R. R. Troutman

VLSI Device Phenomena in Dynamic Memory and Their Application to Technology Development and Device Design

Ever-increasing density poses significant challenges to the device designer, who must relate an integrated technology to the numerous electrical characteristics required for successful memory design. Success of a VLSI technology depends as much on the extensive design of small devices as on the sophisticated lithography with which to fabricate them. Several dimensional limitations arise from the electrical characteristics both of intentionally switching devices and of possible parasitic devices. Account must be taken of threshold dependence on both channel length and width. Furthermore, any isolation scheme must not introduce leakage from the storage node, such as parasitic subthreshold and low-level punchthrough currents. Hot electron emission depends on both horizontal and vertical dimensions and must be minimized to guarantee the requisite long-term device behavior. This paper will briefly discuss the physical origins of the above fundamental device phenomena, their influence on SAMOS device design, and implications for future memory technologies.

1. Introduction

Today, the highest-density dynamic memory produced commercially uses a memory cell consisting of an insulated gate field-effect transistor (IGFET) connected in series with a capacitor. As simple as that configuration may seem, there are more device phenomena associated with high-density memory than can be covered in this paper. The phenomena to be discussed here were chosen because of their applicability to all dynamic memory designs and because of their clear and direct relation to device dimensions. Even within this restricted list, space limitations preclude covering all small-dimensional effects.

Such phenomena are important because they can impose fundamental lower limits on the dimensions of devices used for high-density dynamic memory. To illustrate these relationships, examples are drawn from the development phase of SAMOS (Silicon and Aluminum Metal Oxide Semiconductor) memory technology, a technology that IBM has been using to manufacture 64K-bit RAMs for several years. Specifically, the device phenomena to be discussed are as follows: 1) reduction of threshold voltage as channel length is reduced; 2) increase of threshold voltage as channel width is reduced; 3) low-level punchthrough current between two closely spaced diffusions; and 4) increased thermal emission of hot electrons into the gate insulator as channel length is reduced. Items 1 and 3 are closely related, each being a different manifestation of drain-induced barrier lowering. This paper will cover the physical origin of each phenomenon, its design consequences, and its effect on SAMOS memory technology.

The following section briefly reviews the one-device dynamic memory cell and explains the importance of threshold voltage in dynamic memory design. Section 3 introduces the SAMOS technology and discusses the reasons behind various process steps. Then in Section 4 the details of threshold design are used to illustrate the interaction of process and device modeling during the development of the SAMOS technology. Section 5 shows how functional chip designs incorporate the effect of threshold shifts caused by hot electron emission and subsequent trapping in the gate insulator. A summary and conclusions are given in Section 6.

2. The one-device memory cell

Before turning to the device phenomena, let us consider the one-device memory cell pictured in Fig. 1. Information is stored in the form of charge on a capacitor. Access

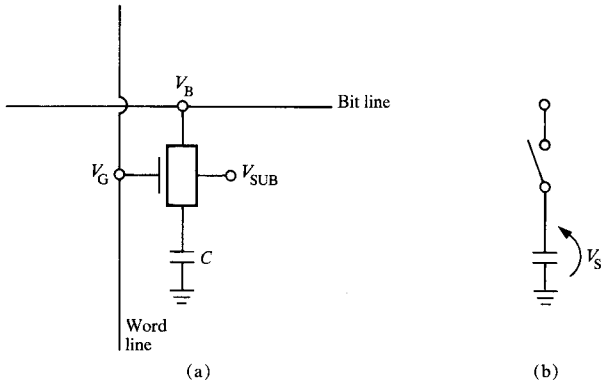

Figure 1 (a) Schematic representation of the one-device memory cell. (b) Functional representation of the one-device memory cell. For $V_{GS} < V_{T,OFF}$ the switch is open, while for $V_{GS} > V_{T,ON}$ the switch is closed. The impedance of the switch increases as V_{GS} decreases from $V_{T,ON}$ to $V_{T,OFF}$, typically a range of several hundred millivolts. The number of electrons stored $N_S = CV_S/q$ is of the order of a million.

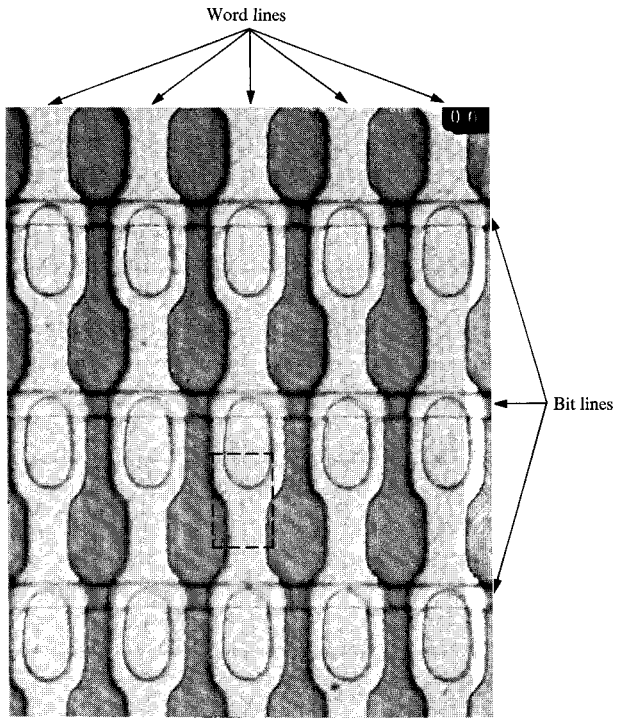

Figure 2 SAMOS memory cells on the 64K-bit RAM chip. See Section 3 for a description.

to the capacitor occurs through the IGFET, which acts as a switch. When the gate-to-source voltage V_{GS} exceeds the threshold voltage V_T, the switch is "ON," and a highly conductive path exists between the source and drain diffusions. If the bit line voltage V_B is high during the time that $V_{GS} > V_T$, the capacitor is charged, while if V_B is zero, the capacitor is discharged. For dimensions typical of today's designs, the amount of charge stored in the first case amounts to only a few million electrons and is proportional to $(V_{GS} - V_T)$. Clearly it is advantageous to have V_T as small as possible in order to maximize the signal stored.

Because of the small amount of charge stored, the "OFF" impedance of the switch must be very high to prevent charge from leaking back through the switch during storage time. Indeed, an important design criterion is to ensure that the minimum threshold is high enough to guarantee that this subthreshold current is less than the thermal leakage from the storage node [1, 2]. Because threshold voltage is bounded on both the high and low sides, it must be carefully selected to maximize the stored charge.

