Richard A. Larsen

A Silicon and Aluminum Dynamic Memory Technology

The Silicon and Aluminum Metal Oxide Semiconductor (SAMOS) technology is presented as a high-yield, low-cost process to make one-device-cell random access memories. The characteristics of the process are a multilayer dielectric gate insulator (oxide-nitride), a p-type polysilicon field shield, and a doped oxide diffusion source. Added yield-enhancing features are backside ion implant gettering, dual dielectric insulators between metal layers, and circuit redundancy. A family of chips is produced using SAMOS, ranging from 18K bits to 64K bits. System features such as on-chip data registers are designed on some chips. The chip technology is merged with "flip-chip" packaging to provide one-inch-square modules from 72K bits through 512K bits, with typical access times from 90 ns to 300 ns.

Introduction

In October 1978, a new semiconductor memory technology was introduced by IBM Corporation's General Technology Division. This new technology is used to produce a family of chips with densities of 18K-, 32K-, 36K-, and 64K-bits per chip [1]. The Silicon and Aluminum Metal Oxide Semiconductor (SAMOS) process is a metal gate technology which provides a distinct productivity leap over previous IBM 2K- and 4K-bit products through the combination of one-device memory cells [2, 3] and new process concepts. Among these are 1) backside ion implant for leakage control, 2) doped oxide self-aligned diffusion source, 3) multidielectric gate insulator, 4) p-type polysilicon field shield for isolation and cell capacitance, 5) lift-off aluminum metallurgy, and 6) quartz-polyimide passivation between metal layers. The use of new processes poses new challenges in device physics. In particular, the use of a nitride layer in the gate dielectric prevents further oxidation of the gate during polysilicon oxidation and provides low defect densities due to the dual dielectric. At the same time, the hot electron [4] effects were intensified by the use of nitride, and they had to be understood, characterized and brought under routine process control. These effects are not unique to SAMOS, however, and will be encountered by all technologies of small dimensions.

Since the objective of the SAMOS technology is to provide large-volume, low-cost memory bits suitable for use in computing systems, the capabilities of the semiconductor process must be merged with the needs of the systems. A number of tradeoffs can be made during process and chip design among yield, performance, density, function, flexibility, and manufacturability. In SAMOS, a design philosophy is adopted to optimize yield and reliability, and circuits are added on the chips to further enhance yield (redundancy [5]) and provide systems functions (on-chip register). Different requirements for performance and density are met by the family of chips.

In order to cover the above topics, this paper is divided into a number of sections. First, the process is shown in a skeletal outline, and the development history is reviewed. Next, the process is presented in detail, divided into sections through and after first metallization, while the third section is devoted to the technology challenges associated with SAMOS, and their solutions. The final section discusses the tradeoffs made in SAMOS to maximize reliability and productivity, with specific discussion of redundancy, on-chip register, and the chip family.

SAMOS process overview

The SAMOS process and the related technology issues will be covered in the sections that follow. For reference, a skeletal outline of the process through first-level metallization will be given here. Figure 1 provides a pictorial summary.

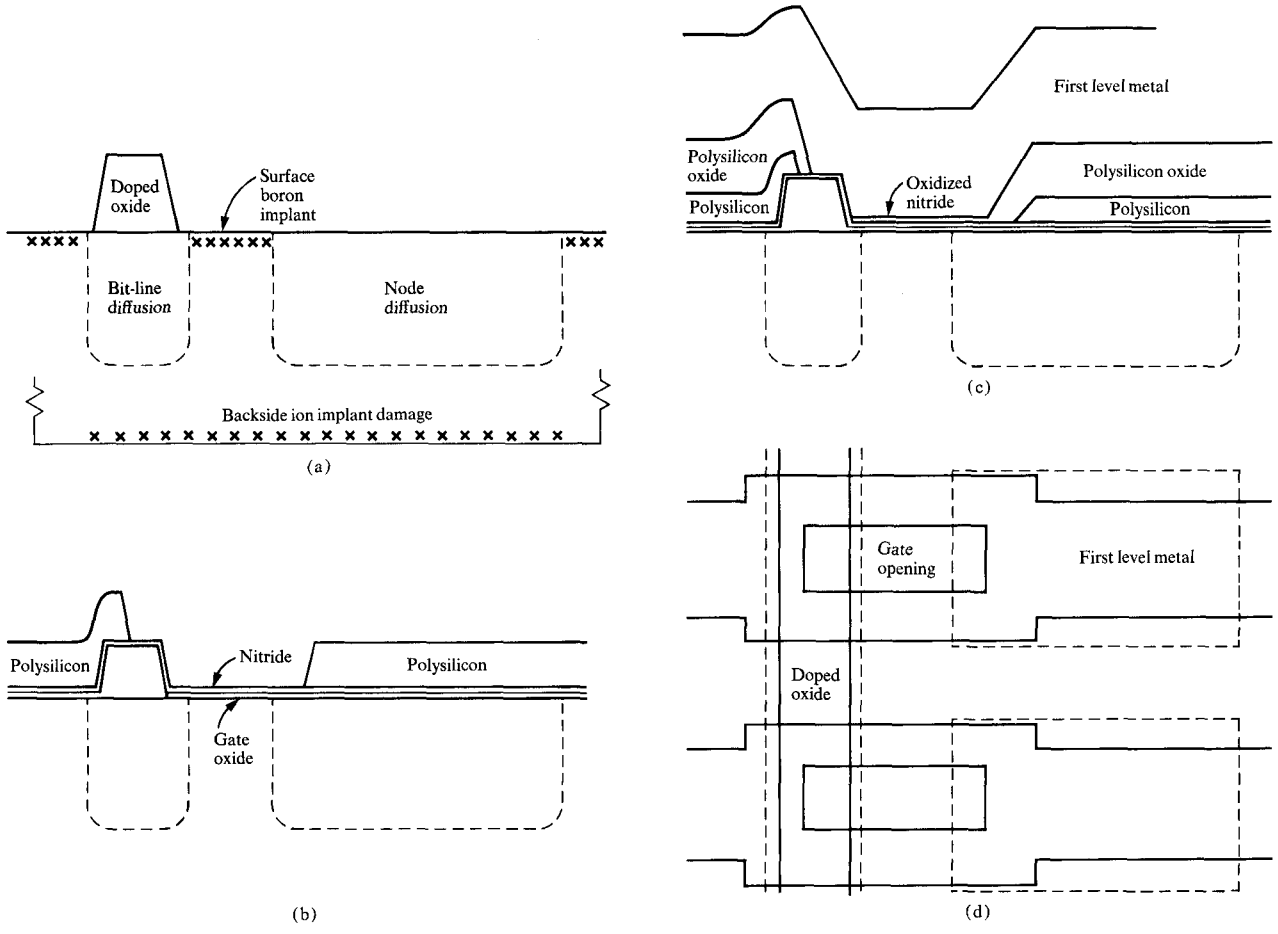

Figure 1 Views of the SAMOS memory cell. Figure (a) shows a cross section immediately prior to gate oxidation, while (b) is the same cross section after the polysilicon patterning mask step. The structure after first-level metallization is shown in (c). A top view of two nodes on a bit line is shown in (d), with the extent of n^+ diffused areas being shown by dashed lines.

1. Ion implant argon (wafer backside)
2. Deposit arsenic-doped oxide/cap oxide
3. Pattern doped oxide
4. Drive in arsenic diffusions
5. Implant boron for surface tailoring
6. Remove doped oxide on nodes [see Fig. 1(a)]
7. Oxidize gate silicon (gate dielectric)
8. Deposit nitride/polysilicon
9. Pattern polysilicon [see Fig. 1(b)]
10. Oxidize polysilicon
11. Open contact holes
12. Deposit and pattern first metal [see Figs. 1(c) and (d)]

Historical background

The SAMOS technology was developed at the IBM Burlington laboratory, Essex Junction, Vermont. The physical structures needed for this field effect transistor memory were proposed by W. Smith [6], with a suitable fabrication process provided by W. Smith and R. Garnache [7]. As proposed, the process used four mask steps to first metallization and could produce a chip of $\approx 200\ \mu m^2$/bit with an access time of 1800 ns. The process was basically similar to the presently used process, with one clear difference in concept. After doped oxide was patterned on the wafer and diffusions driven in, all doped oxide was stripped from the wafer. The result was that the diffused bit lines were separated from the polysilicon field shield by thin oxide and therefore exhibited a large bit-line capacitance which limited performance. The process was developed in a piloting facility, and a 32K-bit, 4.09-mm × 5.1-mm chip was designed with the above performance and density. Redundant lines were used on this chip, as on future chips. The first functional 32K-bit chip, fixable by redundancy, was obtained in November 1972. The first perfect chip followed in April 1973.

In order to meet more aggressive density/performance goals, SAMOS evolved into a five-mask version. The fifth

Table 1 Leakage data—diffused monitors on product wafers.

