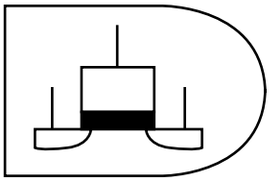
ISOLATION

The definition for "electrical isolation" in a typical encyclopedia/dictionary would be something like "physical and electrical separation of the parts of an electrical circuit to prevent interference currents within or between the parts." While this definition is correct, isolation is a highly complex subject that needs to be described within the context of the technology being discussed. In this article we will consider isolation within the world of microelectronics, more specifically in the area of silicon very large scale integration (VLSI).

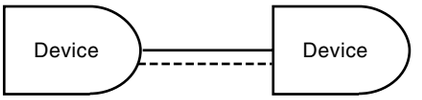
The integrated circuit (IC) was invented by Kilby (1) in 1958. The first ICs were phase shift oscillators and flip-flops, fabricated in germanium substrates. The individual devices in these circuits were laterally isolated mesa structure, which is no longer used. Over the subsequent 40 years, isolation technologies were developed, and integrated circuits evolved into complex electronic circuitry containing millions of devices on a single piece of silicon, routed using many layers of interconnect metalization.

Within VLSI, isolation has various meanings depending upon the interaction between the different components. As sketched in the following, we will begin discussing isolation with individual devices and then proceed through the different components as the level of integration increases up to and including isolation issues in complex systems such as mixed digital/analog circuits.

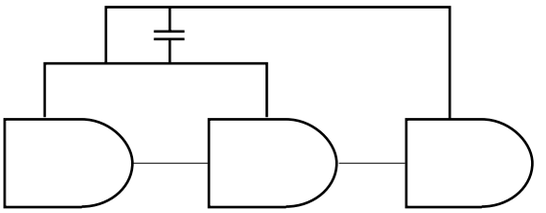
1. *Isolation Within a "Device."* In individual devices, typical isolations are as follows: *gate dielectric,* which is an insulator, yet its role is regulating the channel between the source and the drain contacts, and *junction isolation,* whose isolation is created by a reverse bias voltage.



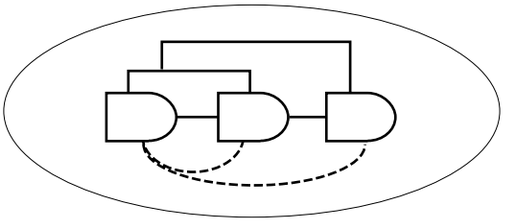
2. *"Device–Device" Isolation.* Normally referred to as lateral isolation, the *field oxide* controls surface leakage effects; that is, it stops the leakage paths between active components.



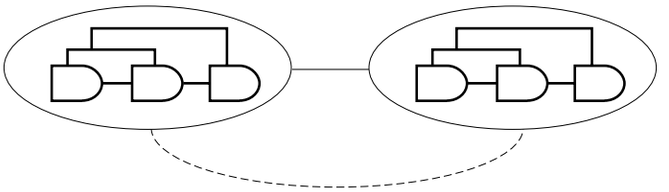
3. *"Device Interconnect" Isolation.* Metal lines which connect the devices to make up electrical circuits need to be isolated, which means no current should flow between them, but one cannot eliminate the capacitive coupling.



4. *Isolation Within a "Circuit."* Unwanted coupling of the devices through the substrate within a tub or between two tubs.



5. *Isolation Between "Circuit Blocks."* The coupling between sub-circuit blocks fabricated on the same substrate can give rise to noise in the more sensitive components of a circuit.



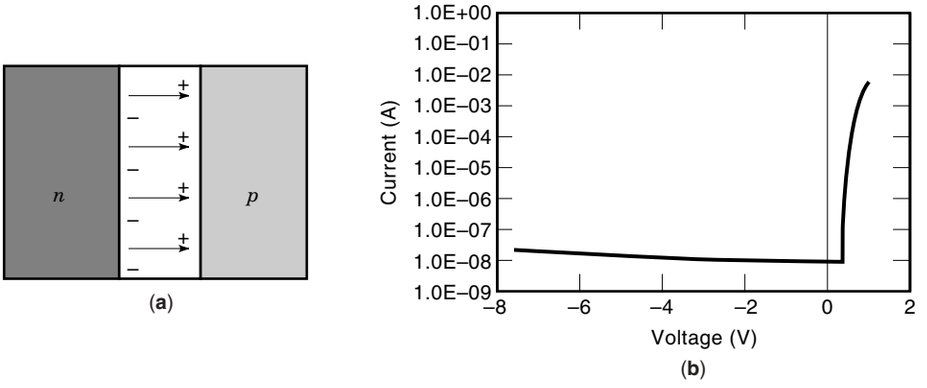


Figure 1. (a) Schematic representation of a p–n junction. (b) Junction I–V characteristics, illustrating forward and reverse bias.

Isolation in integrated circuits are of *insulating, p–n junction,* or a *mixed* type. The gate electrode in a metal-oxide-semiconductor field-effect transistor (MOSFET) and the various levels of interconnect (wiring) are isolated by insulating dielectrics/oxides. The source/drain junctions are isolated from the substrate by the depletion layer formed between two oppositely doped regions. The lateral isolation between devices, on the other hand, is provided by the combination of an oxide and a chemical doping underneath the oxide to prevent formation of an inverted surface and its electrical path.

To visualize the importance of isolation, consider a VLSI circuit that has one million transistors. If each transistor produces 1 μA leakage, the total power dissipation for the entire chip would increase on the order of a few watts.

We will cover the different types of isolation and how they will apply to the different areas of IC technology in the following order: (1) intradevice, (2) interdevices, (3) between metal lines, (4) within subcircuits, and (5) between subcircuits. These will be discussed in the field of Si ICs. The extrapolation to III–V materials is straightforward.

ISOLATION IN DEVICES

Diode/*p–n* Junction Isolation

Most semiconductor devices incorporate both p- and n-type regions. The n region, a semiconductor doped with electron *donors,* is characterized by the fact that it contains a large number of electrons. Similarly the p-region, doped with electron acceptors, contains a large number of holes. At the junction between two oppositely doped semiconductors, as shown schematically in Fig. 1(a), electrons and holes flow out; establishing an electric field which keeps the holes in the p-region and the electrons in the n-region. The region thus formed is depleted from charge carriers.

It is the junction between these regions that leads to the useful electrical characteristics of diodes. Diodes permit the passage of electric current in only one direction. As indicated in Fig. 1(b) when a negative voltage is applied to the n region a current begins to flow (forward bias). In contrast, when a positive voltage is applied to the n region, the current drops by 6 orders of magnitude. Eventually when the positive voltage is large enough, current does begin to flow again (junction breakdown). *The reverse bias situation is what provides the isolation.*

MOSFET Isolation

Complementary metal–oxide–semiconductor (CMOS) technology, with its inherently low static power consumption, has become dominant in VLSI applications.

MOSFETs are built in Si substrates. A MOSFET is a four-terminal device including gate, source, drain, and substrate (Fig. 2). Voltage applied to the gate controls the charges in the channel region. The gate electrode is isolated from the channel (substrate) by a dielectric/oxide through which no current should flow.