Even for a well-designed cell thermal leakage slowly discharges the storage capacitor so that the information state corresponding to a charged capacitor gradually changes to the information state of a discharged capacitor, causing an error. This thermal leakage necessitates periodic refreshing of the binary states to avoid errors. It is for this reason the cell is referred to as dynamic.

Another important detail of the one-device memory cell is that as the capacitor is charged, the reverse bias between the source diffusion and substrate increases. This has the undesirable effect of raising the threshold voltage. Thus, one goal of memory cell design is to minimize this "substrate sensitivity" (also referred to as "body effect" or "back gate bias effect").

For devices with wide and long channels, threshold depends on the gate insulator thickness, the gate insulator and electrode materials, the level of fixed charge in the insulator, the total substrate doping, and the substrate bias. For the small devices commonly used in memory arrays, threshold also depends, to some extent, on channel length and width and on drain voltage [3–11]. Because all these parameters are essentially random variables distributed about their nominal values, the threshold voltage of all devices in a memory array is expected to have a process-dependent distribution. Not only must threshold be carefully selected in a successful design, it must also be tightly controlled when initially fabricated and over the lifetime of the device.

By now it should be quite clear that any parameter affecting threshold is important in dynamic memory design. This point is examined in more detail in Section 4 with

regard to threshold tailoring via ion implantation. Let us turn now to a process sequence that results in small memory cell area.

3. SAMOS—An example of a VLSI technology

A top view of several SAMOS [12, 13] one-device memory cells on the 64K-bit chip is shown in Fig. 2. The three horizontal rails are the diffused bit lines, and the undulating vertical lines are the metal word lines. Holes have been etched through the polysilicon field shield in the oval-shaped areas, forming the gate of the transfer device. The metal gate overlaps the bit line at one end of the gate and the diffused storage node at the other. Because of the intervening field shield, the rectangular area of the storage node is barely perceptible and has been enhanced for one cell by the dashed line. Total cell size is 9.25 μm by 17.5 μm.

Figure 3 shows the details of this structure for the memory cell region. The n^+ diffusion on the right forms the bit/sense line while the n^+ region on the left forms one plate of the storage capacitor. They are connected by an aluminum gate switching device.

The SAMOS process is presented in a companion paper [14], but a brief description is included here to illustrate the influence of device properties on technology development. A process outline is given in Table 1.

High resistivity was chosen for the starting substrate to minimize junction capacitance and substrate sensitivity. The memory cell itself is a source follower, and a threshold increase while charging the storage capacitor results in less charge being stored. High resistivity also helps reduce emission of substrate hot electrons.

A pyrolytically deposited oxide highly doped with arsenic is used to diffuse sources and drains. The etch bias between mask and wafer causes diffusion widths to shrink and the spacing between diffusions to grow. (This contrasts with source/drain diffusions fabricated by diffusing or implanting the dopant through an etched opening.) The etch bias favors high-density cells because channel lengths at the wafer level, for both active devices and parasitic thin oxide devices, are longer for the same mask dimensions than in other technologies. These longer channel lengths on the wafer reduce drain-induced barrier-lowering, giving better threshold control for active devices and better isolation for parasitic devices.

A second mask is used to etch the doped oxide after the diffusion step so that it can be selectively removed. Leaving it over the bit line reduces the capacitive coupling to aluminum lines and the polysilicon field shield, giving a better charge transfer ratio. It is removed over the storage node, which consists of thin oxide capacitance between the n^+ diffusion and field shield in parallel with the junction capacitance to the substrate.

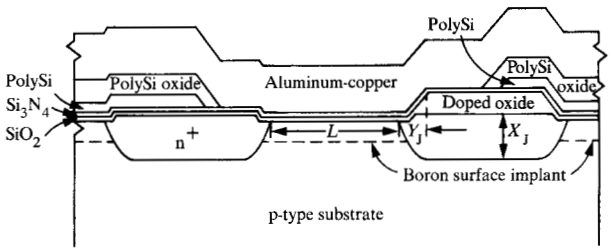

Figure 3 Cross-sectional representation of the SAMOS memory cell.

Table 1 SAMOS process outline.

Process steps	*Threshold tailoring ion implant option*
1. Backside ion implant	
2. Deposit doped oxide/cap oxide	
3. Pattern doped oxide	
4. Arsenic drive-in and oxidation	
← A	Dose = 7.5×10^{11} atoms/cm^3 R_p^* = 70 nm σ = 50 nm
5. Selectively remove doped oxide	
6. Thermal gate oxide	
7. Deposit CVD nitride	
← B	Dose = 7.5×10^{11} atoms/cm^3 R_p^* = 70 nm σ = 50 nm
8. Deposit CVD polysilicon	
← C	Dose = 10^{12} atoms/cm^3 R_p^* = 70 nm σ = 75 nm
9. Pattern polysilicon	
10. Oxidize polysilicon	
← D1	Dose = 6×10^{11} atoms/cm^3 R_p^* = 100 nm σ = 70 nm
← D2	Dose = 6×10^{11} atoms/cm^3 R_p^* = 700 nm σ = 100 nm
11. Open contact holes	
12. Deposit and pattern first-level metal	

*Measured from Si-SiO_2 interface.

The source/drain junction depth of $X_J = 0.8$ μm is a compromise depth. It should be shallow to minimize drain-induced barrier-lowering at the source [15, 16], yet deep enough to avoid aluminum shorting to the substrate and avalanche junction breakdown at the voltages to be used. Tapering of the etched doped oxide makes the lateral junction gradient more gradual than an implanted arsenic junction.

An unmasked boron implant increases the surface concentration of the lightly doped substrate. This implant is used for tailoring the threshold of the switching devices in the memory cell and of the parasitic devices between storage nodes. It increases the boron doping only near the surface so threshold is increased without unduly raising the substrate sensitivity [2, 17]. Details of the design and processing tradeoffs associated with threshold tailoring are given in the next section.

The gate insulator is made thin to provide high transconductance and to reduce drain-induced barrier-lowering. Its dual film layer of oxide and nitride provides a high degree of protection against gate shorts due to pinholes, so that the thin gate insulator has a high yield and very good operational reliability.

A polysilicon field shield is used to provide good electrical isolation between adjacent diffusions and to screen the substrate from aluminum wiring. With the field shield, there are no parasitic field oxide devices. The polysilicon is doped p-type to decrease the surface potential and is biased to the same potential as the substrate. Not only does this deter adjacent bit disturbs, but it also reduces surface leakage. Use of a substrate bias also reduces junction capacitance, substrate sensitivity, subthreshold voltage excursion, and charge pumping effects.

