General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

INVESTIGATION OF NEW CONCEPTS OF ADAPTIVE DEVICES

by

H. A. R. Wegener

INTERIM SCIENTIFIC REPORT
Contract No. NAS 12–570
September 1968

Prepared for
ELECTRONICS RESEARCH CENTER
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
CAMBRIDGE, MASSACHUSETTS

by

SPERRY RAND RESEARCH CENTER
SUDBURY, MASSACHUSETTS 01776

Dr. John Hopkins
Technical Monitor
NAS 12–570
Electronics Research Center
575 Technology Square
Cambridge, Massachusetts 02139

Requests for copies of this report
should be referred to:

NASA Scientific and Technical Information Facility

P.O. Box 33
College Park
Maryland 20740

SRRC–CR–68–43

INVESTIGATION OF NEW CONCEPTS OF ADAPTIVE DEVICES

by

H. A. R. Wegener

INTERIM SCIENTIFIC REPORT
Contract No. NAS 12–570
September 1968

Prepared for

ELECTRONICS RESEARCH CENTER
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
CAMBRIDGE, MASSACHUSETTS

by

SPERRY RAND RESEARCH CENTER
SUDBURY, MASSACHUSETTS 01776

ACKNOWLEDGEMENT

This report was prepared by H. A. R. Wegener. However, there were many individual contributors to different aspects of the work. Each contributor is identified below according to the major section(s) to which he contributed.

Section	Title	Contributors
IIC	Background	H. A. R. Wegener, A. J. Lincoln, H. C. Pao, R. Oleksiak, M. O'Connell
IIIB	Exploration of Device Characteristics	H. A. R. Wegener, H. C. Pao, M. O'Connell
IIIC	Exploration of the Memory Mechanism	
IIID	Theory of Memory Device (and Appendix)	H. A. R. Wegener
IIIE	Experimental Confirmation of Theory	F. A. Sewell
IIIF	Equivalent Circuit and Its Application	E. Lewis
IIIG	Memory Device Fabrication	M. B. Doig, H. A. R. Wegener
IIIH	"Student-001," An Adaptive Circuit Using the MNS-VTT	M. Cohn, A. J. Lincoln

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (cont.)

LIST OF ILLUSTRATIONS (cont.)

LIST OF ILLUSTRATIONS (cont.)

INVESTIGATION OF NEW CONCEPTS OF ADAPTIVE DEVICES

by

H. A. Richard Wegener
Sperry Rand Research Center

SECTION I

SUMMARY

This report describes the work done during the first year of this contract, which deals with a new memory device developed at the Sperry Rand Research Center. This memory device is an IGFET whose threshold voltage can be increased or decreased reproducibly by applied voltage pulses. The amount of change depends on the amplitude and duration of the pulse, and the direction of the change depends on the polarity of the pulse. This property makes this transistor the first semiconductor device that permits electrically alterable, nonvolatile storage of information.

The investigation of this device has led to the establishment of typical operating parameters of the storage function: roughly a $\pm$ 5 V threshold change for writing voltage pulses of the order of $\pm$ 30 to $\pm$ 60 V, typical write-in times of the order of 10^{-2} to 10^{-4} seconds (with 10^{-6} seconds possible), and storage times of the order of several months observed (and many years storage expected from extrapolations of data).

The theory of operation is based on the model of a dielectric that has two regions of different conductivity. The conductivity of one of these regions must be highly nonlinear with respect to applied field. This requirement makes silicon nitride highly useful in this structure. This proposed model has been tested in its most important features; it successfully explained the experimental data obtained in special structures built for this purpose.

Based on this model, the equivalent circuit of the storage element has been determined and applied as an analog computer. This analog circuit has been successful in measuring details of current transients during charging and in the location of the charges stored on the dielectric of the memory element.

In a separate study, different means for forming the memory elements were studied. These results are also reported.

Finally a description of the memory element in an adaptive circuit, a "Learning Machine" called "Student-001", is presented.

In all, this work confirms the initial expectation that the new memory element would be a unique nonvolatile, linearly adaptive, digital storage element that, moreover, can be incorporated in large-scale integrated circuits.

It is recommended that future work be done on the factors that influence the current density and field relationships of the thin film dielectrics used in the gate of this device. The properties of the surfaces occurring in this multilayer structure will also require further attention. The determination and control of the distribution of charges in the dielectric is another important area of future study. Finally, the use of the new devices as a sensor of light is recommended.

SECTION II

INTRODUCTION AND BACKGROUND

A. STATEMENT OF WORK

The contract work statement reads as follows. The contractor shall supply the necessary personnel, facilities, services, and materials to accomplish the work set forth below:

Item 1 - Investigate and determine the feasibility of a proposed new adaptive device, which is to permit switching between two or more sets of electrical characteristics, and to develop such devices in a form compatible with semiconductor devices and integrated circuits. While two-terminal devices are acceptable at present, the possibility of expansion to three-terminal (active) structures is desirable (e.g., devices switchable between two or more voltage-current characteristics). Constraints on the means of switching between modes are simply that it be consistent with simple circuitry; plus the basic consideration that device applications are enhanced to the degree that switching can be accomplished in a short time with low power. No power should be consumed in maintaining any of the possible states of the device. These states should be stable over as wide a range of current and voltage as possible. It is also desirable that device operation be relatively unchanged over a substantial frequency spectrum.

Although operation in the range 300^{o} - 400^{o}K is ultimately desired, early feasibility studies at other, more convenient temperatures may also be useful.

Item 2 - Construct, test, and characterize such devices, particularly in terms of impedance and transfer functions (as appropriate), current-voltage characteristics, and mode-switching parameters—time, current, etc. Both theoretical and experimental studies of temperature sensitivity, radiation tolerance, and noise behavior should be conducted.

Item 3 - Relate device characteristics to potential circuit applications, indicating those areas in which the adaptive device should be most competitive.

Item 4 - Predict the ultimate practical and theoretical parameters and limitations of the devices, indicating possible compromises between different characteristics. This is to include consideration of optimum fabrication processes.

B. CHOICE OF DEVICE

At the outset of this contract a number of devices had been reported which had characteristics suggestive of their use as adpative devices. However,

they all possessed limitations which barred them from consideration for real electronic systems. Three representative examples are the memistor, the ferroelectric field-effect device (FEFED) and the flexode.

The memistor makes use of an electrochemical plating process to alter the resistance of a conductive bar. The device is slow, poorly suited to circuit integration, and deteriorates with time.

The FEFED consists of a ferroelectric layer over a semiconductor, where remanent surface polarization of the ferroelectric material modulates the channel depth in the semiconductor and thereby controls the two-terminal resistive properties. Due to the refractory nature of known ferroelectric materials, reported efforts to make FEFED's require that the devices be structured with the ferroelectric material as the substrate and that poorly controlled semiconductors, such as tellurium, be deposited thereon. These FEFED's exhibited fairly rapid decay of stored information at normal temperatures ($300^{o}K$), and destructive decomposition at temperatures not far above this. Other drawbacks of the FEFED are its lack of suitability for eventual three-terminal operation and a fabrication process which differs markedly from standard processes used in the manufacture of integrated circuits.

The flexode utilizes impurity ion drift to control junction parameters. This drift is obtained by the application of an electric field. Although the construction seems suited to manufacture by processes similar to those used for standard integrated circuits, the time constant associated with the control element is long and would allow only limited use in applications of interest.

A new type of adaptive device had been developed concurrently at the Sperry Rand Research Center as an outgrowth of a program on the development of a silicon nitride-silicon planar device technology. This new device, which we designated the MNS variable threshold transistor (MNS-VTT), offered numerous advantages in performance, fabrication ease, and general utility over any device previously considered.

The MNS-VTT is an insulated-gate field-effect transistor fabricated by silicon nitride planar techniques. The distinguishing feature of the device is that it possesses an electrically alterable threshold voltage. The alteration is achieved by the application of a voltage pulse of short duration to the transistor gate dielectric, a specially prepared silicon nitride layer. In effect, the alteration in threshold voltage changes the electrical operating point of the device in a controllable way.

Our data indicated that the altered threshold value persisted for times of the order of months, both under normal electrical operating conditions and at elevated temperatures. Neither electrostatic potential nor power expenditure was required to maintain the stored information. Both positive and negative shifts in threshold voltage could be reversibly obtained within fractions of a second, and switching between extreme threshold values for several million cycles had been obtained without device degradation.

The device had also been used for both digital and analog circuits. Thus, the MNS-VTT was the first nonvolatile semiconductor memory element that

could be incorporated on the same chip with other semiconductor devices. As such, it was recognized as an extremely important addition to the catalog of semiconductor devices and integrated circuit techniques.

The characteristics of the new MNS-VTT developed at the Sperry Rand Research Center indicated quite clearly that it was superior in electrical performance to all previously described devices intended for adaptive applications. Its importance in the digital field was potentially even greater. It was therefore natural that all efforts under this contract should be concentrated on further advancing the understanding and the technology of the MNS-VTT. The starting point of this work, the state of knowledge at the beginning of the contract, is described in the next section.

C. BACKGROUND

1. MNS Capacitor Behavior

By way of introduction, it will be convenient to consider the voltage behavior of a metal-nitride-silicon (MNS) capacitor structure. The MNS structure is shown in Fig. 1. A qualitative picture of how the capacitance varies with the bias applied to the metal electrode may be had by considering the behavior of the majority and minority carriers. For simplicity, assume that the semiconductor is p-type silicon. If a large negative bias is applied to the metal field plate, holes will be attracted to the silicon surface, which, with the large accumulation of majority carriers, will then behave much like a metal. The capacitance measured is that of the insulator layer alone. Now if a small positive bias is applied, holes are repelled, forming a region at the silicon surface that is depleted of majority carriers. The effective width of the dielectric will now be increased by this depletion region, and the measured capacitance will begin to decrease. A further increase in positive bias will further deplete the surface region and cause an additional decrease in capacitance. However, at some large positive bias there will be an appreciable accumulation of minority carriers at the interface. At this point the depletion region width will approach a maximum, and electrons will begin to form an inversion region at the interface. If the lifetime of the minority carriers is such that they can follow the applied frequency, the capacitance will rise, eventually approaching the original capacitance due to the insulator only. If the minority carriers cannot follow the frequency of the applied signal, then the capacitance will approach a minimum value and finally become independent of voltage. This behavior is shown in Fig. 1(b). In practice, it was found that MNS capacitors could be made which conformed to the ideal capacitance-voltage (C-V) behavior. However, under a special set of processing conditions, an unusual hysteresis effect could be obtained.

For these MNS capacitors, with their special processing history, we found that successive traces on a C-V plotter showed the curve displaced from immediately preceding ones. Holding the voltage on the capacitor at either the positive or the negative extremes of the voltage ramp for some time resulted in maximum displacements of the C-V plot. These plots were quite reminiscent of polarization and other instabilities that had been reported for MOS structures. By further investigation and control of the process technology, we were able to produce MNS structures that displayed controllable amounts of hysteresis.

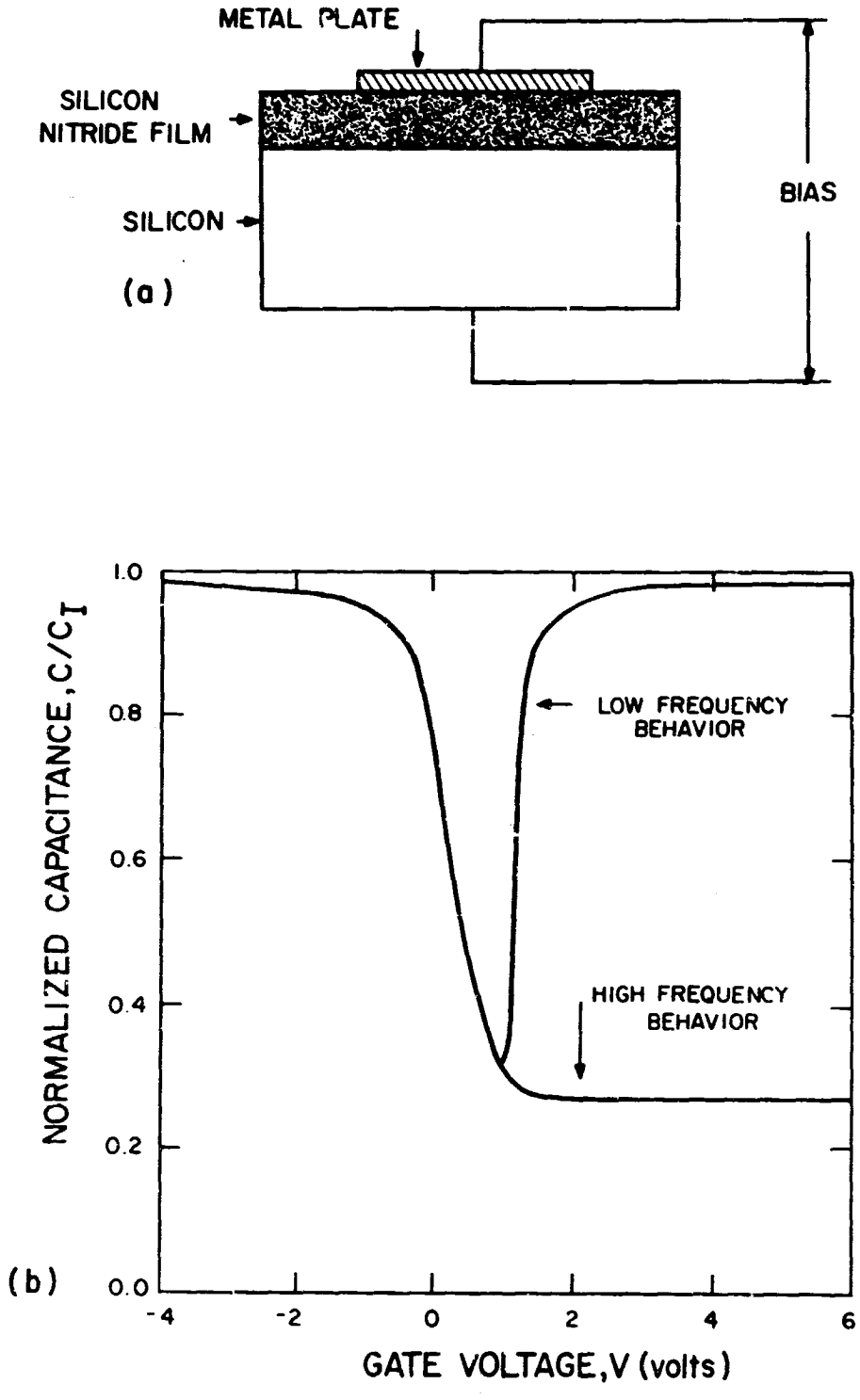


FIG. 1 (a) Metal-insulator-silicon system.
(b) Theoretical C-V curves.

In Fig. 2, we show the X-Y plot of three such MNS structures, all using 10 Ω-cm n-type silicon, but processed in different ways. It is clear that the MNS capacitors in Fig. 2(a) show no hysteresis at all. This device is therefore suitable when stable characteristics are desirable. In Fig. 2(b), we show the intermediate case of a C-V plot of another MNS capacitor which shows some small amount of hysteresis. Finally, in Fig. 2(c), we find a large hysteresis curve indicating the potential sensitivity of the effect. Such hysteresis is the basic characteristic of a device with long-term information storage capabilities. As such, it is similar to the B vs H curve of a magnetic material or the polarization curve of a ferroelectric. It is clear that in the MNS capacitors shown in Figs. 2(b) and (c) we have memory devices that can be operated electrostatically.

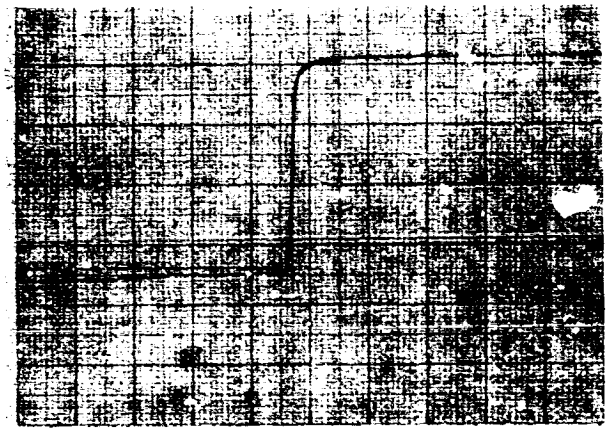
The C-V hysteresis plots can be divided into two types: In the first category are MNS structures which show shifts in characteristics which are monotonically related to the difference between the applied voltage and the inversion voltage that the device happens to have at the moment. In the second type of structure, shifts in the C-V plot only occur when some threshold level of voltage has been exceeded. The first characteristic is most interesting if the device is to be used in analog applications. However, its sensitivity to noise and small shifts in signal power level make it of dubious value for digital applications. The second type of characteristic is clearly of greatest importance in the digital field. In our prior work the digital-type device has been most extensively investigated. In the following section, we describe in some detail the results that have been observed.

2. MNS Capacitor Storage Phenomena

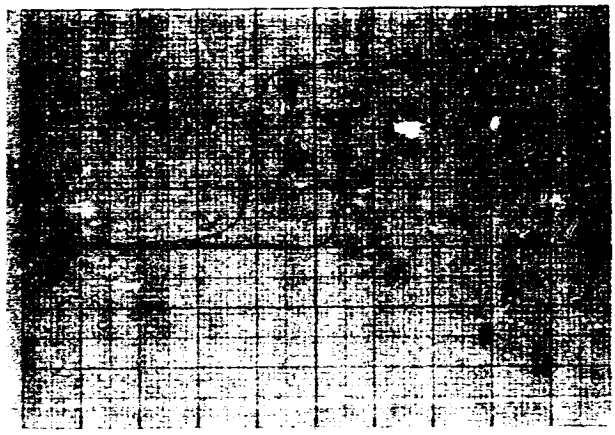
These capacitors were fabricated on n-type silicon wafers with resistivities of either 3 or 10 Ω-cm. The slices were first cleaned in organic and inorganic solvents. Then they were chemically polished in a mixture of hydrofluoric and nitric acid. A layer of silicon nitride approximately 2000 Å thick was deposited pyrolytically on each of the cleaned wafers. Aluminum was then evaporated through a metal mask with 10 mil diameter openings to form the field plates. The back contact was formed by another aluminum evaporation. The capacitors were often diced and mounted on headers, but in most cases the measurements were made on the full slices by means of a probe.

The initial study of these structures utilized voltage ramps or steps of indefinite duration to produce data similar to that shown in Fig. 2. However, it was obvious that more detailed information could be obtained through the use of voltage pulses whose amplitude, duration, and number could be varied. Accordingly, C-V characteristics were measured on an X-Y recorder following a controlled pulse treatment. A typical set of measurements is shown in Fig. 3. As a measure of the effects, the variables of the pulse were plotted against the "inversion voltage" of the MNS capacitors. Physically, this is the voltage at which minority carriers begin to dominate the charge density at the insulator-silicon interface; in a C-V plot it is the voltage at which the curve starts to saturate at the minimum capacitance level.

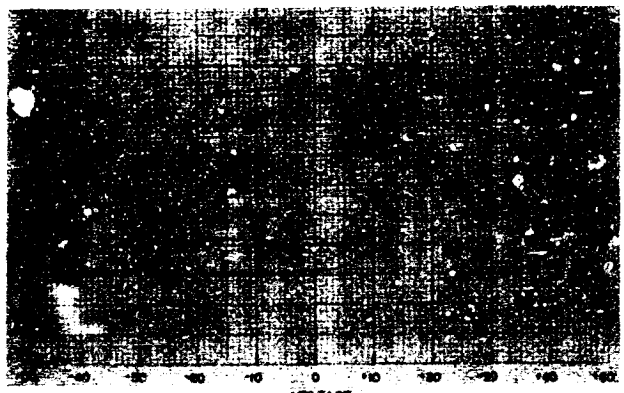
In Fig. 4(a) a plot of pulse duration against the change of the inversion voltage is given. The parameter is the voltage amplitude of the positive



(a)



(b)



(c)

FIG. 2 C-V plots of MNS capacitors produced by holding ramp voltages at +50 V and -50 V for one minute before plotting.
(a) Stable MNS capacitor: no hysteresis.
(b) Variable MNS capacitor: some hysteresis.
(c) Variable MNS capacitor: large hysteresis.

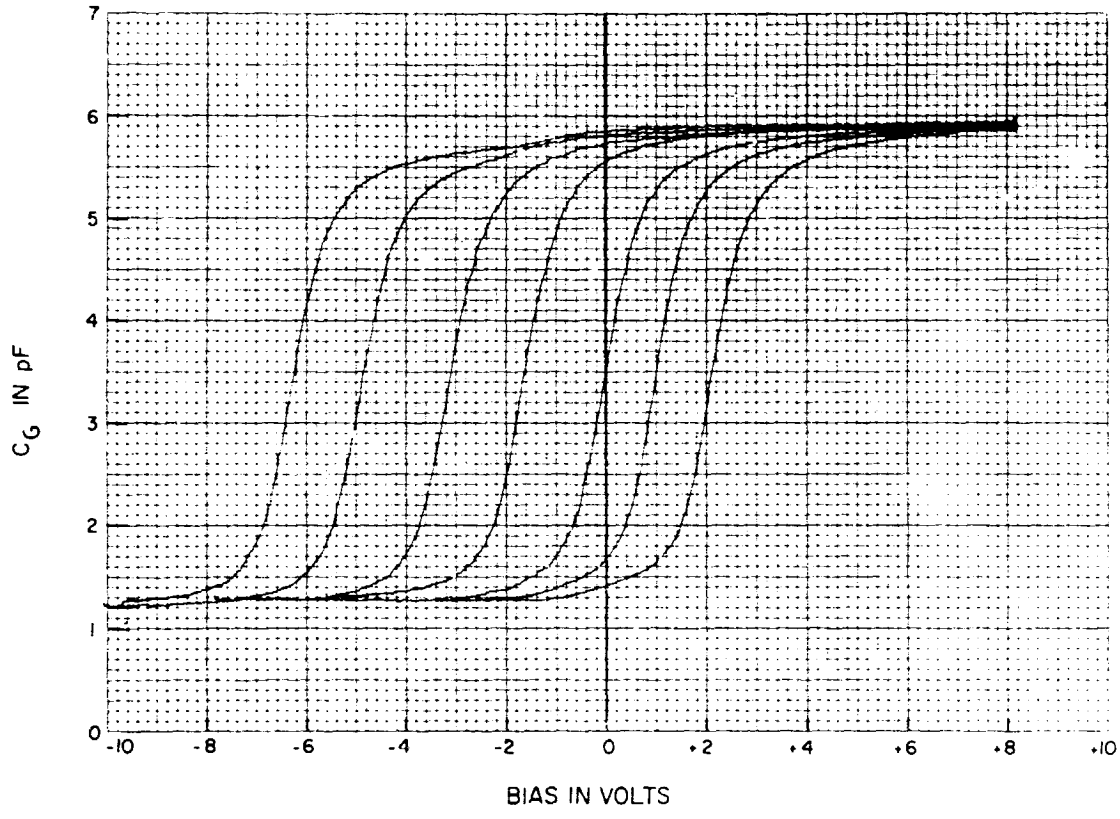
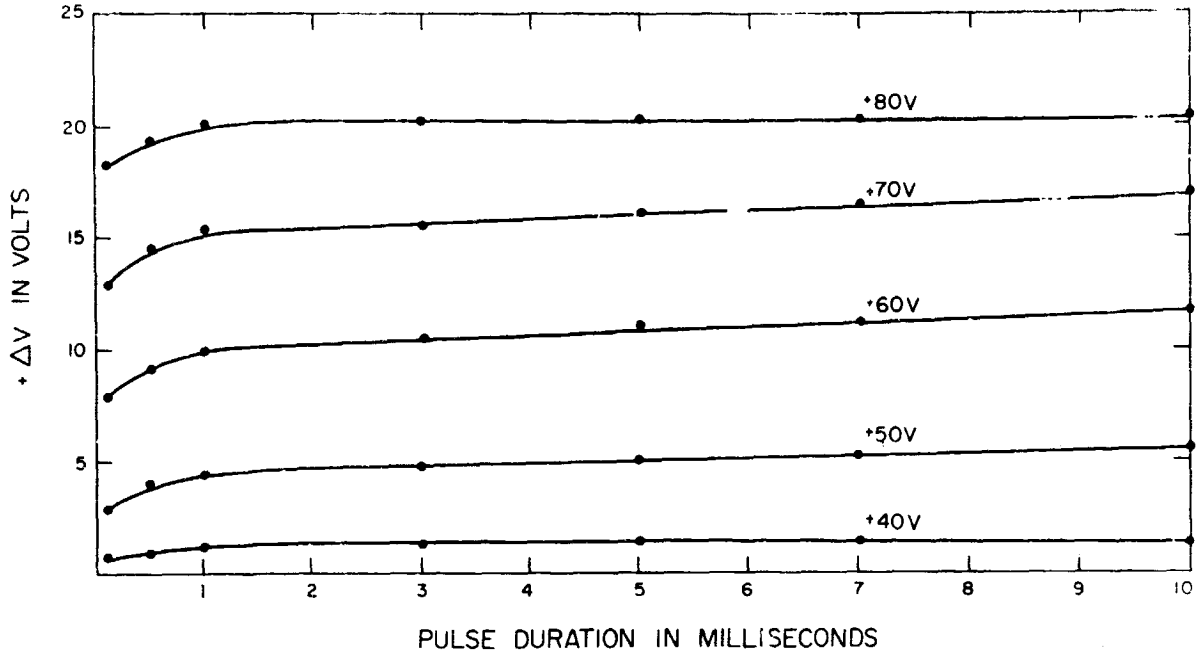


FIG. 3 Typical changes in C-V plots recorded alternately with applied pulses.



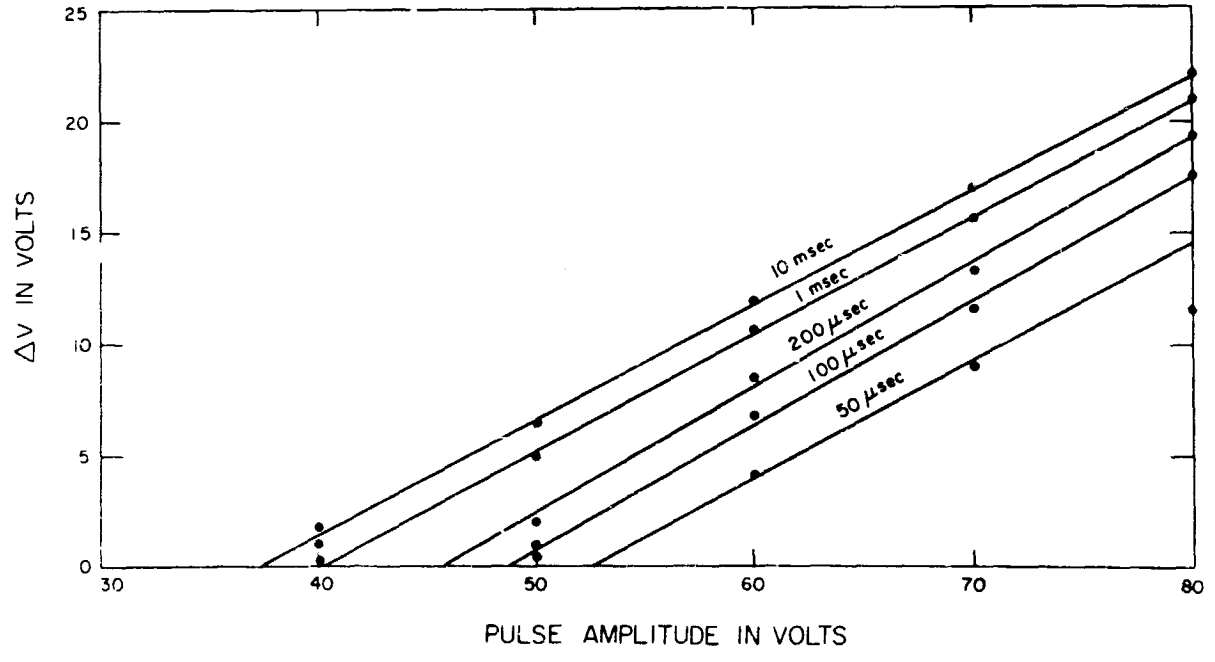


FIG. 4 (a) Change in inversion voltage ΔV of MNS capacitor as a function of positive pulse amplitude and length.
(b) Change in inversion voltage ΔV of MNS capacitor as a function of the width and the amplitude of positive pulses.

pulse used. It is apparent that the positive pulse has caused a shift towards a more positive inversion voltage. For the n-type semiconductor this means that the inversion voltage has become less negative. It is also clear that the shift has reached a saturation level after about 1 msec pulse duration. Essentially, no change was induced by pulses of less than +40 V, so that a threshold level for shifts appears to lie somewhere between +30 and +40 V. By replotting these data, as in Fig. 4(b), the threshold pulse value necessary for any voltage shift in the capacitor is brought out more clearly.