Time period	*Percentage of defective sites*	
	Argon BSII	*Nongettered*
Period 1	5	20
Period 2	0.4	0.5
Period 3	1.2	19.6

mask protected the bit-line doped oxide, while that over the storage node was etched away. Using the five-mask process, a family of product offerings ranging from 18K-bit, 90-ns access to 64K-bit, 300-ns was designed. The 18K-bit chip was designed in IBM Deutschland Lab, Boeblingen, Germany, while three other designs originated in the Burlington laboratory. The first functional 64K-bit chip was produced in the pilot facility in August 1976, and volume production on a manufacturing line started in January 1978. Computer systems using the SAMOS technology were first announced in October 1978.

Process through first metallization

The SAMOS process begins with an argon backside ion implantation of dose and energy sufficient to form an amorphous silicon layer. Backside ion implantation [8–12] provides a particularly simple and clean process for leakage control, an important aspect of VLSI technology, since the charge stored on a cell is on the order of only 200–300 $\times$ 10^{-15} coulombs. Implantation gettering provides a convenient method to ensure low leakage levels on a large-volume manufacturing line. Table 1 shows typical leakage data for diffused monitors on product wafers. The monitors accurately predict leakage-limited yield through a modeling technique [13]. The data illustrate the general SAMOS experience that, in times of optimum initial wafer quality and process cleanliness, gettering does not improve yield, but that when substrate or process problems do occur the backside gettering provides a protection against major yield falloff. The gettering effect results from the formation of 1/2⟨110⟩ dislocations during the regrowth of the amorphous damage region [11, 12]. Metallic impurities precipitate at the dislocation and do not end up near diffusion junctions on the wafer frontside. Initially the SAMOS backside implantation was done using boron, but a switch was made to argon since it is as effective as boron, does not require as high a dose [11], and can be produced with higher flux in the implanter.

Backside ion implantation is followed by three process steps which, taken together, constitute the SAMOS diffusion technology. An arsenic-doped oxide and an undoped capping oxide are deposited in a continuous chemical vapor deposition reactor. The process uses a mixture of silane, arsine and oxygen to produce an As_2O_3–SiO_2 glass as the doped layer. This blanket film is patterned by the first mask level, then driven in. A thin thermal oxide is grown during the drive-in cycle.

Solid state arsenic diffusion sources have been studied for a number of years [14–17]. A blanket doped oxide layer can be applied to the wafer using either arsine-silane oxidation or a spin-on arsenosilicate glass. If a doped oxide is deposited after a masking nitride or undoped oxide has been deposited and patterned, the resulting diffusion line width can be characterized by

$$W_T = W_p + 2 \times 0.7X_j + 2B,$$

where W_p is the photo mask image width, B the etch bias of the masking layer, and X_j the junction depth. A similar equation is obtained if a diffusion is produced by capsule diffusion or by an arsenic ion implantation. In contrast, the SAMOS doped oxide is applied directly to the entire wafer, then patterned and removed subtractively at the first mask level. The etch bias decreases the line diffusion width. In addition, the edge of the doped oxide does not make an efficient source for arsenic, so the apparent edge for diffusion is back under the doped oxide by some apparent amount (δ). The line width becomes

$$W_T = W_p + 2 \times 0.7X_j - 2B - \delta.$$

For the SAMOS process, the empirical result is that the diffusion line width is essentially equal to the photo mask image width. Narrow diffused lines can be made without using very small lithographic lines. A second advantage of the subtractive-etch doped-oxide process is that a thick oxide layer is left self-aligned to the diffusions. Depending on whether this doped oxide is removed or left intact in later processing, the diffusion is selected to exhibit either high or low capacitive coupling to the overlaying conductors. In order to utilize the advantages offered by the doped oxide diffusion source, the doped oxide deposition, patterning, and drive-in steps must be integrated to prevent doping of the exposed silicon areas by arsenic counterdoping and to provide proper sheet resistance (R_s) and junction depth under the doped oxide.

Arsenic counterdoping is controlled by use of 1) low mole percent arsenic in the doped glass ($\approx$2%), 2) an undoped capping oxide over doped oxide, 3) HCl in the drive-in process to react with free arsenic, and 4) an oxidizing atmosphere in the first phase of drive-in to oxidize the bare silicon. In addition, the use of arsenic instead of phosphorus in the doped glass is in part based on the assumption that arsenic has a lower vapor pressure at drive-

in temperature. Even with these steps, careful monitoring must be maintained to detect possible counterdoped conditions and allow early process correction. The monitoring techniques utilized are discussed in a later section.

The junction depth for SAMOS is 0.8 μm, which represents a tradeoff between competing effects. A shallower junction would mean less threshold reduction at short channel lengths and smaller overlap capacitance of the diffused bit line to polysilicon field shield. A deeper junction would have a lower injection rate of channel hot electrons into the dielectric, and would have a lower incidence of metallurgy spiking through the diffusions at contact holes. In order to achieve the 0.8-μm value for X_j, arsenic is preferred, since it will not undergo appreciable further diffusion later in the process. A value of 25 $\Omega/\square$ is chosen for diffusion sheet resistance as the optimum tradeoff between process capability and design requirements. The R_s and X_j values are obtained by using a sequential drive-in O_2/N_2 ambient followed by N_2. During the drive-in the As_2O_3 in the oxide diffuses to the silicon and the arsenic is reduced by

$$2\ As_2O_3 + 3\ Si \rightarrow 4\ As + 3\ SiO_2.$$

This reaction in a nonoxidizing atmosphere would lead to formation of an arsenic layer at the oxide-silicon interface [18], which would impede further reaction. Oxygen is used to prevent the formation of the arsenic layer. The relative amounts of O_2/N_2 in the initial zone strongly affect R_s, as shown in Fig. 2 [19]. The R_s is high with low O_2 percentage because of arsenic buildup, and rises again at high O_2 percentage because of the oxidation of the silicon surface under the doped glass. After the proper amount of arsenic is introduced into the silicon in the O_2/N_2 cycle, a pure N_2 ambient is used to drive in the arsenic to the proper depth.

During the drive-in oxidation an oxide is formed on the wafer surface. This oxide is used as a screen for a boron implant of $\approx 1 \times 10^{12}$ ions/cm^2. As discussed in a companion paper [20], the dopant control inherent in implantation is far superior to the intrinsic doping variability of the wafer itself. In addition, the implant leads to a final doping profile of boron $\approx 1.8 \times 10^{16}$ atoms/cm^3 at the wafer surface, but only $\approx 1.1 \times 10^{15}$ atoms/cm^3 in the silicon bulk. Threshold voltages with substrate bias (V_{SUB}) of -2.2 V are then around 1.3 V, high enough to limit subthreshold leakage of array devices with the gate off. The rise of threshold as V_{SUB} becomes more negative is minimized by the low background doping, allowing more signal to be written into the cell, which operates in a source follower mode. Within limits, the implant can be used to maintain constant threshold voltage, being raised or lowered to counter long-term trends in flatband voltage. The implant

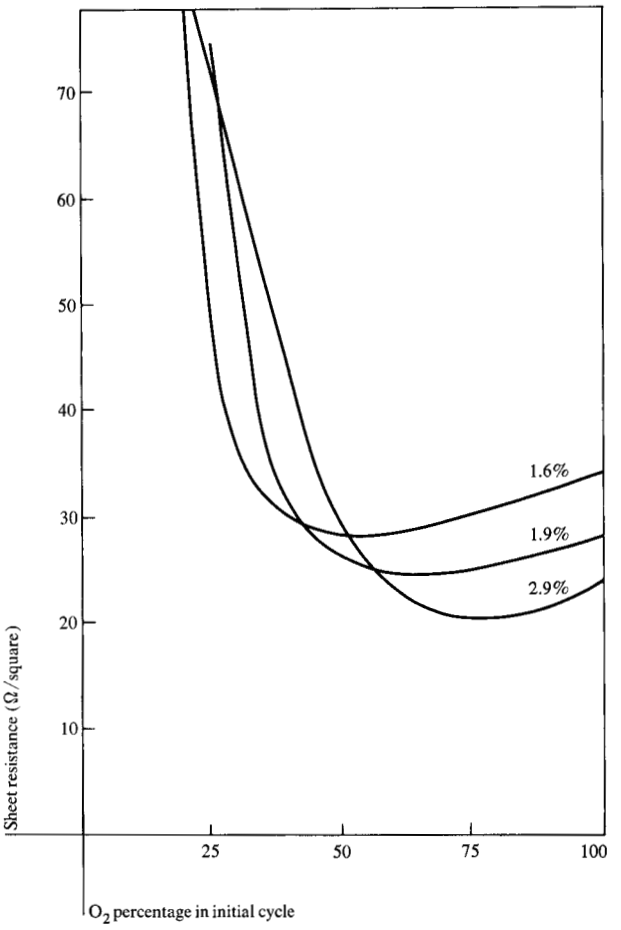

Figure 2 Sheet resistance as a function of O_2 percentage in the oxidizing phase of drive-in, for various mole percentages As_2O_3 in the doped oxide (from Ref. [19]).

level is restricted on the high side by substrate hot electrons, a phenomenon which will be discussed in a later section.

Following implantation, the wafer is masked for a second mask level, the step added in five-mask SAMOS. This photolithographic step protects the doped oxide over the bit line from etchant, but allows the doped oxide to be removed from the storage node. Following photoresist removal, a dip etch is performed to remove the thin drive-in oxide. At this point the cell is as shown in Fig. 1(a). The result is that the storage node will eventually have only thin dielectric between the diffusion and field shield for maximum capacitance, while the doped oxide remains over bit lines to minimize the parasitic capacitance of diffused line to field shield.