The field shield is thermally oxidized to insulate it from subsequent wiring. Because of the nitride in the active gate region, only a very small amount of oxide forms there, and no additional mask is required to define the device regions before aluminum is deposited.

Aluminum-copper metallurgy is used for the first level of metal and is applied using a lift-off technique to maximize wiring density. In this technique, metal is deposited after a pattern has been formed in the underlying photoresist film. Only metal deposited where the photoresist was absent remains after the photoresist is stripped away. Use of a dual layer of quartz and polyimide to insulate the first level of metal from the second ensures high yield because of the small probability of pinhole overlap. Via holes are etched through the dual dielectric to first-level metal pads. Then a layer of chrome-copper-gold is deposited and etched. Finally, a layer of polyimide is added as a seal for the chip.

Since SAMOS was designed to be a high-density memory technology, one would expect some reduction in power supply voltage from that used in previous memory technologies. Indeed, the 8.5-V supply is 2 V less than that used in the 8K-bit RAM chip [18], but careful design and judicious tradeoffs made any further reduction unnecessary for the SAMOS 18K-, 32K-, and 64K-bit RAM chips [19].

It is beyond the scope of this paper to discuss the entire SAMOS design. Instead, two important areas have been selected because they are fundamental to all high-density dynamic memory designs and because they illustrate the integrated nature of device design. The first, threshold modeling, blends a knowledge of device physics, circuit requirements, and process characteristics. The second, channel hot electron emission, involves a relatively new area of device physics and related techniques of incorporating long-term reliability into the initial design of product chips.

4. The interaction of process and device modeling

In the last section we saw how device considerations influenced the formulation of a new technology. This section focuses on 1) how to determine the optimum threshold voltage for the one-device memory cell, and 2) how to adjust the implantation step to achieve this optimum for a given memory cell technology. It also illustrates the interdependence of process and device modeling.

Although the use of ion implantation for threshold tailoring is common today, it was not so at the beginning of SAMOS development. The short channels contemplated for SAMOS required an increase in surface doping level over previous technologies to reduce threshold dependence on channel length. Simply raising the bulk substrate doping has the obvious disadvantages of increasing source follower threshold sensitivity and junction capacitance. Studies comparing dynamic one-device cell designs on uniformly doped substrates with those on surface-implanted, lightly doped substrates indicated that the latter provided superior designs. In one example, the latter resulted in either a 10% reduction in cell size coupled with a 25% power reduction or a 42% power reduction for the same cell size. In both cases, the access time was nearly halved for the same circuit layout. Additional benefits of the surface implant were 1) to compensate for boron depletion during gate oxidation, 2) to reduce threshold variation resulting from bulk doping variation, 3) to adjust threshold precisely (*i.e.*, other process changes could be accommodated more easily), and, with

extra masking, 4) to provide multiple thresholds in a single process (giving, for example, both enhancement and depletion mode devices).

The first question to be settled was at what step in the process the boron implant was to be done. The four options are illustrated in Table 1 and are as follows: immediately following (A) the drive-in oxidation, (B) nitride deposition, (C) polysilicon deposition, or (D) polysilicon oxidation. For the first three a single, unmasked implant would suffice, while for (D) both a shallow and a deep implant would be required because the surface was no longer planar. The shallow implant would provide correct surface tailoring for the active gate, for which the total film layer was approximately 50 nm, and the deep implant would provide correct surface doping under the polysilicon field shield where the total film thickness above the silicon was approximately 500 nm. The shallow implant would not reach the silicon under the field shield, and the deep implant should be deep enough under the active gate region that the substrate sensitivity curve would remain unaffected.

Each of these options was first simulated using a computer model. The simulations were then checked by *C-V* measurements of the implant profile at different stages of the processing steps. An example of each for a specific dose is shown in Figs. 4(a–d).

Agreement of simulated and measured results is quite good in Figs. 4(b and c) and reasonably good in Fig. 4(d). Because there is only one short diffusion step (to electrically activate the implanted boron), the total elapsed time is especially important in this last case. It appears from Fig. 4(d) that the 15 minutes at 900°C does not adequately characterize actual furnace conditions during the anneal since the measured profile appears a bit more spread out than the simulated profile.

The results of Fig. 4(a) were the most surprising. The measured profile shows about half as much boron left in the silicon as the simulated curve, and it appears to have diffused more. The key factor in this case was that the substrate was oxidized after the boron was implanted. Although the simulation program included a segregation coefficient to account for boron depletion, varying this value alone could not explain the measured results. During the oxidation step, there was enhanced diffusion of the boron in the silicon and possibly through the oxide as well since the oxidation was performed in the presence of HCl. Enhanced boron diffusion during oxidation has also been observed in subsequent work [20].

The important point here is that an apparently "anomalous" experimental profile was recognized for what it