The nearly linear increase in inversion voltage with voltage pulse amplitude indicates a potential analog behavior for this form of device. Another form of analog behavior is shown in Fig. 5. Here the inversion voltage change (ΔV) is plotted against the number of positive pulses applied. The result is a monotonic increase in ΔV until a saturation value is reached. The data of Figs. 4 and 5 were obtained with positive pulses. Similar data were obtained using negative voltage pulses, except that the shift in inversion voltage was in the negative direction.

For any type of storage or adaptive use, the longevity of information storage is of concern. To study this effect, a capacitor was shifted to saturation by four pulses of 1 msec duration at -120 V amplitude. The value of the inversion voltage was then measured as a function of time. Typical results are shown in Fig. 6. In this particular instance, the voltage decreased from about -16 V to a value of -14 V over the period of about a day.

All the data are representative of the behavior of capacitors from a given slice. The basic characteristics differed only from slice to slice in a way which depended on the details of slice processing. The data obtained with our MNS capacitors indicate strongly that this device is useful even now as a memory element that stores information in a nonvolatile manner, that is, removal of electrical power will not result in loss of information impressed on the capacitor. However, the simplicity greater flexibility and utility of three-terminal active devices prompted our decision to transfer our technique of making MNS capacitor memory units to the fabrication of MNS variable threshold voltage transistors (MNS-VTT).

3. MNS-VTT's: The Variable Threshold Transistors

A silicon nitride insulated-gate field-effect transistor (IGFET) can be regarded as an interacting combination of an MNS capacitor and two reverse biased p-n junctions. In particular, the same phenomena which determine the inversion voltage of the capacitor also determine the turn-on or threshold voltage of the transistor. It was therefore expected that the hysteresis and storage effects observed in the MNS capacitors would also occur in the MNS transistors. This was indeed found to be the case.

The transistor structure employed was exactly the same as that used in the fabrication of our standard fixed threshold (i.e., stable) transistors. Typical dimensions and characteristics are given in Fig. 7. The device is a fairly large p-channel enhancement transistor with a transconductance near 1500 μmho and a threshold voltage of -5 to -6 V. The threshold voltage

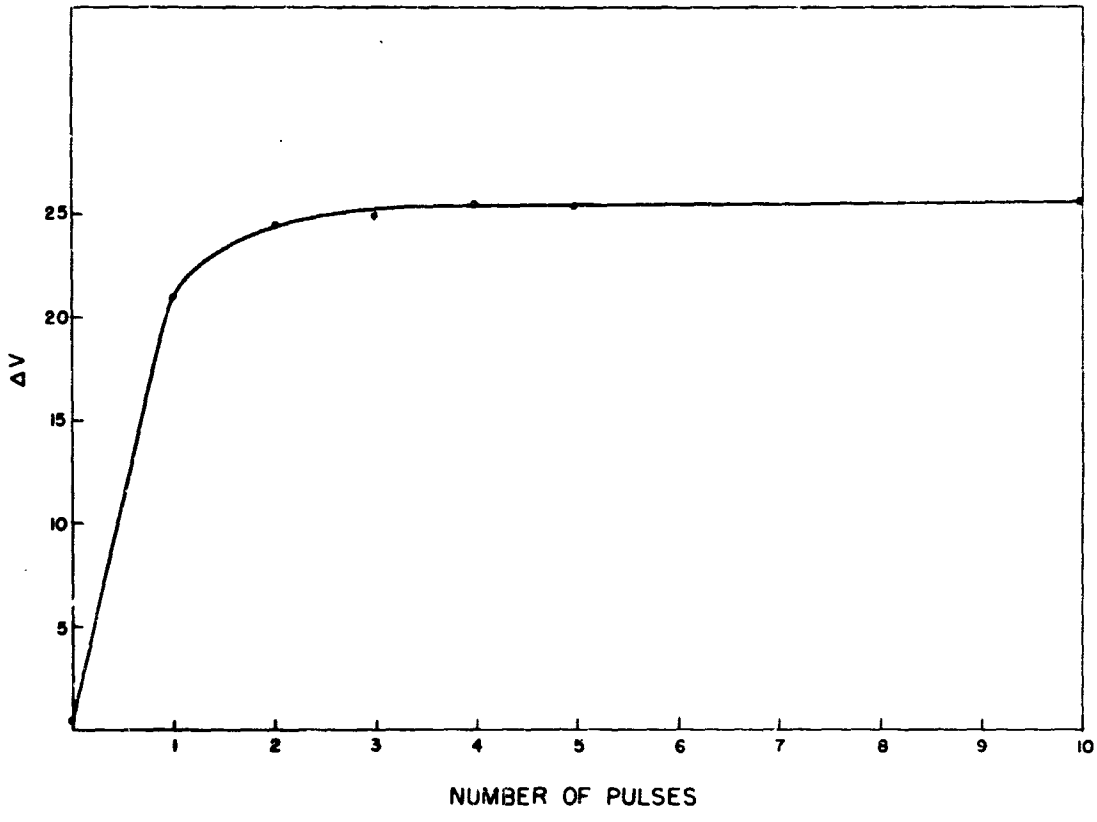


FIG. 5 Change in inversion voltage ΔV of MNS capacitor with the number of positive 10 msec pulses applied.

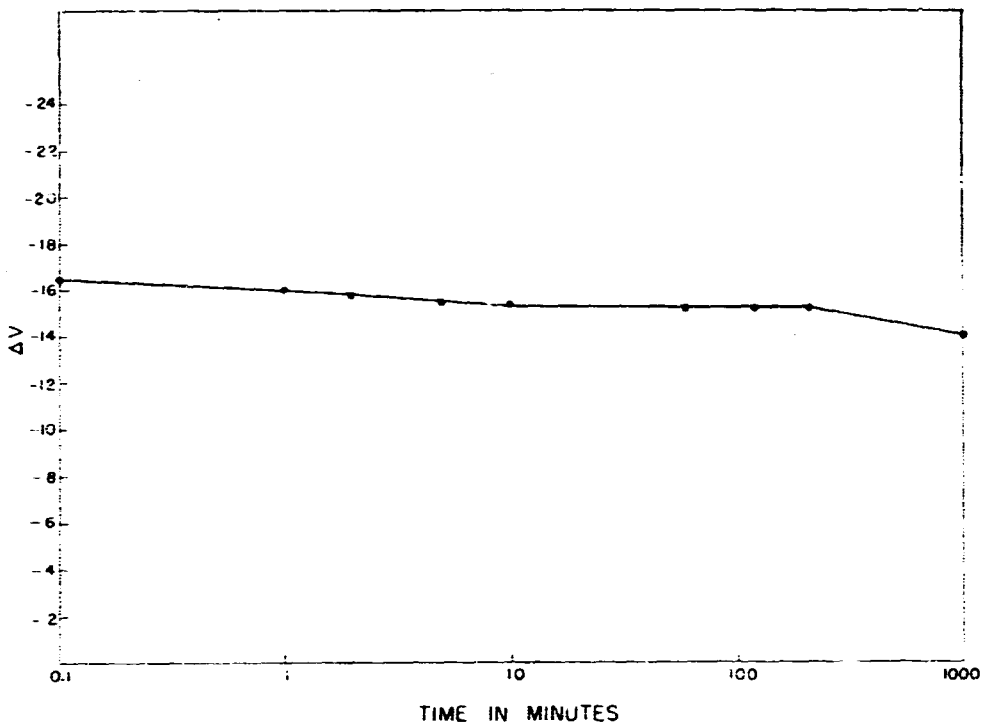


FIG. 6 Persistence of stored information in MNS capacitor dielectric as measured in V_T vs time.

Electrical Characteristics

Symbol	Characteristic	Value			Units	Conditions
		Min.	Typical	Max.		
V_T	Turn-on voltage	-3	-4	-5	V	$V_{DS} = V_{GS}$ $I_D = 10$ μA
G_m	Transconductance	1300	1500	1800	μmhos	$V_{DS} = V_{GS}$ $I_D = 10$ mA
BV_{DSS}	Breakdown voltage	70	72	75	V	$V_{GS} = 0$ $I_D = 10$ μA
I_{GSO}	Gate leakage			100	nA	$V_{DS} = 0$ $V_{GS} = -20$ V
I_{GDO}	Gate leakage			100	nA	$V_{GS} = 0$ $V_{DS} = -20$ V
I_{DSS}	Junction leakage			100	nA	$V_{GS} = 0$ $V_{DS} = -20$ V
V_{on}	Gate voltage	15	17	20	V	$V_{DS} = V_{GS}$ $I_D = 10$ mA

FIG. 7 (a) Standard MNS transistor.
(b) Device specifications of standard MNS transistor.

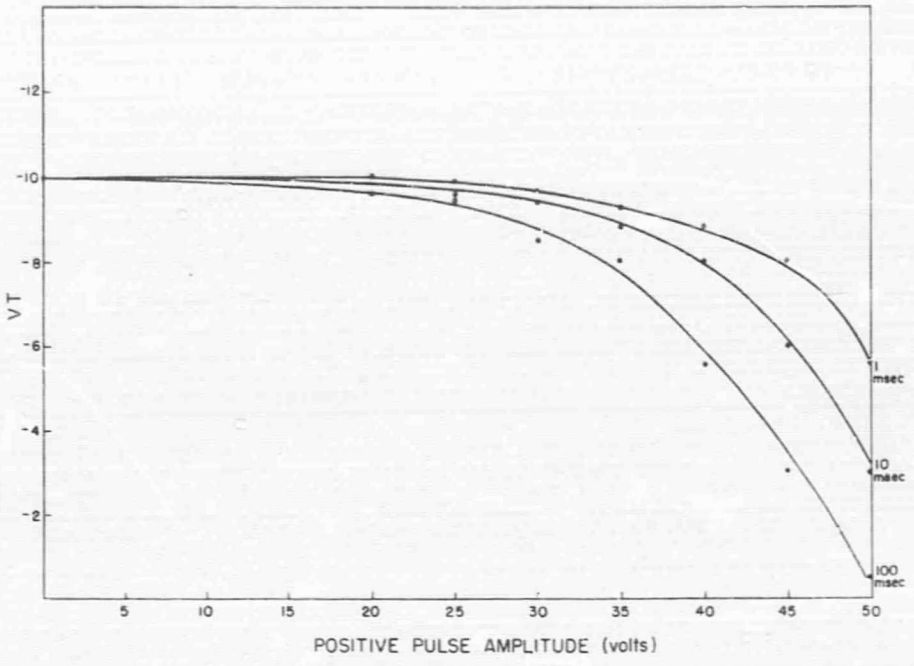
characteristics of the transistors were investigated as a function of various pulse conditions. The threshold voltage (V_T) was measured either by a C-V plot of the gate-to-substrate MNS capacitor, or, alternatively, by the turn-on characteristic of the transistor with the gate electrode shorted to the drain lead. The value of V_T was defined as the voltage at the minimum capacitance in the C-V plot or, more or less equivalently, the voltage at which the drain current had reached a value of 10 μA.

A plot of voltage pulse amplitude vs induced threshold voltage, V_T, is given in (a) and (b) of Fig. 8. In Fig. 8(a), the positive amplitude was varied between 0 and +50 V. The positive pulses were applied after the threshold voltage of the transistor had been set at -10 V by a negative pulse before each measurement. It is clear from the data that the minimum pulse level for a change in V_T is near +10 to +20 V. A saturation in the shift occurred for a pulse length near 100 msec. In Fig. 8(b) the results of similar experiments with negative pulses are given. Here, resetting was done with a standard positive pulse before each negative one. The minimum voltage for the change in this direction appears to be near -5 to -10 V. A plot of V_T vs the number of 1 msec pulses of +50 V amplitude is shown in Fig. 9. The device had been set for a large negative threshold voltage by a pulse of -50 V.

Preliminary life tests were taken in order to check how well the device would stand up under numerous switchings from a high to a low V_T. For this purpose, a squarewave generator was connected to the gate of a VTT device and a stripchart recorder was connected in series with the drain of the VTT. A squarewave voltage of ± 50 V and about 50 Hz (≈ 10 msec pulse length) was employed. A double-pole, double-throw switch was actuated every five minutes to connect in the drain and fixed gate supplies and switch out the pulse generator. The stripchart recorder then recorded for 30 sec the drain current for a large fixed bias on the gate. This current depended on the difference between the last pulse-impressed threshold voltage and the applied gate voltage. After the current measuring cycle, the switch reconnected the squarewave generator and cut out the gate and drain bias supplies. In this way, over a twenty-four hour period maximum and minimum threshold voltages were impressed on the VTT under test. It was found that, in general, the transistors changed slightly for the first hour and thereafter maintained a constant mode of switching. This assures that at a more advanced level of fabrication and testing technology these devices will have long life.

Another set of tests performed concerned the persistence of the state impressed on the gate. For this purpose several transistors were switched to an extreme state by 100 msec pulses of either +50 V or -50 V. They were then stored at room temperature or at 125°C with either zero bias or a -5 V bias on the gate. (These two bias states would be typical of those experienced by a digital memory device.) Periodic tests were made to ascertain the time dependence of the threshold voltage of the VTT. In Fig. 10, some of these data are summarized. After a change of less than 1 V during the first two hours, the threshold voltages changed only by minute amounts over the next 120 to 140 hours. There was no clearcut indication that either bias or the 125° temperature, or the combination, had any effect on the persistence of the impressed threshold voltage. From these data, it could be extrapolated that the digital information stored in these devices would last for a least 100 days.

(a)



(b)

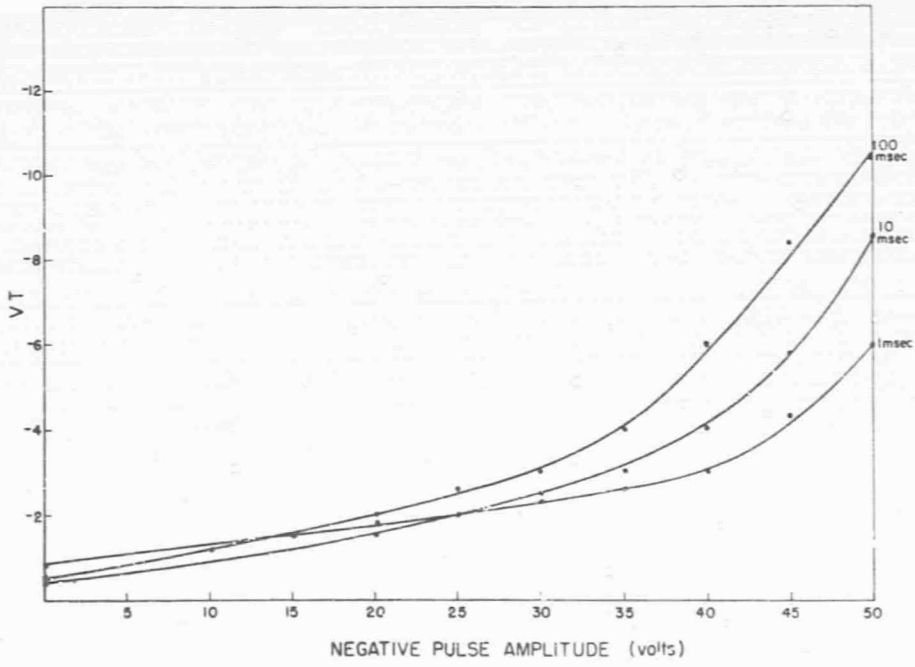


FIG. 8 (a) Threshold voltage V_T of MNS-VTT as a function of positive pulse amplitude and length.
(b) Threshold voltage V_T of MNS-VTT as a function of negative pulse amplitude and length.

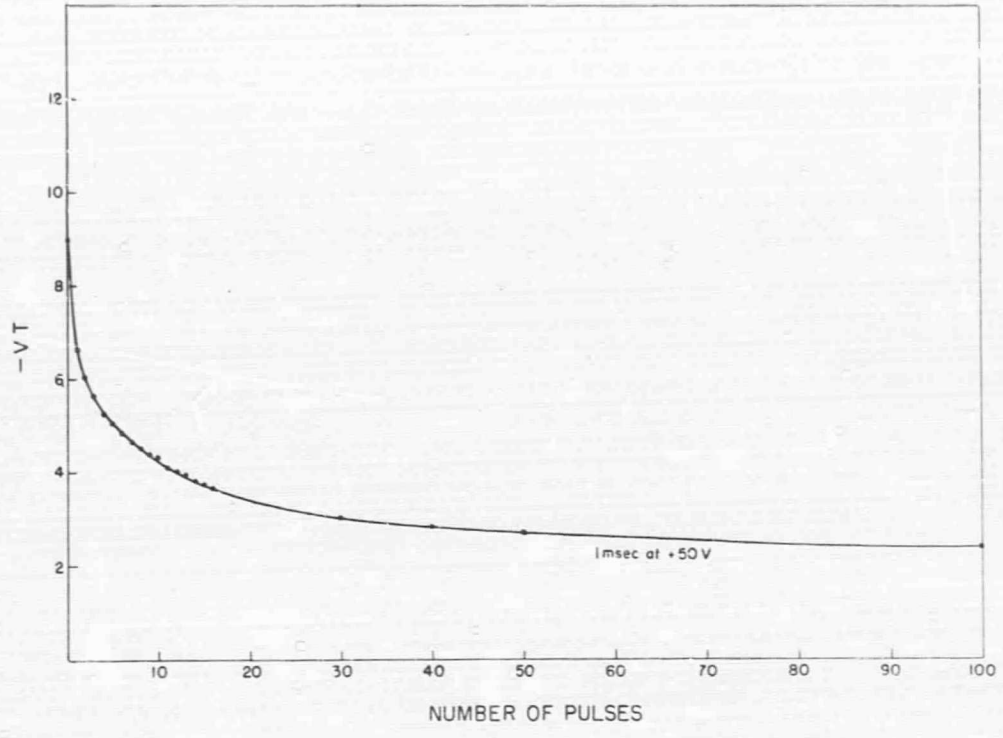


FIG. 9 Threshold voltage V_T of MNS-VTT as a function of the number of positive pulses of +50 V amplitude and 1 msec length.

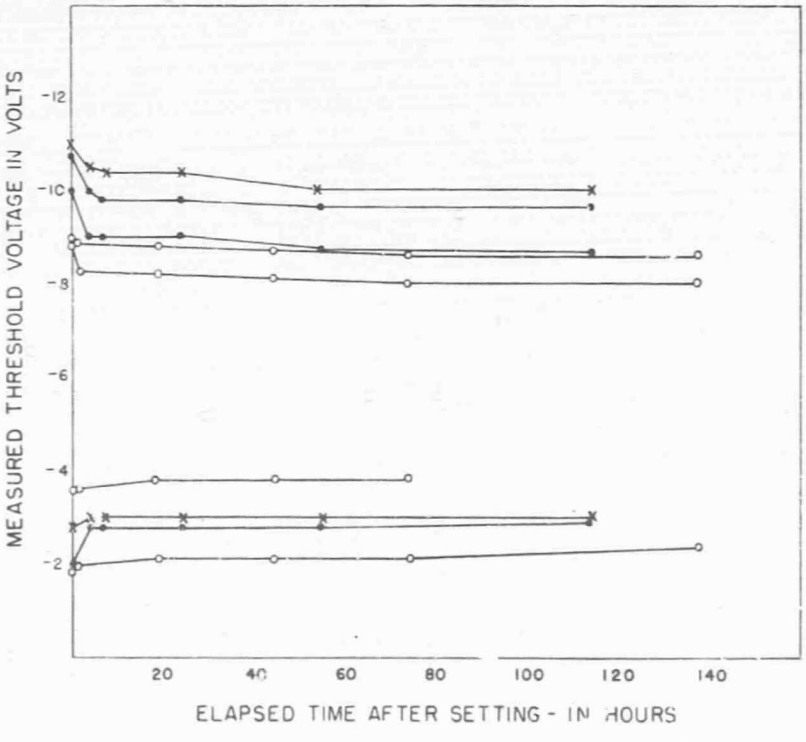


FIG. 10 Persistence of stored information: o 25°C, no bias on gate; x +125°C, no bias on gate; • +125°C, constant -5V bias on gate.

4. Prototype Circuit Use

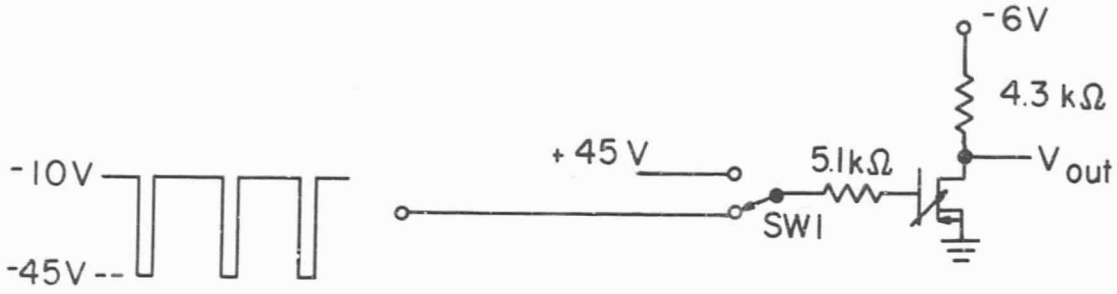
The MNS-VTT had been used both in incrementally adaptive (analog) and digital circuit configurations. In the former application, we demonstrated the ability to change the threshold in small steps in order to change the level of the output voltage in the circuit, as shown in Fig. 11(a). The threshold is first driven toward 0 V by the application to the gate of a +45 V step. The gate is then connected to a negative, high voltage pulse train which drives the threshold in the negative direction. A -10 V bias is maintained at the gate. This causes the device to have initially a low resistance compared to the 4.3 kΩ load resistor. As the threshold is driven toward -10 V, the resistance increases until the device no longer conducts. This occurs when the threshold voltage is more negative than the applied -10 V gate bias. The output voltage waveform shown in Fig. 11(b) was obtained with a 20 pps, 2 μsec, -45 V pulse train. Clearly, the process is reversible, and with the application of a positive pulse train the sequence of output characteristics will be reversed.

Another illustration of the MNS-VTT operating in an analog mode is provided in Fig. 12. These curves show the I-V characteristics of the device in the linear region of operation. Each curve in the family was obtained by pulsing the gate threshold voltage to a different level and then sweeping the source-to-drain voltage over a range while measuring source-drain current. Throughout all of the sweeps the gate-to-source voltage was maintained at a constant bias level. The device, operating in this manner, may be used as an adaptive two-terminal element.

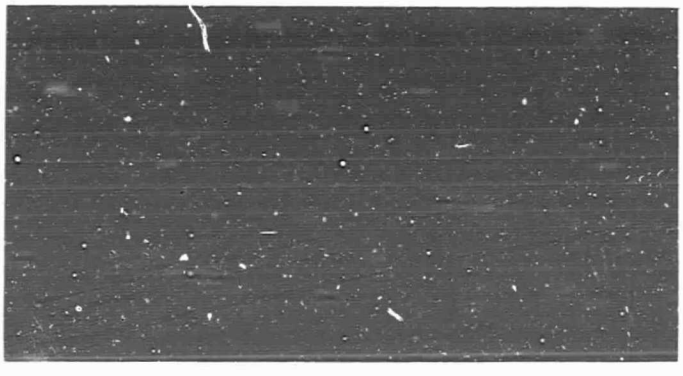
In a digital application the MNS-VTT has been used as the storage element in an electrically-alterable nondestructive read-out storage system. In this application, the bit in storage is one or zero, depending on whether the threshold has been set above or below some reference level. Operation may be either as a two- or three-terminal device.

A breadboard four-bit storage element using this principle has been constructed at SRRC. The circuit is shown in Fig. 13. In the setting mode, all switches are closed to the left and the appropriate setting voltages are applied to the control gates. The switches are then returned to the right and the appropriate cell is interrogated by applying a negative pulse to its I input. The output point is common, so that the information appearing at the output at the time a particular cell is interrogated indicates its contents.

An oscilloscope photograph of the four-bit memory immediately after being set to the pattern 0101 is shown in Fig. 14(a). Figure 14(b) shows the output of the memory after this pattern had been stored for eight hours without reset at 25°C. During this time the memory was continuously interrogated at a 200 kHz rate. The memory was then put into a temperature-power cycle in which it was elevated to 125°C with 200 kHz interrogation for a period of eight hours, after which power was removed and the temperature was lowered to 25°C for a period of 16 hours. This cycle was repeated three times, still without resetting the stored information. The output at the end of this period is shown in Fig. 14(c). It is clear from these results that undegraded storage has been obtained over an extended period of time.



(a)



(b)

FIG. 11 (a) Circuit used to demonstrate incremental threshold variation.
(b) Output voltage waveform with incremental threshold variation.
Vertical scale: 2 V/div.; horizontal scale: 0.2 sec/div.

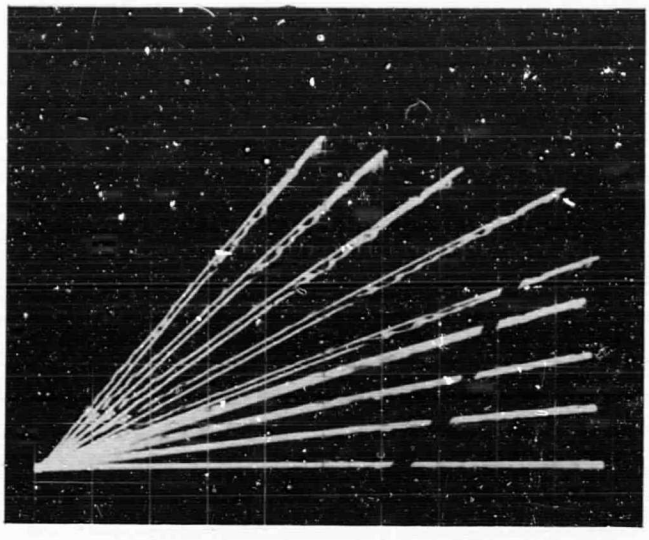


FIG. 12 I_D vs V_{DS} at different thresholds.
Vertical scale: 0.01 mA/div.;
horizontal scale: 0.02 V/div.