The source and drain (S/D) terminals make the connection to this conducting channel and provide the input and output terminals. The S/D areas are regions doped heavily opposite to the substrate. They are isolated from the substrate by junction isolation. Likewise the channel area is isolated from the substrate by the depletion region formed between the channel and the substrate. The substrate is the fourth terminal through which a bias can be applied. The gate oxide and the S/D junctions are the main isolation areas within a MOSFET.

Gate Oxide. Perhaps the most critical element in a MOSFET is the gate dielectric, which in silicon technology is often silicon dioxide (SiO_2). SiO_2 layers on Si can be formed by various methods, such as deposition, but the most often used technique is thermal oxidation of Si. SiO_2 is a very stable oxide.

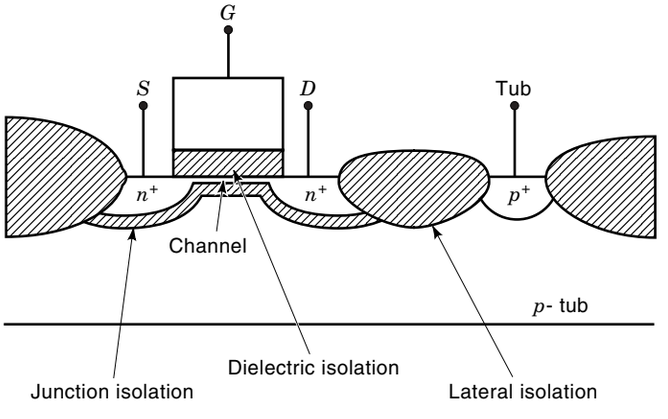


Figure 2. The basic elements of a MOSFET, illustrating the various isolation areas: gate dielectric, S/D and channel depletion regions, and the lateral isolation.

It provides the basic feature for the planar Si technology. The advent of thermally oxidized silicon increased the feasibility of fabricating field-effect transistors (2), which is responsible for the fact that metal-oxide-semiconductor (MOS) transistor has become the most important device in ICs. SiO_2 gate oxide, along with its interface with Si, has been studied and characterized in great detail (3). They are still the subject of intense studies because the fabrication of thinner oxides is necessary for device scaling (4).

In order to improve the device performance for advanced technology, gate oxide has been scaled aggressively. While gate oxide thickness for a 1.0 μm technology is about 15 nm, it needs to be reduced to 3.0 nm or smaller for the 0.13 μm technology. The ultrathin gate oxides are more susceptible to increased leakage, tunneling, and reliability issues. They are also sensitive to process-induced damage, because of the electrical connection to interconnects, and charging up during various processing steps (5).

There are two major challenges as gate oxide thickness decreases: (1) The gate leakage current through the gate oxide increases significantly, and (2) boron penetration from p^+-poly gate used in surface-channel p-MOSFETs increases significantly.

Nitrided oxides have been proposed as an alternative gate dielectric for VLSI technologies. Nitrogen incorporation retards the oxidation rate, making the growth of thin oxides at standard oxidation temperature possible. Also their resistance to boron penetration is an important scaling benefit compared to conventional oxides. Nitrogen can be incorporated in the oxide in a number of ways—for example, by growing the oxide in a nitrogen-containing atmosphere (4,6), or by implanting N into the Si prior to oxidation (4).

In general, gate oxide yield and reliability is the most important issue for the VLSI circuits.

Source/Drain Junctions. The source and drain regions of transistors are isolated by the depletion layer between the doped junctions and the substrate, which is of the opposite charge. Their isolation mechanism is much like the junction isolation in diodes, which was already discussed in a previous section.

In VLSI, junctions are formed by implanting dopants into the S/D areas. A thermal cycle is then used to activate the dopants and to drive the junctions beyond the area damaged by the implantation. Device miniaturization requires shallow junctions. In the case of shallow junctions, if proper care is not taken, then the processing damage (dislocations, defects, and impurities), will lie within the depletion layer. This gives rise to leakage currents which can be detrimental to circuit operation.

Fabrication of shallow p^+–n junctions is particularly difficult, because of boron channeling and fast diffusion. Various techniques such as silicided junctions and raised junctions are being used to address the shallow junction problem (4).

Bipolar Isolation

Bipolar transistors were used in the first integrated circuits. A bipolar is also a four-terminal device: the emitter, the base, the collector, and the substrate. It is made of two p–n junctions (the emitter–base and the collector–base junctions) in very close proximity, with the region between the two being the base. External voltages and currents can be applied to each of the terminals. In an n–p–n bipolar transistor, shown in Fig. 3, the base region consists of (1) a p-type region created in a lightly n-doped epitaxial layer and (2) a n^+-doped layer beneath the n-epi layer. Junction isolation is the principle isolation mechanism in a bipolar. It is used for isolating the three (emitter, base, and the collector) regions from each other.

High-performance bipolar n–p–n transistors require a heavily doped buried n^+-layer, or subcollector, to lower the collector series resistance underneath the thin epitaxial collector region. This collector region requires additional junction isolation which consists of a p-type region formed in an n-type substrate, surrounding the device region.

INTERDEVICE ISOLATION

Lateral isolation has to provide sufficient device-to-device isolation within specific circuit requirements. The key challenge in the IC technology is reducing dimensions, which makes the lateral isolation very critical.

Isolation processing needs to fulfill the following criteria: (1) It has to be scaled properly in order to take full advantage of the device miniaturization, and (2) the isolation process should be manufacturable and have high yield.

The isolation processing, which takes place prior to the device processing, should have no negative impact on the device properties. A serious consideration, which is not readily obvious, is the impact of the isolation processing on the silicon where the thin gate oxide will be grown. It is imperative that the gate-oxide integrity be ensured. The gate oxide along the borders of the active zone is particularly important (7). The transistor subthreshold double-hump effect is another problem with some of the isolation technologies; an additional

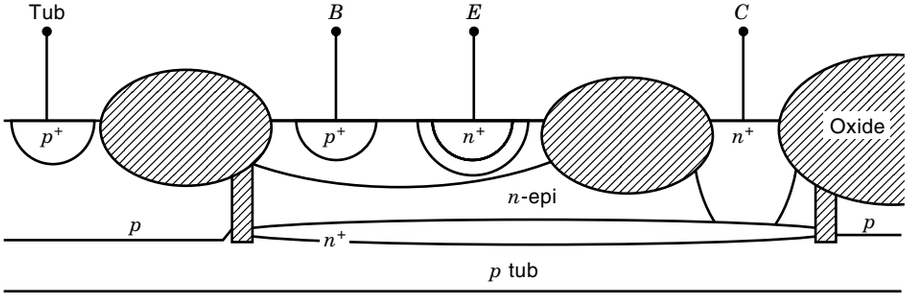


Figure 3. The basic elements of a bipolar, illustrating the oxide and junction isolations.