At this point, the silicon doping is complete. The next series of steps defines the insulators and conductors used

to control the state of the silicon surface. First, a thermal oxide is grown on the wafer, resulting in a thickness of 31 nm over lightly doped p regions and 38 nm over the n^+ diffusion pockets. The gate oxidation is followed by a chemical vapor deposition step which provides 20 nm of Si_3N_4 and a layer of p-doped polysilicon, sequentially. The polysilicon is then removed in areas to be used for gates and for contacts to the silicon, as shown in Fig. 1(b). Next, the wafer is oxidized, producing a layer of oxide ≈300 nm thick on the polysilicon, and converting ≈7 nm of nitride to oxide in the gate and contact openings. Contact holes are then simultaneously etched to polysilicon, silicon substrate, and diffused lines. The simultaneous etch of thick polysilicon oxide and thin oxide-nitride-oxide is achieved by adjusting the temperature and the relative concentrations of buffered HF and water in the etchant. The etch opens the polysilicon oxide first, with contacts opening shortly thereafter. The first metallization layer is then applied and patterned.

The nitride layer in conjunction with polysilicon provides the most distinctive characteristic of SAMOS. Except for bit lines, the entire chip surface is covered by thin dielectric. The nitride provides an oxidation barrier such that when a thick thermal oxide is formed to insulate the polysilicon from metal, no appreciable oxide growth occurs in the gates. The nitride also provides enhanced dielectric integrity. Where polysilicon is not etched from the wafer, the surface characteristics are determined by a silicon-nitride-oxide-silicon (SNOS) structure. Where polysilicon is removed, a metal-oxide-nitride-oxide-silicon (MONOS) structure controls the surface. The understanding and control of SNOS and MONOS properties are crucial to the successful implementation of SAMOS. Items relative to SNOS behavior are discussed here, while the crucial technology challenges relative to MONOS are presented in a later section.

The p-type polysilicon layer serves a twofold purpose—to form a ground plane above the storage node capacitor, and to act as a field shield and shut off surface leakage. The polysilicon layer is shorted to the p-doped substrate using first-level metal. The p-doped polysilicon has somewhat higher resistivity than n-doped, but it provides the important advantage of higher SNOS threshold voltage and therefore, for a given charge level in the dielectric, shuts off the silicon surface more completely [21]. In addition, the intentional shorting of polysilicon to silicon substrate makes any possible SNOS pinholes in this field region inconsequential, and prevents any voltage stressing of the SNOS structure. Potential SNOS threshold instability [22] is of no concern, since SAMOS does not use any SNOS switching devices.

Over the diffused nodes, the SNOS structure provides ≈85% of the capacitance for cell charge storage; near the bit line, however, the SNOS structure results in a parasitic capacitance. The region of thin dielectric capacitance is seen in Fig. 1(b). The diffusion extends out from under the doped oxide and in this region is separated from either polysilicon or metal by only the gate oxide/nitride. A major portion of the total bit-line capacitance results from this region of thin dielectric, and as such it represents an important factor in SAMOS performance.

Dielectric integrity is required over the storage nodes, since a defect will short the diffusion to the polysilicon and therefore to the substrate. The dual dielectric structure can be superior to oxide in defect density, since defects must be coincident in both layers to cause a short. When defects are random due to poor oxide growth or nitride deposition, the likelihood of coincident defects in the layers will still be low, and dielectric fails should not occur. The nitride covers up possible holes in the oxide, and it also can prevent their formation. Areas of weak oxide would normally begin to conduct under high electric field, leading to heating, increased conduction, and eventual breakdown of the oxide in a thermal runaway condition. With the nitride trap density, however, any electrons which start to flow are trapped above the localized problem area, lowering the electric field across the oxide and shutting down the conduction before it can run away.

The dual dielectric clearly offers advantages over oxide alone if the oxide has pinholes or weak areas. SNOS failure can also be caused by particles with dimensions on the order of tenths of microns to microns. The particles cause coincident disruption of both layers in the same spot, so the oxide-nitride structure offers no advantage over a single oxide against particle-caused defects. If an SNOS short should occur over an array storage node, however, the SAMOS structure minimizes any adverse consequences. Since the field shield is at substrate potential, no current flows through the pinhole and the polysilicon plate/field shield usefulness is not impaired; only an individual node is rendered inoperative. With the polysilicon plate operation unaffected by the defect, the defective cell can be straightforwardly replaced by word line/bit line redundancy.

As noted previously, after the SNOS structure is formed, the polysilicon is patterned and then oxidized. Contact holes are opened, and the wafer is given first-level metallization. This is done by forming a "lift-off" structure on the wafer, followed by metal evaporation. The use of lift-off as opposed to subtractive etch results in improved ground rules, as the desire is to produce mini-

mum spacing (W_w) between metal lines. The difference is shown in Fig. 3. With subtractive etch, metal is evaporated on the entire wafer and covered with photoresist. The minimum developed image in the photoresist is given by W_p. When the metal spaces are created by subtractive etch, appreciable over-etch must be ensured to clean out the space and eliminate shorts. The spacing on the wafer becomes

$$W_w = W_p + 2B,$$

where B, the bias, is typically 0.6 μm for metal 700 nm thick.

In a lift-off metallization [23a], a stencil structure is produced before metal evaporation. The stencil is typically a resistant material, either a metal film [23b-d] or a photoresist layer [24a, b] on top of an organic polymer or photoresist layer. Once the stencil is patterned and the underlying, less resistant layer developed, a characteristic lift-off structure results. In general, there will be an overhanging lip of stencil material with the photoresist removed under the lip, as shown in Fig. 3(b). The spacing between the overhanging edges of adjacent lift-off lands is essentially the photolithographic image spacing W_p. Next, the metal is evaporated onto the wafer at low pressure, providing metal coverage on top of the silicon layer and within the wells. Because of the overhang, the walls of the lift-off structure are not covered with metal. When the photoresist is stripped, the metal on top of the photoresist lifts cleanly from the wafer. Note from the figure that the bias which occurs during metal evaporation leads to wider metal lines, and therefore smaller spaces. The minimum metal space obtainable with lift-off metal is given by

$$W_w = W_p - 2B,$$

where B for SAMOS is approximately 0.3 μm.

The metallurgy used must be compatible with the lift-off process, in that it must be applied at normal incidence without overheating the photoresist on the wafer. The metallurgy must give low contact resistance to the diffusions without junction penetration, must be amenable to deposition by rf heating, since electron-beam evaporation could cause large V_{FB} shifts due to radiation damage in the nitride, and must resist electromigration [25], the thinning and possible opening of current-carrying lines. All of these requirements are met by using a thin layer of aluminum, a layer of copper, and a thick aluminum layer [26]. The thin aluminum layer is needed for bonding to the silicon. It must be thick enough for good process control, but thin enough so that Cu diffusion can occur quickly throughout the entire layer. A barrier intermetallic layer of Al_2Cu is formed as the wafer is sintered, retarding the

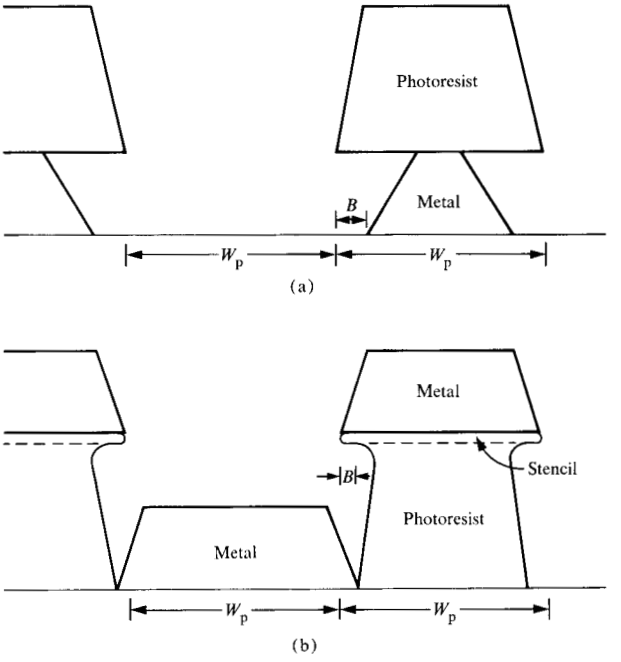

Figure 3 Relative metal line width and space for sub-etch (a) and lift-off (b) aluminum. In both cases equal printed image widths and spaces (W_p) have been used. For lift-off, the bias B enhances the metal line width, while for sub-etch it reduces the metal line.

migration of Si into Al and Al into Si. The Al_2Cu layer also serves to provide added resistance to electromigration. A possible alternative to Al-Cu-Al would be Al-Cu-Si, which was tried on SAMOS. The silicon should provide the solid solution of Al-Si and prevent the upward migration of silicon from the substrate. This technique was found to be less acceptable, however, since the solid solubility temperature dependence caused aggregates of silicon to precipitate from the alloy as the wafer temperature was changed, causing erratic and high resistance at diffusion contacts [27].