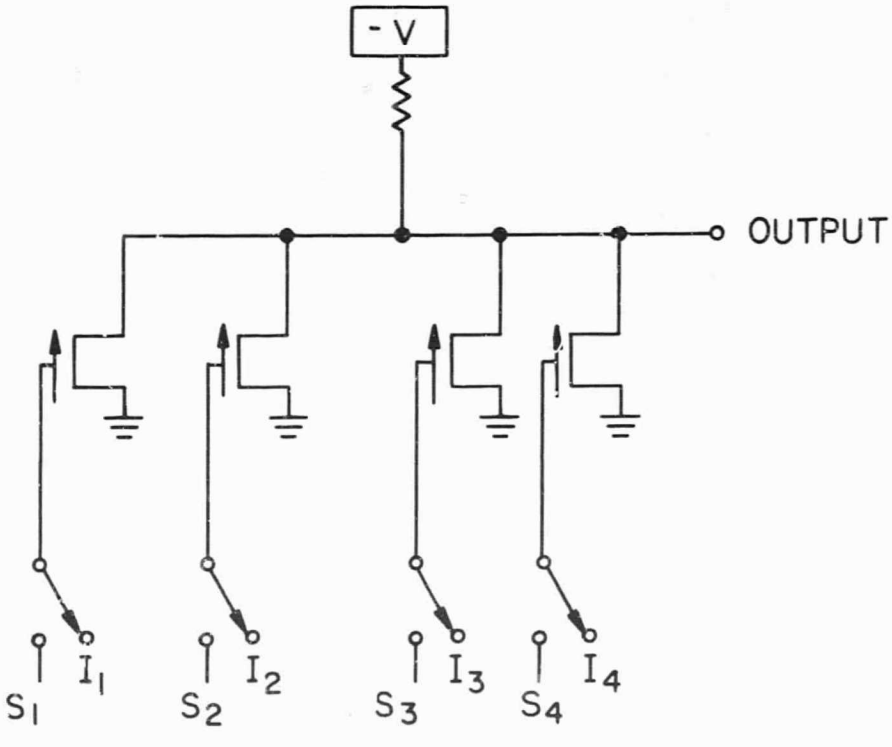
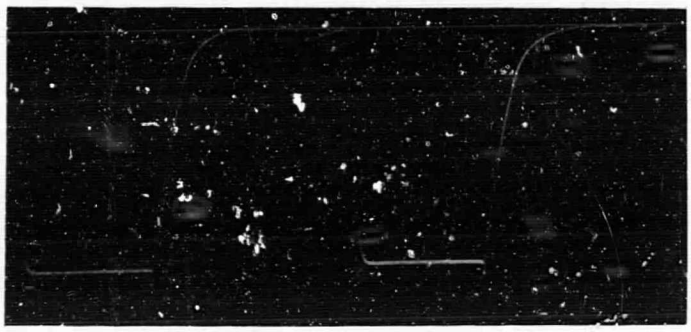
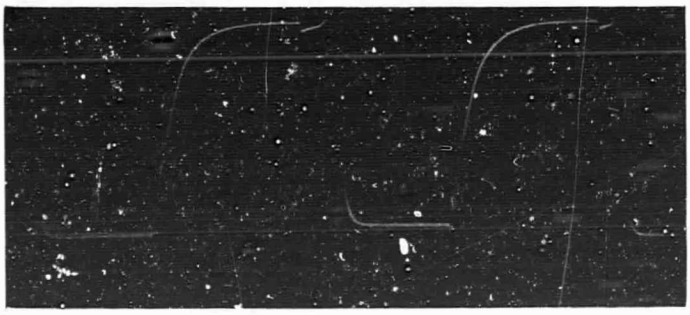


FIG. 13 Circuit of four-bit digital storage unit.

(a)



(b)



(c)

FIG. 14 Output waveforms of digital storage for stored pattern 0101 (see text). Vertical scale: 2V/div.; horizontal scale: 2 μsec/div.

5. Physical Model

The change in the C-V characteristic found in the MNS capacitor structure is clearly due to the change in the charge distribution between its two plates. The characteristic suffers a displacement along the voltage axis only, with no distortion in its shape. Therefore, it can be concluded that surface states on the silicon are not taking part in this change. In silicon oxide (MOS) structures, changes of this sort have been found in connection with the motion of ionic charges such as Na^+ or H^+. The rapidity of the changes in the MNS capacitor makes this unlikely. In phosphorus-glass/oxide structures the cause of change could have been the polarization of $PO_4^- Na^+$ dipoles. However, no combinations of such ionic species exist in silicon nitride, nor are they likely to in this non-ionic material. Finally, the persistence of the polarization in silicon oxide is very short compared to the persistence found in silicon nitride. It is therefore probable that the effects arise from charge distributions resulting from a gain or loss of electrons or holes in the silicon nitride.

In an attempt to determine the location of such charges, a step-etching experiment was conducted in the following way. An MNS capacitor was pulsed into a state of high inversion voltage. The inversion voltage was then successively measured by a probe after small increments of the nitride were removed by etching. In this way, a plot of surface charge density vs thickness revealed that the charges giving rise to the enhanced inversion voltage must be located within about 300 Å of the silicon nitride-silicon interface (Fig. 15).

Since the inversion voltage increased after the application of a negative pulse, either holes were injected from the silicon into the silicon nitride, or electrons were emitted from the nitride into the silicon. The reverse presumably occurred with reversed pulse polarities. Some barrier must be involved also, since the effects were somewhat asymmetric.

The I-V characteristics of MNS capacitors were also studied. They indicated an initial ohmic branch, which at higher voltages developed into a linear relationship between the logarithm of the current and the square root of the voltage. It was not clear whether Schottky emission from the silicon into the insulator or a Poole-Frenkel emission from insulator traps into the silicon was the predominant current mechanism. Both mechanisms result in the same general shape of the I-V relationship (Fig. 16).

There was also no obvious relationship between the trapped charges and the steady state current-voltage characteristic. In this connection there was an observation that appeared significant. It was found that in plotting current vs time across an MNS structure that there was always a higher value at the beginning than at steady state. Thus, there is a good possibility that the charges causing the threshold voltage shifts are due to an initial surge of holes or electrons.

Summing up our physical model, we are certain that there are charged traps within 300 Å of the silicon nitride-silicon interface and that these traps could be charged, discharged, or even produced by the application of a voltage across the MNS capacitor. These traps must be fairly deep energetically in order

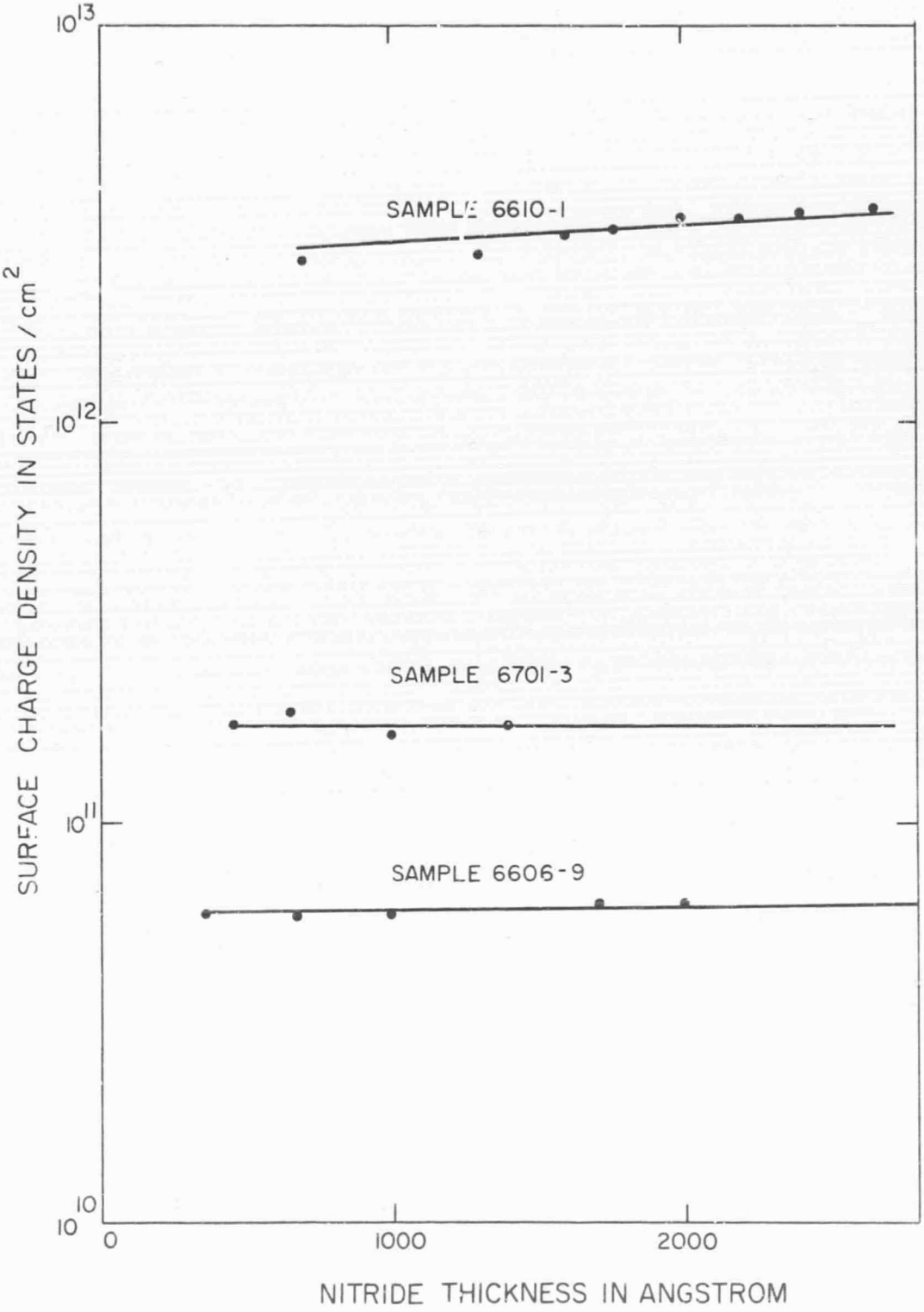


FIG. 15 Plots of surface charge density N_{SS} vs thickness of nitride, as obtained by step-etching experiments.

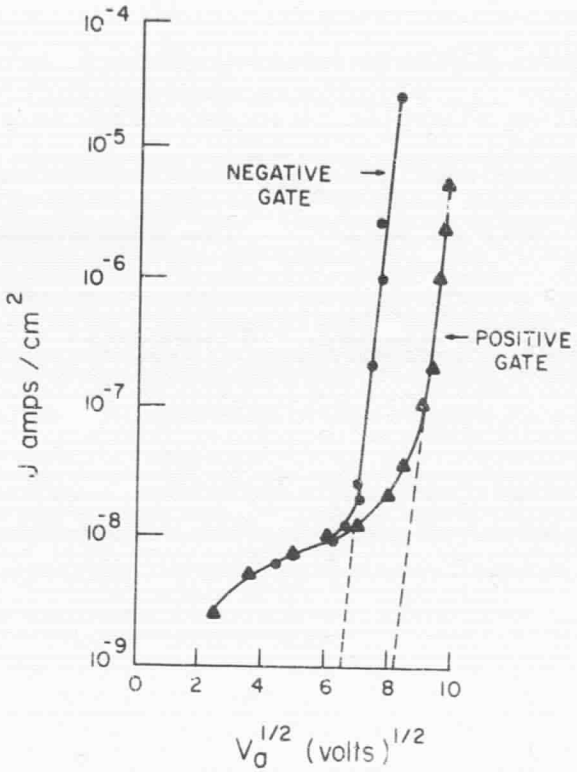


FIG. 16 Current vs voltage plot taken from MNS capacitor. Note the linear portion of log J vs $V^{\frac{1}{2}}$.

to explain the observed long persistence of the added charges at 125°C. Finally, in view of the small distances involved, tunneling may play an important role. Elucidation of the exact mechanism of the high field effects is clearly an important element of a comprehensive device development program.

6. Summary

All these data were interpreted as clear proof that this new device would be useful as a storage element in nonvolatile memories or in various analog applications. The data are representative of hundreds of transistors taken from different slices. The information is impressed on this new storage element by a voltage pulse of the order of tens of volts from a high impedance source. The information can persist for many days under operating bias conditions and at temperatures up to at least 125°C. Life tests have indicated unchanged operation for several million switching cycles. Thus, if one were to use the MNS-VTT in a stored program computer, reprogramming one hundred times a day would permit the use of even the present devices for a minimum time of the order of 200 years.

An important future consideration is that these devices are compatible with integrated circuits, since they are structurally the same elements of which stable insulated-gate field-effect transistor circuits are composed.

SECTION III

DISCUSSION

A. INTRODUCTION

All the data given in the previous section were obtained prior to the award of the contract. They indicated three clear directions in which further work would have to proceed. One was the continued investigation of device characteristics to permit a full empirical description and evaluation of its capabilities. A second direction was the design of experiments to explore those aspects of its operation that would permit a device theory to be established. The third was the theoretical work leading to a physical model of device operation. These lines of attack were pressed in actuality, and their results are given in the first three sections following.

The theoretical model devised in the course of these studies led to areas of work beyond those envisioned at the beginning of the contract. One of them was a detailed experimental proof of the theoretical model, the other a use of the electronic analog of the equivalent circuit of the new device for the refined exploration of its properties. Two other sections that summarize the work performed under this contract deal with the process details of fabricating the devices and with the application of the electrical device properties for the operation of a "Learning Machine."

B. EXPLORATION OF DEVICE CHARACTERISTICS

1. Introduction

The operational feature that makes this IGFET an adaptive device is the possibility of making the threshold voltage larger or smaller in a reproducible manner. Once the threshold voltage has been set to a predetermined value it operates very much like a fixed-threshold IGFET. Thus, all the relationships of IGFET device parameters (e.g., transconductance, etc.) to device geometry are valid. The read-out step is a normal operation of an IGFET. Its speed is therefore determined by the pertinent resistances and capacitances of the device and the interrogation circuit configuration. Given the limits of device geometry, the read-out speed depends on the circuitry employed.

The property of the device that is entirely new is its information storage capability. The phenomena relating to this property are the area of investigation covered by the reported work. The switching characteristic, which is meant to be the relationship of the induced threshold voltage changes with the sign and amplitude of the applied voltage, is the topic of the first section. Next is a discussion of experiments for the measurement of write-in time. The last important memory property, namely storage time, is explored in a section entitled "Persistence."

2. Switching Characteristics

The threshold voltage of the memory device is changed by the application of a voltage pulse between its gate and substrate. The relationship between the

amplitude of the applied voltage pulse, the width of the pulse, and the threshold voltage induced by that pulse is indicated in Figs. 8(a) and 8(b).

The curves in Fig. 8(a) were obtained by setting the device to a low threshold voltage by a positive pulse of 50 V. This was followed by a negative pulse of desired amplitude and width. The threshold voltage obtained from this negative pulse was recorded. Then a +50 V pulse was again applied to the gate, followed again by a negative pulse of another width or amplitude. In this way, all the points in Fig. 8(b) were derived in an analogous manner, except that the signs of the alternating pulses were reversed.

It is clear that the amount of shift depends both on the amplitude and the width of the pulse. There is, however, a saturation pulse width beyond which no further increase in threshold voltage occurs for a given pulse amplitude. This saturation effect is typified by Fig. 9. A plot of the saturation voltage vs both positive and negative applied voltage is shown in Fig. 17. The curve was obtained by applying a dc voltage across the device for as long as ten minutes, so that saturation was achieved. The most noteworthy feature of this curve is the fact that there is a minimum absolute voltage below which no shift in threshold voltage occurs in either direction. As long as this minimum voltage is larger than the interrogation voltage used to determine the content of the storage element, the process of interrogation will not change the information stored in the transistor. This is a necessary requirement for a nondestructive read-out (NDRO) device.

An interesting temperature dependence of the amount of threshold voltage shift was found. First, between 25°C and 100°C no temperature dependence was observed at all. Above this temperature, low pulse amplitudes (50 - 70 V) resulted in an increase in the amount of shift with increasing temperatures. For pulse amplitudes between 70 and 80 V, little change with increasing temperature occurred. For pulse amplitudes above 80 V, the threshold voltage change actually decreased with temperature. All ΔV_T vs °C curves appear to approach a common value near 300°C. The explanation of this behavior requires the assumption of a charge transfer mechanism that is dependent on temperature.

3. Write-In Time

The change in threshold voltage induced in the transistor must have resulted from the injection of charges into the dielectric. An increase in the (negative) threshold voltage must be due to a buildup of positive charges, and a decrease in the (negative) threshold voltage must be due to a loss of positive charges. These charges then must be transported by a transient current that goes to zero as the induced threshold voltage reaches its saturation value.

Our initial attempts to record this current in transistors met with difficulty, since the current was too small to give sufficiently accurate results. However, capacitor structures similarly fabricated, with areas fifteen times larger than the transistor gates, permitted recording of definitive values. An example is given in Fig. 18. In this oscilloscope trace the horizontal scale is in units of 1 μsec per division. The large trapezoid (trace C) is the trace of the applied voltage pulse. Its amplitude is -100 V to produce a readily measurable current. The two other traces are transient currents, at a scale of 50 μA per division. They

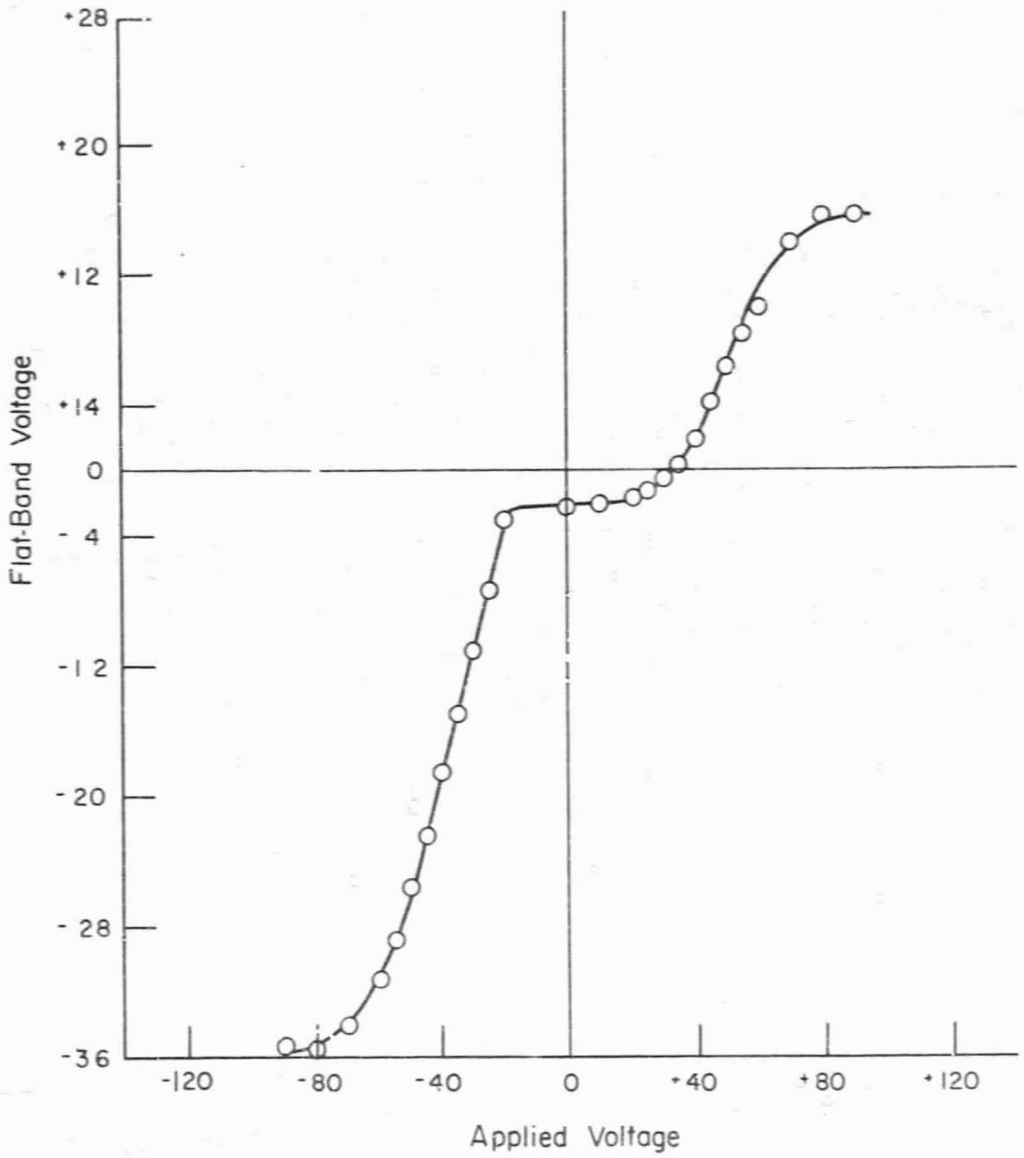


FIG. 17 Saturated values of flatband voltage as a function of applied voltage.

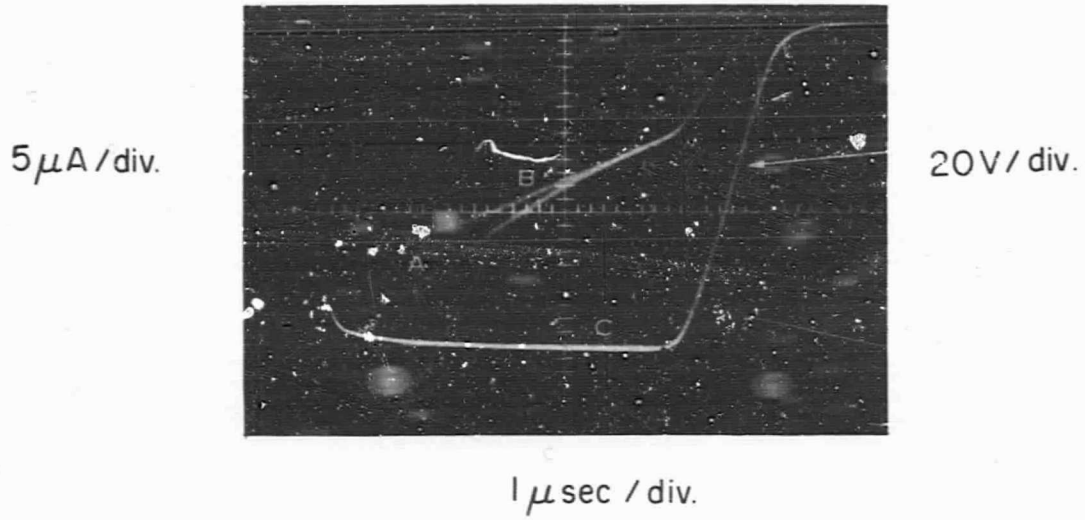


FIG. 18 An oscilloscope trace of: (a) the current pulse which represents the charge going into storage plus the displacement current, (b) the current pulse describing the displacement current alone, and (c) the incoming voltage pulse.

were recorded across a resistor in series with the device. The larger amplitude transient (trace A) resulted when the device was pulsed once after it had been set to a low threshold voltage. The smaller transient (trace B) resulted when the device was pulsed a second time with the same voltage value as the first pulse. Further pulsing resulted in repeated tracings of the lower amplitude transients. Thus, all charge had been transferred during the first pulse, and the saturation pulse width was therefore smaller than the width of the first pulse applied. The first and larger current transient arose from the sum of the currents that transport the charges into the dielectric and the displacement current charging up the device capacitance. The second current transient was due to the displacement current alone.

It is clear from the trace that, for this structure, all charge storage occurred during the time interval when the displacement current was quite large. For the results shown in Fig. 18, the saturation write-in time was about 7 μsec. Since the displacement current is determined only by the capacitance of the structure and its series resistance, the write-in time of a configuration should really depend only on the series resistance in the writing circuit. In a further experiment the values of such resistors were changed from 1 MΩ to 100 kΩ and then to 10 kΩ resistors, and the predicted dependence on the RC constant was confirmed. With some nitride structures, write-in times of the order of 1 μsec were observed. Other structures, made by a different fabrication process, showed very much longer write-in times. A typical example is given in Fig. 19. Here the difference between the first and the second pulses is very small, indicating a small current carrying charge into the dielectric.

4. Persistence of Stored Information

The storage life of the information is a parameter of considerable importance, since it determines the types of application which are possible. Initial tests on transistors were carried out over periods of the order of 200 hours. The transistors were shifted to high and low threshold voltages and then stored at room temperature and at 125°C, with both zero bias and -5 V bias during storage. All transistors exhibited a small change in threshold voltage, ranging between 0.1 and 1 V, over the first two hours. For the remainder of the test period the changes were essentially negligible. Another early test involving four transistors under constant interrogation indicated no important change in the stored information after a period of three months.

Since these early measurements, charge decay has been studied in more detail. An example of a typical experiment is given in Fig. 20. The study was carried out with two identical capacitors that had been shifted to +21 V and -43 V flatband voltages by the application of 140 V, 1 msec pulses with the correct polarity. The original threshold voltage of the capacitors was -10 V. These capacitors were observed over a period of a month. The curves in Fig. 20 exhibit an initial rapid decay that progressively slows down. At the time of measurement, the decay law was not known, so that it was convenient to use a log-log display of the data. Extrapolating from these data indicates that it would take about a year for one-half of either a positive or a negative charge to decay. This is probably a low estimate.

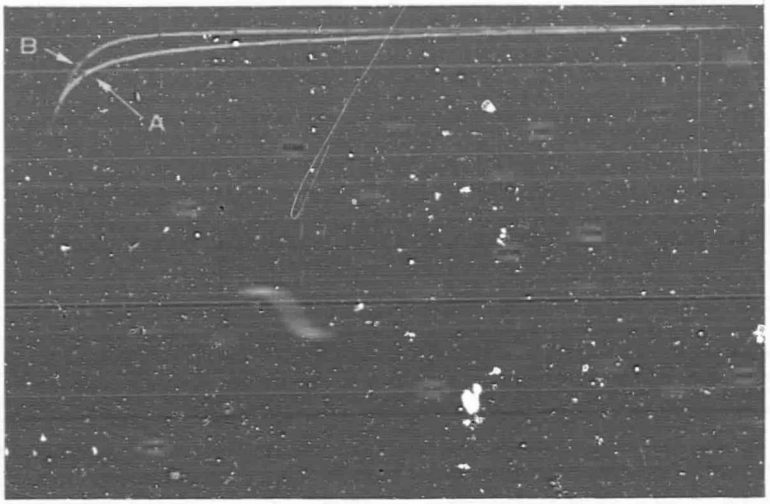


FIG. 19 The difference between trace A and trace B is the amplitude of the conduction current into the dielectric.