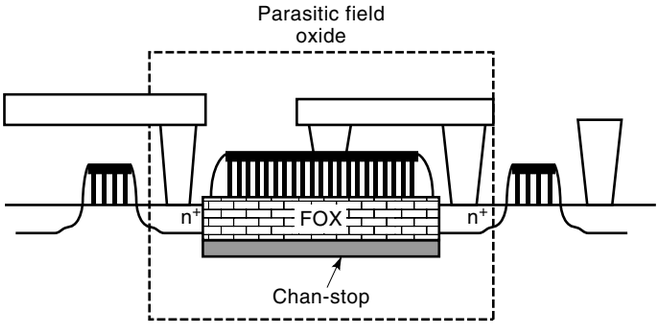


Figure 4. Two n-channel MOSFETs and the field oxide parasitic transistor, made by the polysilicon gate runner above the field oxide.

problem is the presence of the parasitic channels along the "side-wall edges" under the gate which give rise to subthreshold leakage.

In early bipolar integrated circuits, junction isolation was the main isolation scheme for device–device isolation. Excellent lateral isolation can be achieved by using junction isolation; however, junction isolation requires large areas compared with the dielectric isolation, and it introduces large parasitic capacitance. Bipolar technology now employs an oxide-based isolation much like the MOS technology, discussed below.

MOS transistors are isolated as long as the surface of the silicon is inverted under the gate and the S/D are in reverse bias condition. They do, however, consume power through unwanted leakage in their "off" state. In the region between MOSFETs, without adequate lateral isolation, the surface of the silicon can be inverted and provide a leakage path.

In the MOS circuits, isolation between two devices is provided by a thick oxide called the "field oxide" (FOX). The complication that arises in this type of isolation is that often a poly-line runs between two transistors as illustrated in Fig. 4, under the poly-gate between the MOSFETs parasitic transistor forms.

If the threshold voltage of the parasitic transistor is too low, an inversion layer can form between the n^+ regions of the individual transistors and tie them together. The threshold voltage (V_t) and punch-through of the parasitic FOX should be high enough to prevent leakage according to the circuit requirements. This can be accomplished by using a heavily doped Si layer (channel stop) underneath. Two methods are typically used to increase FOX V_t: increasing the FOX thickness and raising the doping beneath the field oxide.

Field Oxide Structure

The field oxide can be formed in a number of ways, some of which are discussed below. For a comprehensive review see Ref. 8.

Oxide Grow-and-Etch. In the first approach (used until about 1970), the oxide is grown to the desired thickness on a flat silicon surface and then etched in the active regions [Fig. 5(a)]. Although this grow-and-etch technique allows the isolation region to be sharply defined, the fact that the field oxide steps are high and have sharp upper corners has prevented it from being utilized in the VLSI applications. The sharp and high corners result in poor step coverage at the subsequent metal-interconnect processing.

LOCOS. The most popular technique for lateral isolation is LOCOS, an acronym for LOCal Oxidation of Silicon (9). Conventional LOCOS and its derivatives have been remarkably successful in meeting device isolation requirements for MOS technologies developed and manufactured over the past quarter century. Reasons for the longevity of LOCOS-based approaches are its use of relatively simple and well-controlled process steps and the fact that the MOSFET scaling has good synergistic requirements with the isolation, such as shallow junctions and high tub doping. LOCOS forms a semirecessed oxide in the field areas of the substrate [Fig. 5 (b)]. In VLSI a thin layer of oxide (about 10 nm to 30 nm) is thermally grown on the silicon wafer, followed by the deposition of a relatively thick layer of silicon nitride (about 150 nm to 300 nm) forming the LOCOS stack. The field regions are defined with a lithographic step; depending on the technology, either photo, deep ultraviolet (DUV), or e-beam lithography is employed.

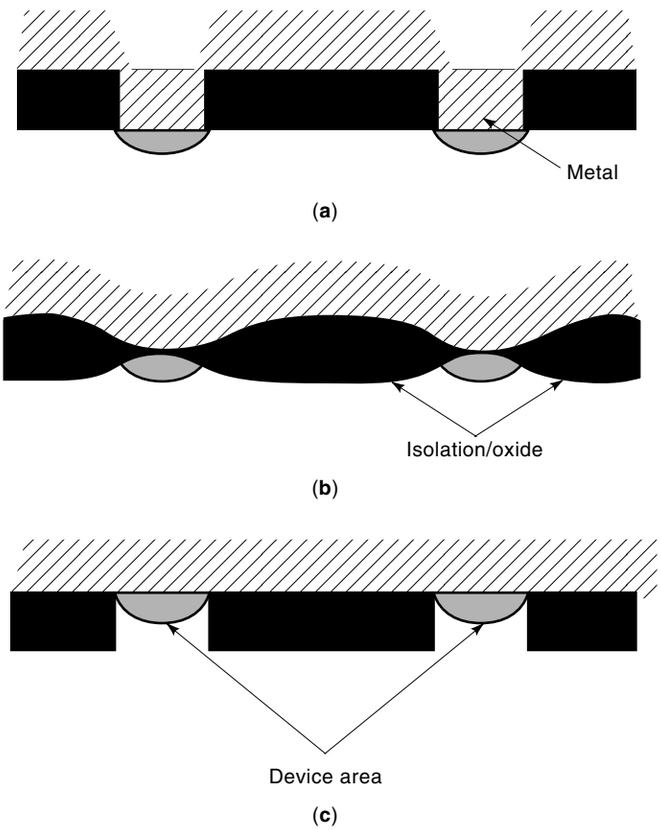


Figure 5. Schematic of three lateral isolation techniques: (a) oxide is grown everywhere and windows are opened; inside the windows the devices are made. (b) Device areas are masked and a local oxidation of the isolation area is performed. (c) Trenches are cut around the device areas and then filled with an oxide.

In the older technologies a channel stop implant was used in the field oxide areas prior to the FOX growth. This had a number of problems associated with it (see section entitled "Isolation Doping".

The field oxide is grown selectively in the field oxide regions (not masked by the nitride layer) at high temperatures of 950°C to 1100°C. Field oxide forms a semirecessed oxide in the Si, with slightly more than half of it protruding above the silicon surface. The SiO_2 above the Si layer is partially removed during the subsequent etch steps, resulting in less topography.

Silicon nitride is an effective oxidation barrier because oxygen and water vapor diffuse very slowly through it, thus preventing oxidizing species from reaching the silicon surface under the nitride. However, since the oxidants can diffuse through the pad oxide at the edges of the nitride, some oxide is also grown under the edges, thereby resulting in the FOX encroachment and "bird's-beak" (BB) formation which results in transistor width loss [see Fig. 5(b)]. In order to reduce the BB, in principle, the thickness of the pad oxide can be decreased. However, because of the stress that the nitride has on the Si, the amount of reduction is very limited.

Another shortcoming of LOCOS is that less oxide is grown in the narrow-field regions, the so-called "FOX thinning" (10), as compared with the wide-open areas. As a result the threshold voltage of parasitic transistors in small geometries is reduced, causing conduction between devices.

After the LOCOS stack removal, a thin oxide is grown in the active device areas. This oxide is used as a sacrificial (clean-up) oxide prior to the gate-oxide growth, and it also serves as the screen oxide for the channel implants.

