interlevel dielectric. Figure 10 shows a cross-section of an exemplary gate oxide structure including an oxide-nitride-oxide (ONO) dielectric region utilizing a multilayered nitride structure. As shown, a silicon substrate 10 includes a field oxide region 12 and a gate oxide structure formed within an opening in field oxide 12. In Figure 10, the gate oxide structure includes a first layer 14 of oxide, a nitride layer 16 and a second oxide layer 18 (ONO multilayer stack). A poly-Si layer 20 forms the upper surface of the gate structure. The multilayer structure inherently includes stress accommodation at the interface between each layer of the multilayer structure. The internal stress relief thus results in an overall improved uniformity of the composite nitride layer. A LOCOS isolation application for utilizing a multilayered nitride layer is shown in Fig. 11. As shown, a silicon substrate 30 includes an isolation region 32, with an ONON structure formed to surround isolation region 32. In particular, the ONON structure includes a first oxide layer 34, a first nitride multilayer region, 36, a second oxide layer 38, and a second nitride multilayer region 40. The utilization of the multilayer structure, as stated above, has been found to improve the uniformity of the overall nitride layer thickness. Figure 12 illustrates an exemplary E^2 PROM application that utilizes an interlevel dielectric including a multilayered nitride region. As shown, oxide layer 56 and multilayered nitride region 58 are disposed between a transfer gate oxide 50 and a gate poly-Si region 54. Again, the use of the multilayering process yields a final nitride layer with improved uniformity of thickness, as well as reduced pinholes and microcracks.

Figures 13 and 14 are useful in understanding the nitride multilayering process. Referring to Fig. 13, the rate of depositing the nitride layer is varied as a function of time. In the illustrative case, two period

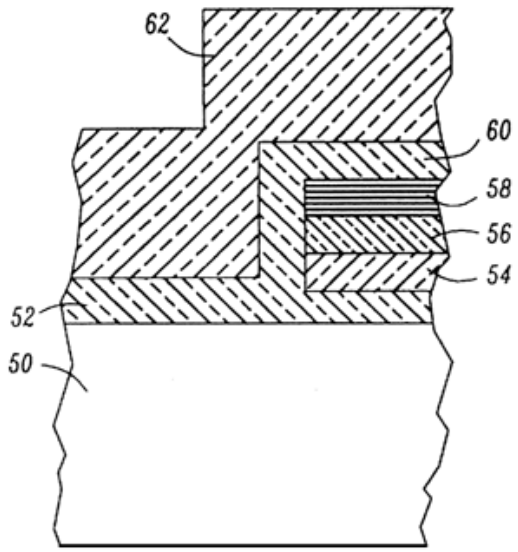

Fig. 12. Schematic diagram illustrating a cross-section of an exemplary flash E^2PROM using an ONO isolation region.

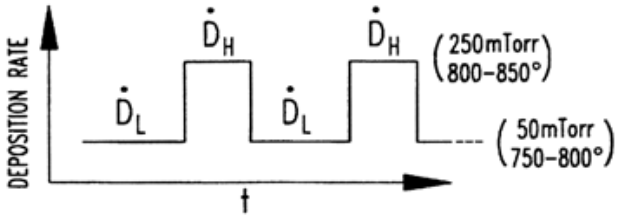

Fig. 13. Schematic illustrating the periodic change in nitride deposition rate as a function of time.

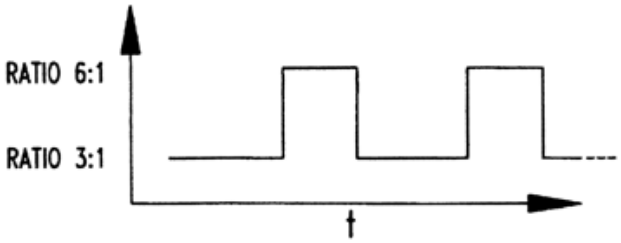

Fig. 14. Schematic illustrating an optional change in reactant concentration during the deposition process.

cycles are shown, producing four sublayers of deposited nitride. However, the number of cycles, and hence the number of sublayers, may be any desired number. The deposition sequence typically begins with a lower deposition rate (D_L), followed by a high deposition rate (D_H), which comprises the first cycle. The lower rate may be, for example, between 10 A/min and 15 A/min and the higher rate may be anywhere between 30 A/min and 50 A/min. The sequence continues with additional cycles. The duration of each portion is controlled so

that approximately the same thickness layer is deposited during each half-cycle. For example, if the higher deposition rate is three times that of the lower rate (e.g., 30 A/min vs. 10 A/min), the deposition at the lower rate will be carried out for a length of time that is triple the length of the high deposition rate half-cycle. The deposition rate variations are typically obtained by changing either the process pressure, or the gas flow rates, or both. Other factors, including the temperature, may also be used to change the deposition rate if desired. For example, the lower rate deposition process may be carried out a temperature between 750° and 800°C; the higher rate process may be performed at a temperature within the range of 800° and 850°C. The result of these changes during deposition is sublayering of the deposited nitride, also referred to as "multilayering". The interfaces between the sublayers provide for stress accommodation within the structure. Other benefits will result, including the reduction of microcrack formation within the nitride structure.