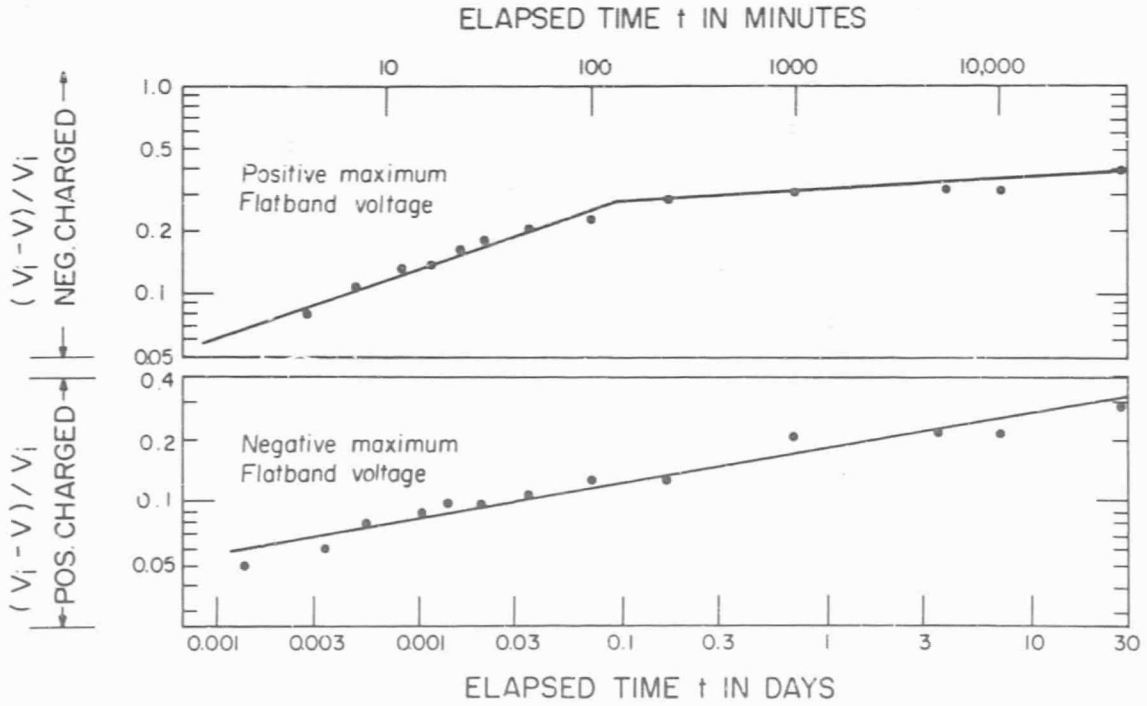


FIG. 20 Typical storage-life test results for a positively shifted capacitor and a negatively shifted capacitor.

The effect of temperature is indicated in Fig. 21. It indicates that at sufficiently elevated temperatures there is an increase in the decay rate of capacitors similar to those used in Fig. 18. If a relationship between time constants of charge decay is assumed to have the form $\tau = \tau_0 \exp(e\varphi_t/kT)$, then Figs. 20 and 21 permit a calculation of an activation energy, or perhaps a trap depth. The value found is 0.25 eV.

5. Summary

The relationship between a voltage applied to the gate of the memory device and the induced threshold voltage can be described briefly by a straight line section denoting no change in threshold for applied voltages between zero and a characteristic minimum shifting voltage, and another straight line section for applied voltages between this minimum shifting voltage and higher values. This last section forms an angle of roughly 45° with the applied voltage axis. The minimum shifting voltage increases monotonically with pulse width, until some saturation pulse width has been achieved. It is this saturation pulse width that is designated "write-in time." The variation in pulse width results only in a parallel displacement of the switching characteristic. In the device described, a positive applied pulse resulted in the location of the 45° section such that the induced threshold voltages became more positive, and negative applied pulses caused the inclined section to indicate increasingly negative induced threshold voltages. The curves shown as Figs. 7 and 8 are representative, but in the many transistors and capacitors examined, minimum switching voltages between 5 and 80 V have actually been observed. Typical values, however, are between 30 and 50 V. The shifting characteristics were independent of temperature up to 100°C.

The write-in time of a device is defined as the pulse width that for a given voltage level results in a complete shift in threshold voltage, with succeeding pulses of the same voltage not causing any further changes in induced threshold voltage. This definition results in a self-consistent index of speed that does not necessarily reflect actual operating speeds. Inspecting Fig. 8 makes it clear that for a desired threshold voltage, increasingly narrower pulses are required, if the applied pulse amplitude is increased. In this report, write-in times between 10 msec and 7 μsec are described. The shortest write-in time observed was about 0.3 μsec. Conditions permitting shorter write-in times than that have generally resulted in the destruction of the device, indicating a rough speed-power product limitation for write-in that needs to be further explored. In general, devices exhibiting write-in times between 10 msec and 100 μsec are typical for present devices. Measurements described in the following section indicated temperature independence up to 300°C.

The decay of the induced threshold voltage follows a relationship characterized by the logarithm of time. Direct observations only up to between one to three months have indicated that the change in the induced threshold voltage over these periods is small enough to be unimportant in digital applications. Extrapolation from such data has indicated storage periods from several years up to several powers of ten years. The decay of the induced threshold voltage is accelerated by elevated temperature. At 150°C it was found that the storage time index had been reduced by about a factor of three.

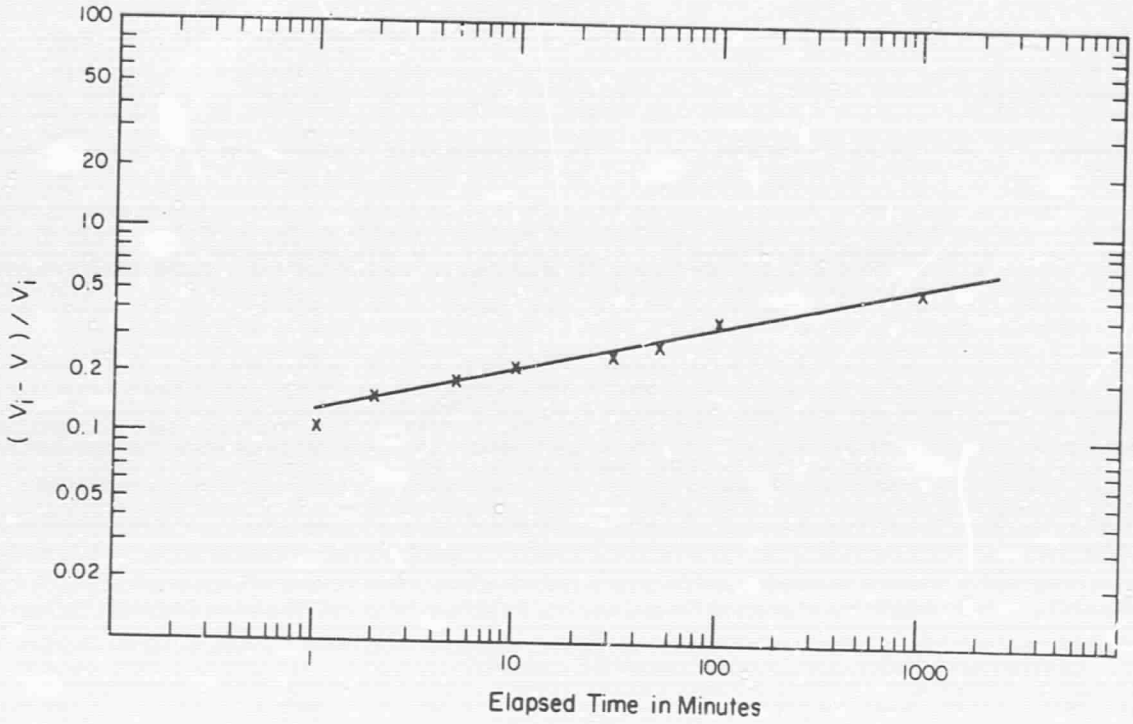


FIG. 21 Decay of stored charge at 150°C; V_i = - 16 V.

C. EXPLORATION OF THE MEMORY MECHANISM

1. Introduction

It is the induced change in threshold voltage which contributes he memory effect. This change is due to the storage of charge in the gate dielectric of the memory IGFET. In this section, experiments will be described that will associate this charge and its storage with special gate structures and gate materials. The location and amount of charge will be described in another study. An experiment to indicate a particular mechanism of charge transfer will be described next, and finally, the range of possible charge transfer mechanisms described in the literature will be explored.

2. The Effect of the Structure of the Dielectric

It was initially surprising and, to most investigators, disturbing to find changes in capacitance-voltage characteristics occurring at room temperature, when MIS capacitors containing silicon nitride were studied. This effect had not been common in MOS capacitors, so that this so-called instability was associated with some mysterious material property of the silicon nitride dielectric. However, it was recognized quite early that aside from a difference in the material of the dielectric, there was also a difference in the macroscopic structure of MOS and MNS capacitors. This difference arises from the fact that the best MOS capacitors are formed by thermal oxidation with the oxide always microscopically generated at the silicon oxide—silicon interface. Thus, MOS capacitors have a good chance of having a single homogeneous dielectric between the silicon and the evaporated metal capacitor plates. In contrast, the dielectric in MNS capacitors is generally formed by depositing silicon nitride on top of a silicon surface that has been treated in various ways before the deposition. In fact, in many instances no treatment is given to the silicon wafer at all. It is well known that silicon spontaneously forms an oxide layer in air, ranging in thickness from 20 Å to 50 Å. Only elaborate treatments, such as heating in vacuum, in hydrogen, or in hydrogen chloride, at temperatures near 1250°C, are capable of yielding an oxide-free silicon surface.

Thus, any MNS capacitor formed by normal methods can be expected to have a thin oxide layer and a thick nitride layer for its dielectric. The memory behavior could therefore be associated with the presence of the thin oxide layer in particular or the existence of a two layer dielectric in general. In order to test the first alternative, capacitors were formed with 1000 Å of silicon nitride on oxide layers of varying thicknesses. These layers were formed by short oxidation at different temperatures. From published data, the oxide thicknesses were calculated to range from about 30 Å to 200 Å. The threshold voltage change for a 100 V pulse of 1 msec width was inversely related to the oxide thickness. Thus, the presence of the oxide layer clearly played a role. Its thickness was also important, but there did not seem to be a critical thickness involved.

The effect of the type of silicon nitride was tested in another series. A number of capacitors were formed having a silicon-oxide layer of constant thickness and a 1000 Å thickness of nitride. The concentration of oxygen in the nitride was varied from 0 to about 30%. Figure 22 shows a plot of the flatband voltage

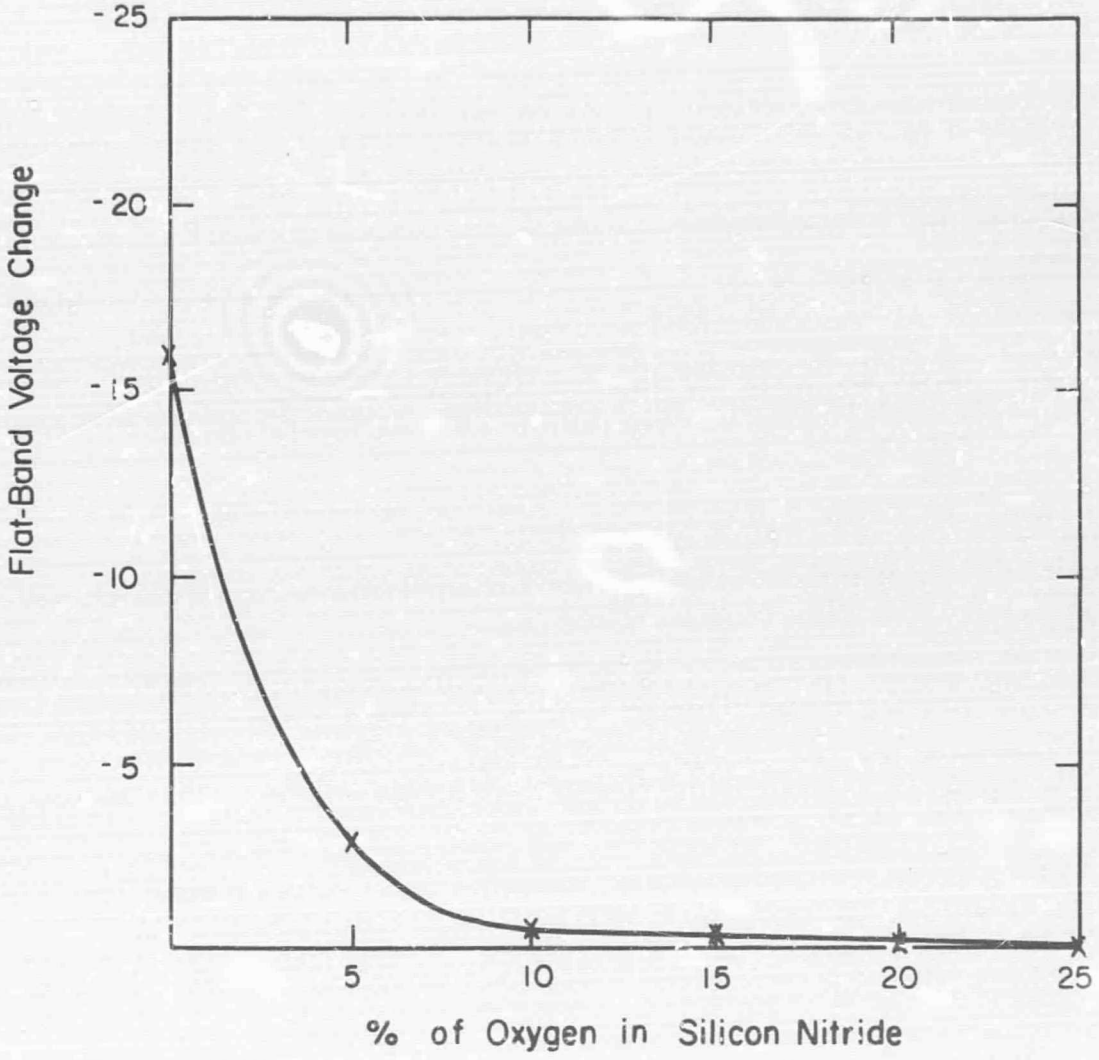


FIG. 22 Shift in flatband voltage for capacitor structure of 100 Å silicon oxide and 1000 Å silicon nitride, as a function of oxygen concentration in silicon nitride. Voltage pulse: - 100 V, 1 msec.

change for a constant pulse amplitude and width as a function of the oxygen concentration of the silicon nitride. Obviously, the amount of shift is critically dependent on the oxygen concentration.

3. Location and Number of Charges

It was clear that the charges stored in the dielectric were giving rise to changes in threshold voltage in the memory device. The exact location of these charges was of importance for the understanding of the device and for the control of its fabrication. In order to determine this parameter, MIS structures were made and shifted to high negative and positive flatband voltages. The metal plates were then removed and the insulator was removed in small increments by etching. After each removal the flatband voltage of the structure was measured with a mercury probe. In this way, a relationship between flatband voltage and insulator structure was established. From both sets of values, the surface charge density in the dielectric was calculated and plotted as a function of insulator thickness. The data in Fig. 23 indicate that down to about 100 Å, the smallest thickness that could be estimated, there was no discernible change in surface density. The charges were therefore located within 100 Å of the insulator-silicon interface. Since the oxide layer in these capacitors was near 50 Å, it appeared clear that these charges were located somewhere near that layer.

It will be recalled that in a previous section the transfer of these charges into the dielectric had been associated with write-in time, and measured as conduction current. A point of considerable importance was whether the conduction current observed during these experiments really was completely converted into stored charge, or whether some of it had passed through the dielectric structure. In order to test this, a pulse of a given amplitude was placed across a memory capacitor and the current vs time trace was recorded on an oscilloscope. Then the amount of stored charge was measured by a C-V plot. That was followed by another pulse of the same amplitude which was recorded on the same photograph as the first pulse. The difference in area between the two traces was integrated as $\int I dt$ to obtain the charge transferred into the dielectric by conduction. This value was plotted against the value actually measured by the C-V plot. The results are shown in Fig. 24. The 1:1 correspondence indicates that within the limits of accuracy of the experiment, all the charge observed was transferred by conduction current, and vice versa. Thus a study of current characteristics assumed greater importance.

4. Experimental Study of the Charge Transfer Mechanism

In order to gather information on the mechanism of charge transfer, the transient current produced during the pulse that changed the flatband voltage of a capacitor was recorded as a function of temperature of that capacitor. Table I shows the results, which indicate that the charge transfer mechanism is independent of temperature, at least between 25°C and 100°C. This suggests a mechanism involving either tunneling of conduction electrons through a barrier or tunnelling of electrons from isolated states into a conduction band. It is necessary to assume that electrons are doing the tunneling from the dielectric into the silicon. The transfer of holes from silicon into the dielectric appears unimportant, since during the charge transfer the hole concentration at the

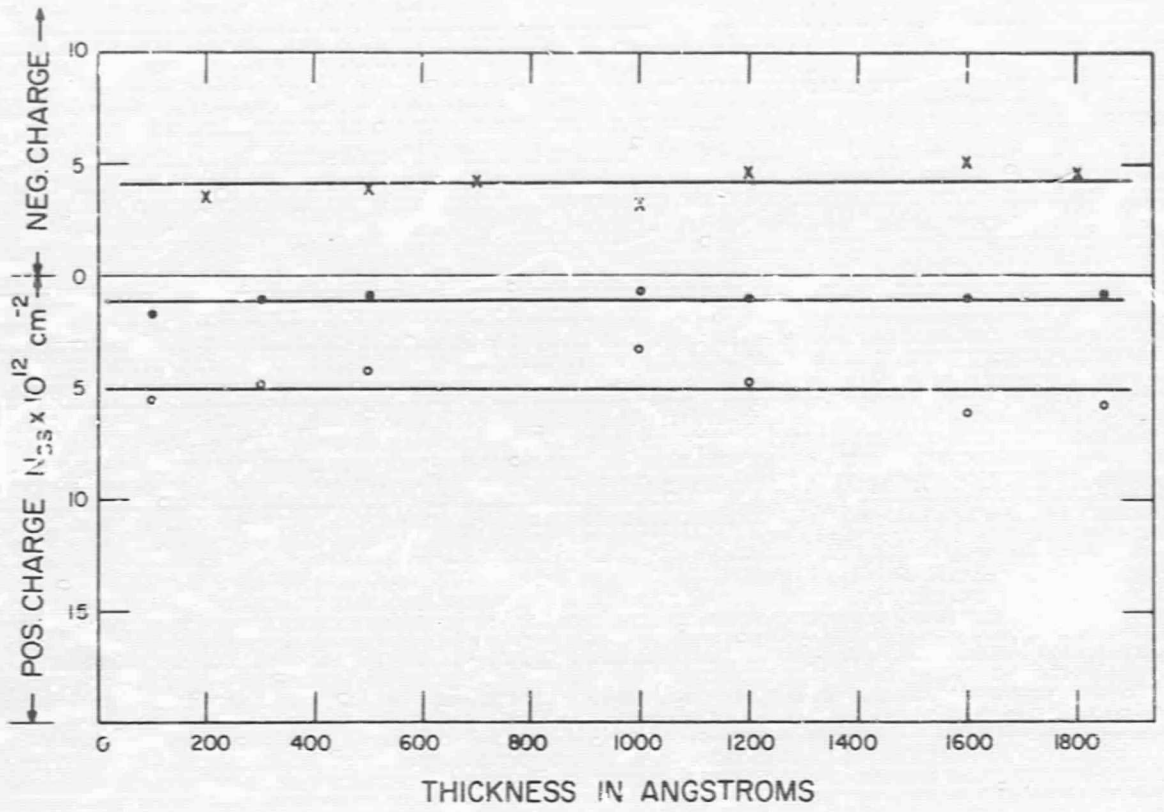


FIG. 23 Step etch results for capacitors: (x) subjected to positive pulses prior to test; (●) not subjected to pulses; and (o) subjected to negative pulses.

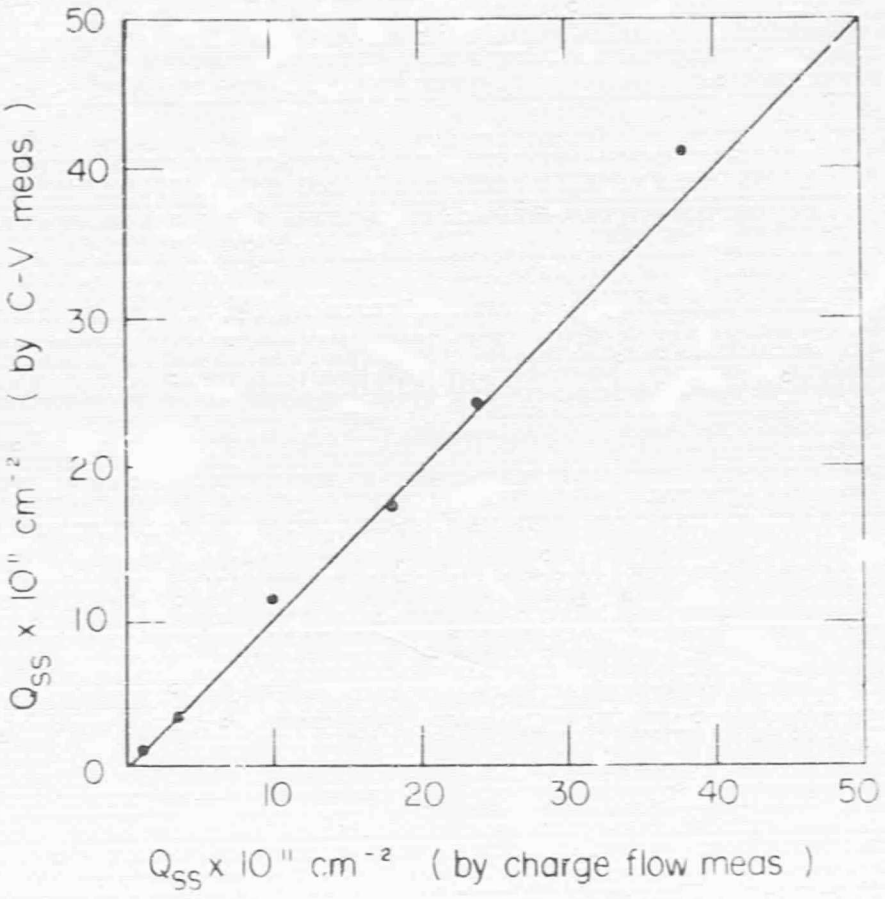


FIG. 24 Plot of charge stored in dielectric as measured by C-V plot vs charge transferred into the dielectric by conduction.

TABLE I

Effect of Temperature on Duration and
Peak Current of Charging Transient

Temperature	Transient Time	Peak Current
25°C	1.2 μsec	2.0 mA
100°C	1.3 μsec	1.9 mA
150°C	1.2 μsec	1.8 mA
200°C	1.2 μsec	2.0 mA
300°C	1.4 μsec	1.6 mA

dielectric-silicon interface is still of the order of that of the minority carriers, namely $\approx 10^{20}/10^{15}$ or 10^5 holes/cm^3. The temperature independence of the current pulse is further evidence that the positive charges in the dielectric were formed by electrons leaving neutral states. The density of neutral states in the dielectric is clearly independent of temperature. In contrast, the density of holes in silicon at 300°C would be much higher than that at 25°C. As described before, the magnitude of the change of the threshold voltage was temperature dependent. Therefore, there must be several processes at work simultaneously, with only one of them temperature independent.

5. Conductance Laws

At least two charge transfer mechanisms appear to be associated with write-in, one temperature dependent, and one temperature independent. Thus, it is of interest to see what conductance laws exist in the realm of thin film insulators. Guiding papers in this field are authored by C. A. Mead (Ref. 1), Sze (Ref. 2), and Dahlke and Sze (Ref. 3). The second, particularly, has dealt with silicon nitride films.

Of importance to the present problem are the following facts. First, silicon nitride is an electronic conductor. Second, it can be expected that the conduction current density through the thick insulator formed by silicon nitride is bulk controlled. Third, all three different conduction mechanisms observed are important in different ranges of applied field.

The current density J obtained for an applied field E may be the sum of three current contributions, i.e.,

$$J = J_{PF} + J_{IFE} + J_H$$

The quantity J_{PF} is the current arising from an internal Schottky emission of trapped electrons into the conduction band, known as the Poole-Frenkel effect, and is given by

$$J_{PF} = \sigma_{PF} E \exp\left\{-e\left[\varphi_{PF} - \left(eE/\pi\ \epsilon_0 \epsilon_d\right)^{1/2}\right]/kT\right\}$$

where e is the electronic charge, φ_{PF} is the barrier height for the electron traps, ϵ_0 is the permittivity of free space, ϵ_d is the dynamic dielectric constant, k is the Boltzmann constant, and T is the absolute temperature. The constant σ_{PF} is a function of trap density.

The quantity J_{IFE} is the current arising from the internal field emission of electrons from traps into the conduction band. It is essentially a tunneling effect described by the expression

$$J_{IFE} = SE^2 \exp\left(E_{IFE}/E\right)$$

where S is a function of effective mass and trap density, and E_{IFE} is a constant depending on effective mass and trap depth.

The current J_H arises from the hopping of thermally excited electrons between isolated states. It has an ohmic characteristic, but is dependent on a thermal activation energy φ_H, i.e.,

$$J_H = \sigma_H E \exp\left(-e\varphi_H/kT\right)$$

The thin silicon oxide layer may have bulk conduction properties of the same nature as those described for silicon nitride. It is, however, more probable that tunneling through it occurs. Schottky emission at either of its interfaces is also a possibility. The exact mechanism therefore is hardly predictable and depends on its thickness and many factors arising from the details of its formation process.

6. Summary

These experiments point out features that must be accounted for in a model of the device properties of the memory effect. First, the accumulation of charges in the dielectric is probably associated with the existence of two different dielectric layers, or the presence of a thin oxide layer. The charge is located somewhere near the silicon interface and may be associated with the postulated thin oxide layer. All the charges are transferred into the dielectric with the help of a conduction current. This conduction current has aspects both of temperature dependence and temperature independence. In the recorded literature it has been reported that silicon nitride exhibits electronic bulk conduction which may or may not be temperature dependent, depending on the magnitude of the applied field. The effect of the oxygen content on the memory effect is undoubtedly associated with the effect of the oxygen on the conduction mechanisms of silicon nitride.

D. THEORY OF MEMORY DEVICE

1. Introduction

The accumulating weight of the experimental evidence described in the previous section, and the result of theoretical examinations of potential models can be combined to determine the two basic features of the memory effect. The first is the requirement of a dielectric with at least two regions of different conductivity. This structure results in the storage of charge at the interface regions in any combination of dielectrics. The second requirement is that at least one of the two layers should exhibit a highly nonlinear conductivity, very high at high fields and very low at low fields. This property results in short write-in times during the application of high fields and long storage times under the normal conditions of low field. The consequences of this model are treated in the following subsections. They progress from a discussion of the basic mechanism, and static conditions in the device to the transient conditions during write-in and storage.

2. Basic Mechanism

The basic structure in which all action takes place is that of an MIS capacitor. Its top electrode is a metal plate and its bottom electrode is n-type silicon. Its most important feature is the fact that the insulator consists of two layers with different conductance properties. Keeping close to the structure of the present memory element, the thinner region at the silicon interface (region 1) is designated silicon oxide, and the thicker region at the aluminum interface (region 2) is designated silicon nitride. A one-dimensional sketch of this structure is given in Fig. 25. The current density in region 1, J_1, is controlled by the field across region 1, E_1, according to the conductance law $J_1 = f(E_1)$ appropriate for this insulator. The current density J_2 in region 2 is controlled by the field across region 2, namely E_2, according to the conductance law $J_2 = f(E_2)$ appropriate for the insulator in region 2.

When a voltage V_{appl} is applied across the plates of this MI_2I_1S structure, the displacement vectors D_2 (across region 2) and D_1 (across region 1) are identical. The fields are then $\epsilon_2 E_2$ and $\epsilon_1 E_1$ in those regions, respectively, where ϵ is the permittivity of each region. We assume for the present that there are no charges in either insulator region, so that the fields are constant in each.

Region 2 consists of silicon nitride for which $\epsilon_2 = K_{SiN}\epsilon_o$, where $K_{SiN} = 6$; and region 1 consists of silicon oxide for which $\epsilon_1 = K_{Ox}\epsilon_o$, where $K_{Ox} = 4$.

At the instant after application of the dc voltage step V_{appl},

$$D_1 = D_2 \quad \text{or} \quad E_1\epsilon_1 = E_2\epsilon_2$$

and

$$E_2 = E_1 \frac{\epsilon_1}{\epsilon_2} = E_1 \frac{K_{Ox}}{K_{SiN}} = \frac{2}{3} E_1$$

We can therefore describe the initial field across the MI_2I_1S capacitor by Fig. 26. Immediately after this first instant, the fields in each region will start the conduction of electrons, each independently in its region according to the fixed $J_1 = f(E_1)$ and $J_2 = f(E_2)$ laws. It is clear that given different conductance laws and fields the charges in each region will be transported at different rates. We will assume that transport across the electrode-insulator interfaces does not affect the conductance laws (i.e., these electrodes act as ohmic contacts). Since the current densities affecting charge transport are different in each region, a charge pile-up will occur at the interface between regions 1 and 2.