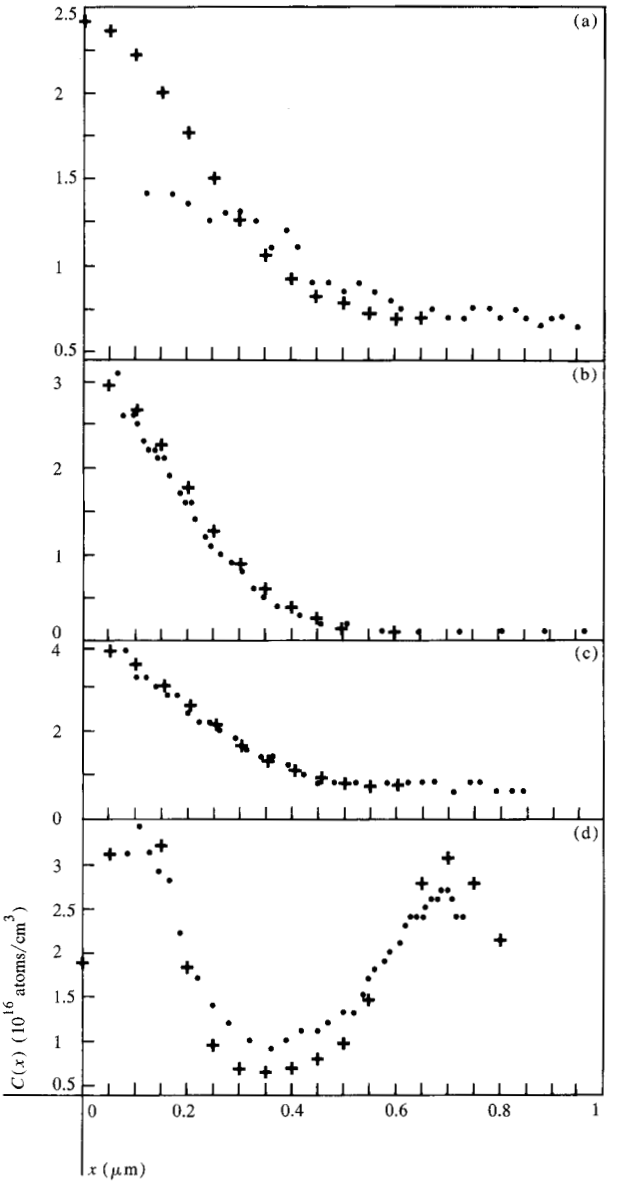

Figure 4 Comparison of measured (○) and simulated (+) profile for (a) Option A, (b) Option B, (c) Option C, and (d) Option D. See Table 1 for details of process steps used in simulation.

was—proof that the physics involved in the simulation was incomplete. Without the comparison of the simulated profile, this might not have been so readily apparent. Moreover, a second phase of experiments was quickly planned to allow for this larger boron loss.

When the substrate sensitivity was calculated for the double implant required for the fourth case, it was found that the depletion width under the switching gate for the correct shallow dose reached the deeper implant for a

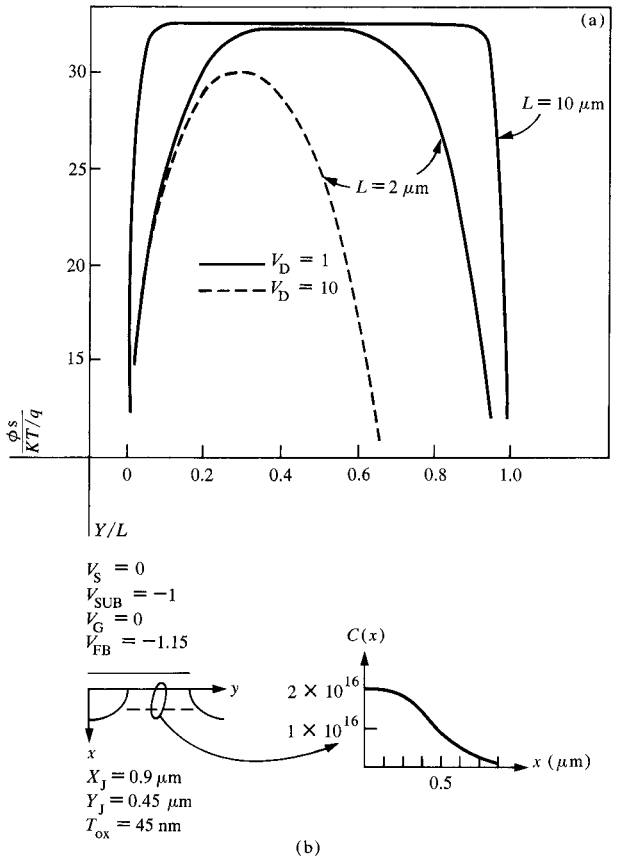

Figure 5 (a) Surface potential distribution illustrating barrier reduction for short channel. (b) Detail of cross section and implant profile used in two-dimensional calculation of surface potential shown in (a). The junction parameters are defined in Fig. 3.

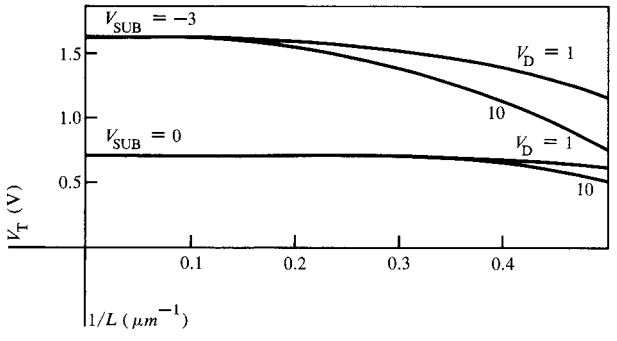

Figure 6 Two-dimensional numerical simulation of the short channel effect on threshold voltage. Figure 5(b) shows the implant profile used.

source-to-substrate voltage of 4 V. Thus, the substrate sensitivity no longer benefits from the lightly doped substrate. Option D was, therefore, discarded.

The remaining three options were all theoretically viable. Each of the final implant profiles was well characterized by a Gaussian curve [21], and the final standard deviation for each was approximately the same because each experienced at least the diffusion caused by the polysilicon oxidation step. Option B had the best overall tolerance control because there was no boron depletion as in A, and the screening film was thinner and better controlled than in C. In both options B and C, however, the implant passes through the gate oxide. Because of the possible long-term deleterious effects, it was felt that option A would provide better long-term stability.

Implantation energy was chosen to place the peak of the implant below the silicon surface by twice the straggle (the initial standard deviation), minimizing the effect of variations in the screening insulator on the fractional dose residing in the silicon. At this energy approximately 98% of the boron ions reach the substrate.

The choice of implantation dose is determined by the threshold desired. An analytical relation between dose and threshold can be written once a valid functional form is found for the implantation profile. A Gaussian function has been found to be an accurate and convenient description for the profile both as it is implanted and as it diffuses during subsequent hot processing. The details of this relation for a long, wide IGFET are given elsewhere [21].

Since the turn-on characteristic of an FET is governed by the potential barrier at the source, any parameter affecting this barrier affects threshold voltage. Figure 5 shows an electron's surface potential plotted from source to drain for two different channel lengths. For $L = 10\ \mu m$ the potential barrier has a long, flat top, and the depletion regions around the source and drain diffusions occupy less than ten percent of the total distance between the metallurgical junctions. In the case of "long channel" FETs the potential barrier is independent of drain voltage, and the relation between potential barrier height and gate voltage is found by solving Poisson's equation in a direction normal to the surface.

For $L = 2\ \mu m$ the flat portion of the barrier occupies a smaller portion of the distance between source and drain. Eventually the top is rounded and the peak is reduced by increasing drain voltage. In the case of "short channel" FETs, the threshold voltage also becomes a function of drain voltage, as well as any other parameter that affects the potential distribution between source and drain, such as the drain diffusion depth and lateral profile.