In addition to providing cyclical changes in the deposition rate, the silicon content in alternate sublayers may be varied by varying the ratio of NH_3/SiH_2Cl_2 between 3:1 (or lower) for the low deposition rate and 6:1 (or higher) for the high deposition rate. This NH_3/SiH_2Cl_2 ratio of 3:1 generates stoichiometric silicon nitride. However, for some applications deliberate nonstoichiometric nitride films are often needed. For Si-rich nitride, a ratio of $<3:1$ is used and conversely, for N-rich nitride, a ratio of $>3:1$ is used. Sometimes silicon nitride stoichiometry is used to control stress and relative transparency to the oxidizing specie to anneal out traps at the oxide/nitride interface. Figure 14 illustrates this variation in silicon content.

The above process will be more fully illustrated by means of the following example. A multilayered silicon nitride film was deposited on an oxide layer. The first nitride sublayer was deposited at a rate D_L of 10 A/min using an NH_3:SiH_2Cl_2 ratio of 3:1 at a reactor pressure of 50 mtorr and reactor temperature of 750°C. The next sublayer was deposited at a rate D_H of 30 A/min using an NH_3:SiH_2Cl_2 ratio of 6:1 at a reactor pressure of 250 mtorr and reactor temperature of 800°C. This sequence is then repeated until the desired nitride thickness is obtained. In general, at least three layers are required to provide the desired stress accommodation results. The nitride thickness variability in the above-described multilayer structure was found to be $<2\%$. The reduction in stress was also found to reduce pinholes and microcracks in the nitride film.

Multilevel Interconnects

Aluminum alloys are the most common materials used in integrated circuits to make electrical contact to device active areas (e.g., as electrodes to the gate, source, and drain of field effect transistors) and to interconnect devices to one another (e.g., as runners between devices on the same level of an IC or as vias between devices on different levels) (20). As device dimensions are scaled down to submicrometer geometries, conventional single-layer aluminum (Al) based metallization shows limitations. Smaller ohmic contacts on the chip can result in high contact resistance, which can degrade circuit performance by increasing the *RC* time constant, cross-talk, and causing higher power consumption. Consequently, multilevel metallization schemes for contacts and interconnects have become a requirement for bipolar and CMOS devices. Such multilevel schemes try to achieve low contact resistance while maintaining junction integrity and consist of salicided contacts, refractory metal barriers (such as titanium/titanium nitride composites), and aluminum alloy interconnects (21).

Typically, these Al-based thin films are deposited as single layers by means of a single-step deposition such as physical vapor deposition (*PVD*) also known as sputtering. Sputtering processes use a plasma of ionized Argon (Ar) gas. Ionized Ar atoms strike a target made of a desired Al alloy and, through momentum transfer, "sputter" atoms off the target. The angular distribution of these sputtered atoms is found to follow a "cosine law" (20) and deposition occurs in an essentially line-of-sight path with a cosine distribution. Such a deposition path results in several major problems when Al alloys are deposited in windows or vias with abrupt surface topography. These processing problems are shown schematically in Fig. 15:

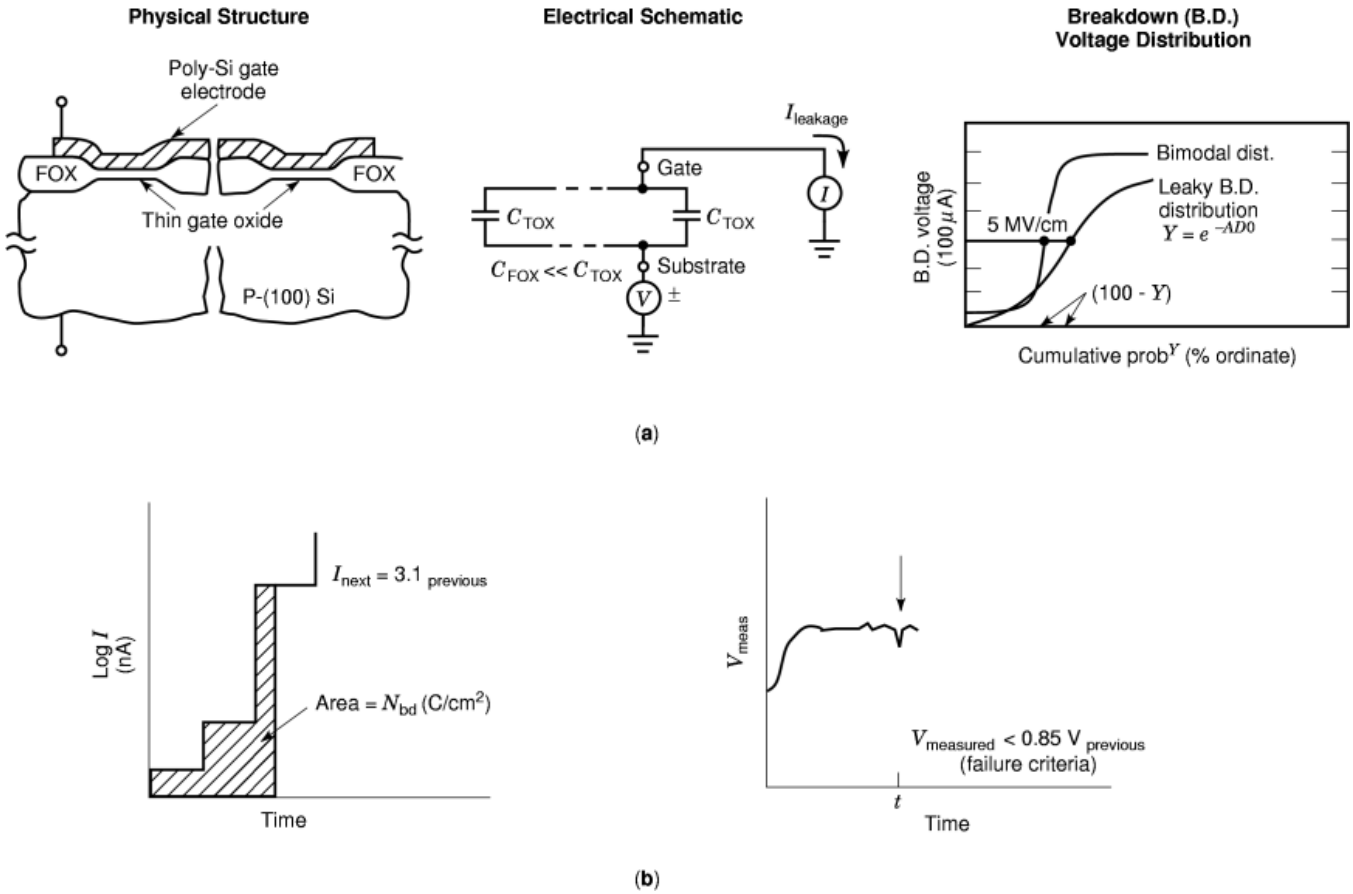

Fig. 15. Schematic diagram illustrating the lack of Al step coverage with pinch-off, overhang and missing Al.

(1) Poor step coverage (ratio of material thickness on opening sidewall to thickness in open field areas) in the opening
(2) Pinch-off or a lack of material deposition along a sidewall of the opening
(3) Overhang or excessive metal above a top edge of the opening

Any one or more of these problems may adversely affect IC reliability. Thus, poor step coverage and overhang often produce voids (an absence of metal) in the windows or vias that reduce the ability of the interconnect to carry current. Pinch-off, on the other hand, causes electron crowding, and hence undesirably high current densities, in the sidewall regions where metal is absent.

Lack of adequate step coverage is often found in multilevel structures in devices with vertical or near-vertical-walled openings, especially those with aspect ratios (ratio of opening diameter to height) greater than unity (22). Such a problem is often addressed by using tapered (i.e., sloped) sidewalls, such as shown schematically in Fig. 16. However, a taper technology results in a larger pitch. Because a larger pitch means increased spacing between parallel metal conductors (e.g., runners), any significant taper is likely to render this technology unsuitable for design rules of, say, 0.5 μm or less. A variety of deposition schemes and material

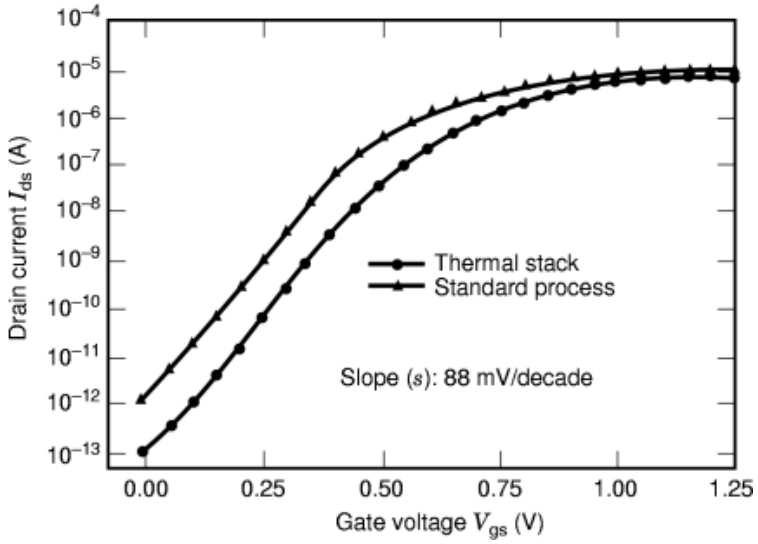

Fig. 16. Tapered opening shows good Al step coverage.

combinations are considered here in order to circumvent the issues listed above and will be discussed in the next section.