Process through module level

The SAMOS process after first-level metallization is designed to passivate the first-level metallurgy, insulate it from second-level metal, and protect the entire chip surface from the outside environment. The ideal passivation/insulation layer should provide an environment for semiconductor device and interconnection metallurgy which protects them from degradation in their useful life. It should provide good coverage for all edges and steps in topology, dielectric strength and freedom from pinholes, and protection against chemical contamination. For SAMOS, this is accomplished by a dual layer consisting of sputtered quartz covered by polyimide [28]. The dual layer provides distinct advantages in defect densities and in mechanical properties.

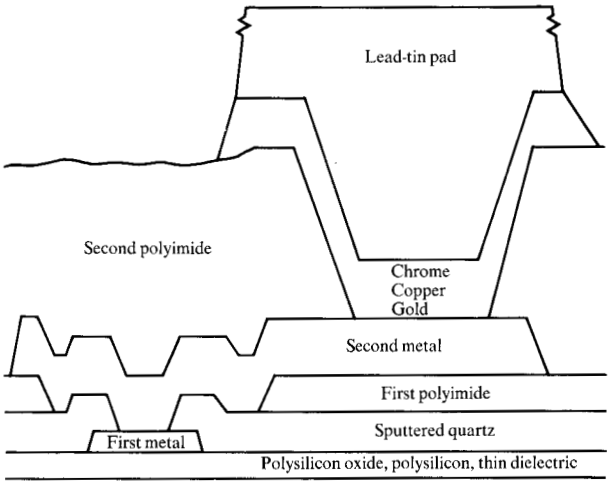

Figure 4 General view of metallization and passivation layers used in SAMOS. The structures are roughly to scale, with the exception of the much-reduced lead-tin pad height.

In the case of a single-layer insulator, a defect in the insulator itself (quartz flakes, polyimide nonhomogeneity) or in the photoresist layer during insulator etch can lead to at least three potential problems. Yield can be reduced by interlayer shorting, or reliability can be impacted by either of two problems. Ionic contamination can reach the dielectric under the first-level metal, leading to threshold instability, or leakage-induced metal migration can occur, resulting in metallic "whisker" growth and eventual shorting. With the dual layer, both quartz and polyimide are put down as blanket films over the entire wafer. Via holes are etched in the polyimide using one mask level. The photoresist developer additionally serves as the polyimide etch. After photoresist strip, a second photoresist layer is applied, and vias are etched through the quartz within the polyimide vias. Through use of two dielectrics and two mask steps, an extraneous hole will occur only in the unlikely event that dielectric or photoresist random defects are coincident. The use of the polyimide cushion on the rigid quartz layer also provides improved mechanical properties, resulting in better integrity of metal lines and metal-metal contacts.

The via etching is followed by a co-deposited aluminum-copper layer, again using rf evaporation to avoid potential radiation damage to the nitride. During evaporation the wafer is kept heated to desorb organics and provide lower contact resistance between first and second metal layers. As in first metal, the copper in second metal inhibits electromigration. Second-level metal serves a number of distinct functions in SAMOS. It is used to bring the first-level metal pads out to terminals, and it provides an extra wiring capacity to carry signals within the chip. The second-level metal is also patterned to form the fuses for SAMOS redundancy [29].

After second metal the wafer is tested for functionality. Chips which are non-perfect but fixable have the failing addresses written into the second-level metal by blowing the redundancy fuses. Next, the final chip seal is provided by a second polyimide layer, the thickness of which is sufficient to cover all metal exposed during the fuse-blowing operation. The polyimide is removed from the wafer only between chips and above the second-level metal pads. A molybdenum mask is aligned to the polyimide vias and the wafer is dc sputter-cleaned for good contact resistance. Next, five metal layers are evaporated through the mask sequentially to provide 1) sealing of via holes, 2) good conductivity, 3) corrosion resistance, and 4) a strong, reliable seal to a solder connection.

A thin chromium layer is evaporated first to provide complete sealing of the via holes by virtue of good adhesion to both aluminum and polyimide. The chromium also protects the underlying aluminum metallurgy from solder. Before chromium evaporation ends, evaporation of a thick copper layer is initiated. Copper provides an inexpensive, highly conductive layer. Gold is used to cover the copper and protect it from oxidation. The basic rationale of the sputter clean and Cr-Cu-Au layer has been discussed previously by Totta and Sopher [30]. The chrome-copper-gold evaporation is followed by a lead-tin evaporation for solder connections. After the molybdenum mask is removed, the structure is heated to reflow the solder [31]. During reflow, the Sn-Cu layers form an intermetallic compound. The mechanical mixing of copper and chromium prevents this intermetallic compound from breaking away from the chromium. The structure resulting from these processes is shown in Fig. 4.

After this terminal metallization, the wafers are diced and the individual chips tested. Usable chips are mounted to one-inch-square ceramic substrates by the IBM "flip-chip" method with a maximum of four chips mounted on each substrate. In the case of the 64K-bit chip, modules with as many as 524 288 bits are produced through use of stacked substrates with a total of eight chips. The ceramics have a basic array of 9 × 9 pins. For any given application, all pins under the chips are removed. Typical modules have 40–60 usable I/O pins.

SAMOS technology challenges

In any semiconductor technology, potential problem areas must be recognized and understood, and the process and products must be developed either to prevent the problem occurrence or to reduce it to an in-

consequential level. For the doped oxide and nitride SAMOS process, the major technology challenges are listed below, along with a discussion of the appropriate solutions.

1. Movement of dielectric charge when gate is stressed positively (stability).
2. Injection of channel electrons or leakage-generated electrons into the dielectric under stress conditions (channel and substrate hot electrons).
3. Parasitic leakage from source to drain with gate off (sidewalk).
4. Control of fixed charge in dielectric (flatband voltage).
5. Prevention of arsenic doping into gate regions from the doped oxide (counterdoping).

The stability, hot electron, and sidewalk effects are all reliability-oriented, since the device characteristic changes in use and can lead to failure of an initially good chip. The flatband voltage and counterdoping impact the threshold voltage at the time of initial test, and as such are potential yield detractors.

• *Stability*

Threshold instability under positive gate bias results from movement of charges within the insulator. Instability can occur in MOS structures due to dielectric contamination with mobile ions, usually alkali ions [32]. This mechanism could occur on SAMOS due to contamination of either oxide layer in the MONOS structure, but it is essentially eliminated by the usual precautions of line cleanliness coupled with the use of HCl in the oxidizing ambients during gate growth and polysilicon oxidation. In addition, any mobile contamination which may be present in the oxidized nitride layer is prevented from moving appreciable distances, since the nitride layer acts as a barrier to alkali ions [33].

While mobile ion migration is reduced by the nitride, electron conduction is enhanced, resulting in a second possible instability mechanism which must be understood. In any multilayer dielectric, the differences in conduction between the layers will lead to charging of the interfaces between the dielectrics [34]. For the MONOS dielectric at the gate voltages used in SAMOS, the 31-nm oxide is essentially nonconductive, while electrons move through the nitride by a number of mechanisms, with Frenkel-Poole conduction predominating [35]. The electrons move across the upper oxide by tunneling. The result is an apparent positive charging of the lower oxide-nitride interface. In SAMOS, these problems were solved by developing a suitable nitride process. The process is hot (925°C), uses an H_2 carrier gas, and uses a large excess of ammonia in the ammonia-silane reaction. The H_2 carrier is thought to inhibit gas phase decomposition of NH_3 or SiH_4 by forcing the reverse reaction. With this process the MONOS dielectric is quite stable under positive gate stress, with data provided by J. Franz showing an expected V_T shift of only ≈20 mV at 100 kh at 125°C with 10-V gate stress.

• *Channel hot electrons*

Channel hot electrons present a second mechanism for threshold shift under stress. During a strongly "on" condition with gate and drain voltages high, electrons flowing from source to drain can be scattered by the lattice and injected over the potential energy barrier at the oxide-silicon interface [4, 36, 37]. This problem will be present in all VLSI technologies as shorter channels are approached, although it will lessen with lower power supplies. The problem is exacerbated in a nitride technology, however, since the nitride acts as a nearly perfect trap for the injected electrons [38]. The amount of V_T shift can be reduced by keeping the nitride thin to minimize charge imaging in the silicon, and by keeping the boron doping low and the drain junctions deep to minimize the vertical fields near the drain. These factors all had to be considered in the design of the process. Once the process was adjusted to minimize hot electrons, the residual effects were carefully characterized and modeled for various use conditions and included as part of the SAMOS design criteria [20].

• *Substrate hot electrons*

Hot electrons can also be generated by leakage current, the generation of electrons within the depletion region of an "on" gate, or the drift of electrons into that region. This mechanism is known as leakage-induced threshold shift (LITS), or as "substrate hot electrons" [4, 39-41]. The effect will also be present in any technology, but again it is intensified in SAMOS due to the nitride trap efficiency. Substrate hot electron injection occurs if the electric field near the semiconductor surface is strong enough so that electron energy within a mean free path of the surface is comparable to the oxide-silicon potential barrier [42] (see Fig. 5). LITS is controlled on SAMOS by keeping the electric fields low. A high-resistivity wafer is used and the ion implant doping is strictly maintained below certain levels. Added protection is provided by keeping wafer leakage as low as possible and by operating the chips in the dark.