Modified LOCOS Isolation Schemes. For use in more advanced technologies, many variations of LOCOS have been suggested, to limit the field oxide encroachment and/or reduce FOX topography, but few have been accepted in production environment. Some of these LOCOS derivatives are as follows:

1. Poly buffer LOCOS (PBL) (11), perhaps the most widely used derivative of LOCOS, is a modified LOCOS isolation in which a thin poly layer is inserted between the pad oxide/nitride stack. This helps reduce the FOX encroachment (by allowing the use of a thinner pad oxide and/or thicker nitride), without increasing the stress of the nitride on the silicon.
2. Fully recessed LOCOS achieved by an anisotropic overetch into the Si at the active area definition is also an attractive alternative, because of the resulting reduced topography at the gate lithography, but it results in significant FOX encroachment.
3. In the sealed-interface local oxidation (SILO) technique a thin layer of nitride is deposited on the Si surface before depositing the pad oxide in order to seal the lateral diffusion of the oxidants under the BB. This technique has the possible drawback of increasing the gate oxide defects at the isolation edge.

With the scaling of device dimensions, and in particular the supply voltage, the required FOX thickness has also been reduced. With a thinner field oxide thickness, BB is shorter and the FOX thinning is reduced, making the "standard" LOCOS isolation scheme a viable isolation technique again (12,13). FOX cannot, however be made too thin since it can increase the capacitance of the polysilicon runners on the field oxide to the substrate.

Trench Isolation. Shallow trench isolation (STI) combined with chemical mechanical polishing (CMP) planarization is becoming an important isolation scheme for 0.20 μm devices and smaller, mainly due to its scalability and planar topography (14). The trench isolation scheme is more complex than LOCOS, but it offers less topography and smaller BB [Fig. 5(c)].

In the trench/refill process, after defining the active device areas, as described above for LOCOS, the field areas are filled with glass/oxide, instead of thermally growing the oxide on Si. In this process, trenches (0.3 μm to 3.0 μm deep, depending on the application) are anisotropically etched into the silicon substrate in the field areas. This is followed by a mini-LOCOS oxidation, which rounds the top of the trench corner. The mini-LOCOS oxide is etched away and a CVD oxide is deposited onto the wafer surface. Following this, the topography of the CVD oxide is removed by a chemical mechanical polish.

The dielectric material used to fill the trenches is one of the key parameters in making STI successful. The material criteria are as follows: gap-filling of narrow trenches, HF etch rates in field and gaps, moisture absorption, film stress and shrinkage due to thermal cycles, and CMP polish rates. The trench corner engineering to achieve a smooth top corner is another important factor (e.g., with sloped trench etch and corner rounding oxidation) for gate-oxide integrity and to prevent transistor threshold double hump.

Isolation Doping

For a high enough V_t to sustain the isolation, the doping beneath the field oxide should be raised. In the older technologies, ion implantation (channel stop implant) prior to the FOX growth was used to accomplish this. The disadvantage of this technique is the lateral diffusion of the dopants under the field oxide, due to the long, high-temperature cycle for the FOX growth which gives rise to narrow width effect.

In addition in a CMOS process, an extra lithography step is needed to shield either *n*- or *p*-tub regions from the wrong implant, adding to process complexity.

With the scaling of devices the thickness of the FOX has also been reduced, making it possible for the channel stop dopants to be implanted after the FOX growth (15), resulting in improved device isolation (16). This can be done in the standard diffused tubs at the same lithography as the threshold voltage adjust implant or by employing the high-energy implant (HEI) tub process described below.

Profiled Tubs Using High-Energy Ion Implantation. The primary advantage of the HEI tub is that the tub and channel stop dopants do not experience the large tub drive and FOX growth thermal cycles. Conventional tubs are formed by implanting dopants and diffusing them to the desired depth. However, the dopants diffuse laterally as well as vertically, which reduces packing density. If a high-energy implant is used to place the dopants at the desired depth without further diffusion, much less lateral spread will occur (16–18). This

tub engineering results in a profiled tub, instead of a uniformly doped tub, with optimized latchup immunity, isolation, and transistor characteristics.

In terms of processing ease, the HEI tubs are much superior to the conventional tubs. A conventional tub may require up to four additional levels of lithography, along with a tub drive which not only is costly, but has risks associated with it.

Because the dopants are implanted in the correct depth, the HEI-tub process does not require a long, high-temperature tub drive. Also the isolation process is done prior to dopant implants, and thus the dopant profiles are not affected by the isolation thermal cycles. The implants are activated with a moderate anneal cycle which minimizes dopant diffusion. Often the gate-oxide growth thermal cycle is sufficient for the dopant activation.

Isolation Simulations. Process simulations have contributed to a better understanding of device physics and to the development of new processing techniques. This is also true for isolation, but to a lesser degree. These simulations have been most commonly performed for LOCOS due to its process simplicity and a greater applicability in IC technology. The aspects of LOCOS, which are useful to simulate, are the mechanical features such as oxide thickness (19), shape of the BB and stress in the Si (20), and the dopant profile.

The oxide growth simulation is based on the known physical effects such as diffusion, oxidation, viscoelastic flow of oxide, and so on, as summarized below (21).

In thermal oxidation, oxidant from the gas phase diffuses, in the form of O_2 molecules, through the SiO_2 network toward the interface to react at the interface and form new SiO_2 material. This growth is accompanied by a large volume increase. At sufficiently high temperatures, the reaction is aided by viscous flow of the oxide film toward the surface. For temperatures below 960°C, oxides exhibit viscoelastic behavior (22), while at temperatures below 600°C the elastic model holds.

With device sizes shrinking, more sophisticated process simulations such as done in Ref. 23 are required to predict the accurate shape of the oxide, the stress distribution, and the three-dimensional effects, such as center effect and mask-lifting effect. However, in order to ensure optimal control of the technological oxidation process, yet more accurate and robust oxidation models are needed.

INTERCONNECT ISOLATION

Isolated devices fabricated within the substrate need to be connected through specific electrical paths to form the electrical circuits. The connection is done using high-conductivity thin-film structures which are fabricated on an insulator above the Si surface and which make contact to the devices through small openings (contact holes) made in the insulator. In most circuits, more than one level of interconnect is used, up to five or six levels. These levels are separated from one another by dielectric layers called interlevel dielectrics (ILDs).

With the continued scaling of the devices, the interconnect technology becomes more compact and more complicated. One of the consequences of the higher packing density is the increase in the line capacitance which increases both power and delay. While the total circuit delay is dominated by the intrinsic gate delay for metal widths and spacings greater than 0.3 μm, it is largely determined by delays in the metal interconnects for metal features below about 0.2 μm. It is essential to reduce the interconnect capacitance in order to maintain the trend of low delay time, power consumption, and noise. The metal interconnect delays are associated with dramatic increases in line-to-line capacitance as the spaces between lines is decreased in the sub-quarter-micron regime. The total line capacitance is the sum of C_v (vertical or interlevel capacitance), and C_l (lateral or interline capacitance). It is worth noting that while C_l grows as the design rules shrink, the decrease in the ILD thickness (hence increase in C_v) is slower because of processing and reliability issues.

The standard ILD material is deposited SiO_2, with dielectric constant of 3.9. Lowering the dielectric constant of the ILD can be a solution to this problem. Doping with fluoride lowers dielectric constant considerably (~3.5). Also organic, low-k dielectric materials such as polyimides (which can have dielectric constant below 3) and porous material (with dielectric constant below 2) have attracted much attention (24).