Multilayering of Al Alloys. Multilayering is a method used to fill high aspect ratio openings (contacts or windows and/or vias) using multiple-step metallization processes. In principle, an Al-alloy nucleation layer is deposited at a relatively low temperature (<100°C) and the remainder of the metal thickness desired is deposited as the temperature is ramped to allow for planarization of the metal layer. The maximum wafer temperature during the ramp cycle may be between 400° and 500°C. The high surface mobility of Al at these elevated temperatures causes significant diffusion of Al to allow filling of windows or vias. Several variations of the multilayering technique can be found in the literature; a few common practices are as follows:

(1) The two steps are carried out in a single deposition module where the nucleation layer is deposited at a low temperature (typically <100°C) without backside heater argon, and later the argon is turned on to allow adequate heating of the wafer while the metal is being deposited. This is commonly called a single-chamber cold/hot deposition sequence (23,24).
(2) The two steps are carried out in a single deposition module where the entire thickness of the Al-alloy layer is deposited without backside heater argon. Later the heater argon is turned on to allow adequate heating of the wafer and to cause the Al to "flow" and fill openings in the dielectric. This is commonly called a single-chamber cold/flow deposition sequence (25).
(3) The entire desired thickness of metal is deposited in the process chamber at a low temperature (typically <100°C). The wafer is then heated in another chamber to high temperatures (typically 400° to 550°C) to allow Al planarization. This method is referred to in the literature as a two-chamber cold/reflow deposition sequence (26). If the wafer temperature during the second Al-alloy deposition is ∼400° to 450°C, the process is referred to in the literature as a two-chamber cold/hot deposition sequence (27).

Each of the three methods discussed above has unique advantages and drawbacks that need to be weighed for technology integration issues, dielectric materials, thermal budgets, window or via dimensions, barrier integrity, Al-alloy microstructure, vacuum conditions, wafer surface condition and contamination control prior

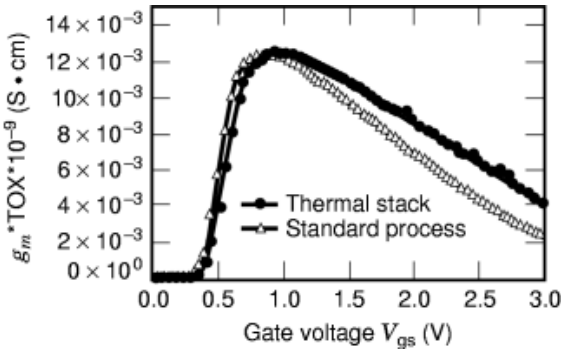

Fig. 17. Schematic diagram illustrating the various steps for multilayered Al planarization.

to Al deposition, throughput, ability to manufacture, and so on. Electrical tests (26) with fully processed CMOS devices with multilayered Al-alloy structures show no degradation of contact resistance and junction leakage current. Improved electromigration resistance in such planarized structures, over that observed for Al-alloy films deposited using conventional single-step techniques, has been attributed by several workers [see, e.g., (26)] in the field to a larger grain size. In general, these Al planarization techniques have found applications in the fabrication of memory devices.

A single-chamber cold/hot/hot deposition sequence, a variation of technique 1 above, will be discussed here. It is assumed that the deposition process described below is carried out in a PVD cluster tool equipped with isolated, high vacuum (better than 2×10^{-8} torr) chambers capable of multilevel metallization processes (28). The following steps, as shown schematically in Fig. 17, are typical of the process sequence for such a deposition:

Step 1. A barrier layer consisting of a titanium/titanium nitride (Ti/TiN) stack of ~20/100 nm, respectively, is first deposited using conventional PVD techniques. The barrier stack thickness can be significantly reduced if a collimator (29), a CVD method (30), or any ion-assisted PVD (31) techniques are employed. At the contact level, a postbarrier-layer deposition anneal is common. This heat treatment (generally at ~700°C in a rapid thermal annealer or forming-gas atmosphere furnace) reduces the contact resistance of the opening. Porosity and pinholes in the TiN layer may also be "stuffed," and an overall relaxation in stack stress occurs. It is important for increased reliability that the Ti layer is essentially [0002] oriented and the TiN has the [111] orientation. Highly textured films are desirable (32); other random structures must be minimized through selective processing parameters. Typical deposition temperatures for the Ti/TiN layer range between 25° and 400°C. If a postdeposition anneal is carried out, the barrier layer can be deposited at lower temperatures.