The effects of the sign of the applied voltage and the relative levels of the instantaneous currents in region 1, namely j_1, and in region 2, namely j_2, are sketched in Fig. 27. Figure 27(a) provides the labels for the different regions shown. Figure 27(b) indicates the effect of a negative voltage on the metal plate for $j_1 > j_2$. Since more electrons are leaving region 1 than are entering it, a positive charge is left behind at the boundary between the insulators. Figure 27(c) depicts the case for $j_1 < j_2$ at a negative applied voltage. Now fewer electrons are leaving region 2 rather than entering it, so that a negative charge is left at the insulator-insulator boundary. Figure 27(d) shows the condition for a positive applied voltage and $j_1 > j_2$. Fewer electrons are leaving region 1 than are entering it, again leaving a negative charge at the insulator-insulator boundary. Finally, in Fig. 27(e) the condition is given for a positive applied voltage and $j_1 < j_2$. Since more electrons are leaving region 2 than are entering it, a positive charge is left behind. Thus, depending on the individual conductance laws in each region, positive and negative applied voltages can result in both positive and negative charge accumulation at the insulator-insulator interface.

Whatever the sign of the charge accumulating at the interface, it will give rise to a field which opposes that in existence in the more highly conducting region. The field due to the accumulating charge will at the same time add to that in existence in the less highly conducting region. All charge accumulation will stop when the effective field across the high conductance region has been lowered to such an extent that its current density is equal to that passing through the low conductance region, whose field has been increased by accumulated charge. The end point of charge accumulation is therefore obtained when the conduction currents across both regions are equal.

The same model is valid if we have the same insulator I throughout the dielectric instead of an insulator I_1 different from I_2. But then there must be a (probably thin) region 1 where conduction is controlled by a different mechanism, giving rise to a current law $J_1 = f(E_1)$ in a thin region of characteristic

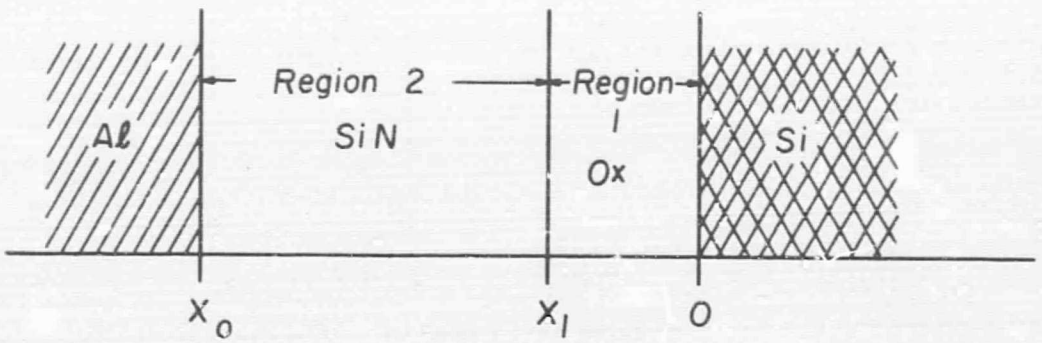


FIG. 25 Structure of memory capacitor.

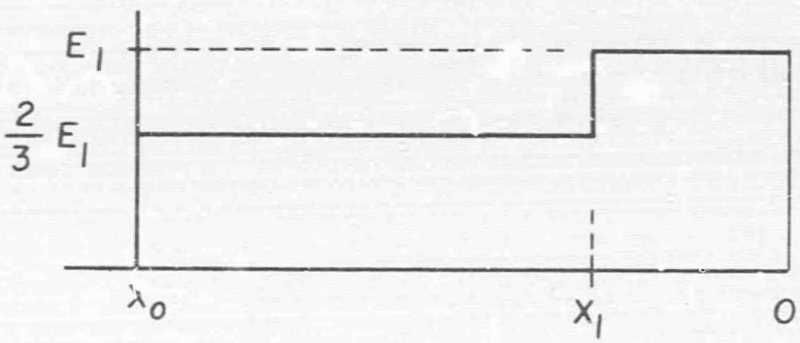


FIG. 26 Initial field distribution in memory capacitor.

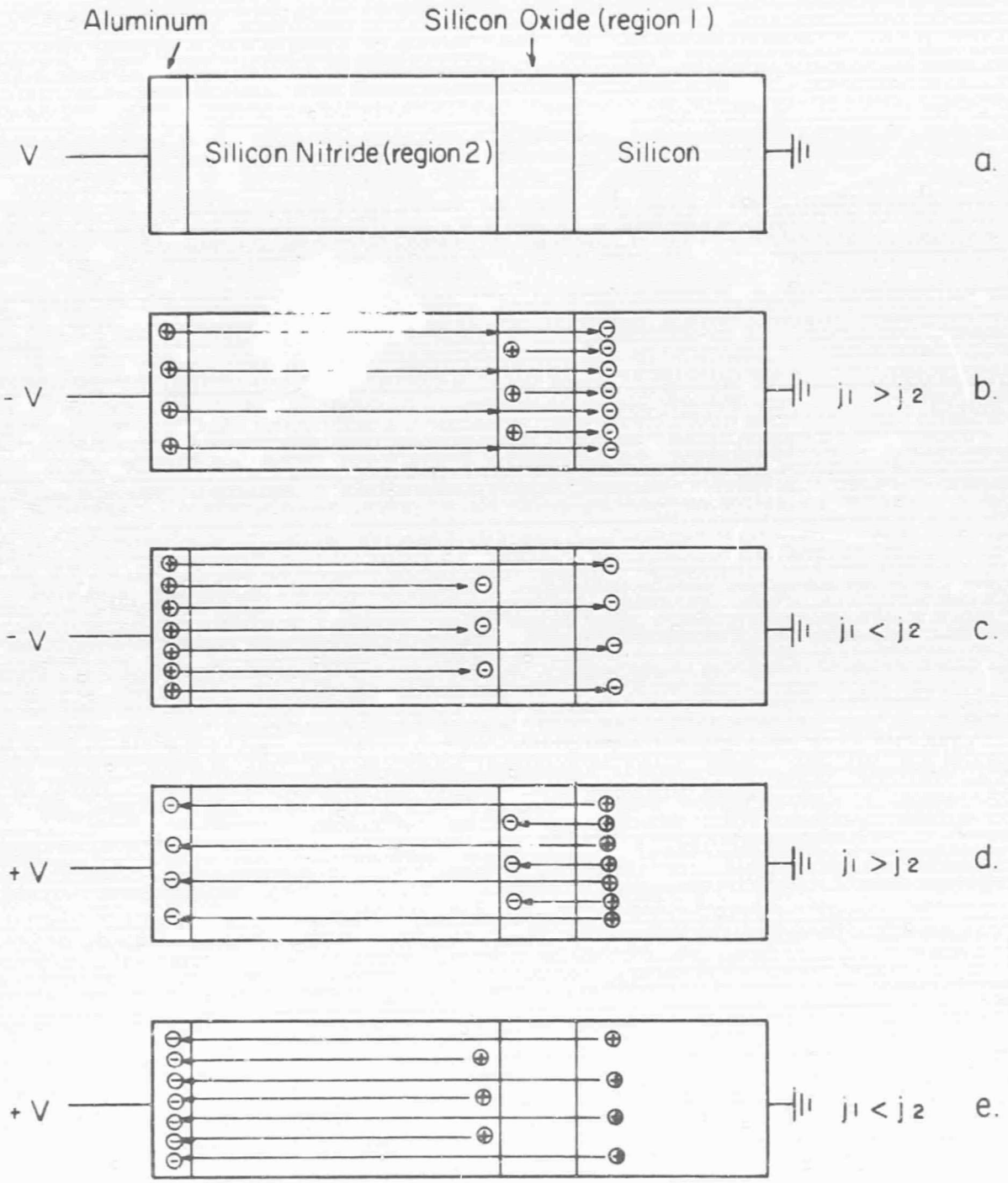


FIG. 27 Effect of the polarity of the applied voltage and the relative magnitude of j_1 with respect to j_2 on the polarity of the charge sheet built up at the insulator-insulator interface.

thickness x_1 having a field E_1 across it. Such a situation will exist when an interface or barrier controlled current (emission limited) crosses the metal- or silicon-insulator interface while the conduction in the insulator is controlled by bulk limited mechanism $J_2 = f(E_2)$.

3. Static Characteristics

In this section we will examine the relationships between applied voltage, internal fields, and accumulated charges. In order to yield a maximum of insight for a minimum of manipulation, it is assumed that the accumulated charge is spread over an infinitesimal distance Δx , so that it may be treated as a delta function distribution. This assumption appears well justified in view of the fact that step-etch experiments have established that the stored charge must lie within about 100 Å of the silicon-insulator interface.

At steady state, the field distribution in the insulator is given in Fig. 28. This is expressed by Poisson's equation

$$\nabla \cdot D = \rho \tag{D3.1}$$

where the displacement vector $D = \varepsilon E$, and ρ = volume charge density in coulombs/cm^3. Integration with respect to x yields

$$D_2 - D_1 = \int_{\cdot}^{\Delta x} \rho dx \quad \text{or} \quad \varepsilon_2 E_2 - \varepsilon_1 E_1 = \sigma \tag{D3.2}$$

where σ is the sheet charge density in coulombs/cm^2.

The division of the applied voltage V_{appl} across the insulator is given by

$$V_{appl} = E_1 x_1 + E_2 \left(x_o - x_1 \right) \tag{D3.3}$$

In this formulation, the voltage drop across the narrow region Δx is neglected. Eliminating E_2 from (D3.3) with the second form of (D3.2) results in

$$V_{appl} = E_1 \left(\frac{\varepsilon_1}{\varepsilon_2} \right) \left[x_o - x_1 + \left(\frac{\varepsilon_2}{\varepsilon_1} \right) x_1 \right] + \frac{\sigma}{\varepsilon_2} \left(x_o - x_1 \right) \tag{D3.4}$$

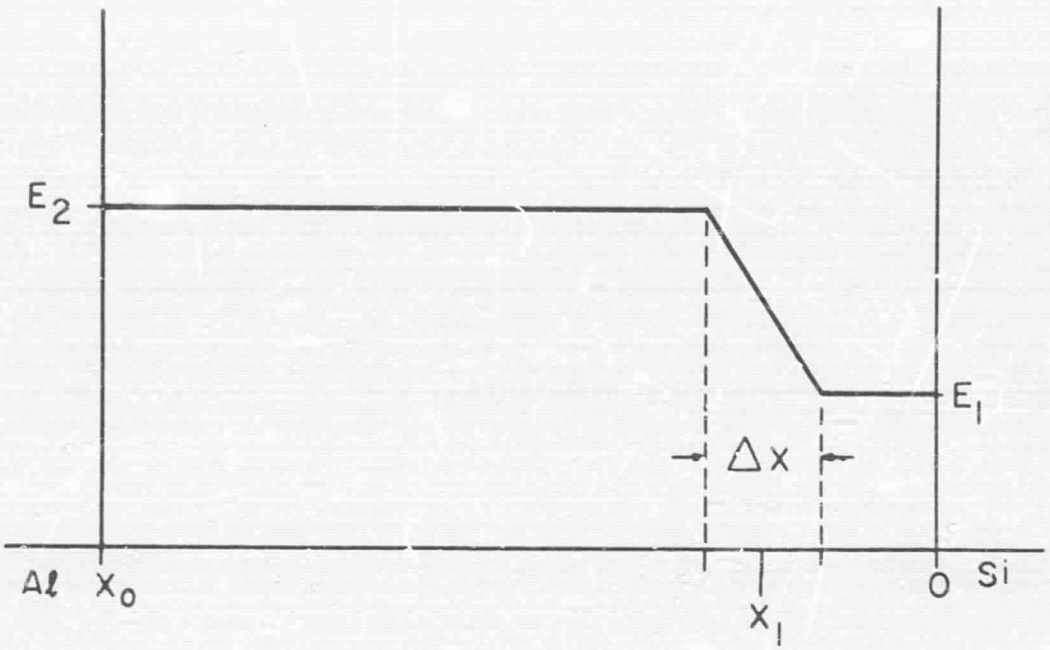


FIG. 28 Steady-state field distribution in memory capacitor.

The applied voltage that causes the field E_1 at the insulator-silicon interface to go to zero is called the flatband voltage V_{FB}, i.e., $V_{appl} = V_{FB}$ when $E_1 = 0$. This causes (D3.4) to become

$$V_{FB} = \frac{\sigma}{\varepsilon_2}\left(x_o - x_1\right) \tag{D3.5}$$

Substituting V_{FB} in Eq. (D3.5) for σ in Eq. (D3.4) results in

$$E_1 = \left[\left(V_{appl} - V_{FB}\right)\left(\frac{\varepsilon_2}{\varepsilon_1}\right)\right] \Big/ \left[x_o - x_1 + \left(\frac{\varepsilon_2}{\varepsilon_1}\right)x_1\right] \tag{D3.6}$$

If we now eliminate E_1 from Eq. (D3.3), again using (D3.2) expressed in terms of $\varepsilon_1 E_1$ and $\varepsilon_2 E_2$ and going through the same manipulations to arrive at E_2, we find

$$E_2 = \left[V_{appl} + V_{FB}\left(\frac{\varepsilon_2}{\varepsilon_1}\right)\left(\frac{x_1}{x_o - x_1}\right)\right] \Big/ \left[x_o - x_1 + \left(\frac{\varepsilon_2}{\varepsilon_1}\right)x_1\right] \tag{D3.7}$$

In order to establish a relationship between V_{appl} and V_{FB}, we divide (D3.7) by (D3.6) and solve for V_{FB}/V_{appl}. This results in

$$\frac{V_{FB}}{V_{appl}} = \frac{\dfrac{\varepsilon_2 E_2}{\varepsilon_1 E_1} - 1}{\dfrac{\varepsilon_2 E_2}{\varepsilon_1 E_1} + \left(\dfrac{x_1}{x_o - x_1}\right)\dfrac{\varepsilon_2}{\varepsilon_1}} \tag{D3.8}$$

From this expression we find that if

$$\frac{\varepsilon_2 E_2}{\varepsilon_1 E_1} = 1, \; V_{FB} = 0 \text{ (i.e., no charge is stored)}$$

and if

$$\frac{\varepsilon_2 E_2}{\varepsilon_1 E_1} \gg 1, \; V_{FB} \approx V_{appl} \tag{D3.9}$$

that is, for large amounts of stored charge, the slope of V_{FB}/V_{appl} approaches unity. The value of $\varepsilon_2 E_2/\varepsilon_1 F_1$ depends, of course, very much on the previous history of the capacitor structure.

A short postscript should be added with respect to the characteristics of real MI_2I_1S structures before any charges are introduced into the dielectric by conduction. They have a flatband voltage that is due to the presence of surface states, ionic charges in the dielectric, and the semiconductor-metal work function difference. It is clear that the ionic charges will act much like the injected charges, except that their location is not as clearly defined. It can be assumed, however, that their effect simply adds to or subtracts from the induced charges. The effect of the work function difference is operative regardless of the number of injected charges, and it generally is so small that it can be disregarded altogether. Surface states, however, may alter the whole picture. These states are charged at the silicon-insulator interface, so that they disturb the field distribution in region 1 to a great extent. They undoubtedly also affect the conduction mechanism in region 1 very severely, so that it is important to reduce them to a level much below that of the charges to be stored in the dielectric.

4. Transient Conditions During Write-In

In this section, a short survey of the treatment of transient conditions will be given. It will be assumed that a positive voltage is applied to MI_2I_1S capacitors so that the space charge region in the n-type silicon does not have to be taken into account. This and many other details are discussed in Appendix A.

The most important tool in this treatment is the current continuity equation:

$$\frac{dD_1}{dt} + j_1 = \frac{dD_2}{dt} + j_2 = j \qquad \text{(D4.1)}$$

where d/dt represents the derivative with respect to time, D and j with their subscripts have the meanings established previously, and j is the current density in the external circuit.

Integration of (D4.1) gives the important result that

$$D_2 - D_1 = \int \left(j_1 - j_2\right) dt$$

$$= \sigma = \left(\frac{\varepsilon_2}{x_o - x_1}\right) V_{FB} \qquad \text{(from (D3.1))} \qquad \text{(D4.2)}$$

This expresses the fact that the accumulated charge density can be expressed as the time integral of the density of the conduction current in the dielectric layers. In order to describe the transient conditions during write-in, we equate the potential drop across a memory capacitor and a series resistor with the applied voltage:

$$V_{appl} = E_1 x_1 + E_2(x_o - x_1) + IR \tag{D4.3}$$

From (D3.2), $E_2 = (\epsilon_1/\epsilon_2)E_1 + \sigma)$; from (D4.2), $\sigma = \int (j_1 - j_2)dt$; I can be transformed into current density j by the use of the MIS capacitor area A; and j in turn can be replaced by (D4.1). When all substitutions have been made, we are left with the equation

$$V_{appl} = E_1\left[x_1 + \frac{\epsilon_2}{\epsilon_2}(x_o - x_1)\right] + \frac{(x_o - x_1)}{\epsilon_2}\int (j_1 - j_2)dt + AR\left(\frac{\epsilon_1 dE_1}{dt} + j_1\right) \tag{D4.4}$$

Differentiation with respect to time leads to the form

$$\epsilon_1 AR\frac{d^2E_1}{dt^2} + \frac{dE_1}{dt}\left[x_1 + \frac{\epsilon_2}{\epsilon_1}(x_o - x_1) + AR\left(\frac{dj_1(E_1)}{dE_1}\right)\right] + \frac{(x_o - x_1)}{\epsilon_2} j_1(E_1) = 0 \tag{D4.5}$$

This equation cannot be solved at present, but it can be made to yield the two limiting cases for the write-in time. The first requires the assumption of a j_1 so small, that during the time the MI_2I_1S capacitor charges up, only a negligible amount of charge is transferred into the dielectric. Under these conditions, all terms containing R can be neglected and the write-in time follows the equation

$$\int \frac{\epsilon_2 dE_1}{j_1(E_1)} = -\left(\frac{x_1 + \frac{\epsilon_2}{\epsilon_1}(x_o - x_1)}{(x_o - x_1)}\right)t \tag{D4.6}$$

The second requires that j_1 is so large, even at small E_1, that the charge is transferred into the dielectric almost simultaneously after it arrives at the MI_2I_1S capacitor. In this case, the RC constant of the series resistor and the MI_2I_1S capacitor limits the write-in time.

To sum up, the maximum write-in speed may be limited on the one hand by the RC constant of the MI_2I_1S capacitor and its series resistance combination. Given a small enough series resistor, such that j_1AR is negligible, the limit

is the current density j_1 under the conditions of applied field. Offhand it could be expected that the larger j_1, the shorter the maximum write-in time. However, it is expected that power dissipation in the dielectric sets a maximum speed limit for this case.

5. Charge Decay

The decrease in an existing charge density can be brought about by a migration of those charges away from their original site or by the removal of charges through neutralization by carriers of the opposite sign. The rate of charge density decrease has the dimensions of current density. There can be two limiting mechanisms that control this rate. Assuming purely electronic conduction, the first is that the rates of negative charge removal by conduction away from the trap, or positive charge removal by electron currents toward the trap, are slower than the rate of thermal activation of negative charges into the conduction band or the rate of recombination of conduction electrons with positively charged traps. For this case, the $J=f(E)$ laws should determine the rates of carrier transportation. If, instead, thermal activation of negative charges and electron-trap recombination of positive charges are rate-determining, the decay laws will depend on the respective activation energies and on the density of charged traps. The effective activation energies would depend on the charged trap density due to the field set up by the charges, and the recombination rates of the positive traps would depend on charge density in controlling the density of electrons available for recombination.

If current transport is the rate limiting step, the whole problem is really a special aspect of the equations derived in Sections 3 and 4. The only difference is that V_{appl} is either zero or, at any rate, very low. Recalling Eqs. (C3.6) and (C3.7) and setting $V_{appl}=0$, we find that

$$E_1 = -V_{FB}\left(\frac{\epsilon_2}{\epsilon_1}\right)\Big/\left[x_o - x_1 + \left(\frac{\epsilon_2}{\epsilon_1}\right)x_1\right] \qquad (D5.1)$$

and

$$E_2 = +V_{FB}\left(\frac{\epsilon_2}{\epsilon_1}\right)\left(\frac{x_1}{x_o - x_1}\right)\Big/\left[x_o - x_1 + \left(\frac{\epsilon_2}{\epsilon_1}\right)x_1\right] \qquad (D5.2)$$

Since E is a vector quantity, E_1 and E_2 must have opposing signs because they originate from the same charge distribution at x_1, but point in opposite directions.

From Eq. (D4.2) we can write

$$-\frac{d\sigma}{dt} = j_1 + j_2$$

where the sign of one j is changed since both now go in opposite directions. The negative sign of the charge density change is required to signify a decrease. This leads to

$$-\frac{dV_{FB}}{dt} = \left(\frac{x_o - x_1}{\epsilon_2}\right)\left[j_1(E_1) + j_2(E_2)\right] \qquad \text{(D5.3)}$$

with the help of (D3.5). Since E_1 and E_2 (D3.6) and (D3.7) are functions of V_{appl} and V_{FB} only, and V_{appl} is assumed to be either constant or zero((D5.1), (D5.2)), we can write

$$\int \frac{dV_{FB}}{j_1(V_{FB_1}) + j_2(V_{FB_2})} = -\left(\frac{x_o - x_1}{\epsilon_2}\right)t \qquad \text{(D5.4)}$$

Careful attention must be paid to the net direction of the field for given V_{appl} and V_{FB}, since it is conceivable that a situation exists when one side of the double layer capacitors causes electron movement towards the I_2-I_1 interface, and the other side causes a movement away from it.

It is clear that one or the other j_i in (D5.4) may be neglected if they are different by orders of magnitude. This simplifies the solution of the integral. A final remark should be made as to the behavior of a real charge distribution. The basic model assumes a delta function distribution independent of distance at the I_2-I_1 interface. For positive charges of essentially zero mobility, this model continues to hold as the charges are neutralized by an electron current. For negative stored charges, this model will hold only during the initial part of the decay. As the electrons move away under the influence of their own field, the electrons will spread out thus changing the field as a function of distance, and with that, the decay law indicated in (D5.4). This extended distribution, incidentally, should also have an effect on the periodic recharging of the dielectric, since then the field distribution is changed for that situation also.

Considering a decay mechanism that is limited by the recombination rate for electrons with positively charged traps, and by the rate of escape of electrons from negatively charged traps, we can make the following predictions. If the recombination rate of conduction electrons with positively charged traps is rate limiting, then the rate of transport of electrons to the traps is faster

than their recombination. This means that even unrecombined, local charge neutrality will become established at the I_2-I_1 interface at the rate of electron conduction. The net effect of this on the flatband voltage (the field established at the I_1-insulator interface) is the same as if the electrons had combined already. Therefore, the conduction mechanism described before is valid for that situation, as far as charge decay is concerned. If the escape of electrons from specially placed traps is the rate limiting factor, then a Poole-Frenkel conduction law is still valid. It may not be exactly identical to that for either layer, but it will have local modifications for trap depth and density, and the initial decay can be treated again in the form given by (D5.4).

6. Summary

In the preceding sections the details of the memory characteristics of the variable-threshold-voltage IGFET have been explained. They are based on the elements of an MIS capacitor whose dielectric has two coplanar regions of different conductivity. This difference may arise from the fact that there are two different materials, or it may be due to two conduction mechanisms operating in the same material, one at the surface and one in the bulk. The theoretical model is the same for either structure. As a matter of fact, in the course of experimentation both have been realized. When a voltage is placed across this MI_2I_1S capacitor, both regions transport electrons, but at different rates, so that charge accumulation must occur at the interface. This accumulating charge lowers the field across the more highly conducting region, so that its conduction progressively decreases. Charge accumulation ceases when the currents in both regions have become equal. The current laws in thin film insulators all require the existence of traps so that the storage of the accumulated charge in traps does not require additional assumptions. The rate of decrease of the stored charge may be controlled simply by the charge transport mechanisms active in the structure at the low fields generated by the stored charge. It may also be controlled by an activation process even slower than charge leakage. In that case, additional deep traps, due to doping or the existence of an interface, must be postulated. The possibility of this slight complication will be ignored to give unity to the following discussion.

All memory characteristics of the device, steady-state, transient, or storage characteristics, are dependent on the relative relationships of the two conductance laws $j_1(E_1)$ and $j_2(E_2)$. The high field characteristics determine the write-in speed, the relative magnitude of j_1 and j_2, and the amount of charge stored, while the low field characteristics determine the rate of charge decay. In order to obtain the observed write-in time of about a microsecond and a charge decay time of the order of a few months, the current density range must extend over thirteen orders of magnitude for applied fields ranging from 5×10^5 to 5×10^6. This will be illustrated in the following examples. As an aid, idealized plots of currents j_1 and j_2 have been drawn as a function of field in Fig. 29. In order to ease the discussion, it is assumed that $x_1 = 0$ and $\epsilon_1 = \epsilon_2$, so that, from Eq. (D3.7), $E_2 = V_{appl}/x_0 = E_{appl}$, and from Eq. (D3.6), $E_1 = (V_{appl} - V_{FB})/x_0 = E_{appl} - E_{FB}$. At the instant when E_{appl} appears across region 1, a large current j_1 flows until the accumulating charge density at the interface between region 1 and 2 has built up an opposing field E_{FB}. When $j_1 = j_2$, no further change in field occurs. When the switching field E_{appl} is

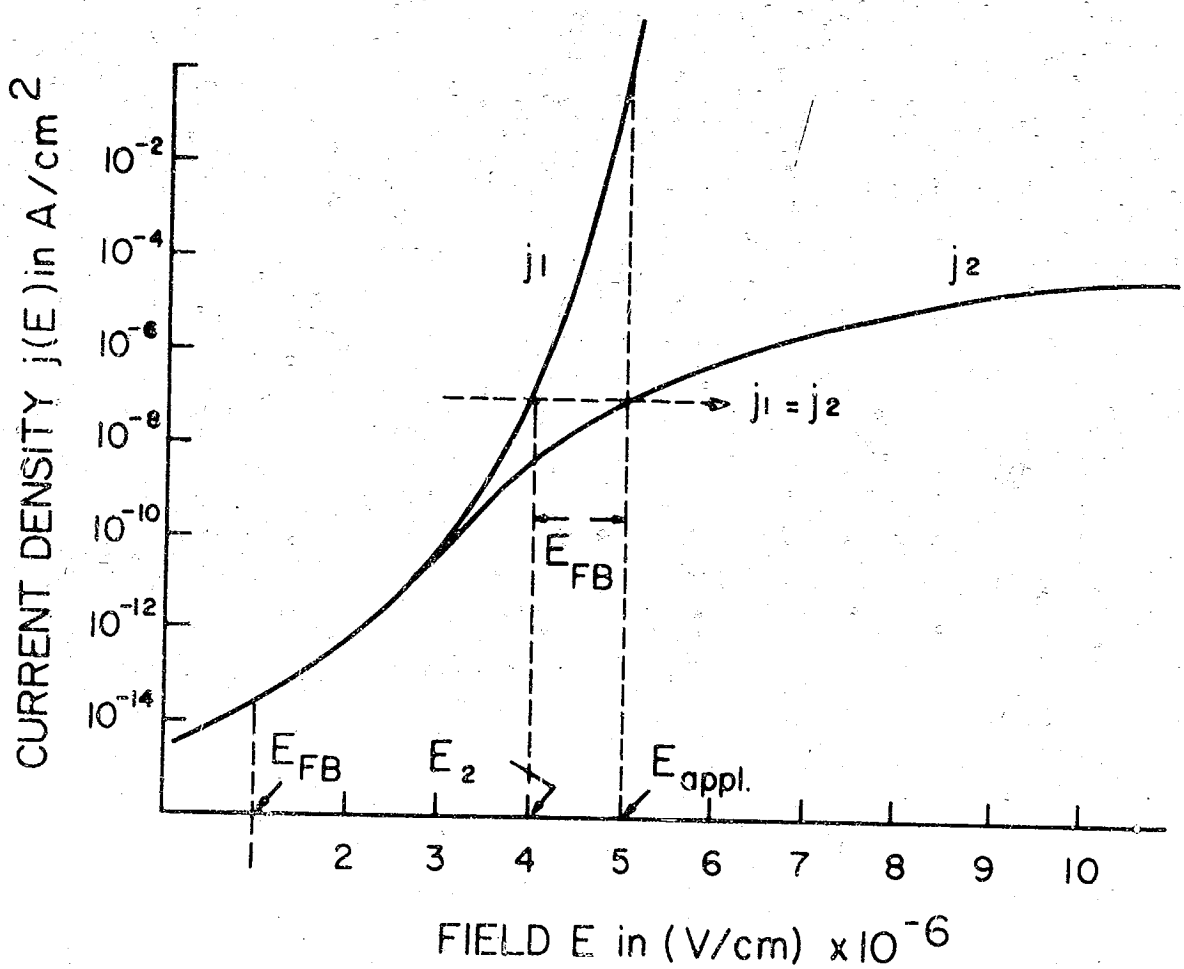


FIG. 29 Plot of idealized current density vs field for dielectrics I_1 and I_2.

removed, only the field E_{FB} due to the interface charge remains, and it starts to decay (assuming conduction mode decay) at a rate given by the current at that field.