These dielectric layers must have low capacitance, low leakage, and high dielectric breakdown field strength. In addition to the performance considerations, they must be compatible with standard IC fabrication steps in terms of adhesion to the substrate, thermal budget and stability, and various chemical etches used to make via holes or strip the resist. The deposition temperature, step coverage and gap fill capability, CMP properties for planar architecture, and contaminants and defects which affect the processing yield are also critical issues.

It is also worthwhile to note that while for the second and higher-level dielectrics the presence of Al limits the processing temperatures to below 450°C, the first-level dielectric which isolates the devices from the first-level metal can be deposited or flowed at relatively high temperatures (800°C). Hence, the material and the processing for the first ILD can be different from others. Also note that using certain barriers deposited between the metal and the junction is needed to prevent junction spiking.

INTEGRATED CIRCUIT ISOLATION

n-MOS and *p*-MOS

The first MOS ICs were built in the mid-1960s using p-channel transistors. This was despite the problems with p-MOS devices such as (1) a V_t of -4 V which required a V_{DD} of -12 V and (2) the low mobility. NMOS circuits with higher mobilities and a V_t of about 1 V, which was compatible with the existing 5 V bipolar technology, were not made until the mid-1970s.

The two major problems, which needed to be resolved before n-MOS ICs could be manufacturable, were both dielectric- and isolation-related. One was the positive or fixed charges beneath the gate oxide which caused inversion even at zero applied voltage, and the other was boron depletion under the field oxide, making it difficult to isolate n-channel devices. Understanding the gate oxidation issues, along with the introduction of ion implantation in MOS technology for the channel stop, was critical in NMOS IC development.

Another obstacle for the MOS IC technology was the lateral isolation topography inherent to the oxidation-etch technique, combined with the difficulty that it caused for the subsequent metalization. The advent of LOCOS in the mid-1970s with the smoothly tapered edges and smaller topography helped the device integration enormously.

MOS lateral isolation considerations are as follows: parasitic FOX transistor leakage and threshold voltage, FOX spacing, FOX encroachment, transistor subthreshold kink effect, and gate-oxide integrity.

CMOS

Complementary MOS (CMOS) circuits, by virtue of integrating both *n*-MOS and *p*-MOS transistors and requiring tub-to-tub isolation, make the isolation more difficult. However, even though NMOS has a better packing density, the power dissipation concerns makes CMOS a technology of choice when the number of transistors in an IC are in the millions. This is because NMOS logic gates draw direct-current (dc) power during one of the inverter states. However, in a CMOS gate, one of the *n*- or *p*-MOS transistors is always "off" when gate is in either of the logic states. Since no current flows onto the gate terminal and there is no dc current path from V_{DD} to V_{SS}, the resultant quiescent (steady-state) current—and hence power dissipated—is minimal.

In terms of lateral device isolation the various technologies exhibit different attributes with respect to minimum device spacing, isolation depth, surface planarity (topography), process complexity, and defects generated as a result of the isolation processing. When selecting an isolation technology for a particular circuit application, tradeoffs among these characteristics may be needed.

In addition to the tub-to-tub isolation, other difficulties with CMOS circuits are latchup and crosstalk as described in later sections.

Bi-CMOS

Bi-CMOS technology integrates large drive current bipolar transistors and low-power, high-density CMOS transistors in a single chip to yield speed–power–density performance unattainable in either technology.

The demands placed upon isolation increase as one goes from *n*-MOS to CMOS to Bi-CMOS. In the case of bipolar devices especially, device area is a strong function of isolation technology, and minimizing this area is the driving force behind the improvement of isolation even at the expense of greater processing complexity.

The deep n^+-collector contact requires a deep phosphorus implant followed by a thermal anneal which causes lateral diffusion of the dopants all the way through the epi layer. For a better packing density the polysilicon plug process has been developed. A trench is etched in the silicon down to the n^+ buried layer. This trench is refilled with heavily *in-situ*-doped polysilicon to form a low-resistance plug to the layer, while the side walls are covered with a dielectric to prevent lateral diffusion (see Fig. 3).

To further save space, in addition, a BPSG trench-filled isolation is used (25) instead of the *p*-tub moat around the device, discussed in the "Junction Isolation" section.

ISOLATION BETWEEN SUBCIRCUIT BLOCKS

In addition to being a mechanical support, the substrate is the "insulating" barrier between the different components in an IC. However, not being a true insulator, if it is not properly designed and processed, then the substrate will become a medium for excess ac and dc current flow. It could give rise to problems of leakage, latchup, crosstalk, and so on. Devices with dynamic nodes, most notably dynamic random access memories (DRAMs), are particularly sensitive to excess leakage.

Conventional bulk silicon substrate is doped lightly (about 10^{15} atoms/cm^3) to minimize junction capacitance and to reduce the transistor threshold voltage sensitivity to the back gate (substrate) bias. In addition, it needs to be free from impurities (in particular, midgap impurities such as iron, gold, etc.) to prevent junction leakage. Such material possesses a long minority carrier lifetime, with a diffusion length which could approach the wafer thickness and which makes the minority carriers available from the entire wafer thickness. This is in conflict with the fact that in order to suppress the reverse-bias leakage, the availability of the minority carriers should be limited.

In addition to that, the VLSI CMOS circuits require high substrate doping to avoid undesirable feedback such as IR biasing of junctions (parasitic bipolar breakdown) and coupling between circuits. The coupling is via resistive voltage drops induced by ground currents. The process takes place by injection of electrons from a grounded junction. This junction becomes forward-biased due to the resistive potential drop in the substrate (26). Both higher packing density and miniaturization cause higher substrate current, aggravating the holding-time degradation problem (27). The traditional solution to circumvent this problem has been to use the so-called epi wafers described below.

Epitaxial Silicon

The use of epi material was first introduced in 64K DRAMs in 1977 (28). The two primary advantages of a p^+/p^- substrate are damping of substrate voltage transients and efficient collection of minority carriers injected into the substrate.

An epi substrate is typically a thin few-micron p layer, doped with boron to a concentration of 2×10^{15} atoms/cm^3, grown on top of a p^+ substrate doped to about 1×10^{19} atoms/cm^3. The advantage of fabricating devices (both bipolar and MOS) in an epitaxial layer is that the doping concentration of the device can be successfully controlled and that the epi layer can be made free of unwanted impurities. At the same time, the highly doped epi substrate serves as an excellent ground plane damping substrate noise/crosstalk.

Many hot-electron-induced problems such as overloading of on-chip substrate bias generator, threshold variation due to changing substrate potential, drain-source breakdown, excess minority carrier, photon generation, DRAM refresh time, and device degradation are directly due to the amount of substrate current generated in the circuit.

The epi-layer thickness is set by the technology and the amount of thermal budget which a wafer experiences. During the wafer fabrication, thermal cycles drive the boron from the heavily doped substrate into the epi region, thereby reducing

the effective epi thickness. The HEI-tub process, where the tub drive is eliminated, allows the use of a thinner epi (29).