Step 2. Next a thin (~20 nm to 30 nm) Ti layer is deposited in a separate chamber at temperatures between 25° and 100°C, preferably ~50°C. As stated earlier, the Ti layer can be deposited using conventional, collimated, CVD or ion-assisted PVD techniques. The Ti layer serves as a wetting layer for subsequent Al film deposits discussed in steps 3 to 5 below. For openings of aspect ratio = 1, the Ti wetting layer may not be necessary; however, for submicron devices and technologies with openings <0.5 μm in diameter, the layer is considered essential. At the via level (second metallization level or higher), the Ti/TiN or Ti/TiN/Ti stack may be substituted by a single thick Ti layer alone. This single Ti layer not only reduces cost associated with the two- or three-layer stack, but also reduces large defect densities common with TiN deposition.

Step 3. As soon as the Ti wetting layer has been deposited, the wafer is cycled into a separate chamber where an Al "seed" layer is deposited. The process sequence is such that there is no deliberate delay between steps 2 and 3, and the Al-nucleation layer is put down at the maximum possible deposition rate. The module-heater backside argon (Ar) is turned off while the nucleation layer is being deposited. Typically the seed layer is deposited at 50° to 100°C at very high power (9 kW to 20 kW); the seed layer constitutes 30% to 50% of the desired metal thickness. The nucleation layer should be continuous to allow adequate adhesion of Al atoms that are deposited in step 4.

Step 4. Immediately (no delay) upon complete deposition of the Al seed layer, the chamber-heater backside Ar is turned on and the wafer is allowed to ramp to high temperatures of 460° to 500°C (typically up to 475°C). At the same time, ∼30% to 50% of the remainder Al film is deposited at a slow rate, which ensures good window-, via- and trench-fill as well as planarization in the field areas. Typically this step takes about 120 s to 240 s; process conditions are chosen such that a plasma is sustained at low power levels supplied to the sputtering source.

Step 5. The remainder thickness of the Al film is now deposited at a very rapid rate, again at 9 kW to 20 kW, while the wafer is held at constant temperature. This step allows further planarization and cuts down on the time required for the entire Al deposition process.

Step 6. If required, especially for high aspect ratio openings, the wafer may be allowed to sit for an extended period in the chamber at elevated temperatures to allow for complete planarization. The stack is now capped with a TiN or Ti/TiN antireflection coating in an adjacent chamber. In some cases, a poststack-deposition sinter anneal of 400° to 425°C for 1 h to 2 h is common.

Planarized multilayered Al-alloy structures are preferred over conventional tungsten (W) plugs for the following reasons:

(1) An all-Al-alloy structure shows a much lower resistance than a W-plug/Al-alloy interconnect structure because of the inherent lower resistivity of Al-alloys (∼3 $\mu\Omega$·cm to 3.5 $\mu\Omega$·cm) versus W (∼7 $\mu\Omega$·cm to 10 $\mu\Omega$·cm).

(2) The plug deposition and the interconnect are accomplished in one singular deposition sequence as the plug and interconnect are all made of the same material. In the conventional W-plug case, the CVD-W deposition is followed by an etch-back or a chemical–mechanical polish process to form W-plugs. Subsequently, an interconnect stack is then deposited; separate processing steps are required for W-plug formation and Al-alloy interconnect. In an all-Al solution, a singular step is required to form the plug and the interconnect, thus reducing cycle time and costs during manufacture.

(3) All CVD W-plugs depositions have to be preceded by a Ti/TiN adhesion/barrier stack. The Ti film allows the stack to adhere to the contact or via dielectric, and the TiN serves as a nucleating layer for CVD-W deposition. While a Ti/TiN stack is required at the contact level for multilayered planarized Al-alloy structures, at the via level only a Ti wetting layer is necessary, thus further reducing the overall wafer fabrication cost.

(4) A cost comparison for an optimized planarized Al-alloy process shows significant advantages over the conventional W-plug deposition and etch-back processes. At the via level, this cost reduction is estimated to be between $8 and $10 per level for 200 mm diameter wafers.

(5) The via chain resistance and individual interlevel contact resistance with planarized Al-alloy structures are significantly lower than those obtained with conventional W-plugs.

(6) Low processing temperatures associated with multilayered Al-alloy structures make these films ideally suited for low glass transition temperatures of advanced low-ε dielectrics, which are becoming increasingly popular in the industry [see (33)].

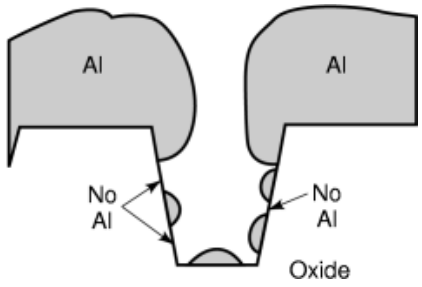

Fig. 18. Normalized via chain resistance for low pressure (LP) Al and CVD-W filled via holes with various via sizes.