Assuming that $x_o = 1000$ Å, we can calculate all pertinent characteristics, since $V = E \times 10^{-5}$. A plot of V_{appl} vs V_{FB} is derived by finding the value of j_2 at E_{appl} and then finding the field E_1 at the same current density on the j_1 curve. The difference between E_{appl} and E_1 is E_{FB}. If we assume $V_{appl} = 50$ V, then Fig. 29 indicates that the final E_2 results in $V_2 = 40$ V, and a V_{FB} of 10 V.

An estimate of the switching time can be obtained by utilizing a simplified form of Eq. (D4.1). It is obtained by setting $j_2 = 0$ and otherwise utilizing $\epsilon_2 = \epsilon_1$ and $x_1 = 0$ (which makes $\Delta E_2 = 0$). Then,

$$\Delta t = \epsilon_1 \Delta E_1 / j_1 = \epsilon_1 E_{FB} / j_1$$

Using $\epsilon_1 = 3.5 \times 10^{-13}$ and j_1 at $E_{appl} = 1$ A/cm^2, we find Δt for write-in is 0.4×10^{-6} sec.

For the decay time we can use the same expression except that ΔE_1 is now ΔE_{FB}, which we will assume to be 5×10^5 V/cm and j_1 at $E_{FB} - 1/2\ \Delta E_{FB} \approx 10^{14}$ A/cm^2. The decay time Δt for loss of one-half the flat-band voltage is calculated from this to be 1.8×10^7 sec, or about seven months.

This exercise shows clearly that all steady-state and transient charge build-up relationships can be derived from a knowledge of the j vs E relationships in both regions of the insulator. It is quite sufficient that this knowledge is only empirical, although theoretical relationships will naturally be preferable since they permit the explicit prediction of the complete device behavior.

The behavior of the device characteristics as a function of temperature can also be predicted on the basis of j vs E curves taken at the appropriate temperatures.

E. EXPERIMENTAL CONFIRMATION OF THEORY

1. Introduction

The previous sections give a detailed theoretical analysis of the charge storage behavior of an MI_2I_1S device, as well as the results of certain qualitative experiments. In this section we present the results of several experiments, in particular J-E curves, which substantiate the theory. We will show that it is possible to predict the relationship between the applied bias to a device and the charge stored in the device (hereafter referred to as bias-storage curves). From the same data we can also quantitatively establish the charging and discharging times. These three pieces of information characterize the memory behavior of the MI_2I_1S device.

2. Experimental Structure

In order to determine the J-E characteristics of each layer of a two insulator device separately, a sample similar to that shown in Fig. 30 was prepared. A 10 Ω-cm n-type silicon wafer was chemically cleaned and placed in the boat in the reactor tube. The reactor was brought up to 900°C and a layer of silicon nitride approximately 1000 Å thick was deposited. This layer was "pure" silicon nitride formed by the pyrolysis of extremely pure ammonia and silane with less than 2 ppm of oxygen. Since there was no deliberate injection of O_2 or compounds of O_2, the film is designated 0%. The 0%, or pure silicon nitride, is characterized by a 12 μ IR absorption, a dielectric constant of about 7, an etch rate in buffered HF of about 10 Å/min, and a low insulation resistance.

After this deposition, the silicon wafer was removed from the reactor tube and the silicon nitride was chemically etched away over an area covering about one third of the wafer. The rest of the surface was masked with black wax. The sample, with the wax removed, was then placed in the reactor again and another layer was deposited. This time the film was about 850 Å thick. During deposition N_2O was injected at a rate of 15% of the overall flow rate, forming an oxynitride (designated 15%) with an IR absorption of 10.6 μ, a dielectric constant of about 5, an etch rate in buffered HF of about 150 Å/min, and an insulation resistance considerably higher than the 0% film. After deposition the slice was removed and a portion of the 15% film was removed, leaving the 0% film beneath intact (see Fig. 30).

This procedure resulted in separate layers of different conductivity films as well as a composite film. Aluminum dots of 10 mil diameter were evaporated through a contact mask. Then the slice was diced and the individual capacitors were mounted on headers. Voltage vs current plots were taken on these samples using a Cary Vibrating-Reed Electrometer, Model 31. The voltage was applied with a battery, and before each current reading was taken enough time was allowed to elapse so that the current became stable with time. So far, curves have been taken only with a positive bias on the gate electrode. For the n-type substrate we are using, the positive bias on the gate electrode results in an accumulation layer of electrons.

3. Test of Bias-Storage Curves

These individual J-E curves and the respective dielectric constant completely characterize the two materials used in fabricating the two insulator layer device shown in Fig. 30. From this information, and from the known thickness of the layers, it is possible to predict or estimate the three parameters that describe the memory type behavior of this device, i.e., the bias-storage curve, the charging time for a given charge, and the discharge time for a given stored charge. The bias-storage curve is simply the relationship between the voltage applied to the MI_2I_1S device and the charge stored at the interface between the two insulators. The latter quantity we express as a flatband voltage, that is, the applied voltage necessary to cause the field at the semiconductor-insulator interface to go to zero. This expression, derived in Eq. (D3.5) is

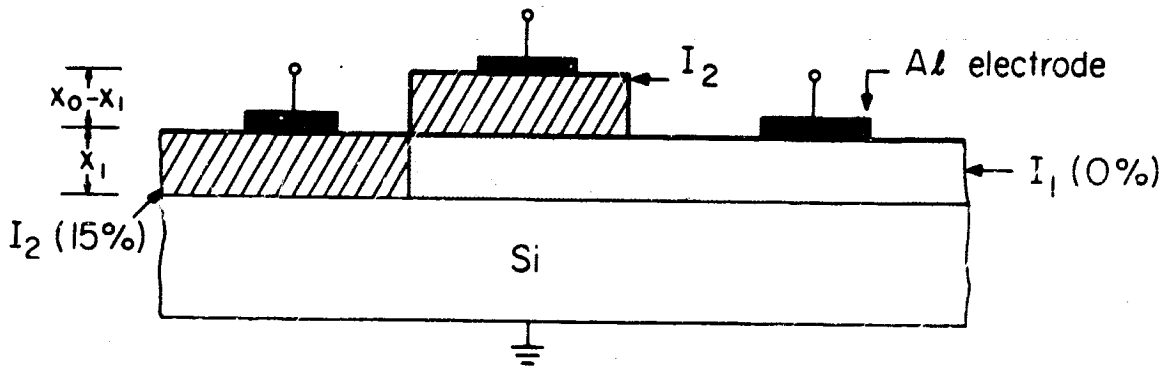


FIG. 30 Schematic representation of device fabrication.

$$V_{FB} = \sigma(x_o - x_1)/\epsilon_2$$

where σ is the stored charge in coulombs/cm^2, ϵ_2 is the dielectric constant of the top layer (I_2), and $x_o - x_1$ is the thickness of the top layer. V_{FB} is a convenient measure of the stored charge because it is approximately equal and related to the voltage necessary to turn on a field effect transistor whose gate is a MI_2I_1S structure with stored charge σ.

Now to predict the V_{appl} vs V_{FB} relationship we proceed as follows. The condition for the equilibrium situation in the device is that the voltage be applied for a long enough time for all displacement currents to vanish. Initially, when a voltage is applied, the continuity condition

$$\frac{dD_2}{dt} + j_2 = \frac{dD_1}{dt} + j_1 = j$$

is applicable. However, when sufficient charge builds up at the interface between I_1 and I_2,

$$\frac{dD_2}{dt} = \frac{dD_1}{dt} = 0$$

and

$$j = j_1 = j_2$$

If j flows in an insulating film, then the field in the film, E, must be given by the J-E curve to the material. Therefore, for any steady state current flowing through both films, the existing fields in the films E_1 and E_2 may be determined from the J-E plots in Fig. 31. We may therefore take sets of E_1 and E_2 for successively higher values of steady state current and from these sets construct the V_{appl} vs V_{FB} relationship. For a given set E_1, E_2 the accumulated charge is

$$\epsilon_2 E_2 - \epsilon_1 E_1 = \sigma$$

from which $V_{FB} = \sigma(x_o - x_1)/\epsilon_2$. The V_{appl} corresponding to this set is just

$$V_{appl} = E_2(x_o - x_1) + E_1 x_1$$

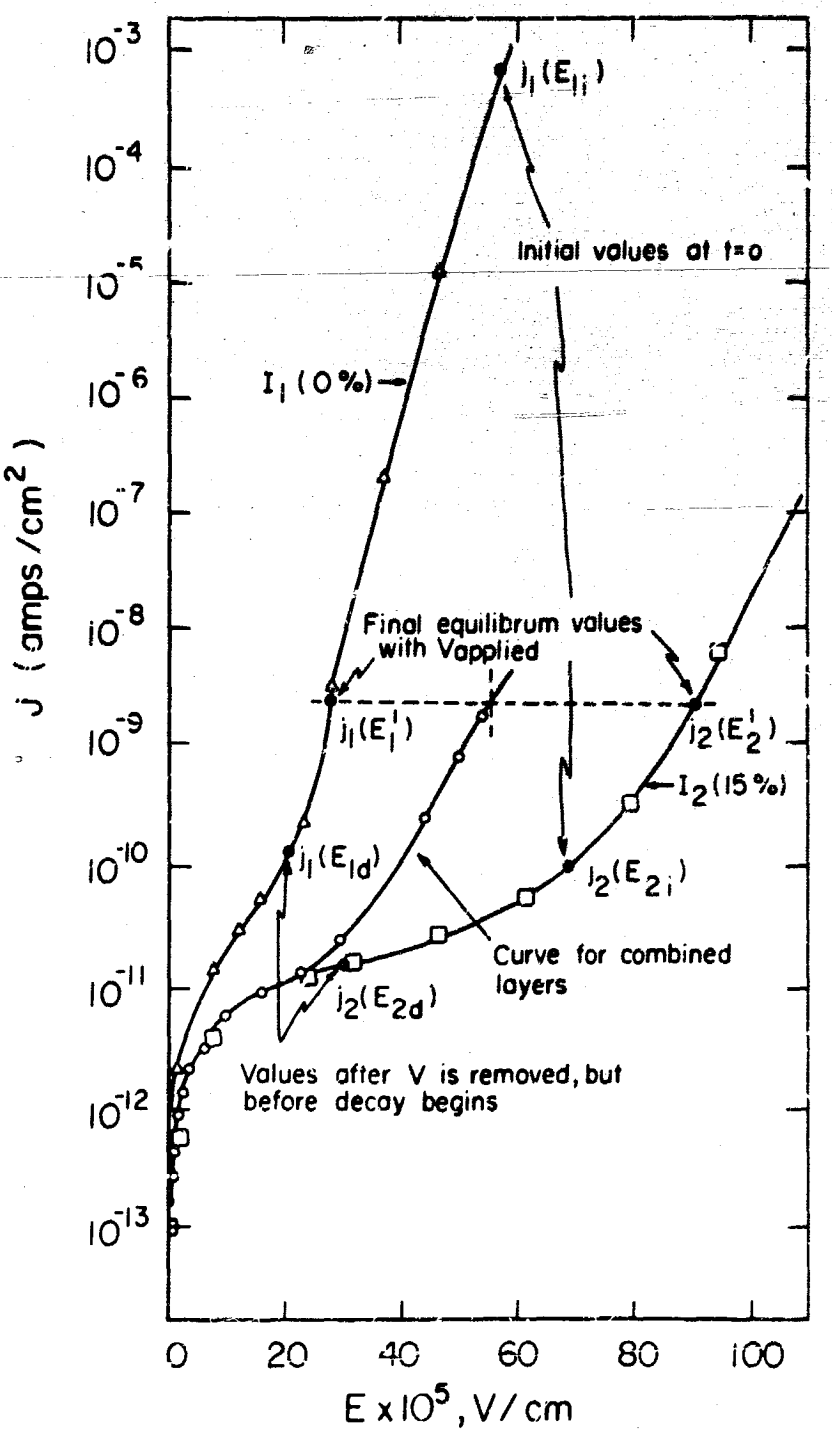


FIG. 31 Experimental J-E curves.

These calculations can be made for successive sets of E_1 and E_2 until the complete V_{appl} vs V_{FB} plot is constructed.

In Fig. 32 we show the calculated and experimental bias-storage curves for the combined layers, and in Fig. 33 we show the experimental curves for the individual layers. A few words of explanation are in order for these graphs. First, the positive bias and negative bias curves were taken on separate capacitors for all three configurations (MI_1S, MI_2S, MI_2I_1S), and since there were variations in the initial C-V offset voltage there is a slight discontinuity at the $V=0$ line. First, consider the effect on V_{FB} of a positive bias. Both single layer samples (Fig. 33) exhibit a negative charge storage (positive V_{FB}) not considerably different from that of the composite layer (Fig. 32). Some storage effects in single layers of silicon nitride are to be expected because there is an observable gradation of certain nitride properties near the interface; in particular, the etch rate increases, implying the possibility of a silicon-rich nitride layer with a higher conductance than that of the remaining film. However, for a positive V_{appl} the combination film (Fig. 32) does not show significantly more storage than the individual layers. At present this phenomenon is still unexplained. For a negative applied bias, however, there is a pronounced difference between the combined layer and the single layers, indicating charge storage at the interface between I_1 and I_2. (The calculated curve showing this charge is also shown in Fig. 32.) This was determined as previously described and then corrected by adding to the combined layer, the V_{FB} that had been determined experimentally for the layer I_1, using the appropriate E_1. The agreement with the experimental curve is fair, considering the approximations and assumptions that were made. The fact that the J-E curves were taken with a positive bias and the charge storage measurements were made with a negative bias may lead to some discrepancy, since we are neglecting the space charge effects mentioned earlier and also since there could be interface injection effects that depend on polarity.

4. Test of Charge Location

Another experiment was performed independently to determine the location of the stored charge in the double layer system. This work involved storing charge in the system by application of a large negative bias, thinning the sample in steps by etching, and then measuring the C-V offset (V_{FB}) after each step. A control sample was also biased so that any decay effects with time could be accounted for. Figure 34 shows a plot of V_{FB} measured vs the total thickness x_o of the dielectric. It is evident that a distinct change in slope occurs at a distance from the silicon that is equal to the thickness of the layer I_1. This change in slope indicates that the charge stored at the I_1-I_2 interface was localized, and therefore could be removed in one etching step. Thus, the assumed distribution has, at most, a width of the thickness of this step, or about 200 Å. This justifies the simple delta function distribution of stored charge used in the development of the theory. Assuming a charge density σ at the interface of I_1 and I_2 and a charge density of σ_{ss} at I_1 and the silicon, the relations between V_{FB} and x_o and x_1 are given by

$$V_{FB} = \sigma(x_o - x_1)/\epsilon_2 + \sigma_{ss}\left[\epsilon_2 x_1 + \epsilon_1(x_o - x_1)\right]/\epsilon_1\epsilon_2 \qquad x_o > x_1$$

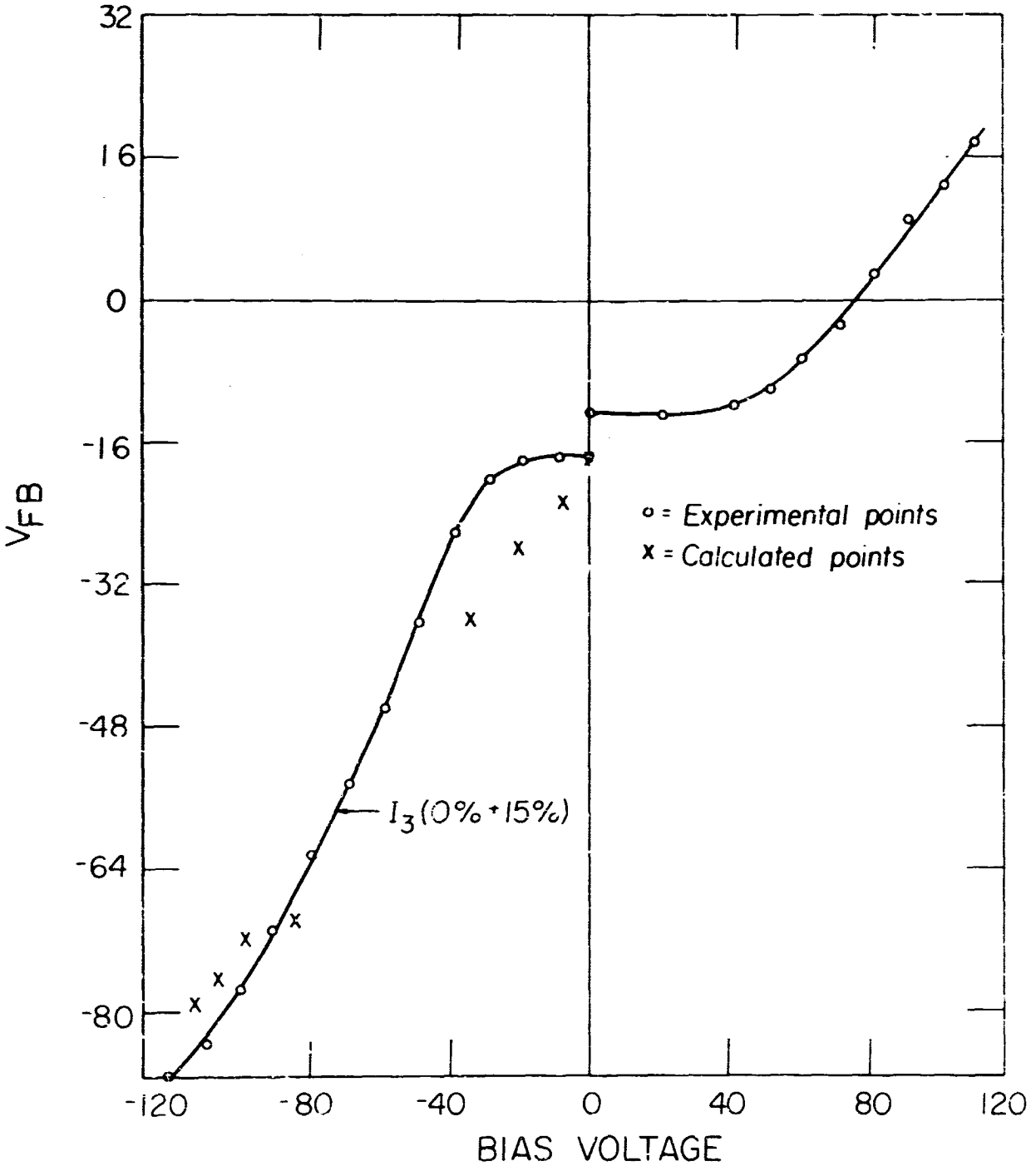


FIG. 32 Storage-bias curves for MI_2I_1S device.

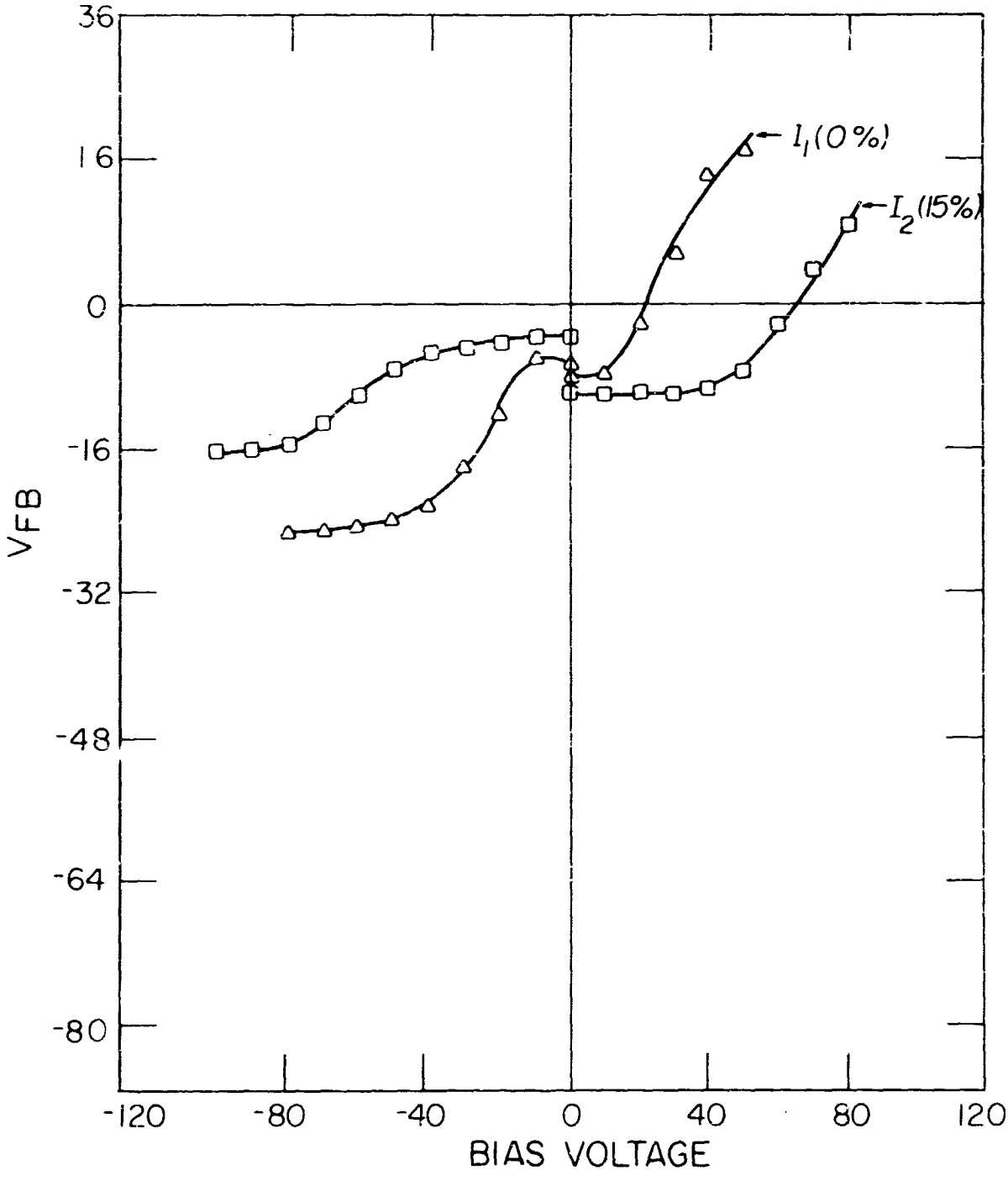


FIG. 33 Storage-bias curves for separate layers I_1 and I_2.

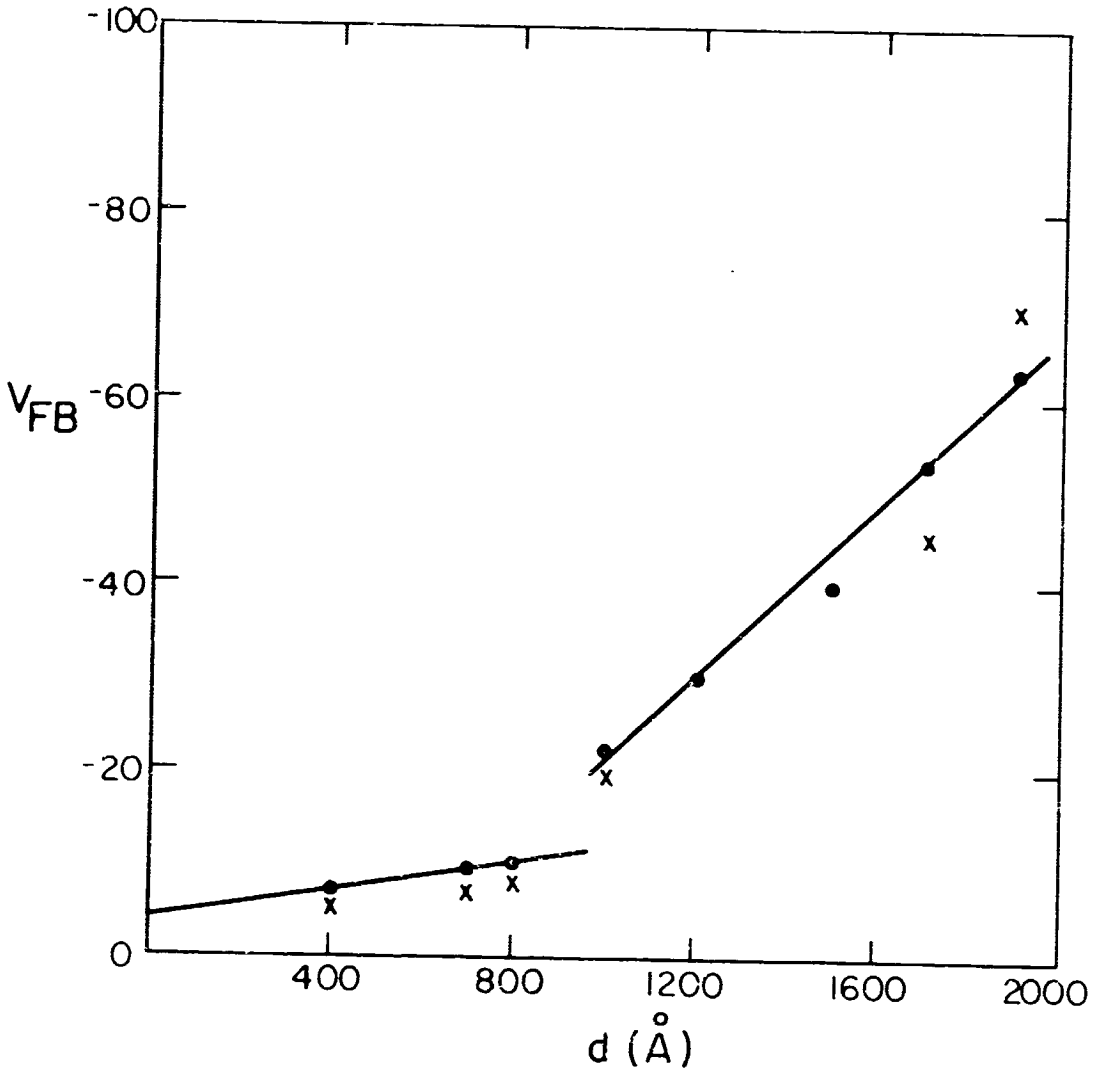


FIG. 34 V_{FB} vs thickness for step-etched sample.

and

$$V_{FB} = \sigma_{ss} x_o / \epsilon_1 \qquad x_o < x_1$$

The slope M_2 is the differential dV_{FB}/dx_o of the first expression, the slope M_1 that of the second. In terms of the slopes M_1 and M_2, the ratio of the stored charges becomes

$$\frac{\sigma}{\sigma_{ss}} = \frac{M_2\epsilon_2 - M_1\epsilon_1}{M_1\epsilon_1} \approx \frac{M_2 - M_1}{M_1} = 5$$

Approximately the same estimate is obtained from the storage-bias curves.