The consequences of this additional threshold dependence in the "short channel" case can be seen in Fig. 6,

which shows the measured threshold *vs* inverse channel length. As expected, there is more threshold reduction at a given channel length as either the drain or substrate voltage is increased. For a fixed set of voltages a *decrease* from the nominal channel length by a given amount causes a larger change in the threshold than does an *increase* from the nominal by the same amount. Note also that for a 2-μm channel the threshold measured at V_D = 10 and $V_{SUB} = -3$ is only 50 mV greater than the long channel threshold at $V_{SUB} = 0$, while for $L \gtrsim 10\,\mu$m the increase is 900 mV.

The preceding discussion assumes that electron injection from the source to the channel occurred uniformly over the width of the gate. Because of the laterally varying surface potential at the edges of the channel, this assumption is not strictly valid. The change in surface potential from the field to the active gate region cannot occur abruptly, and usually a portion of this transition occurs under the gate. Injection then occurs first in the central portion of the channel because barrier reduction is greatest there. If the surface potential varies laterally over a significant fraction of the total channel width, the resulting channel current is reduced, and the threshold voltage is effectively increased. The effect of device width on threshold for the SAMOS structure is shown in Fig. 7. For larger source-to-substrate biases the transition region extends further under the gate. Comparing Figs. 6 and 7 reveals that for comparable dimensions the narrow channel effect on threshold is less than the short channel effect.

Figure 8 shows the effect of all processing tolerances (and operation over a temperature range of 22 to 85°C) on the substrate sensitivity curve for a particular chip design. Along with the effects of high-temperature operation, channel length variations around the nominal cause an asymmetrical tolerance in the threshold. As mentioned previously, the lower limit in threshold is imposed by the need to minimize subthreshold leakage through the switching FET when a word line is returned to ground. The nominal threshold voltage must be higher by an amount that includes channel length variation. For the channel length tolerance indicated, the threshold variation due to channel length is comparable to its variation due to all other process variations. Since the device in Fig. 8 was intended for operation with a nominal substrate bias of one volt, it is clearly acceptable for use in a dynamic memory cell.

We have seen that a structure designed to be an FET, even when it operates as an FET, has a rather complicated threshold behavior as device dimensions diminish. But does that structure continue to operate as an FET regardless of its size? In particular, are two closely spaced diffusions under the field shield adequately isolated from each other?

In none of the first three options presented above does the boron implant reach a depth of more than approximately half the source-drain junction depth. Because of

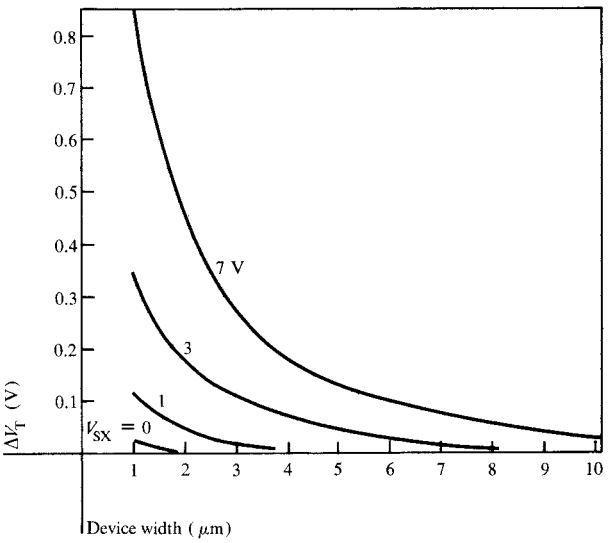

Figure 7 Increase in threshold caused by narrow channel effect as a function of source-to-substrate bias.

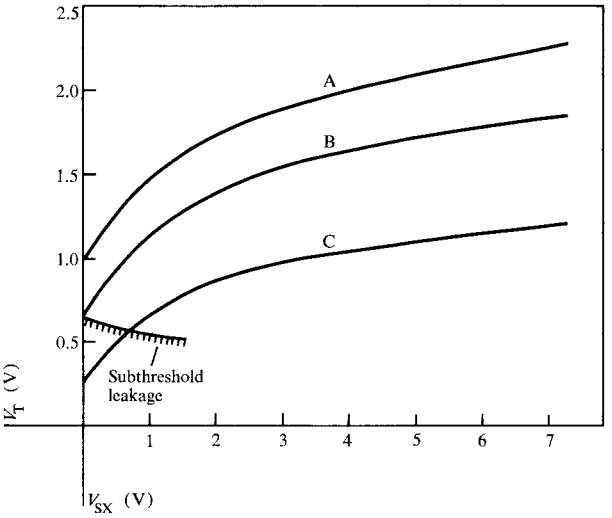

Figure 8 Short channel substrate sensitivity curve for nominal and 3σ worst-case high and low conditions: Curve A, high condition (22°C); Curve B, nominal condition (22°C); Curve C, low condition (85°C). For this particular device, channel length and width are $L = 3.5 \pm 1.0\,\mu$m and $W = 5.0 \pm 1.0\,\mu$m.

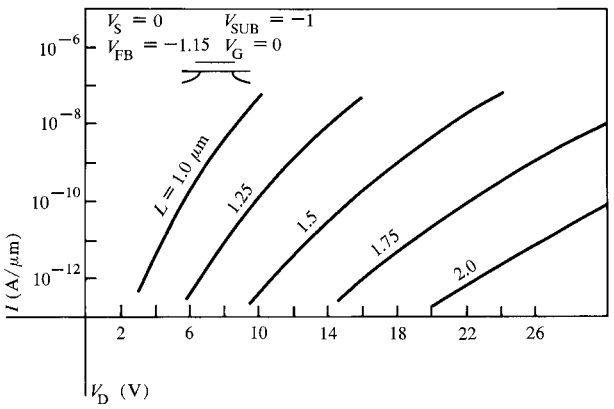

Figure 9 Two-dimensional numerical simulation showing the dependence of low-level punchthrough on diffusion spacing. Relevant bias voltages are shown. The implant profile is the same as that in Fig. 5 with a bulk doping of $C_B = 8 \times 10^{14}$ cm^{-3}. The equivalent oxide thickness of the composite insulator is 45 nm, and the junction parameters $X_J = 0.9$ μm and $Y_J = 0.45$ μm.