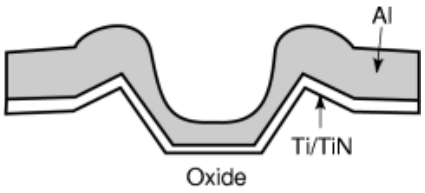

Fig. 19. SEM Cross-Section of 0.35 μm vias with an Al thickness of 350 nm using the low-pressure Al planarization process.

(7) Multilayered planarized Al-alloy films lend themselves to structures fabricated using damascene or dual-damascene techniques.

In order to improve device performance, the introduction of low-dielectric-constant (ε) interlevel dielectrics (*ILD*) that can reduce interline capacitance has to be considered [see (33)]. The introduction of low-ε materials depends very strongly on the allowed process thermal budget. At present, a typical CVD-W process exposes the ILD to temperatures of $\sim$400° to 470°C. Current low-ε materials are not compatible with these temperatures. For device manufacturers to consider planarized Al, in combination with low-ε materials, equipment suppliers have to develop Al planarization processes that (a) are compatible with the above-mentioned thermal budgets, (b) do not cause photolithography alignment problems, (c) are capable of filling high aspect ratio via holes, and (d) offer full-face deposition (entire wafer) in order to maximize die yield.

For reasons just described, there is a drive in the industry to reduce Al-alloy deposition temperatures. Recently, two papers have discussed depositions at low wafer temperatures (<400°C). In the study by Biberger et al. (34), the Al alloy was also deposited at low pressures ($\sim$0.7 mtorr instead of 2 mtorr commonly used) to give adequate fill for sub-0.5 μm openings. At low pressures, atomic collisions between sputtered Al atoms and Ar ions are reduced so that directional deposition into the opening is enhanced. Moreover, small hardware changes, such as properly configured sources and magnet designs (34) or the deployment of a floating shield to maximize electron utilization in the plasma (35), allow the plasma to be sustained at such low pressures. Layered low-stress Al films have been fabricated using the above low-pressure process with adequate grain growth (34). Figure 18 shows the normalized via chain resistance as a function of via size for the two cases of low-pressure Al planarization and conventional W-plugs. The graph demonstrates the advantages of Al over CVD-W: The average via chain resistance is, by a factor of 2.5, lower for Al than for W in the case of 0.35 μm via diameter. A typical scanning electron micrograph of a series of filled via openings using the low-pressure two-step planarization process is shown in Fig. 19. In the study by Wang et al. (35), the integration of a relatively low wafer temperature (<400°C), a floating shield, and low pressure ($\sim$0.35 mtorr) allows filling of via with 0.25 μm to 0.5 μm widths and aspect ratios of 2.4 to 4.8:1 with multilayered planarized Al.

The Al-fill process can also be accomplished using an integrated CVD-Al + PVD-AlCu planarization process (36). CVD of Al films has been demonstrated in previous publications (37), but several problems have

precluded its use: (a) the lack of Cu-doping for improved resistance to electromigration due to incompatibility of using both Cu and Al precursors for codeposition, (b) high precursor cost, (c) rough surface morphology that makes lithography difficult, and (d) void formation in the opening that results in failure to completely fill the window or via. These several issues, especially the formation of voids, have prevented acceptance of a fully planarized CVD-Al structure since entrapped gases or other materials in such voids—or, in the worst case, seams—pose a large reliability risk. Consequently, a CVD–PVD integrated process has been proposed (36), consisting of the deposition of a thin layer of CVD-Al, followed by PVD Al–Cu deposition and an in situ planarization at wafer temperatures $<400^{\circ}$C. A very thin (<60 nm), highly conformal CVD-Al layer that serves as a nucleation layer is first deposited using dimethyl-aluminum hydride (*DMAH*) at temperatures $<250^{\circ}$C at a deposition rate of 6 nm/s. Subsequently, a 500 nm Al–Cu alloy film is deposited using conventional PVD techniques and allowed to planarize in situ for 3 min at a wafer temperature of less than 400°C. Such planarized CVD-Al–PVD Al–Cu integrated reflow structures are reported to exhibit significant improvements in electromigration resistance and contact and via resistance over conventional W-plug and Al-interconnect technology.

As conductor dimensions approach the film thickness, electron microscopy studies also show the complete obliteration of the "layered" structure into a "bamboo" or "near-bamboo" Al-alloy microstructure. The overall Al-alloy grain size in planarized structures is large (1 μm to 5 μm is not uncommon), with very little incidence of small grains that can cause mechanical reliability problems often associated with films fabricated using conventional single-layered Al deposition methods (38).

The Ti wetting layer in multilayered Al-planarization processes acts to promote Al diffusion over its entire surface during Al-alloy deposition. Solid-state interdiffusion and high surface mobility, driven by capillary forces, are responsible for Al mobility during the "fill" process. Additionally, energy transfer from the bombarding atoms and the available latent heat of condensation during deposition also assist in obtaining a filled opening. Moreover, the Ti wetting layer of step 2 transforms to $TiAl_3$ as a result of a reaction between the Ti and Al films during steps 3 to 6. A continuous layer of $TiAl_3$ acts as a shunt and alternate current path, improves electromigration behavior (39) and reduces "hillock" formation (40). Significant improvements in electromigration lifetimes have been reported in the literature when the Ti layer has been used (41). Improvements in mechanical properties, as seen by lack of stress-induced void formation, have also been reported (43).