Latchup in CMOS Circuits

A major concern for VLSI CMOS devices is latchup between transistors in neighboring tubs. The result of latchup is a state of heavy conduction (a short) between V_{DD} and V_{SS} power lines. Once latchup occurs, it does not stop without shutoff of the power supply, which erases the stored information and possibly cause damage to the chip itself.

Latchup is a well-understood phenomena, and it can be dealt with using processing and/or design methodologies; however, with the constant reduction of spacing, latchup immunity is a continuing challenge.

As shown in Fig. 6, the S/D junctions of transistors in neighboring tubs form parasitic bipolar structures, that is, the p^+ junction, n tub, and p substrate produce vertical *pnp,* while the n^+ junction, p substrate, and n tub form lateral *npn* devices. By being in close proximity of each other, these parasitic bipolars can interact electrically to form *npnp* or *pnpn* diodes. For latchup to occur, the parasitic bipolar structures have to be triggered by injecting a current into either an *npn* or a *pnp* emitter, and the holding state must be maintained.

The initial current can be induced under abnormal (yet frequently occurring) operating conditions such as fluctuation of the supply voltage or high-energy particles from outside the chip. If this current and/or the resistance of the tub or the substrate are high enough, a sufficient IR drop can be developed across them to forward bias the emitter–base junctions. Then the collector current of one parasitic bipolar supplies base current to the other, in a positive feedback arrangement, resulting in a sustained current.

The relevant device components in latchup, in addition to the parasitic bipolars, are the resistance of the substrate in both the n- and p-tub regions. Hence, an effective technique for suppressing latchup is to reduce the tub and substrate resistances. With small enough shunt resistors, a sufficient IR drop cannot be developed across them to forward bias the emitter–base junctions. Substrate resistance can be reduced by either using an epi substrate (30) or a retrograde tub, using high-energy implants (22), or using both (31).

Substrate Crosstalk

Because of the increase in the level of integration in ICs, the combination of analog and digital functions on the same chip has become increasingly common. This type of integration, which started as early as the mid-1970s (32), now includes multimedia applications such as switched capacitor filters, analog-to-digital (A/D) and digital-to-analog (D/A) converters, mobil communication products, voice-band coder–decoder (CODECs), and video digital-signal processors (DSPs).

These applications require large-scale integration of highly accurate, high-performance *analog* circuits with many high-speed *digital* circuit gates. The coupling through the substrate, however, is an important limiting factor in mixed-mode high-frequency integrated circuits. In such mixed-signal systems, fast-switching transients produced in the digital circuit can couple into sensitive analog components, thereby limiting the analog precision that can be achieved. The problem becomes more severe as click rates increase, circuit features shrink, and applications demand greater precision from the analog circuitry.

Several methods such as a low-inductance substrate, physical separation, or guard rings have been proposed for reducing substrate noise (33). Substrate coupling problems in BiCMOS technologies have also been studied (4).

Separation of the analog circuits from the digital circuits on the substrate can be accomplished by fabricating them on separate substrates, by the SOI process (34). This approach is effective, but not always practical because of the complex fabrication and high cost.

INSULATOR AS THE SUBSTRATE

The original driving force behind the Silicon-on-Insulator (SOI) technology was to produce a new radiation hardened material to replace Silicon-on-Sapphire (SOS). SOS uses sapphire (Al_2O_3) as both the substrate and the insulator (35). SOI devices are made in a thin Si layer isolated from the wafer substrate by a fairly thick layer of SiO_2.

MOSFETs in the SOI geometry (sketched in Fig. 7) do not need special precautions in terms of junction isolation. The parasitic FOX transistor does not exist because the field oxide extends through the Si film and joins the buried oxide. Together with the lateral isolation, SOI constitutes a complete isolation.

The thickness of the Si layer can be chosen so that the MOSFET will operate in the fully depleted (FD) or partially

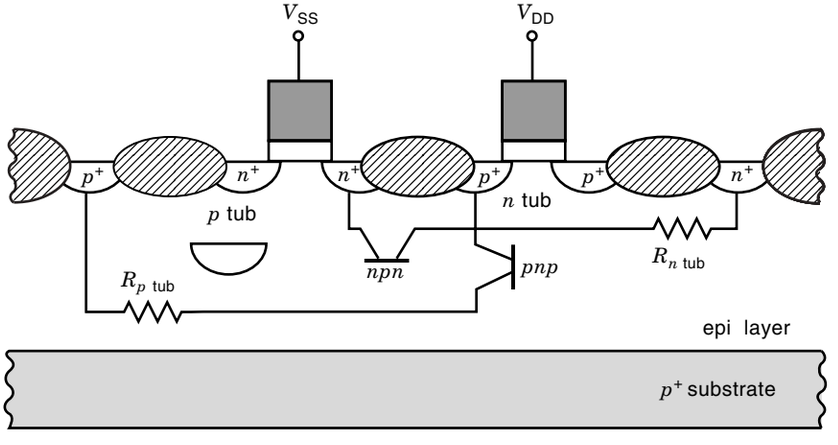


Figure 6. Two transistors in neighboring tubs, and the parasitic *npn* and *pnp* bipolars which cause latchup.

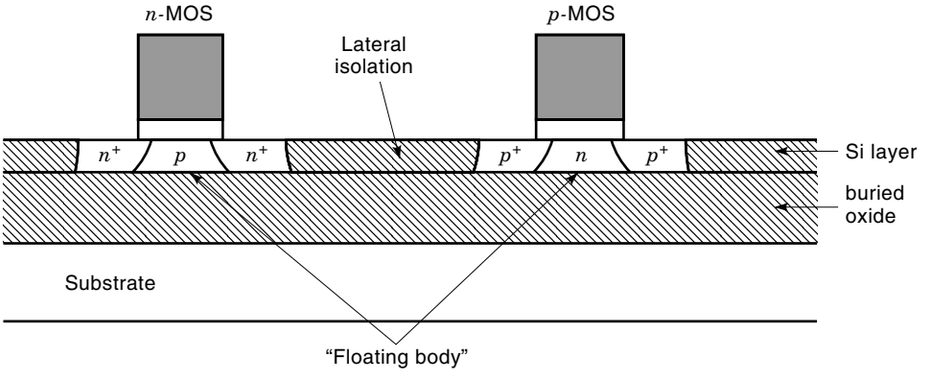


Figure 7. n- and p-MOSFETs on SOI structure, illustrating Si layer, buried oxide, lateral isolation, and the floating body.

depleted (PD) regimes. Fully depleted is when the channel depletion region extends through the entire thickness of the Si layer. FD MOSFET operates faster than bulk because of the increase in drain current, along with a reduction in the dynamic gate capacitance.

In the case of circuits, SOI dielectrically isolates the components, making the fabrication of smaller, denser, and faster circuits possible. It avoids the parasitic effects, which add to device and circuit capacitances or cause latchup, alpha-particle events, and the IC crosstalk

SOI material promises a new branch of the silicon-based industry thanks to the device and circuit advantages. It is particularly attractive for low-power applications in the ultra large silicon integration (ULSI) chips required for the portable systems. The capability of SOI circuits to operate at 1 V or below even in the case of DRAMs has been demonstrated (36).