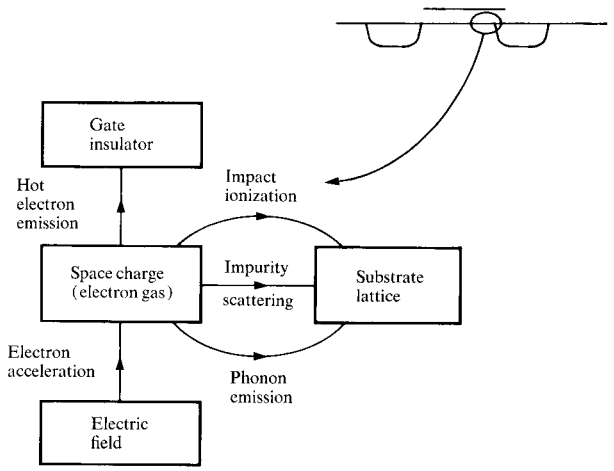

Figure 10 Schematic representation of the energy flow near the drain of an IGFET.

the high-resistivity substrate, an important part of the design was to ensure good punchthrough control. Strictly analytical models that are generally valid for an FET structure do not exist. The problem is inherently two-dimensional and is complicated by the curvature of the source/drain junction and by the nonuniform doping of the substrate. For this reason, a two-dimensional numerical FET model [22] was used to simulate punchthrough behavior.

Because one of the diffusions in the memory cell is a storage node holding less than a few million electrons, punchthrough simulations must be performed down to the thermal leakage level, which is typically one picoampere or lower. Figure 9 gives an example of these calculations for various diffusion spacings. Parameters for this FET gated by the field shield are chosen to represent the worst-case situation for punchthrough, and the spacing L is the distance between adjacent metallurgical junctions as measured at the oxide-silicon interface.

These simulations show, for example, that if the voltage drop between two adjacent diffusions is 16 V with less than 1 pA of current flowing, the spacing must be 1.75 μm. If this distance is reduced to 1.50 μm, the maximum allowed voltage drop is reduced to 10.5 V.

Note the nearly exponential increase of current in Fig. 9. This is characteristic of currents limited by injection over a potential barrier. The increase in slope for decreasing L indicates that the field lines from the drain are more efficient in lowering the barrier as diffusions are moved closer together.

5. Reliability as a design consideration

In addition to providing the required electrical performance at the time a part is first tested, a good design should continue to provide this performance throughout the lifetime of the part. If deterioration mechanisms are well characterized, steps can be taken in the original design to guarantee long-term reliability. One example of a deterioration mechanism is the channel hot electron effect, in which threshold voltage tends to increase with time as electrons accumulate in the gate insulator [23–26].

Figure 10 depicts the source of these hot electrons. Channel electrons are accelerated by the electric field, which increases sharply in the vicinity of the drain. (A second source of hot electrons is the substrate itself. For a given vertical device structure the emission probability of these substrate hot electrons actually decreases at shorter channel lengths. See Ref. [26] for a discussion of substrate hot electrons and their relation to device design.)

At low fields, these electrons lose energy to the lattice by scattering and emitting acoustic phonons. Over the range of field values for which phonon emission is important, energy is also transferred to the substrate by defect and impurity scattering. At larger field values, the electrons gain enough energy to emit optical phonons as well. The electron velocity in the direction of the field then saturates since the lattice can quickly dissipate the locally generated optical phonons. The emitted phonons diffuse through the lattice, which is another way of saying that heat generated near the drain is conducted through the silicon substrate.

At still higher field strengths, the electron gains enough energy so that when it collides with the lattice it can ionize a bound electron and create a hole-electron pair (impact ionization). The secondary electrons are collected at the drain, and the secondary holes are collected in the substrate, generating a substrate current.

When fields are large enough to cause measurable impact ionization, a small fraction of electrons gain enough energy to surmount the potential wall at the Si-SiO_2 interface. These electrons, if they have the proper direction, are thermally emitted into the insulator where they can be trapped. For SAMOS devices the trapping probability for electrons is essentially unity at the oxide-nitride interface of the composite gate insulator. Conceptually, one can view this process as "evaporation" of electrons from the silicon into the SiO_2 where they "condense" at discrete sites. In time the accumulated charge becomes large enough to distort the device characteristics, particularly if the stressed diffusion is subsequently used as a source (since the surface potential barrier is a function of the local charge level).

One way to limit channel hot electron emission is to restrict the electric field near the drain. Figure 11 shows the calculated longitudinal field near the drain for two IGFETs of different channel lengths (but similar in all other respects) biased at the same voltages. For a machine life of many years, fields in excess of approximately 2×10^5 V/cm can pose a problem. Thus, limiting the field implies operating at lower voltages as channel length decreases. Restricting the longitudinal electric field on this basis alone is overly pessimistic, however.

Any consideration of voltage limits should also take into account the effective increase in the emission barrier when $V_D > V_G$ and, especially in the subthreshold region, the level of channel current.

A more general procedure for evaluating the voltage limits imposed by channel hot electrons is to calculate (or measure) threshold shift as a function of time under various stress conditions. Then, specifying the threshold shift that can be tolerated after a given length of time, a locus of terminal voltages can be produced that yields such a shift. These voltages then represent the maximum allowable voltages. Figure 12 depicts such a set of voltages for a 10-mV threshold shift in 3000 hours as a function of channel length. At very low currents, such as in the subthreshold region, the field dependence on channel length is weak. Because the amount of charge emitted into the oxide does depend on current, however, drain voltage must be reduced as subthreshold current increases. At gate voltages greatly exceeding threshold, the current

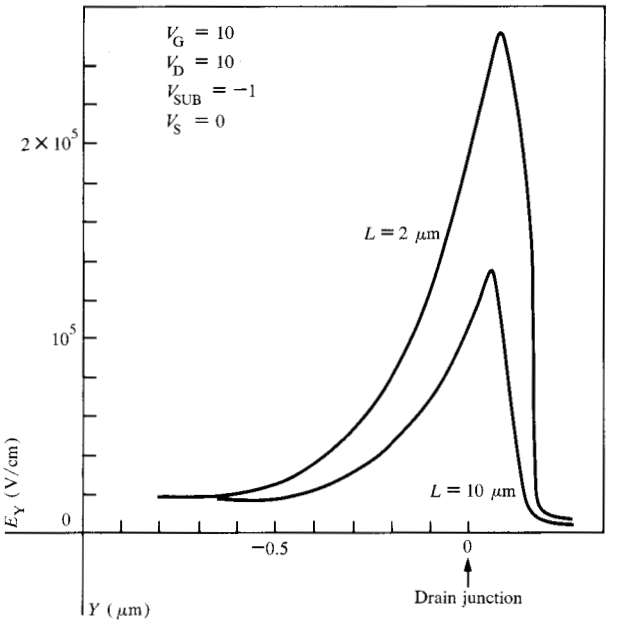

Figure 11 The longitudinal electric field near the drain. Source junction is at $Y = -L$.