• *Sidewalk*

A fourth reliability mechanism (sidewalk) exists, which, unlike the mechanisms discussed earlier, does not involve vertical charge movement within or into the gate dielectric, but rather the development of a parasitic channel between source and drain after positive gate stress. Sidewalk is caused by mobile charged contaminants near the

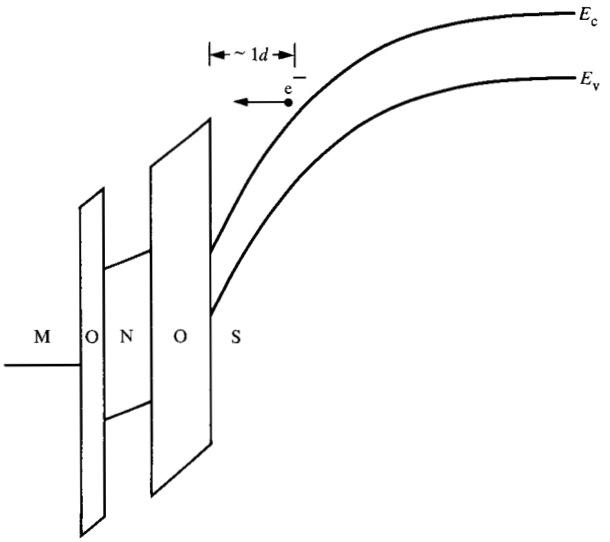

Figure 5 Band diagram for triple dielectric metal-oxide-nitride-oxide-semiconductor (MONOS) structure depicting electrons most likely to be injected over the oxide barrier and into nitride traps.

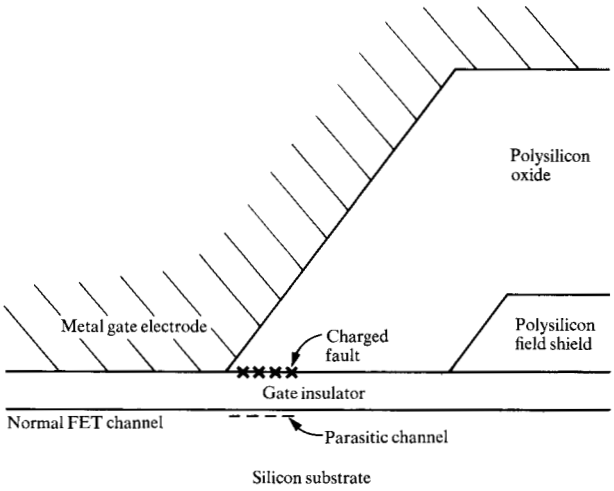

Figure 6 View of the gate sidewall, showing the region into which charged contaminants could be forced, resulting in a parasitic sidewalk current (from Ref. [43]).

edge of the metal gate, in conjunction with recessed areas into which the contamination can be driven. Repeated positive bias on the gate forces the charged ions into the region between the metal gate and field shield. The situation shown in Fig. 6 can develop, where the ions are not in contact with either conductor. Some of the charge is imaged in the silicon, resulting in a parallel FET not controlled by the metal gate. The device will then exhibit subthreshold current characteristics, as shown in Fig. 7. With the lift-off aluminum technology it has proven difficult to eliminate the ions, so the recesses have been eliminated. The polysilicon oxidation temperature was adjusted to eliminate lifting of the polysilicon oxide, and the etched polysilicon edges have shallow slopes to prevent the formation of cusps at the polysilicon oxide edge. Most important, SAMOS was changed from an MNOS gate to a MONOS gate. Originally, the oxidized nitride was removed from the gate region by a brief dip-etch. This etch created or aggravated crevices at the polysilicon oxide-nitride interface, providing a site for contamination. By leaving the oxidized nitride intact, this mechanism was eliminated [43].

• *Threshold voltage control and monitoring*

The foregoing represented reliability mechanisms which needed to be understood and controlled by SAMOS processing. A yield-related challenge, threshold voltage (V_T) control, also arose due to 1) charges in the multilayer insulator (V_{FB}) and 2) arsenic contamination of the gates affecting surface doping (counterdoping). On SAMOS product wafers, the first problem is sorting V_{FB} from doping effects, since both impact V_T. Troutman [44] has provided analytic solutions for the SAMOS gaussian doping profile, giving V_T as a function of V_{FB}, oxide thickness, and doping magnitude. J. Coady has demonstrated that, by measuring device FET characteristics to give oxide thickness and to give V_T as a function of substrate bias, V_{FB} and counterdoping effects can be individually extracted and used for large-scale SAMOS line characterization.

The actual V_{FB} behavior found for MONOS is much more complex than that of simple oxides, due to the existence of three layers and two interior interfaces for possible charge buildup. The amount of charge is influenced by oxidation precleans, oxidation anneals, nitride conditions and metallization conditions. (For example, electron-beam-evaporated aluminum is found to be unacceptable for SAMOS due to radiation damage of the nitride.) The fixed charge level of the MONOS gate is even affected by processing after first metallization, as quartzing anneals some of the charge. An interesting observation in SAMOS is that certain metallic impurities can "ride" on top of the growing oxide. No effect is seen on MOS V_{FB}, since the charge is all imaged in the metal gate. Once nitride is deposited, however, the charge is separated from the gate and alters V_{FB}.

The doping behavior is equally complex, since arsenic can be deposited in the bare silicon surface either during drive-in, during doped oxide etch, or during chemical cleaning steps. The problem occurrence can be limited by frequent changes of etchants and cleans, use of HCl during oxidation, and control of temperature profile for quick

oxidation at drive-in. If counterdoping does occur, it can be quickly detected using electrical measurements capable of detecting arsenic contamination in the 1×10^{10} atoms/cm^2 range. The techniques are employed on MOS capacitor monitor wafers processed with product through drive-in. A measurement of surface potential as a function of capacitor "gate" voltage is made [45], from which arsenic contamination level can be deduced [46]. Alternatively, a measurement of the slope of the usual capacitance-voltage curve can be made near the flatband voltage, with the slope becoming steeper in a quantitative manner as counterdoping occurs.

The various facets of threshold are controlled by in-line MOS monitoring. Additional process leverage in controlling V_{FB} and doping is obtained on product wafers themselves by pulsed capacitance measurements immediately after polysilicon patterning on capacitors etched in the polysilicon. The wafer doping and flatband voltage can be accurately predicted, providing process feedback without waiting for measurements after first metallization.

Yield-density-performance tradeoffs

While the description thus far has been of the semiconductor process, the ultimate goal is a semiconductor product which can meet system needs for performance and reliability, doing so at the highest possible productivity and lowest possible cost. A number of decisions must be made to determine the appropriate technology, the general design philosophy, and the possibility of added on-chip functions. An optimum tradeoff is then selected among yield, reliability, density, and performance.

The SAMOS technology was chosen for 1) high reliability, and 2) yielded bits per wafer. SAMOS is basically a simple, manufacturable process, producing 64K-bit chips with only five masks through first metallization. As will happen with any new technology, certain challenges were raised by the SAMOS process. Phenomena related to the use of nitride in the gate dielectric required learning and careful characterization, but they are not fundamental problems. Hot electrons, stability, and sidewalk are cleanly overcome through routine process control.

The second decision is the design philosophy relative to horizontal layout ground rules and parametric assumptions. Each layout ground rule is characterized by a standard deviation σ which is based on the statistical variations of the components which contribute to the ground rule. For each ground rule, a limit is set at the wafer dimension which results in a yield or reliability exposure to the chip. A guard band, measured by $n\sigma$, is set between the wafer nominal dimension and the ground rule limit. A

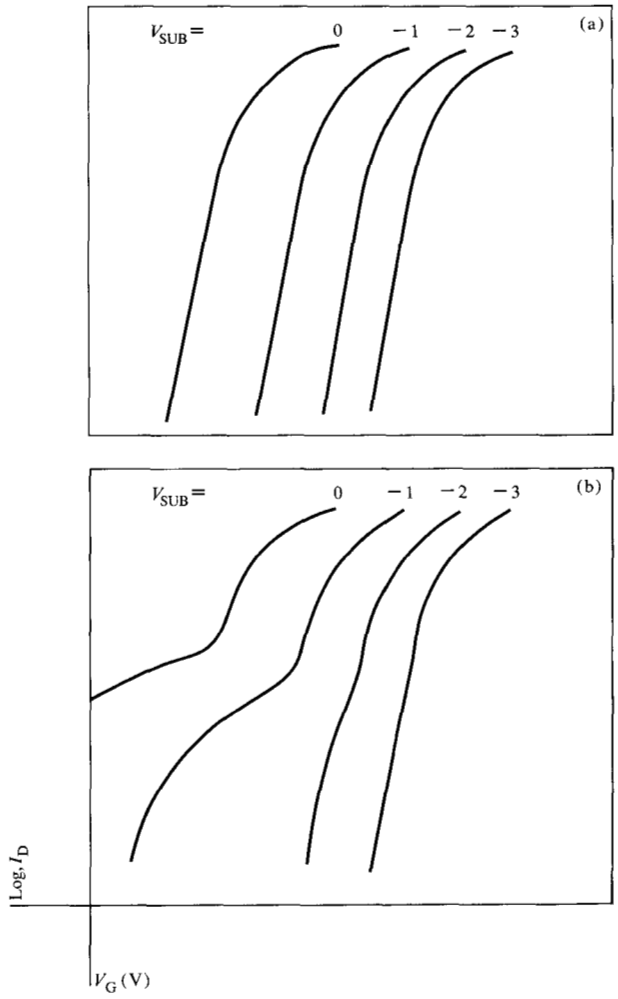

Figure 7 Subthreshold leakage resulting from sidewalk contamination. The drain current I_D is shown as a function of gate-source voltage V_G for a number of source-substrate voltages V_{SUB}, (a) prior to stressing and (b) post-stressing (from Ref. [43]).

large value of n reduces the probability that the limit will be exceeded but results in a penalty in chip area. Yield and reliability are traded directly against density. Similar considerations apply to parametric assumptions. A chip can be designed to function at large variations in threshold, transconductance, channel length, etc., but a density/performance penalty must be paid in such a design. For SAMOS, both layout and parametric rules are highly weighted to maximize yield and reliability.