Yao (44) has suggested a model that describes Al-planarization and the subsequent fill process. Results show that the fill mechanism varies with the dimensions of the opening. Large openings tend to fill from the bottom up, and small openings tend to fill from the top down. The material flux into the opening is dictated by the radius of curvature at the corner of the opening and the opening diameter itself. Simulation studies on Al-planarization have been published by Dew et al. (45).

The need for temperature hierarchy during wafer fabrication and the absence of an adequate barrier against Al spiking and junction leakage has prompted the use of W-plugs at the contact level and has multilayered all Al-alloy structures at via levels (46). The reader is invited to examine the several papers cited in this section [also, see (47)] to compare the advantages and disadvantages of using planarized multilayered Al-alloy structures for submicron IC manufacture.

BIBLIOGRAPHY

1. K. E. Bean, W. R. Runyan, Dielectric isolation: Comprehensive, current and future, *J. Electrochem. Soc.*, **124** (1): 5, 1977.
2. P. K. Roy, Multi-layering process for stress accommodation in deposited polysilicon, US patent 4,742,020, May 3, 1988.
3. P. K. Roy, Technique for reducing substrate warpage springback using polysilicon subsurface strained layer, US patent 4,631,804, December 30, 1986.
4. P. Rosser, G. Tomkins, Silicides for VLSI interconnects, *Vacuum*, **34**; 419, .
5. S. P. Murarka, Refractory silicides for integrated circuits, *J. Vac. Sci. Technol.*, **17**; 775, 1980.

6. R. S. Rosler, Low pressure CVD production processes for poly, nitride, and oxide, *Solid State Technol.*, **20** (4): 63–70, 1977.
7. B. S. Meyerson, M. L. Yu, Phosphorous-doped polycrystalline silicon via LPCVD, *J. Electrochem. Soc.*, **131**: 2361, 1984.
8. S. Chittipeddi, P. K. Roy, Silicide formation on polysilicion, US patent 5,147,820, September 15, 1992.
9. J. R. Radosevich, P. K. Roy, Forming a device dielectric on a deposited semiconductor having sub-layers, US patent 5,298,436, March 29, 1994.
10. W. Kern, D. A. Puotinen, Cleaning solutions based on hydrogen peroxide for use in silicon semiconductor technology, *RCA Rev.*, **31**: 187, 1970.
11. T. Watanabe *et al.*, A 100A thick stacked $SiO_2/Si_3N_4/SiO_2$ dielectric layer for memory capacitors, *Proc. Int. Reliability Phys. Symp.*, 1985, p. 18.
12. P. K. Roy, A. K. Sinha, Synthesis of high-quality ultra-thin guide oxides for ULSI applications, *AT&T Tech. J.*, **67** (6): 155–174, 1988.
13. E. A. Irene, Thermally prepared SiO_2 films for VLSI, *Semicond. Int.*, **8** (6): 91–97, 1985.
14. I. Ohdomari *et al.*, The structural models of the Si/SiO_2 interface, *J. Non-Cryst. Solids*, **89** (1–2): 239–248, 1987.
15. R. T. Howe, R. S. Muller, Stress in polycrystalline and amorphous silicon thin films, *J. Appl. Phys.*, **54** (8): 4674–4675, 1983.
16. P. K. Roy, V. C. Kannan, Characterization of Si/SiO_2 interfaces using TEM lattice imaging and X-ray microdiffraction techniques, *MRS Fall Symp.*, **138**: December 1988, pp. 367–372.
17. H-H. Tseng *et al.*, A comparison of CVD stacked gate oxide and thermal gate oxide for 0.5-μm transistors subjected to process-induced damage, *IEEE Trans. Electron Devices*, **40**: 613, 1993.
18. M. Delfino, J. A. Fair, S. Salimian, Thermal Nitriduation of Silicon in a Cluster Tool, *Appl. Phys. Lett.* **60**: 341, 1992.
19. Y. Ma, S. M. Merchant, P. K. Roy, Layered Silicon Nitride Deposition Process, US patent 5,981,403, November 9, 1999.
20. S. Murarka, Metallization, in S. Sze (ed.), *VLSI Technology*, New York: McGraw-Hill, 1988, p. 375.
21. D. Pramanik, V. Jain, Barrier metals for ULSI: Deposition and manufacturing, *Solid State Technol.*, **36** (1): 73–76, 78–79, 82, 1993.
22. D. Pramanik, A. Saxena, Aluminum metallization for ULSI, *Solid State Technol.*, **33** (3): 73–79, 1990.
23. C. J. Tracy *et al.*, Multistep metallization process, US patent 4,970,176, November 13, 1990.
24. Z. Xu *et al.*, Contact, via and trench fill by Al planarization process, *VLSI Metallization Conf.*, 1994, p. 158.
25. C. H. Ting *et al.*, Process optimization of planarized Al via for sub-half micron ULSI interconnects, in R. Blumenthal and G. Janssen (eds.), *Advanced Metallization for ULSI Applications in 1994*, Pittsburgh: Materials Research Society, 1995, p. 517.
26. C. S. Park *et al.*, AL-PLAPH: Aluminum-planarization by post-heating process for planarized double metal CMOS applications, *VLSI Metallization Conf.*, Santa Clara, 1991, p. 326.
27. R. B. MacNaugton, Method of forming an aluminum metal contact, US patent 5,374,592, December 20, 1994.
28. E. Ong, Method of filling contacts in semiconductor devices, US patent 5,371,042, December 6, 1994.
29. D. Liu *et al.*, Properties of Ti and Al thin films deposited by collimated sputtering, *Thin Solid Films*, **236** (1–2): 267–273, 1993.
30. T. Mandrekar *et al.*, The Endura advanced liner tool—integrated Ti/TiN and barrier technology, *Appl. Mater. MPD Prod. Technol. Publ.*, **1** (1): 8, 1997.
31. J. Forster, Vectra IMP source, *Appl. Mater. HP PVD Prod. Technol. Publ.*, **3** (3): 2, 1996.
32. M. Kageyama, K. Hashimoto, H. Onoda, Formation of texture-controlled aluminum and its migration performance in Al-Si/TiN stacked structures, *29th Annu. Proc. Reliability Phys., IEEE*, 1991, p. 97.
33. M. T. Bohr, Interconnect scaling—the real limiter to high performance ULSI, *Solid State Technol.*, **39** (9): 105–106, 109, 111, 1996.
34. M. Biberger *et al.*, Low pressure aluminum planarization for sub-0.5 mm via holes, in D. Patel and M. Graef (eds.), *Multilevel Interconnect Technology*, Bellingham, WA: SPIE, **3214**: 1997, p. 72.
35. H. Wang *et al.*, Warm Al for via planarization, *Appl. Mater. MPD Prod. Technol. Publ.*, **1** (1): 3, 1997.
36. A. Konecni *et al.*, Use of integrated CVD Al–PVD AlCu reflow, *Future Fab Int.*, **1** (3): 211, 1997.
37. T. Ohta *et al.*, New structure of selective-CVD-Al filling direct on Al for submicron plug applications and process using UHV cluster tool, *VLSI Metallization Conf.*, Santa Clara, 1994, p. 329.
38. V. Ryan, S. M. Merchant, Enhanced stress-migration reliability for ULSI interconnect: An insight into the perils of screening Al depositions based on grain size, *J. Electron. Mater.*, **24** (8): 969–974, 1995.