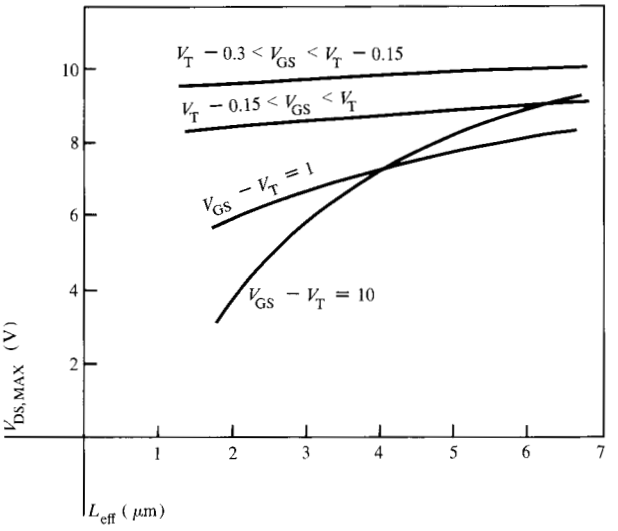

Figure 12 Maximum allowable voltage to limit threshold shift to 10 mV in 3000 hours.

density is large compared to the background doping, and the increased space charge near the drain introduces a channel length dependence—as was depicted by the increase in peak electric field shown in Fig. 11.

The crossover behavior in the curves of Fig. 12 is a consequence of this increased space charge at short chan-

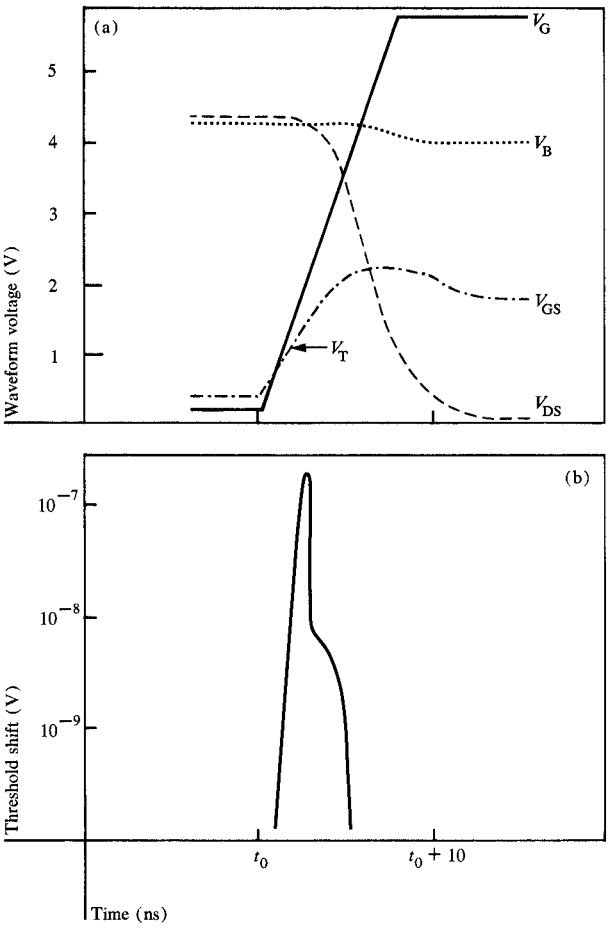

Figure 13 (a) Simulated voltage waveforms for the charging of an advanced one-device memory cell; (b) the resulting incremental hot electron threshold shift during the first charging cycle.

nels. Because of velocity saturation the electrons contribute to the effective space charge near the drain, and the longitudinal electric field increases monotonically with increasing gate voltage. At long channels much higher drain voltages are required to produce velocity saturation. An increase in gate voltage now reduces the longitudinal electric field as the device is driven into the linear region of operation. Thus, low drive conditions result in the lowest allowable drain-to-source voltage V_{DS}. For lower gate voltages a larger V_{DS} is allowed because channel current is greatly reduced, while for higher gate voltages a larger V_{DS} is allowed because the longitudinal heating field is reduced. The exact location of a crossover depends on the threshold shift judged to be tolerable.

Such curves are useful for initial, rough estimates, but they can be difficult to apply to actual circuit designs. Since different circuits can tolerate different amounts of threshold shift before their performance is affected, many sets of curves similar to Fig. 12 are often needed. Also, because emission current exhibits a sensitive dependence on terminal voltages, the equivalent time-invariant voltage stress is often difficult to estimate.

For actual chip designs, therefore, an FET degradation model is used to simulate circuit performance after any length of operating time, say 50 000 hours. This degradation model uses the channel hot electron emission and insulator trapping characteristics pertinent to a given FET technology to predict threshold shift as a function of time.

As an example, the pertinent waveforms from simulations of an advanced dynamic memory cell design are shown in Fig. 13(a). The bit line has been charged through another FET so its value is a threshold voltage below the maximum level of 5.8 V. The effective gate-to-source and drain-to-source voltages are shown along with the word line and bit line waveforms. Charging of the word line begins at t_0, and threshold is reached approximately 2 ns later, after which V_{DS} begins to fall.

The corresponding incremental threshold shift is shown in Fig. 13(b). It peaks at approximately $(t_0 + 2.6)$ ns, then rapidly diminishes since V_{DS} is falling. The total hot electron threshold shift during a single charging operation of the storage node capacitor is given by the area under the curve of Fig. 13(b) and amounts to 0.28 μV. Extrapolation to multiple cycles can be done once the correct time dependence (generally sublinear) is known.

By using the degradation model, a given circuit design can be tested against its unique performance goal. Should the simulation indicate that its performance degrades by an intolerable amount before the end of its expected life, design modifications can be made and tested by further simulations. Often a slight increase of channel length on selected devices is all that is required to ensure that performance goals continue to be met over the operating life of the chip. A detailed discussion of the channel hot electron degradation model is given elsewhere [27].

6. Summary and conclusions

Because of its inherent small size and operational simplicity, the dynamic one-device cell will continue to be an important memory element for many years. How successfully it is utilized in the VLSI era depends partly on how well fundamental device phenomena are appreciated and accounted for in future designs.