The process description and design ground rules define one set of inputs to chip design. Another input must be the system needs in terms of performance, data organization, and off-chip supports. The chip designer must merge these capabilities and needs in order to maximize the number of usable bits per wafer. In SAMOS, the needs of

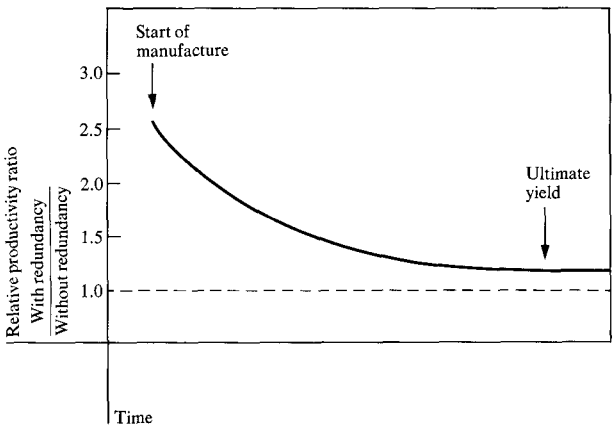

Figure 8 Yield improvement with redundancy, modeled for the 64K-bit chip (from Ref. [47]).

different systems and applications are met by a family of chips, each trading performance against density in a different manner, varying from a chip of 18K-bit, 90-ns typical access to 64K-bit, 300-ns typical access. Each member of the chip family also contains (in addition to the normal array, timing chain, decoders and I/O circuits) specialized circuits to enhance productivity or provide system flexibility. Two examples will be discussed here: the use of redundancy and the use of an on-chip register. The chip family will be discussed in the final section.

Redundancy

Redundancy to increase yield is presently used on all SAMOS chips. The redundancy actually exists in two levels. The first is the use of two-island chips. For instance, the 64K-bit chip has two 32K-bit islands, each capable of standing alone should the other be nonfunctional. The second level is the use of redundant bit lines and word lines within each island. The two-island approach and redundant lines both require extra circuits and result in increased chip size. The 64K-bit chip pays a penalty of ≈20% in chip area due to the two-island approach, and another 10% for word and bit redundancy [47]. The use of redundancy, either as dual islands or as fixable word and bit lines, is always based on an optimization of productivity. For each chip design, critical areas and defect densities are input to the yield model [13], and the projected number of usable bits per wafer calculated with and without redundancy. For all four chip designs thus far, some word or bit redundancy has resulted in a higher modeled productivity, while the dual-island approach has been incorporated on both the 64K-bit and 32K-bit chips. Any of the redundancy techniques will result in smaller improvements as manufacturing experience is gained and defect densities reduced. For dual islands, a crossover point will be reached eventually when the yield improvement does not compensate for the area penalty. Similarly, word or bit redundancy will become less important. Figure 8 shows the productivity improvement obtained by redundancy as a function of time, reflecting the real and projected SAMOS experience.

Even though the array represents less than 50% of the chip area, array redundancy is especially important because the cell node charge is so small and the array is so tightly packed. Defects which are of no consequence in support circuits will cause node failures. An example is leakage, for which a typical support will still function at more than 10× the level sufficient to cause node failure. Typical SAMOS history is that the large majority of defects on leakage monitors are of the "cell node fail" type and are not of a sufficiently high leakage level to cause a fail in a support circuit. Most leakage defects are therefore fixable by redundancy if they occur in the array and are inconsequential if they occur in other areas.

SAMOS array redundancy is implemented using what can be described as a "write-once read-only memory." Fusible links in the second-level metallurgy are used in a scheme which has minimum impact on chip performance. Redundancy is provided in both the bit-line and word-line directions. After second-level metal testing, the failing addresses of "fixable" chips are stored for fuse blowing. When the fuses are blown electrically, the address of a failing line is written into the fuses, and an additional enable fuse is blown to activate the redundant line. The circuit details of how the redundant lines are used are treated in a companion paper [48]. Whenever any redundancy is used in SAMOS, the chips are mounted into modules such that the redundancy is transparent to the user. A module made of "single-island good" chips will have twice as many chips inside as one made from perfect parts, but all module inputs and outputs will be identical.

On-chip register

The on-chip register is used on the 64K-bit chip to provide extra flexibility to the system. It is included at a penalty in chip area of ≈7%. The chip is organized in two islands, each with 128 word lines and 256 bit lines. When the ith bit line (i = 1, 32) is addressed, eight bit lines $i + 32n$, where n = 0 through 7, are decoded. All eight bits are read into the register in parallel. Eight input pins called data gates are available, and any number of these (1-8) may be sequentially addressed in any desired order at a 10-MHz rate. The stored bits are brought off the chip on a single pair of output pins. This chip organization has been used at the card level to achieve data rates higher than the cycle time of the chip. For example, the card used in the IBM 4300 processors is capable of reading a 4-

Table 2 Summary of chip characteristics.

	18K-bit	*32K-bit*	*36K-bit*	*64K-bit*
Chip organization	2K by 9	8K by 4	4K by 9	64K by 1
Cell type	Twin-cell	Single	Single	Single
Cell size (μm)	9.3 × 42	9.3 × 32.7	11.5 × 22	9.3 × 17.9
Chip size (mm)	4.46 × 5.52	5.9 × 6.1	5.0 × 5.0	5.95 × 6.15
Redundant word lines	2	8	4	8
Redundant bit lines	0	8	4	4
Typical access (ns)	90	190	900	300
W.C. access/cycle (ns)	140/256	285/470	1500/2000	440/980
Power; stdby/sel (mW)	10/690	12/612	10/220	40/360
No. I/O	45	31	30	38
Bits/cycle	9	1, 2, or 4	9	8

Table 3 Summary of module characteristics.

	18K-bit	*32K-bit*	*36K-bit*	*64K-bit*
Module organization (max)	8K by 9	32K by 4	16K by 18	128K by 4
Inputs	Hi-level FET	Hi-level FET	Hi-level FET & TTL	Hi-level FET
Data out	Differential current sense	Current sense	TTL	Current sense
No. I/O	59	45	46	51
Supplies	+8.5 V ± 10% +5.0 V ± 10% −2.2 V ± 15%	+8.5 V ± 10% −2.2 V ± 15%	+8.5 V ± 10% +1.5 V ± 15% −1.5 V ± 15%	+8.5 V ± 10% +4.25 V ± 10% −2.2 V ± 15%
Bits/cycle	9	1, 2, or 4	18	32

byte word (40 bits) at a sustained data rate of 150 ns per word, rather than at the chip cycle time specification. With a conventional design approach, this would have required a chip with a 150-ns cycle time.

It should be noted that the 64K-bit chip is used predominantly in large systems that use a single-bit-correction, double-bit-detection error correction code, typically with a 39-bit word. Since each chip supplies only one bit of the word, a chip fail in the system causes only a single-bit error and is hence correctable at the system level.

Chip family

The SAMOS process is used to meet the needs of a multitude of users through a family of four chips, ranging from 18K-bit, 90-ns typical access through 64K-bit, 300-ns access. By design, the different chips are all made identically in the process, with the differences being determined only by the photolithographic masks used. A summary of chip characteristics is given in Table 2 [1]. The 18K-, 32K-, and 64K-bit chips show directly the tradeoff of performance against density among the different designs. The 36K-bit chip appears to be anomalous in this progression, in that it is the slowest chip but not the densest. It was, however, an earlier-vintage, more conservatively designed chip used for early manufacturing learning. The high-performance chips are characterized by larger storage nodes and shorter bit lines resulting in larger signals and simpler sense detection schemes [49]. All chips can be used to provide multiple bits in each cycle, either simultaneously (18K, 32K, 36K) or sequentially (64K). The 18K- and 36K-bit chips are typically used in byte-organized applications where each 9-bit output will be an 8-bit byte plus a parity bit for error detection. The 32K- and 64K-bit chips are bit-organized. Figures 9 and 10 show the 18K-bit and 64K-bit chips, respectively, with areas labeled as to function.