SOI is also appropriate for the "system-on-chip" approach such as smart power, microelectromechanical systems (MEMS), mixed signal, and integrated optics.

SOI Material

Reports on the fabrication of SOI structure go as far back as 1980 (37). After 20 years and many variations on the SOI fabrication, the problems related to the quality of the SOI material and to the availability in large volume are slowly being removed. There are a number of ways to fabricate the SOI structure. Originally it used a fairly thin (200 nm) Si film deposited on a thick (500 nm) SiO_2 which was grown on an Si wafer. The deposited amorphous Si film was then recrystallized by using different heating techniques: laser (37), graphite strip heater (38), or lamp (39). These techniques are cumbersome and yield poor-quality Si layers.

The SOI fabrication now utilizes three other, more promising techniques of SIMOX (Separation by IMplantion of OXygen) (40), BESOI (Bond and Etch-back SOI) (41) wafers, or the Smart Cut Technology (42).

1. In SIMOX technology an oxygen implantation is employed (by a dedicated machine: 100 mA, 200 keV of O^+ ions) to place a high concentration of oxygen underneath the initial silicon surface. This is followed by a high-temperature anneal which reconstitutes the crystalline quality of the silicon layer remaining over the oxide, and it forms the stochiometric oxide buried in the silicon wafer. The thickness of the Si and the SiO_2 layers in SIMOX are typically around 100 nm and 350 nm. However, the new trend in the SIMOX development is the use of low oxygen implantation doses to obtain an improved, low-cost SOI material. This new approach has drastically improved the top silicon film crystalline quality, but also yields much thinner Si and SiO_2 layers.
2. Bonding is an inexpensive technique for manufacturing thick films of both oxide and silicon. Two silicon wafers, at least one with an oxide layer on top, are bonded together using van der Waalls forces with a subsequent anneal to increase the bonding strength. Then one of the substrates is thinned down to 1 μm by mechanical grinding and polishing to within 10% to 30% uniformity. Better uniformity can be achieved with the help of a chemical etch stop. Typically the bonded wafers have thicker yet better-quality Si and buried oxide layers.
3. Smart Cut, the most recent technique, is based on both ion implantation and wafer bonding technologies. A wafer is oxidized to form what will become the buried oxide layer of the SOI structure. A high-dose (5×10^{16} cm^{-2}) H^+ ion implantation through the oxide forms cavities or microbubbles at the implantation range. This wafer is bonded to another wafer using van der Waalls forces. A 500°C thermal activation merges all the cavities, and then the top wafer is cut away. Because of the use of ion implantation for the layer separation, the layer thickness uniformity is improved over that of "bonded" wafers.

SOI Lateral Isolation

The lateral isolation between SOI devices are obtained traditionally by MESA structure and/or by LOCOS (43). In MESA isolation the active device regions are masked to etch the field device areas. The SOI oxide helps as the etch-stop layer, while anisotropic etching allows for efficient isolation scaling. The weak point of this isolation technique is the sharpness of the side wall and its impact on gate-oxide integrity and device subthreshold characteristics.

LOCOS isolation in SOI is much the same as LOCOS in bulk; however, the oxidation kinetics in SOI are somewhat different from the bulk, in particular when the growing oxide reaches the buried oxide. The oxidation times to consume the entire Si film become longer than in bulk, resulting in larger FOX encroachment and transistor width loss.

SOI Advantages

The appealing aspect of the SOI technology has been its compatibility with the standard semiconductor fabrication, while providing many advantages.

1. An SOI MOSFET is one of the best candidates for low-power, high-speed operation (44). Fully depleted MOS/SOI operates faster than bulk because of the increase in drain current, along with a reduction in the dynamic gate capacitance. A sharper subthreshold slope allows an increase in the switching behavior of the MOS devices, along with shrinking of the threshold voltage, thus increasing the drive current at low voltages.
2. An MOS/SOI transistor has a junction capacitance lower than that of a bulk-CMOS transistor. The total capacitance of an SOI circuit is reduced by 15% to 30%, depending on the circuit layout, compared with the same circuit fabricated in bulk. The wiring capacitance due to the existence of the buried oxide beneath the FOX is also reduced. Operation of a multigigahertz prescalar with 50% increase in the speed compared with those fabricated in bulk was reported in 1989 (45). The gain in the speed was shown to be entirely due to the decrease in the capacitance of the junctions. Similar results were also reported in the propagation delay time of CMOS inverters (46).
3. SOI technology provides higher packing density through more aggressive design rules (contact overlap, lateral isolation, and, in particular, the n^+–p^+ distance across tubs).
4. Alpha particles generate photocurrents whose amount depends on the collection volume. SOI reduces the substrate/collection volume, and thus it lowers the soft error sensitivity. SOI technology for bipolar/ECL (emitter coupled logic)-CMOS devices are needed because of the high sensitivity of the ECL peripheral circuits to radiation (47).
5. Dielectric isolation in SOI also helps in decoupling the analog and digital components of a mixed IC, and thus it reduces substrate crosstalk. For high-frequency applications, above 1 GHz, special low-resistivity substrates may be needed (34).

SOI Issues

Although for many years the fabrication of ICs on SOI has been recognized as very attractive, to date, the production has been limited to some niche markets due to a number of SOI issues that have not yet been completely resolved:

1. Unavailability of the SOI material for production.
2. Quality of the Si layer in terms of defects and gate-oxide yield, thickness uniformity, and the resulting V_t uniformity.
3. Quality of the buried oxide in terms of pinholes, along with its interface with the Si layer.
4. The complete isolation from the substrate results in the so-called floating-body effects (FBEs) (Fig. 7). The FBE is one of the most crucial issues of SOI devices that must be taken into consideration when applying SOI devices to actual VLSI circuits. The floating body can lead to circuit instabilities, hysteretic behavior, frequency-dependent delay time, and pulse stretching. FBE can be dealt with by using a body contact for single devices; but generally, this is not an advisable solution, due to the density and the bulk design compatibility issues.
5. Unavailability of SOI circuit design tools, to take full advantage of the packing density, is another serious issue which impedes the introduction of SOI into main stream Si fabrication.

Nevertheless, SOI is still a burgeoning field for research and process development, as well as new circuit design strategies. As an example, see the Proceedings of the International Electron Device Meeting (IEDM), 1996 and 1997.