This paper has discussed how device physics guided the development of SAMOS during both the conceptual

formulation and the development of the technology. Threshold design has served to illustrate the relation between device and process modeling and how both contribute to a timely and accurate design. Discussion of the channel hot electron effect illustrated one way that long-term reliability was incorporated into the design at an early stage.

Further reduction in the size of the one-device cell requires 1) circuit operation at lower voltage levels to limit the electric field at the drain, 2) tighter processing tolerances to maximize the available signal at the reduced voltage level, 3) increasing amounts of device and process modeling of small-dimensional effects, and 4) extensive electrical and physical characterization to optimize design tradeoffs within a given technology.

The device phenomena selected for discussion in this paper are fundamental to small devices, and their relation to dynamic one-device memory design will continue to be important. In addition, they have been, and will continue to be, applicable to static memory and to dynamic memory other than the one-device cell. Not only will they affect future designs and future technologies by their limitations, but, in all likelihood, increased understanding of these device phenomena will lead to new solutions to apparent limitations and even to entirely new concepts in device physics.

Acknowledgment

The author wishes to acknowledge K. M. Cogley, W. P. Noble, and R. R. Seymour for their help in the preparation of this paper.

References

1. R. R. Troutman, "Subthreshold Design Considerations for Insulated Gate Field-Effect Transistors," *1973 IEEE ISSCC Digest of Technical Papers* **16,** 108–109 (1973). See also *IEEE J. Solid-State Circuits* **SC-9,** 55–60 (1974).
2. R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of Ion-Implanted MOSFET's with Very Small Dimensions," *IEEE J. Solid-State Circuits* **SC-9,** 256–268 (1974).
3. H. S. Lee, "An Analysis of the Threshold Voltage for Short Channel IGFET's," *Solid-State Electron.* **16,** 1407–1417 (1973).
4. K. E. Kroell, "Geometry Effects on Field Effect Transistors," presented at the 1973 European Semiconductor Device Research Conference, Munich, Germany.
5. L. D. Yau, "A Simple Theory to Predict the Threshold Voltage of Short-Channel IGFET's," *Solid-State Electron.* **17,** 1059–1063 (1974).
6. K. O. Jeppson, "Influence of Channel Width on the Threshold Voltage Modulation in MOSFET's," *Electron Lett.* **11,** 297–299 (1975).
7. R. W. Swanson and J. D. Meindl, "Fundamental Performance Limits of MOS Integrated Circuits," *1975 IEEE ISSCC Digest of Technical Papers* **18,** 110–111 (1975).
8. K. Kroell and G. A. Ackerman, "Threshold Voltage of Narrow Channel Field Effect Transistors," *Solid-State Electron.* **19,** 77–81 (1975).
9. W. P. Noble and P. E. Cottrell, "Narrow Channel Effects in Insulated Gate Field Effect Transistors," *IEDM Tech. Digest*, 582–586 (1976).
10. R. R. Troutman and A. G. Fortino, "Simple Model for Threshold Voltage in a Short-Channel IGFET," *IEEE Trans. Electron Devices* **ED-24,** 1266–1268 (1977).
11. F. H. Gaensslen, "Geometry Effects of Small MOSFET Devices," *IEDM Tech. Digest,* 512–515 (1977); see also F. H. Gaensslen, "Geometry Effects of Small MOSFET Devices," *IBM J. Res. Develop.* **23,** 682–688 (1979).
12. W. M. Smith, Jr., "Field-Effect Transistor Integrated Circuit and Memory," U.S. Patent 3,811,076, 1974.
13. R. R. Garnache and W. M. Smith, Jr., "Integrated Circuit Fabrication Process," U.S. Patent 3,841,926, 1974.
14. Richard A. Larsen, "A Silicon and Aluminum Dynamic Memory Technology," *IBM J. Res. Develop.* **24,** 268–282 (1980, this issue).
15. R. R. Troutman, "VLSI Limitations from Drain-Induced Barrier-Lowering," *IEEE Trans. Electron Devices* **ED-26,** 461–469 (1979).
16. J. J. Barnes, K. Shimohigashi, and R. W. Dutton, "Short Channel MOSFET's in the Punchthrough Current Mode," *IEEE Trans. Electron Devices* **ED-26,** 446–453 (1979).
17. V. L. Rideout, F. H. Gaensslen, and A. LeBlanc, "Device Design Considerations for Ion Implanted n-Channel MOSFETs," *IBM J. Res. Develop.* **19,** 50–59 (1975).
18. W. K. Hoffman and H. L. Kalter, "An 8 Kbit Random Access Memory Chip Using the One-Device FET Cell," *IEEE J. Solid-State Circuits* **SC-8,** 298–305 (1973).
19. R. R. DeSimone, N. M. Donofrio, B. L. Flur, R. H. Kruggel, H. H. Leung, and R. Schnadt, "FET RAM's," *1979 IEEE ISSCC Digest of Technical Papers* **22,** 154–155 (1979).
20. A. G. Fortino and H. J. Geipel, "Process Design Procedure for IGFET Thresholds," presented at the 8th Semiconductor Interface Specialists Conference, Miami, FL, December, 1977.
21. R. R. Troutman, "Ion-Implanted Threshold Tailoring for Insulated Gate Field-Effect Transistors," *IEEE Trans. Electron Devices* **ED-24,** 182–192 (1977).
22. D. P. Kennedy and P. C. Murley, "Steady State Mathematical Theory for the Insulated Gate Field Effect Transistor," *IBM J. Res. Develop.* **17,** 2–12 (1973).
23. S. A. Abbas and R. C. Dockerty, "Hot Carrier Instability in IGFET's," *Appl. Phys. Lett.* **27,** 147–148 (1975).
24. S. A. Abbas and R. C. Dockerty, "N-Channel IGFET Design Limitations Due to Hot Electron Trapping," *IEDM Tech. Digest*, 35–38 (1975).
25. T. H. Ning, C. M. Osburn, and H. N. Yu, "Effect of Electron Trapping on IGFET Characteristics," *J. Electron. Mater.* **6,** 65–76 (1977).
26. P. E. Cottrell, R. R. Troutman, and T. H. Ning, "Hot Electron Emission in N-Channel IGFET's," *IEEE Trans. Electron Devices* **ED-26,** 520–533 (1979).
27. R. R. Troutman, T. V. Harroun, P. E. Cottrell, and S. Chakravarti, "Hot Electron Design Considerations for High Density RAM Chips," *IEEE Trans. Electron Devices* **ED-27** (1980), accepted for publication.

Received June 2, 1979; revised November 5, 1979

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