The characteristics of the various chips after they are mounted into modules are given in Table 3, with a stacked module of 64K-bit chips shown in Fig. 11. At the module level, the full capability and flexibility of the overall tech-

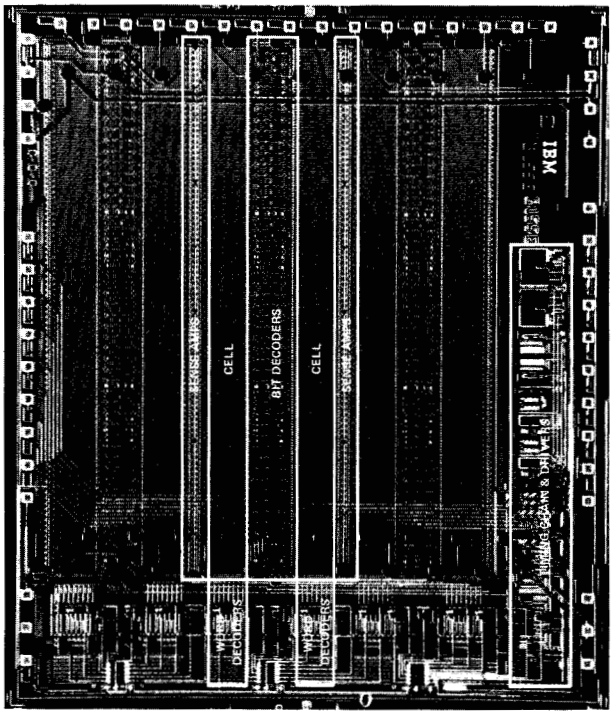

Figure 9 The 18K-bit chip, with areas labeled as to function. Typical access time is 90 ns, worst-case access is 140 ns.

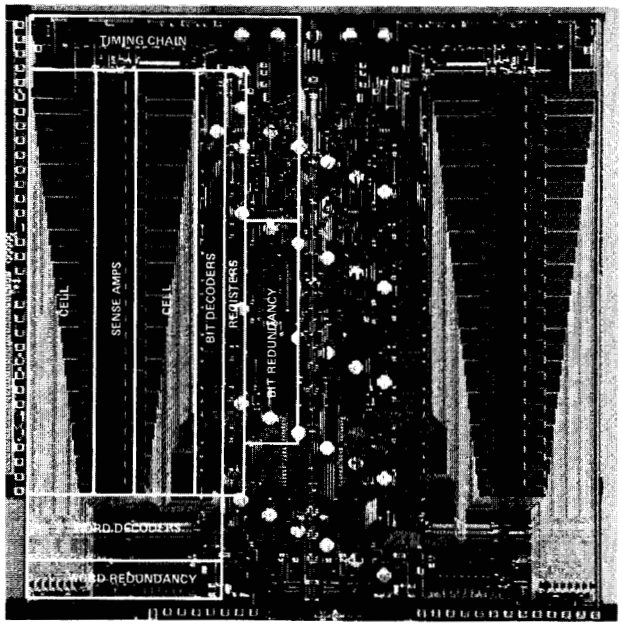

Figure 10 The 64K-bit chip, showing the on-chip shift register. Typical access time is 300 ns, worst-case access is 440 ns.

nology becomes apparent. For high-performance applications a 2.54-cm (one-inch)-square module with 72K-bit storage can provide nine bits of simultaneous output in a 256-ns worst-case cycle. When high density is required, a different 2.54-cm (one-inch)-square module with more than one-half megabit capacity can provide, in a cycle of $\approx$1000 ns, four parallel channels of output, each channel eight bits deep.

Summary

The SAMOS process, the motivation behind the process steps, and the merger of process capabilities and system needs have been discussed in this paper. SAMOS combines the one-device-cell concept and a number of unique process steps, which can be summarized as follows.

1. Backside ion implantation provides a simple, clean technique for leakage control.
2. Diffusions are driven in from a doped oxide layer which was applied to the bare wafer and removed subtractively. The technique allows narrow diffused lines to be made, and the doped oxide can be left above the diffusions where desired to provide a self-aligned oxide for low capacitance to conductive layers.
3. A thin oxide-nitride dual dielectric covers the entire wafer with the exception of contact holes and some diffused areas. The dual dielectric provides low defect density, and the nitride performs the crucial role of preventing oxidation of gate regions in later processing.
4. A conductive polysilicon layer is used as a field shield to shut off surface leakage, and forms the plate of the diffused storage node capacitors.
5. Lift-off aluminum metallurgy is used to provide minimum spaces between aluminum lines.
6. A dual layer of quartz and polyimide separates first and second levels of metallization, providing passivation and insulation with very low defect density.

Because of the use of the nitride in the gate dielectric, a number of yield- and reliability-related phenomena had to be overcome. Potential problem areas included threshold voltage control, threshold stability, hot carrier injection (channel or substrate), and parasitic device leakage. The general technique to overcome these items has been to understand the physics and eliminate processing causes where possible. When the problem was inherent in VLSI devices, it was carefully characterized, and circuits were designed to tolerate the changes expected with time.

The process was merged with a design philosophy aimed at optimal reliability and yield. Guard bands were used so that devices still function at large excursions of the process from nominal. A family of chips was designed for different applications, ranging from high performance (18K-bit, 90-ns typical access) to high density (64K-bit, 300-ns typical access). All chips are produced by the same process, with the only difference being the photo-

lithographic patterns. The 32K-bit and 64K-bit chips use a dual-island approach, and all chips use some word or bit redundancy for enhanced yield. A penalty is paid in chip size in order to optimize the number of usable bits per wafer, and features such as an on-chip register are included where appropriate to enhance system performance. The chips are mounted on one-inch-square substrates with up to four chips on a substrate. Using stacked substrates, a maximum module density of 524 288 bits is obtained.

Acknowledgments

Sincere appreciation must be expressed for the efforts of the large number of colleagues at the IBM facilities in Essex Junction, Vermont, and in Sindelfingen/Boeblingen, West Germany, who developed the SAMOS process and product set, overcame the obstacles of a new technology, and successfully brought the process into manufacturing. In addition, many people made helpful suggestions on this manuscript. The assistance of G. S. Alberts, R. R. Troutman, V. L. Rideout, R. H. Kruggel, R. H. Linton, P. A. Farrar, D. R. Weed, E. P. Thoma, G. H. Parikh, R. M. Quinn, R. R. DeSimone, and others is appreciated.

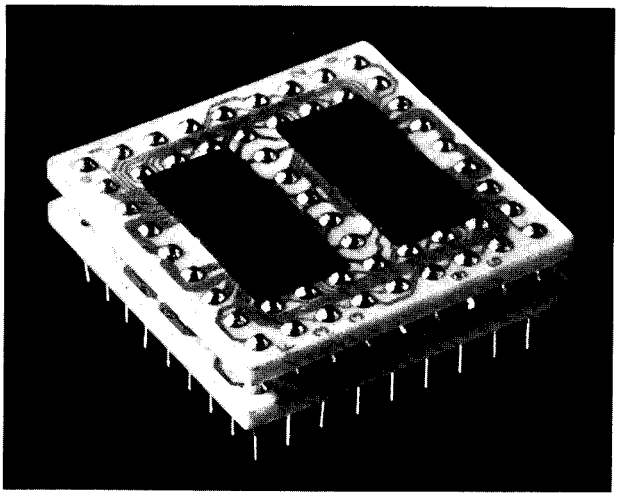

Figure 11 Uncapped view of a module with two decks, each deck having four 64K-bit chips. This module contains 524 288 bits of random access memory in a 2.54-cm × 2.54-cm-square area.