BIBLIOGRAPHY

1. J. S. Kilby, The invention of the integrated circuit, *IEEE Trans. Electron Devices,* **ED-23**: 648, 1976.
2. D. Kahng and M. M. Atalla, Silicon-silicon dioxide field induced surface devices, *IRE Solid State Devices Res. Conf.,* 1960.
3. See, for example, Pantelides (ed.), *The Physics of SiO_2 and Its Interfaces,* Elmsford, NY: Pergamon, 1978.
4. See, for example, *Proc. IEDM (Int. Electron Device Meet.) Conf.,* 1996, 1997, 1998.
5. See, for example, *Proc. Int. Symp. Plasma Process Induced Damage,* 1997.
6. C. C. Chang, A. Kamgar, and D. Kahng, High temperature rapid thermal nitridation of silicon dioxide for future VLSI applications, *IEEE Electron Devices Lett.,* **EDL-6**: 476, 1985.
7. A. Kamgar et al., Gate oxide thinning at the active device/FOX bouondary in submicron isolation, *IEEE Trans. Electron Devices,* **42**: 2089, 1995.
8. S. Wolf, *Silicon Processing for the VLSI Era,* Vol. 2: *Process Integration,* Sunset Beach, CA: Lattice Press, 1990.
9. J. A. Appels et al., *Philips Res. Rep.,* **25**: 118, 1970.
10. J. W. Lutze, A. H. Perera, and J. P. Krusius, Field oxide thinning in poly buffer LOCOS isolation with active area spacings to 0.1 μm, *J. Electrochem. Soc.,* **137**: 1867, 1990.
11. R. H. Havermann and G. P. Pollack, US patent 4,541,167, 1985.
12. I.-C. Chen, M. Rodder, and J. M. Hwang, A comparative study of LOCOS-type isolation for 256 Mbit DRAM, *Proc. 1995 Int. Symp. VLSI Tech. Syst. Appl.,* 1995, pp. 31–35.
13. V. M. H. Meyssen, R. M. D. A. Velghe, and A. H. Montree, Optimized scaled LOCOS isolation scheme for 0.25 μm CMOS, *Proc. 25th Eur. Solid State Device Res. Conf.,* 1995, pp. 679–682.
14. See, for example, C. Chen et al., A novel 0.25 μm shallow trench isolation technology, *IEDM Tech. Dig.,* 1996.
15. R. D. Rung et al., A retrograde p-well for higher density CMOS, *IEEE Trans. Electron Devices,* **ED-28**: 1115, 1981.
16. D. C. Jacobson et al., High energy ion implantation for profiled tub formation in deep submicron CMOS technology, *Proc. 10th Ion Implantation Technol.,* 1994.
17. A. Stolmeijer, A twin-well CMOS process employing high-energy ion implantation, *IEEE Trans. Electron Devices,* **ED-33**: 450, 1986.
18. K. Tsukamoto et al., High energy ion implantation for ULSI: Well engineering and gettering, *Solid State Technol.,* **35** (6): 49–55, 1992.
19. B. E. Deal and A. S. Grove, General relationship for the thermal oxidation of Si, *J. Appl. Phys.,* **36**: 3770, 1965.
20. P. B. Griffin and C. S. Rafferty, A viscous nitride model for nitride/oxide isolation structure, *IEDM Tech. Dig.,* 1990, p. 741.
21. B. R. Penumalli, A comprehensive two-dimensional VLSI process simulation program—BICEPS, *IEEE Trans. Electron Devices,* **ED-30**: 986–992, 1983.
22. P. Ferreira et al., Study of local Si oxidation with calibrated nitride model, *Proc. ESSDERC '94,* p. 259.
23. S. Cea and M. Law, Multidimensional nonlinear viscoelastic oxidation modeling, *Simulation Semicond. Device Processes,* **6**: 139, 1995.

24. See, for example, *Proc. Mater. Res. Soc. (MRS) Symp.* or the *Dielectr. VLSI/ULSI Multilevel Interconnection Conf. (DUMIC),* 1997.
25. M. Suglyama et al., Bipolar VLSI memory cell technology utilizing nBPSG-filled trench isolation, *Tech. Dig. Symp. VLSI Technol.,* 1989, pp. 59–60.
26. B. Eitan, D. F. Bentchkowsky, and J. Shappir, Holding time degradation in dynamic MOS RAM by injection-induced electron current, *IEEE Trans. Electron Devices,* **ED-28**: 1515, 1981.
27. O. Kudoh et al., Influence of substrate current on hold-time characteristics of dynamic MOS IC's, *IEEE J. Solid-State Circuits,* **SC-13**: 235, 1978.
28. J. T. Clemens et al., *MOS Dynamic Memory in a diffusion current limited semiconductor structure,* US Patent 4,216,489, 1980.
29. R. A. Martin et al., A new process for one micron and finer CMOS, *IEDM Tech. Dig.,* 1985, p. 403.
30. R. S. Payne, W. N. Grant, and W. J. Bertram, Elimination of latchup in bulk CMOS, *IEDM Dig.,* 1980, p. 248.
31. S. Voldman et al., Retrograde well and epitaxial thickness optimization for shallow and deep-trench collar merged isolation and node trench SPT DRAM cell and CMOS logic technology, *IEDM Tech. Dig.,* 1992, p. 811.
32. D. A. Hodges, P. R. Gray, and R. W. Broderson, Potential of technologies for analog integrated circuits, *IEEE J. Solid-State Circuits,* **SC-13** (3): 285–294, 1978.
33. T. Tsukada and K. Makie-Fukuda, Approaches to reducing digital-noise coupling in CMOS mixed-signal LSIs, *IEICE Trans. Fundamentals Electron., Commun. Comput.,* **E80-A**: 263, 1997.
34. J. P. Raskin et al., Substrate crosstalk reduction using SOI technology, *IEEE Trans. Electron Devices,* **44**: 2252, 1997.
35. G. W. Cullen, M. T. Duffy, and R. K. Smeltzer, Recent advances in heteroepitaxial Si on sapphire technology, *VLSI Sci. Technol.,* 230, 1984.
36. T. Oashi et al., 16Mb DRAM/SOI technologies for sub-1 V operation, *IEDM Tech. Dig.,* 1996.
37. D. K. Beigelsen et al., *Appl. Phys., Lett.,* **38**: 150, 1981.
38. B.-Y. Tsaur, ZMR SOI technology, *IEEE Circuits Syst. Mag.,* **July**: 12, 1987.
39. A. Kamgar, G. A. Rozgonyi, and R. V. Knoell, Linear zone-melt recrystallized Si films using incoherent light, *Laser Solid Interactions Thermal Process. Mater.,* New York: Elsevier, 1983, pp. 569–574.
40. K. Izumi et al., SIMOX technology for CMOS LSI's, *VLSI Symp. Tech. Dig.,* 10, 1982.
41. S. S. Iyer et al., *Proc. ECS Meeting Semicond. Si,* **PV94-10**: 391, 1994.
42. M. Bruel, Silicone on insulator material technology, *EDL,* **31** (14): 1201–1202, 1995.
43. M. Haond and O. Le Neel, Lateral isolation in SOI CMOS technology, *Solid State Technol.,* **34** (7): 47–52, 1991.
44. A. Yoshino et al., The most essential factor for high-speed, low power 0.35 μm CMOS circuits fabricated on SIMOX substrates, *Jpn. J. Appl. Phys.,* **36**: 6699, 1997.
45. A. Kamgar et al., Ultrafast (0.5 μm) CMOS circuits in fully depleted SOI films, *IEEE Trans. Electron Devices,* **39**: 640, 1992.
46. Kado et al., An experimental full-CMOS multigigahertz PLL LSI using 0.4 μm gate ultrathin-film SIMOX technology, *IEICE Trans. Electron.,* **E76-C**: 562, 1993.
47. T. Nakamura, T. Shiba, and T. Ikeda, Bipolar and BiCMOS on SOI, *ECS Proc.,* **96-3**: 1996, p. 411.

AVID KAMGAR
Bell Laboratories Lucent Technologies