5. Test of Write-In Time

It is appropriate now to discuss the subject of charging and discharging times. It is in this area that the essence of the memory behavior requirements must be met, i.e., a rapid charge time and a slow discharge time.

These requirements are met by using materials with appropriately nonlinear J-E characteristics such as displayed in Fig. 31. We can estimate the charging time as follows.

First, we assume that there is negligible series resistance in the charging circuit, so that at the application of a voltage V_{appl} the external capacitor plates are fully charged before internal charging begins. At $t = 0$, no charge has accumulated at the interface so that the initial fields E_{1i} and E_{2i} are given by the conditions

$$\epsilon_2 E_{2i} - \epsilon_1 E_{1i} = 0$$

and

$$E_{2i}(x_o - x_1) + E_{1i} x_1 = V_{appl}$$

Solving for E_{1i} and E_{2i} gives

$$E_{1i} = \frac{\epsilon_2 V_{appl}}{(x_o - x_1)\epsilon_1 + \epsilon_2 x_1}$$

$$E_{2i} = \frac{\epsilon_1 V_{appl}}{(x_o - x_i)\epsilon_1 + \epsilon_2 x_1}$$

For $V_{appl} = 100$ V , and using the appropriate parameters for the devices, $E_{1i} = 5.81 \times 10^6$ and $E_{2i} = 6.98 \times 10^6$. These values are shown in Fig. 31. Immediately after $t = 0$, charge will begin to build up such that j_1 and E_1 will decrease as given by the J-E curve for I_1 , and j_2 and E_2 will increase according to the J-E curve for I_2 . Finally, the two currents will become equal so that $j = j_1(E_1 f) = j_2(E_2 f)$. Now according to the equation of continuity,

$$\frac{dD_2}{dt} - \frac{dD_1}{dt} = \frac{d\sigma}{dt} = j_1 - j_2$$

$$\Delta t = \frac{\Delta\sigma}{j_2 - j_1} = \frac{\epsilon_2 \Delta V_{FB}}{(x_o - x_1)(j_2 - j_1)}$$

To calculate Δt as a function of stored charge increment, we must know over what range E_{1i} and E_{2i} change to give ΔV_{FB} . Then the appropriate j_1 and j_2 may be estimated from the experimental J-E curves. We assume ΔE , and from this calculate a new E_1 , a new E_2 , and therefore, a ΔV_{FB} . Then we repeat the process and calculate Δt as a function of ΔV_{FB} from the last expression. The results of such a calculation, along with experimental points, are shown in Fig. 35. The initial V_{FB} for this curve is 17.5 V. The agreement is excellent.

6. Test of Charge Decay: Storage Time

Estimating the decay or discharge time presents somewhat more difficulty. Immediately after the removal of the applied voltage V_{appl} , the fields E_1 and E_2 take on new values E_{1d} and E_{2d} , depending on σ and on the conductance laws. The instantaneous values of E_{1d} and E_{2d} after the removal of V_{appl} and before discharge conduction starts are given by

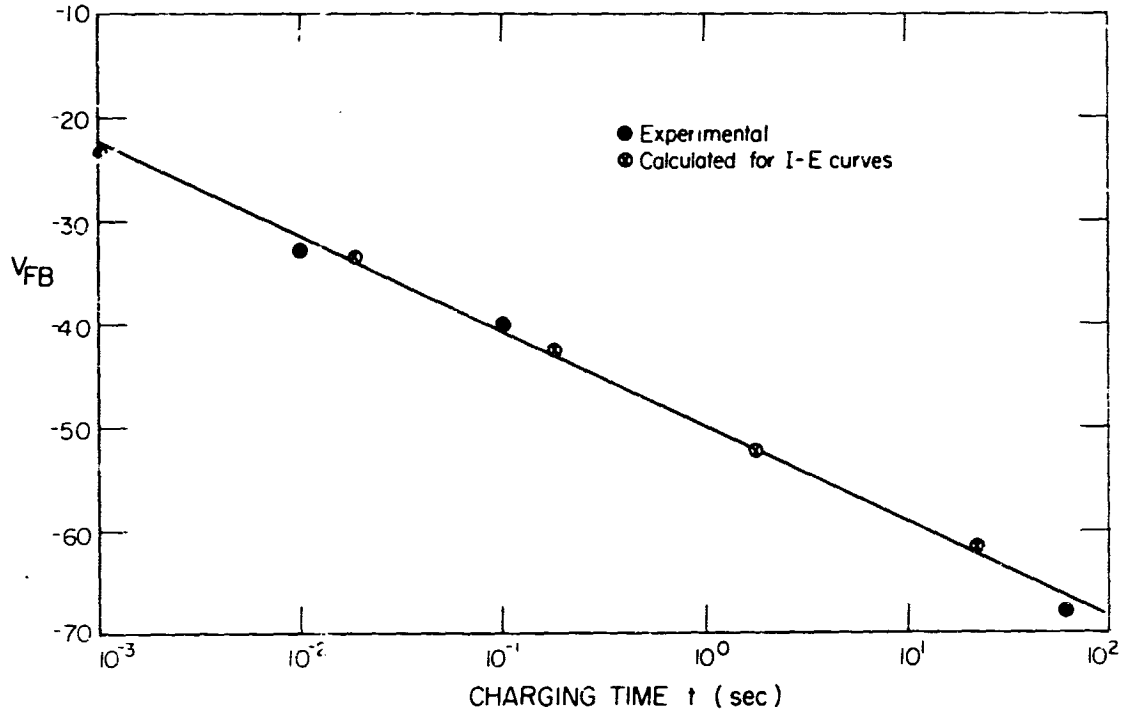


FIG. 35 Charging time vs V_{FB} (for V_{appl} = 100 V).

$$E_{1d} = \frac{-\sigma(x_o - x_1)}{\epsilon_2 x_1 + (x_o - x_1)\epsilon_1} = \frac{-V_{FB}\epsilon_2}{\epsilon_2 x_1 + (x_o - x_1)\epsilon_1}$$

$$E_{2d} = \frac{\sigma x_1}{\epsilon_2 x_1 + (x_o - x_1)\epsilon_1} = \frac{V_{FB}\epsilon_2 x_1/(x_o - x_1)}{\epsilon_2 x_1 + (x_o - x_1)\epsilon_1}$$

The values have been calculated from the σ due to an applied voltage of 100 V, and are shown in Fig. 31 (E_1 is negative, but it is plotted on Fig. 31, and it is assumed that the J-E curves are symmetrical). The stored charge σ effectively will now be shared by two capacitors in parallel, with the charges on one capacitor given by $\epsilon_1 E_{1d}$ and on the other by $\epsilon_2 E_{2d}$. They will dissipate via currents j_1 and j_2, which are, of course, interdependent. With these instantaneous fields established after the removal of V_{appl}, the instantaneous currents are $j_1(E_{1d})$ and $j_2(E_{2d})$. A rough estimate of the time to reduce V_{FB} to $V_{FB}/2$ is given by

$$\Delta t = \frac{\epsilon_1 E_{1d}/2}{j_1(E_1)} \approx 80 \text{ min.}$$

Since $j_1(E_1) \gg j_2(E_2)$, this will be the controlling mechanism. This calculation, of course, assumes a charge removal only by conduction processes.

The experimental discharge curve is given in Fig. 36, covering a range up to 30 minutes. In this time period, V_{FB} loses about one third of its value, agreeing qualitatively with predictions that the charge across I_1 decays to half its value in 80 minutes.

7. Test of Static Characteristics

In the description of the static characteristics of the MI_2I_1S memory capacitor, an expression for the ratio of applied voltage V_{appl} and flatband voltage V_{FB} was derived. This ratio took the form (Eq. (D3.8))

$$\frac{V_{FB}}{V_{appl}} = \frac{(\epsilon_2 E_2/\epsilon_1 E_1) - 1}{(\epsilon_2 E_2/\epsilon_1 E_1) + [\epsilon_2 x_1/\epsilon_1 (x_o - x_1)]}$$

when $\epsilon_2 E_2/\epsilon_1 E_1 = 1$, i.e., no charge is stored, then $V_{FB} = 0$; and as $\epsilon_2 E_2/\epsilon_1 E_1 \to \infty$, i.e., a large amount of charge is stored, $V_{FB}/V_{appl} \to 1$. Thus,

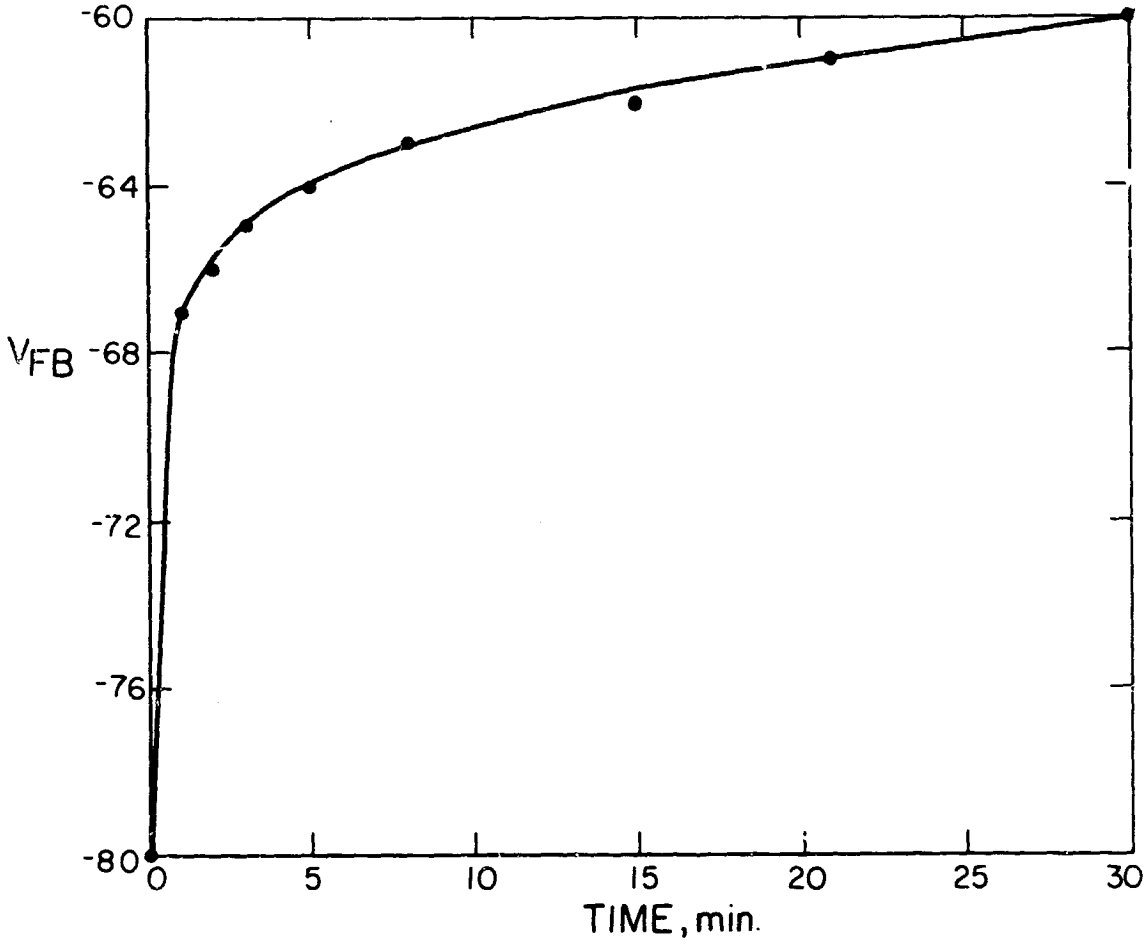


FIG. 36 Decay of V_{FB} with time.

in this approximation, the storage-bias relation has a slope that is approaching unity and becomes increasingly independent of the device dimensions.

A device with a highly conducting silicon nitride layer next to the silicon (I_1) and a less conducting layer of oxynitride on top (I_2) was fabricated with the initial dimensions shown in Fig. 37. The V_{FB} vs V_{appl} relation is plotted for several thicknesses of I_2 obtained by step etching. The curves show a linear relation, with the slope close to unity and independent of the device dimensions. The storage-bias relation was also obtained for another MI_2I_1S device in which the I_2 layer was pyrolytic silicon dioxide instead of oxynitride, as in the previous device. This sample also indicates that the approximation gives a good description of the device performance.

At this point the explanation for the saturation of V_{FB} with increasing V_{appl} is not clear. Also, the curves shown indicate the storage-bias relation for a negative applied voltage only. In the positive polarity, the V_{FB} vs V_{appl} relation is not linear, indicating that refinements are required for this simple approach.

8. Summary

In this section we have tested the major postulates and assumptions of the memory device theory proposed in preceding sections. We have found that experimentally determined J vs E characteristics could be used to predict switching characteristics, write-in times and the leakage of stored charges to a large degree. This confirms the major postulates. In addition, the assumption of a delta function charge distribution at the I_2-I_1 interface was experimentally justified.

F. EQUIVALENT CIRCUIT AND ITS APPLICATION

1. Introduction

The charge storage occurring in the present memory device is the result of nonlinear, unequal charge flows through a two-dielectric gate structure. This charge accumulates at the interface between the two dielectrics. The amount of charge, its accumulation rate, and its persistence can be determined by a detailed knowledge of the current density vs field relations for each dielectric.

Using this physical description of the device structure and behavior it is possible to derive an electrical equivalent circuit that can provide an alternate view of the device and yield dynamic information on the charging characteristics.

In this section such an equivalent circuit and the pertinent charging equations based on this circuit are developed. The solution of the charging equations involved the design and implementation of an analog computer, which is also described.

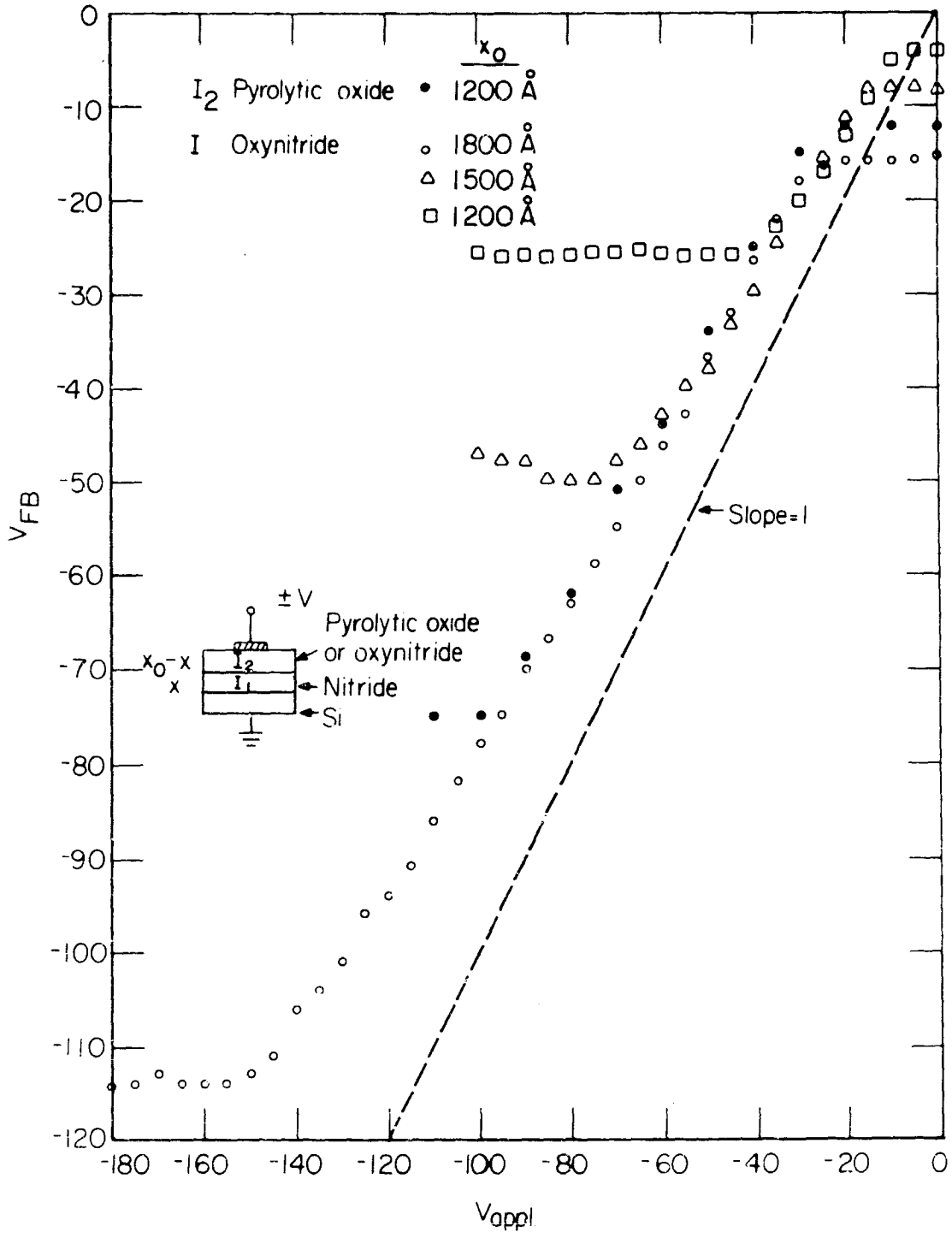


FIG. 37 Confirmation of static characteristics.

2. An Equivalent Circuit of the MI_2I_1S Capacitor with Charge Storage

The physical structure of an MI_2I_1S capacitor is shown in Fig. 38. The unique feature of this structure is that the rates of charge flow through these insulators are not equal. With an external bias it is therefore possible to accumulate excess charge at the interface between the two insulators.

This charging process may be represented by the equivalent electrical circuit shown in Fig. 39. The capacitors C_1 and C_2 represent those of the first and second insulators, respectively. The different rates of charge flow through insulators I_1 and I_2 are represented by the current generators i_1 and i_2, respectively. It is also evident that several additional elements have been included in this generalized equivalent circuit. The capacitor C_{SC} is the space charge capacitance associated with a depletion layer in the semiconductor. The resistor R has been added in series with the MI_2I_1S capacitor so that the net charging current i can be observed via a probe with an input capacitance of C_o.

In the present form, the charging terms i_1 and i_2 cannot be separated. It is necessary to construct a device in which only one of the charging terms is dominant, e.g., i_1. The nonlinear capacitor C_{SC} may also be eliminated by charging the device so that the semiconductor surface is accumulated and thus behaves as a metal plate. If these two conditions are achieved then the equivalent circuit shown in Fig. 40 is applicable.

In this form it is possible to solve for i_1 and v_1 as functions of time in terms of the measurable node voltage v. The circuit equation is as follows:

$$\text{let} \quad C_I = \frac{C_1 C_2}{C_1 + C_2}$$

Then, for $t \geq 0$,

$$C_I \frac{d}{dt}\left[V_{appl} + \frac{1}{C_1}\int i_1 dt - v\right] = \frac{v}{R} + C_o \frac{dv}{dt}$$

From this equation i_1 may be found to be

$$i_1 = C_1\left[\left(1 + \frac{C_o}{C_I}\right)\frac{dv}{dt} + \frac{v}{RC_I} - \frac{dV_{appl}}{dt}\right] \qquad \text{(F2.1)}$$

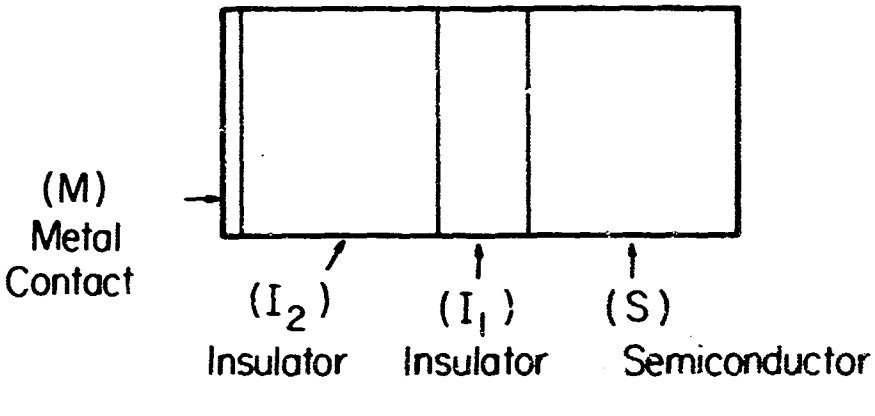


FIG. 38 Physical structure of MI_2I_1S capacitor.

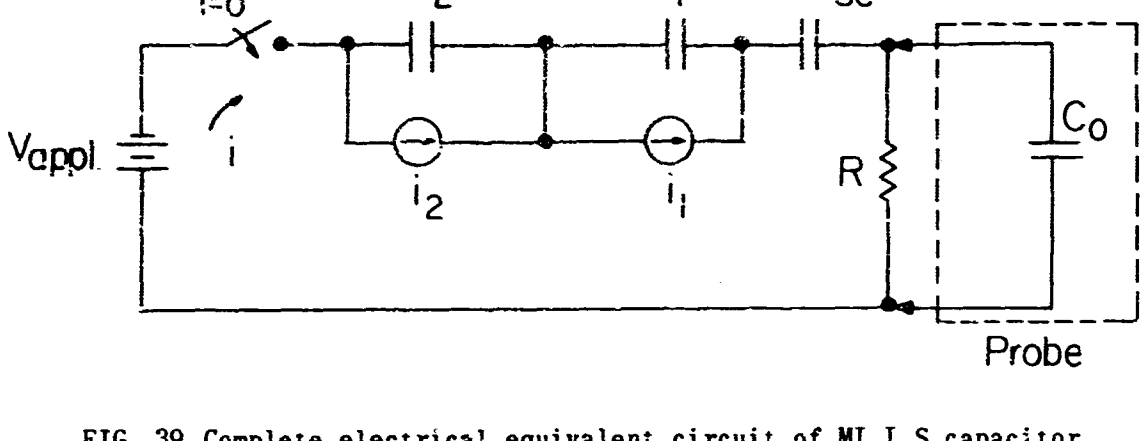


FIG. 39 Complete electrical equivalent circuit of MI_2I_1S capacitor.

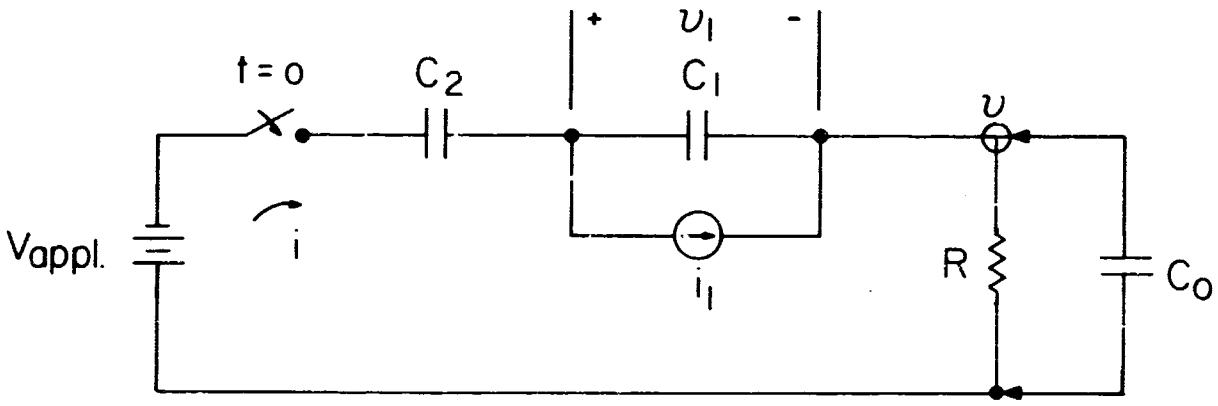


FIG. 40 Simplified electrical equivalent circuit of MI_2I_1S capacitor.

The voltage across insulator I_1 is

$$v_1 = \frac{1}{C_1}\int idt - \frac{1}{C_1}\int i_1 dt \qquad (F2.2)$$

In terms of the circuit elements and the measurable parameter v, this becomes,

$$v_1 = -\frac{1}{R}\left(\frac{1}{C_I} - \frac{1}{C_1}\right)\int vdt - \left(1 + \frac{C_o}{C_I} - \frac{C_o}{C_1}\right)v + V_{appl} \qquad (F2.3)$$

The capacitances C_1 and C_2 may be determined from the device geometry. All that is necessary to display the charging current and voltage is an analog computer that will perform the indicated mathematical operations.

3. An Analog Solution for Transient Current i_1

A system using wideband operational amplifiers was designed that would solve the differential equations for i_1 and v_1 and display them visually on a CRT. This system is shown in Fig. 41.

A generalized operational amplifier arrangement is shown in Fig. 42. The output voltage V_o is related to both inputs V_1 and V_2 by the equation

$$V_o = -V_2\frac{Z_f}{Z_2} + V_1\frac{Z_g}{Z_1 + Z_g}\left(1 + \frac{Z_f}{Z_2}\right) \qquad (F3.1)$$

provided the open loop gain of amplifier A is very large.

The following discussion will outline the functions performed by each amplifier in the analog solution shown in Fig. 41. The first amplifier in the diagram is an input stage that senses the voltage v in the test circuit of Fig. 40. The input impedance of this stage is 33 MΩ in parallel with 6 pF. Therefore, the input resistance presents a negligible load on the test circuit, but the input capacitance is comparable to that of the device. This capacitance is included in the test circuit as C_o. The output of the first stage is

$$v_A = \left(1 + \frac{R_2}{R_1}\right)v \qquad (F3.2)$$

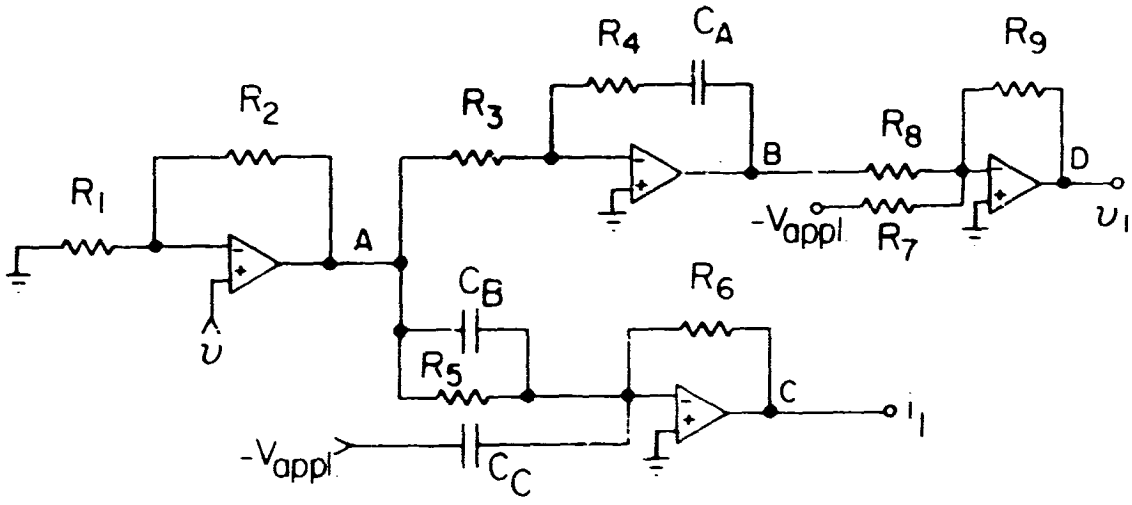


FIG. 41 Analog solution for i_1 and v_1.