39. C.-C. Lee, E. S. Machlin, H. Rathore, Roles of Ti-intermetallic compound on the electromigration resistance of Al–Cu interconnecting stripes, *J. Appl. Phys.*, **71** (12): 5877–5887, 1992.
40. J. K. Kim, Method for manufacturing aluminum wired layer for interconnecting device to device in a semiconductor device, US patent 5,332,693, July 26, 1994.
41. K-Y. Fu *et al.*, Electromigration resistance versus Ti thickness for Al–Si–Cu metallization with Ti barrier and with rapid-thermal-annealed TiN barrier, *VLSI Met. Conf.*, 1992, 407.
42. H. Nishimura, T. Yamada, S. Ogawa, Reliable submicron vias using aluminum alloy high temperature sputter filling, *VLSI Met. Conf.*, 1991, p. 170.
43. M. Kanazawa *et al.*, Stress induced void formation of vias in the Al-based multi-level interconnection system, *VLSI Met. Conf.*, 1991, p. 221.
44. G. Yao, Local thermodynamic analysis for Al planarization, *Appl. Mater. HP PVD Prod. Technol. Publ.*, **2** (4): 10, 1995.
45. S. K. Dew, T. Smy, M. J. Brett, Simulation of elevated temperature Al metallization using SIMBAD, *IEEE Trans. Electron Devices*, **39**: 1599, 1992.
46. J. Y. Lee *et al.*, Simultaneously formed storage node contact and metal contact cell for 1 GB DRAM and beyond, *IEDM Tech. Dig.*, 593, 1996.
47. F. H. Gn, L. J. Liu, M. Guo, A comparison study between tungsten and aluminum plug for submicron contact/via manufacture, in F. Chen and S. Murarka (eds.), *Microelectronics Technology and Process Integration*, Bellingham, WA: SPIE, **2335**: 1994, p. 98.

PRADIP K. ROY
SAILESH M. MERCHANT
Lucent Technologies–Bell Laboratories