References and notes

1. R. R. DeSimone, N. M. Donofrio, B. L. Flur, R. H. Kruggel, H. H. Leung, and R. Schnadt, "FET RAMs," *1979 IEEE ISSCC Digest of Technical Papers* **22,** 154-155 (1979).
2. R. H. Dennard, "Field Effect Transistor Memory," U.S. Patent 3,387,286, 1968.
3. V. L. Rideout, "One Device Cells for Dynamic Random Access Memories—A Tutorial," *IEEE Trans. Electron Devices* **ED-26,** 839-852 (1979).
4. P. E. Cottrell, R. R. Troutman, and T. H. Ning, "Hot-Electron Emission in N-Channel IGFET's," *IEEE J. Solid-State Circuits* **SC-14,** 442-455 (1979).
5. S. E. Schuster, "Multiple Word/Bit Line Redundancy for Semiconductor Memories," *IEEE J. Solid-State Circuits* **SC-13,** 698-703 (1978).
6. W. M. Smith, Jr., "Field Effect Transistor Integrated Circuit and Memory," U.S. Patent 3,811,076, 1974.
7. R. R. Garnache and W. M. Smith, Jr., "Integrated Circuit Fabrication Process," U.S. Patent 3,841,926, 1974.
8. B. Masters, J. Fairfield, and B. Crowder, *Ion Implantation*, F. Eisen, Ed., Gordon and Breach Science Publishers, New York, 1971.
9. T. M. Buck, K. A. Pickar, J. M. Poate, and C. M. Hsieh, "Gettering Rates of Various Fast Diffusion Metal Impurities at Ion-damaged Layers on Silicon," *Appl. Phys. Lett.* **21,** 485 (1972).
10. T. E. Seidel, R. L. Meek, and A. G. Cullis, "Direct Comparison of Ion Damage Gettering and Phosphorus Diffusion Gettering of Au in Si," *J. Appl. Phys.* **46,** 600 (1975).
11. H. J. Geipel and W. K. Tice, "Reduction of Leakage by Implantation Gettering in VLSI Circuits," *IBM J. Res. Develop.* **24,** 310-317 (1980, this issue).
12. H. J. Geipel and W. K. Tice, "Critical Microstructure for Ion-Implantation Gettering Effects in Silicon," *Appl. Phys. Lett.* **30,** 325 (1977).
13. C. H. Stapper, A. N. McLaren, and M. Dreckmann, "Yield Model for Productivity Optimization of VLSI Memory Chips with Redundancy and Partially Good Product," *IBM J. Res. Develop.* **24,** 398-409 (1980, this issue).
14. J. Scott and J. Olmstead, "A Solid-to-Solid Diffusion Technique," *RCA Rev.* **26,** 357 (1965).
15. A. W. Fisher, J. A. Amick, H. Hyman, and J. Scott, "Diffusion Characteristics and Application of Doped Silicon Dioxide Layers Deposited from Silicon (SiH_4)," *RCA Rev.* **29,** 533 (1968).
16. M. L. Barry and P. Olofsen, "Doped Oxides as Diffusion Sources," *J. Electrochem. Soc.* **116,** 854 (1969).
17. M. Ghezzo and D. M. Brown, "Arsenic Glass Source Diffusion in Si and SiO_2," *J. Electrochem. Soc.* **120,** 110 (1973).
18. P. C. Parekh, D. R. Goldstein, and T. C. Chan, "The Influence of the Reaction Kinetics of O_2 and Source Flow Rates on the Uniformity of Boron and Arsenic Diffusions," *Solid-State Electron.* **14,** 281 (1971).
19. This work is attributed to A. S. Bergendahl, D. W. Rakowski, and P. Hazelton, IBM General Technology Division, Essex Junction, VT.
20. R. R. Troutman, "VLSI Device Phenomena in Dynamic Memory and Their Application to Technology Development and Device Design," *IBM J. Res. Develop.* **24,** 299-309 (1980, this issue).
21. S. A. Abbas, C. A. Barile, and R. C. Dockerty, "Low Leakage, N-channel Silicon Gate FET with Self Aligned Field Shield," *IEDM Tech. Digest*, 371-373 (1973).
22. T. L. Chu, J. R. Szedon, and C. H. Lee, "The Preparation and *C-V* Characteristics of Si-Si_3N_4 and Si-SiO_2-Si_3N_4," *Solid-State Electron.* **10,** 897-905 (1967).
23. (a) M. Hatzakis, "Electron Resists for Microcircuit and Mask Productions," *J. Electrochem. Soc.* **116,** 1033 (1969).
    (b) G. Bergasse, "Two-Resist Layers Lift-Off Process," *IBM Tech. Disclosure Bull.* **16,** 2110-2111 (1973).
    (c) E. C. Fredericks, G. C. Schwartz, and L. B. Zielindi, "Polysulfone Lift-Off Masking Technique," *IBM Tech. Disclosure Bull.* **20,** 989 (1977).
    (d) Y. Hom-ma, H. Nozawa, and S. Harada, "A New Lift-Off Metallization Technique for High Speed Bipolar LSI's," *IEDM Tech. Digest*, 54 (1979).
24. (a) B. J. Canavello, M. Hatzakis, and J. M. Shaw, "Process for Obtaining Undercutting of a Photoresist to Facilitate Lift-Off," *IBM Tech. Disclosure Bull.* **19,** 4048 (1977).
    (b) W. D. Grobman, H. E. Luhn, T. P. Donohue, A. J.

Speth, A. Wilson, M. Hatzakis, and T. H. P. Chang, "1 μm MOSFET VLSI Technology: Part VI—Electron-beam Lithography," *IEEE Trans. Electron Devices* **ED-26,** 360 (1979).
25. F. M. d'Heurle and P. S. Ho, "Electromigration in Thin Films," *Thin Films: Interdiffusion and Reactions*, J. M. Poate, K. N. Tu, and J. W. Meyer, Eds., John Wiley & Sons, Inc., New York, 1978.
26. Paul A. Farrar, "Method for Making Integrated Circuit Contact Structure," U.S. Patent 3,830,657, 1974.
27. R. M. Geffken, "Effect of Various Thin Film and Interface Phenomena on Al-Si Contact Resistance," presented at the IEEE VLSI Workshop, New York, May 4, 1979.
28. H. C. Cook, P. A. Farrar, R. R. Uttecht, and J. P. Wilson, "Structure for Improving the Passivation of Semiconductor Chips," *IBM Tech. Disclosure Bull.* **16,** 728 (1973).
29. G. S. Alberts, P. A. Farrar, and R. L. Hallen, "Method for Constructing a ROM for Redundancy and Other Application," U.S. Patent 3,959,047, 1966.
30. P. A. Totta and R. P. Sopher, "SLT Device Metallurgy and its Monolithic Extension," *IBM J. Res. Develop.* **13,** 226 (1969).
31. L. F. Miller, "Controlled Collapse Reflow Chip Joining," *IBM J. Res. Develop.* **13,** 239 (1969).
32. E. H. Snow, A. S. Grove, B. E. Deal, and C. T. Sah, "Ion Transport Phenomena in Insulating Films," *J. Appl. Phys.* **36,** 1664 (1965).
33. J. V. Dalton, "Sodium Drift and Diffusion in Silicon Nitride Films," *J. Electrochem. Soc.* **113,** 165C (1966).
34. D. Frohman-Bentchkowsky and M. Lenzlinger, "Charge Transport and Storage in Metal-Nitride-Oxide-Silicon (MNOS) Structures," *J. Appl. Phys.* **40,** 3307 (1969).
35. P. K. Chaudhari, J. M. Franz, and C. P. Acker, "Electrical Properties of Vapor Deposited Silicon Nitride Films Measured in Strong Electric Fields," *J. Electrochem. Soc.* **120,** 991 (1973).
36. S. A. Abbas and R. C. Dockerty, "Hot Carriers Instability in IGFET's," *Appl. Phys. Lett.* **27,** 147 (1975).
37. T. H. Ning, C. M. Osburn, and H. N. Yu, "Effect of Electron Trapping on IGFET Characteristics," *J. Electron. Mater.* **6,** 65-76 (1977).
38. P. C. Arnett and B. H. Yun, "Silicon Nitride Trap Properties as Revealed by Charge-centroid Measurements on MNOS Devices," *Appl. Phys. Lett.* **26,** 94 (1975).
39. T. H. Ning, C. M. Osburn, and H. N. Yu, "Threshold Instability in IGFETS due to Emission of Leakage Electrons from Silicon Substrate into Silicon Dioxide," *Appl. Phys. Lett.* **29,** 198-200 (1976).
40. R. R. Troutman, "Silicon Surface Emission of Hot Electrons," *Solid-State Electron.* **21,** 283-289 (1978).
41. T. H. Ning, C. M. Osburn, and H. N. Yu, "Emission Probability of Hot Electrons into Silicon Dioxide," *J. Appl. Phys.* **48,** 286-293 (1977).
42. J. F. Verwey, "Nonavalanche Injection of Hot Carriers into SiO_2," *J. Appl. Phys.* **44,** 2681 (1973).
43. This work is attributed to J. P. Wilson, M. D. Potter, R. A. Powlus, R. R. Uttecht, J. P. Stinson, W. J. Ayer, and A. J. Wager, IBM General Technology Division, Essex Junction, VT.
44. R. R. Troutman, "Ion Implanted Threshold Tailoring for Insulator Gate Field-Effect Transistors," *IEEE Trans. Electron Devices* **ED-24,** 182-192 (1977).
45. B. H. Yun, "Inspecting Impurity Concentrations in Semiconductors," *IBM Tech. Disclosure Bull.* **18,** 4226 (1976).
46. This work is attributed to B. H. Yun (IBM Data Systems Division, East Fishkill, NY) and to R. A. Corcoran, W. A. Keenan, A. Kaloustian, and D. Michaelides (IBM General Technology Division, Essex Junction, VT).
47. R. H. Kruggel, presented at the IEEE Computer Elements Committee Workshop, Vail, Colorado, June 24-27, 1979.
48. B. F. Fitzgerald and E. P. Thoma, "Circuit Implementation of Fusible Redundant Addresses on RAMs for Productivity Enhancement," *IBM J. Res. Develop.* **24,** 291-298 (1980, this issue).
49. Kenneth S. Gray, "Cross-Coupled Charge-Transfer Sense Amplifier and Latch Sense Scheme for High-Density FET Memories," *IBM J. Res. Develop.* **24,** 283-290 (1980, this issue).

Received October 3, 1979; revised December 14, 1979

The author is located at the IBM General Technology Division laboratory, Essex Junction, Vermont 05452.