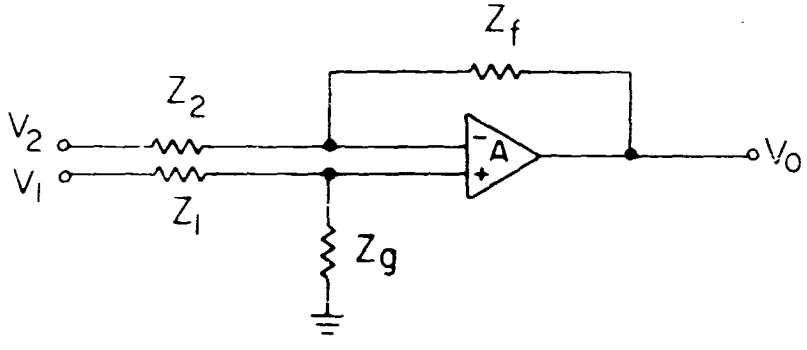


FIG. 42 Generalized operational amplifier.

Resistors R_2 and R_1 may be chosen so as to amplify the signal. This may be desirable, since the remaining operations to be performed will then be immune to noise

The response of the circuit that is generating i_1 is given by the expression

$$v_c = -\frac{R_6}{R_5}\left(1 + \frac{R_2}{R_1}\right)\int v\,dt + \frac{R_9R_4}{R_8R_3}\left(1 + \frac{R_2}{R_1}\right)v + \frac{R_9}{R_7}V_{appl} \qquad \text{(F3.3)}$$

The form of Eq. (F3.3) is equivalent to that of v_1. Again, as for the expression for $-i_1$, these coefficients may be adjusted to give any multiple of v_1.

This analog computer has been assembled with elements that generate $-200\,i_1$ and $2v_1$. The voltage V_{appl} is supplied by a function generator that controls the response time.

4. Device Transient Measurements and Model Correlation

If the device is pulsed with a continuous chain of pulses with the same polarity, then a steady-state situation is reached where the interface between I_1 and I_2 can no longer accept charge. When this condition has been established, the current generator i_1 becomes inoperative, i.e., $i_1 = 0$. For this special case the electrical equivalent may be represented as shown in Fig. 43.

For the case where i_1 is operative, the solution for i_1 as a function of the measurable voltage v_D is

$$i_1 = C_1\left[\left(1 + \frac{C_o}{C_I}\right)\frac{dv_D}{dt} + \frac{v_D}{RC_I} - \frac{dV_{appl}}{dt}\right] \qquad \text{(F4.1)}$$

where

$$C_I = \frac{C_1C_2}{C_1 + C_2}$$

When i_1 is no longer operative, Eq. (F4.1) reduces to

$$\frac{dV_{appl}}{dt} = \left(1 + \frac{C_o}{C_I}\right)\frac{dv_D}{dt} + \frac{v_D}{RC_I} \qquad \text{(F4.2)}$$

In order to minimize the uncertainties in the response of this device, the applied voltage pulse train consisted of well defined truncated ramp functions and the output capacitance was deliberately increased by 100 pF. The solution for v_D with a ramp excitation is

$$v_D = KRC_I\left(1 - e^{-\frac{t}{R(C_o+C_I)}}\right) \qquad \text{(F4.3)}$$

where

K is the slope of the ramp excitation

$K = -0.33 \times 10^6$ V/sec

$R = 10^5$ ohms

The measured response is shown in Fig. 44. The response shown in this figure can be represented by the equation

$$v_D = -0.28\left(1 - e^{-\frac{t}{13.6 \times 10^6}}\right) \qquad \text{(F4.4)}$$

Since R and K are relatively well known it is then possible to determine C_I and C_o using Eqs. (F4.3) and (F4.4).

$$C_I = -\frac{0.28}{KR} = 8.5 \times 10^{-12} \text{ F}$$

$$C_o = \frac{13.6 \times 10^{-6}}{R} - C_I = 127.5 \times 10^{-12} \text{ F} \qquad \text{(F4.5)}$$

The value calculated for C_I corresponds identically to that measured using a capacitance bridge. The value for C_o is higher than the external 100 pF, since it also includes probe capacitance and additional interface capacitance from external circuitry. The primary reason for this calculation is that it establishes the fixed parameters in the model.

By resetting the device with a positive pulse before the application of the negative ramp, it is then possible to observe the effect i_1 has on the measured response v_D. This is shown in Fig. 45. By approximating this response

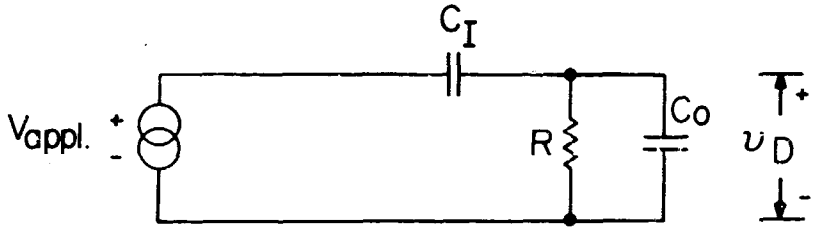


FIG. 43 Electrical equivalent of MI_2I_1S capacitor with i_1 dormant.

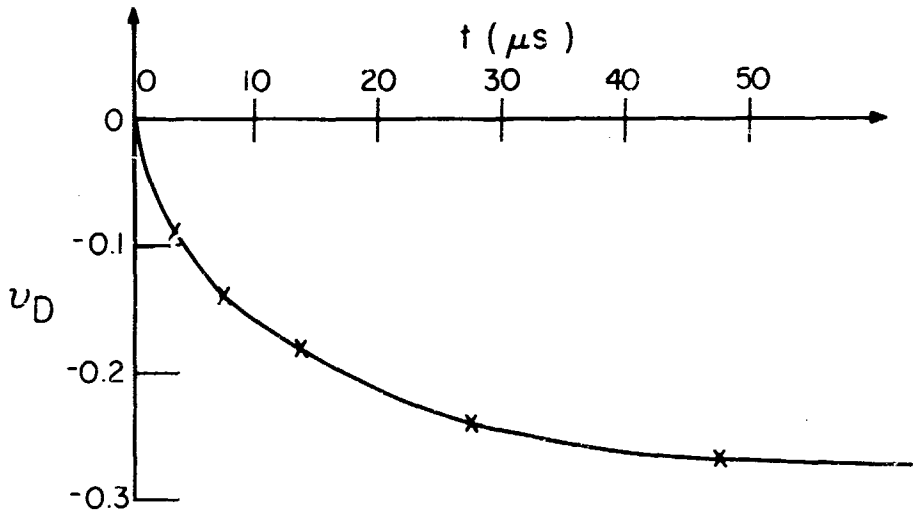


FIG. 44 Measured $v_D(t)$ with a negative ramp input.

by two linear equations it is possible to calculate i_1 and then to compare it with the analog computer solution. For $t < 24$ μsec , v_D may be approximated by

$$v_D \approx -1.13 \times 10^{+4} \qquad \text{(F4.6)}$$

and for $t > 24$ μsec ,

$$v_D \approx -0.27 \text{ V} \qquad \text{(F4.7)}$$

Using these approximations for v_D , i_1 may be calculated using Eq. (F4.1).

$$i_1 \approx 3.06 \times 10^{-6} - 0.266t \qquad t < 24 \text{ μsec} \qquad \text{(F4.8)}$$

$$i_1 \approx -0.26 \times 10^{-6} \qquad t > 24 \text{ μsec} \qquad \text{(F4.9)}$$

A plot of these results is shown in Fig. 46. The $i_1(t)$, as calculated by the analog computer is also shown in this figure. There is a discrepancy between the two curves for $t > 24$ μsec . This could be due to the effect of the approximation made for the mathematical description of v_D and the inexactness of the equivalent components used in the analog computer. In addition, the $i_1(t)$ response from the analog computer was quite noisy.

The analog computer calculation for $i_1(t)$ has involved an equation that contains the time derivative of the transient response $v_D(t)$. When operating in this mode, an operational amplifier is very sensitive to noise. Such noise was quite evident in the analog computer responses observed during the generation of $i_1(t)$. Therefore, it can be assumed that some information has been lost in this analog calculation.

In order to minimize the introduction of noise, all time derivatives should be eliminated from the functions to be performed by the analog computer. In order to accomplish this, $i_1(t)$ must be integrated. This yields a solution for $Q_1(t)$, which should be as informative as $i_1(t)$.

5. The Application of an Analog Computer for the Determination of Charge Centers in the Gate Structure of the Memory Device

Central to the following work is the assumption that the stored charge within the gate structure could be localized in two distinct regions. This is shown in Fig. 47. Here, I_1 represents a very thin insulating layer different from that of either of the thicker insulators, I_2 and I_3 .

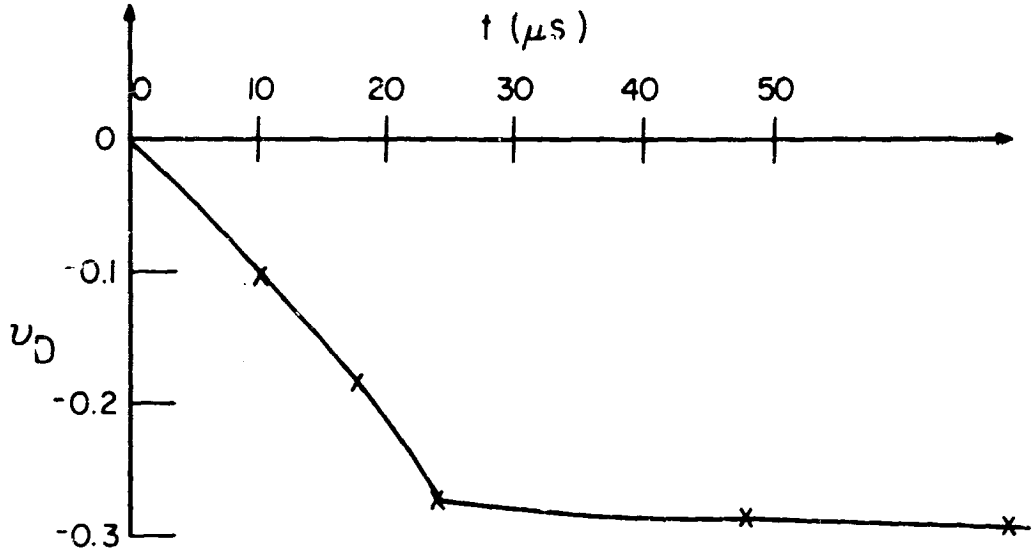


FIG. 45 Response of device to a negative ramp with a positive reset pulse.

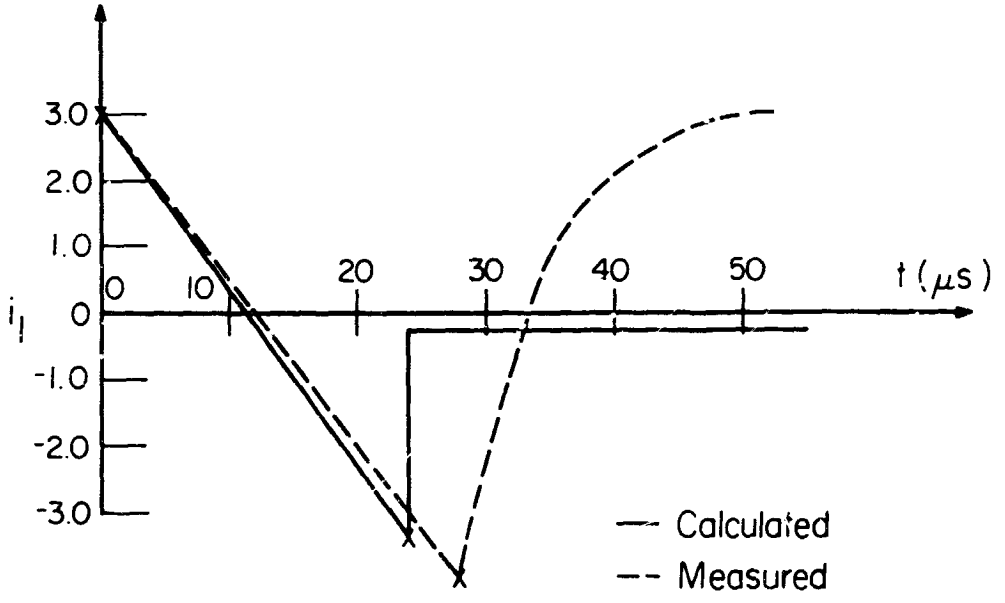


FIG. 46 Calculated and measured i_1 (t).

The following analysis will relate the charge, σ_I and σ_S, stored at the interfaces between these insulating layers to data derived from analog computer measurements and capacitance vs voltage measurements.

In Fig. 47 the insulating layers have been labeled I_1, I_2, and I_3, and the charges accumulating at their respective interfaces are denoted by σ_S and σ_I. This leads to the electrical equivalent shown in Fig. 48.

This electrical equivalent also has loading parameters (R and C_{in}) and a source of excitation V_{appl} included. Each layer is represented by an equivalent capacitance and a charge transfer mechanism (current generators). The semiconductor space charge capacitance has not been included since it was possible to eliminate its effect by the proper interpretation of the test data.

From this circuit it is possible to arrive at a functional relationship between the current generators and the voltage v shown in Fig. 48.

$$\frac{i_3}{C_3} + \frac{i_2}{C_2} + \frac{i_1}{C_1} = \frac{v}{RC_{in}} + \left(1 + \frac{C_{in}}{C_I}\right)\frac{dv}{dt} - \frac{dV_{appl}}{dt} \qquad \text{(F5.1)}$$

where,

$$\frac{1}{C_I} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}$$

If Eq. (F5.1) is integrated, the result is

$$\frac{\Delta Q_3}{C_3} + \frac{\Delta Q_2}{C_2} + \frac{\Delta Q_1}{C_1} = \frac{1}{RC_{in}}\int v\,dt + \left(1 + \frac{C_{in}}{C_I}\right)v - V_{appl}$$

$$\frac{\Delta Q_3}{C_3} + \frac{\Delta Q_2}{C_2} + \frac{\Delta Q_1}{C_1} = f(v) \qquad \text{(F5.2)}$$

The analog computer that was constructed can generate $Kf(v)$, where K is a predetermined multiplying constant. The circuit diagram of this analog computer is shown in Fig. 49.

The information obtained from the output of this analog computer was complemented by a simultaneous measurement of the flatband voltage change when pulsing between two pulse levels.

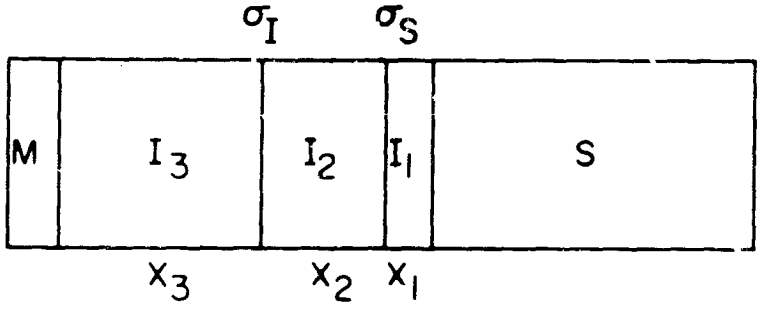


FIG. 47 Three layer model of memory device.

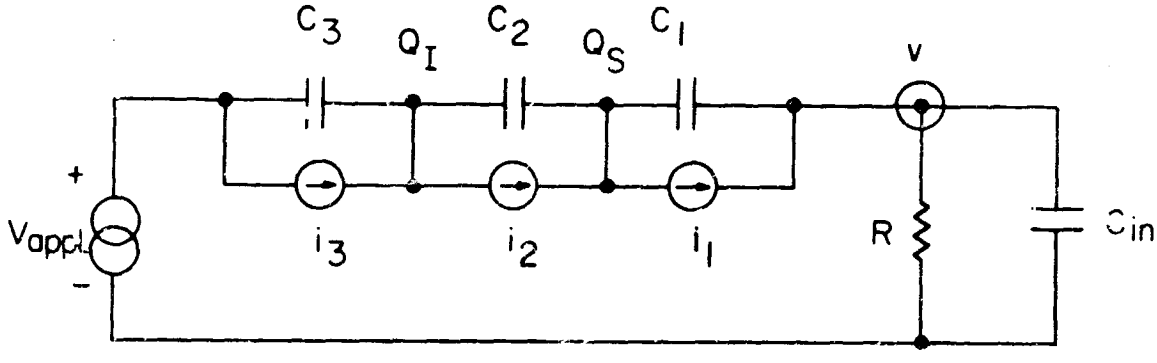


FIG. 48 Electrical equivalent of three-layer device.

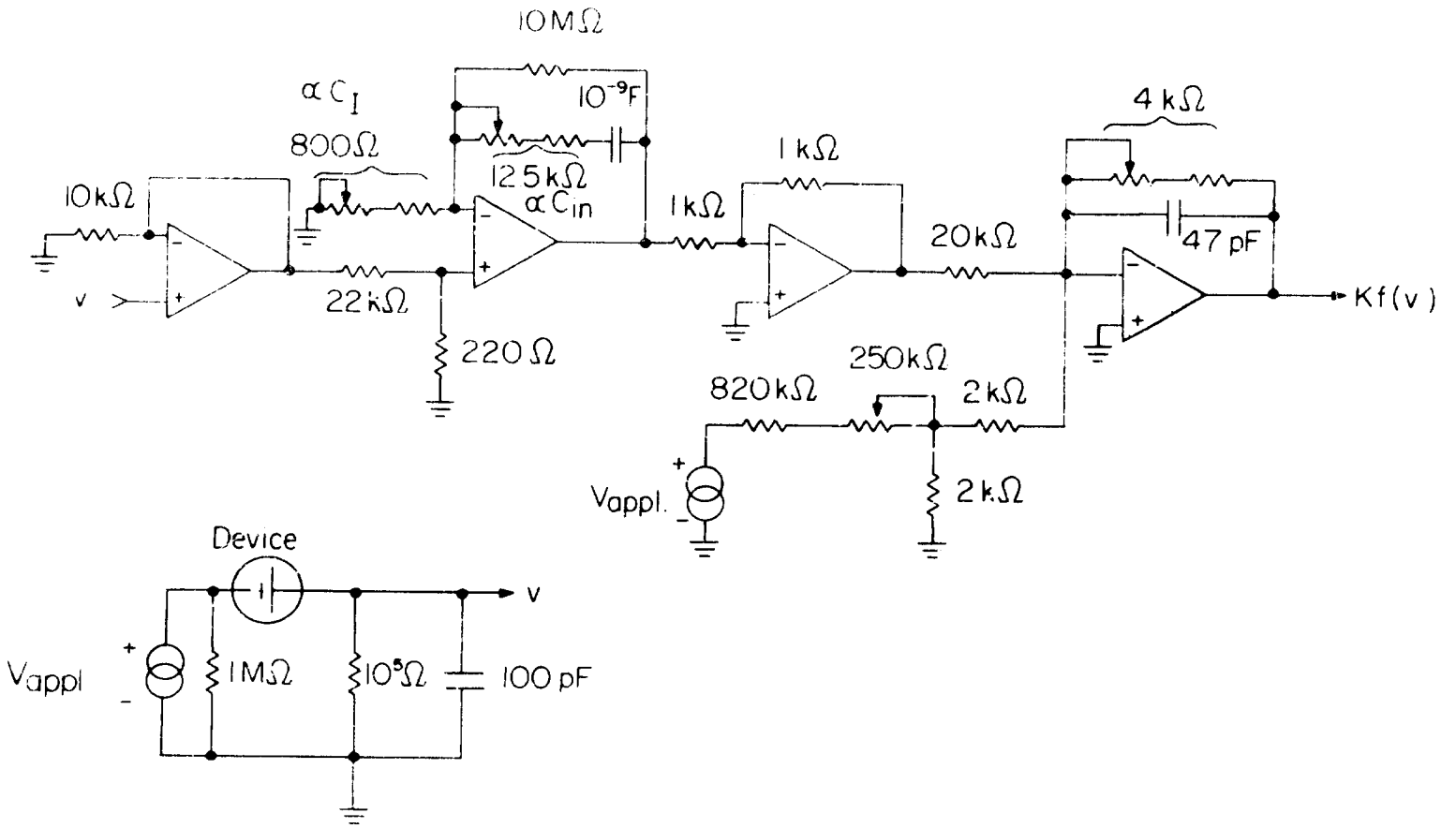


FIG. 49 Analog computer circuit.

$$\Delta V_{FB} = -\Delta Q_S\left(\frac{1}{C_2} + \frac{1}{C_3}\right) - \Delta Q_I\left(\frac{1}{C_3}\right) \qquad \text{(F5.3)}$$

ΔQ_S and ΔQ_I are related to the charge generators by the equations

$$\begin{aligned} \Delta Q_S &= \Delta Q_2 - \Delta Q_1 \\ \Delta Q_I &= \Delta Q_3 - \Delta Q_2 \end{aligned} \qquad \text{(F5.4)}$$

Therefore, ΔV_{FB} may be expressed by

$$\Delta V_{FB} = -\left(\frac{\Delta Q_2}{C_2} + \frac{\Delta Q_3}{C_3}\right) + \Delta Q_1\left(\frac{1}{C_2} + \frac{1}{C_3}\right)$$

$$\Delta V_{FB} = -\left(\frac{\Delta Q_1}{C_1} + \frac{\Delta Q_2}{C_2} + \frac{\Delta Q_3}{C_3}\right) + \frac{\Delta Q_1}{C_I} \qquad \text{(F5.5)}$$

$$\Delta V_{FB} = -f(v) + \frac{\Delta Q_1}{C_I}$$

In order to substantiate the above analysis devices were constructed in which the three layers were deliberately varied to create several special conditions. For each of these cases at least one layer could be classed as an almost perfect insulator. In other words, charge flow occurred through only part of the insulator structure.

The pulsing sequence consisted of setting each device with a prescribed number of -80 V pulses followed by three positive truncated ramps of varying magnitude. A capacitance vs voltage trace was taken after the -80 V pulse sequence and again after the positive pulse sequence. ΔV_{FB} was then determined from these two traces. f(v) was determined from an oscillographic photograph taken from the output of the analog computer while the charging process was occurring during the positive pulse sequence. A typical analog computer response is shown in Fig. 50. Note that most of the charge transfer occurs during the first pulse, since the responses to the second and third pulses overlap each other and are essentially zero. f(v) was measured at the position of maximum deviation, which occurred at the end of the charging pulse.

CASE I - For this condition layers I_2 and I_3 were combined into one layer of silicon nitride. The I_1 layer was formed by a low temperature oxidation. Therefore, the thickness of this layer was estimated to be between 50 and 100 Å. These constraints should yield $Q_2 = Q_3 = 0$, provided no conduction occurs through the nitride.

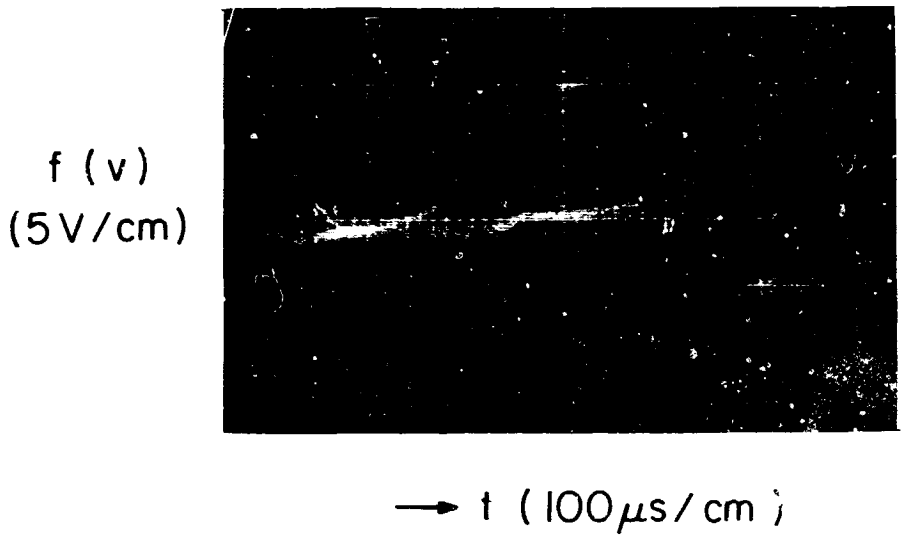


FIG. 50 Typical output response of analog computer.

The results of measurements at three positive pulse levels are tabulated below.

TABLE II

Results for a Case I Device

Pulse	ΔV_{FB}	f(v)	ΔQ_1	x_1
(-80)(+30)	9.7	0.67	8.7×10^{-11}	50 Å
(-80)(+40)	18.2	1.5	16.5×10^{-11}	58 Å
(-80)(+50)	26.6	2.75	24.6×10^{-11}	72 Å

The total device capacitance was measured to be $C_I = 8.4$ pF . From Eq. (F5.5), ΔQ_1 may be calculated.

$$\Delta Q_1 = C_I\left[\Delta V_{FB} + f(v)\right] \tag{F5.6}$$

The results of this calculation are listed in Table II.

From Eq. (F5.2) it was possible to calculate the exact region where the charge is being stored. This is found by solving for C_1 . Since $Q_2 = Q_3 = 0$, Eq. (F5.2) yields

$$C_1 = \frac{\Delta Q_1}{f(v)} = \frac{\epsilon_1 A}{x_1}$$

Then

$$x_1 = \epsilon_1 A \frac{f(v)}{\Delta Q_1} \tag{F5.7}$$

The results of this calculation are also shown in Table II for each pulse level. For the calculation, ϵ_1 has been assumed to be that for silicon dioxide ($K_{\epsilon_1} = 4$); the area of the device is $A = 1.82 \times 10^{-4}$ cm^2.

Therefore, it can be seen that the electrical model corresponds to the physical model. Over a range of three charging levels a calculation of the charge center has yielded values that are within 22 Å of each other. It is

possible to infer from these calculations that the center of charge is tending to move away from the I_1-I_2 interface into the silicon nitride.

CASE II - Devices were fabricated in which I_3 was made nonconductive while I_2 and I_1 were supposedly both conductive. However, the results indicated that I_2 was not as conductive as anticipated. Therefore, a short duration pulse, as was used for these measurements, did not allow enough time for any appreciable charge transfer to occur through this layer. In order to accomplish charging through this layer one could make the I_2 layer more conductive or extend the length of the charging pulse.

The results of the measurements on one of these devices are shown in Table III.

TABLE III

Results for a Case II Device

Pulse	ΔV_{FB}	f(v)	ΔQ_1	ΔQ_2	x_1
(-80)(+50)	7.5	0.75	7.7×10^{-11}	0	63 Å
(-80)(+60)	15	1.5	15.3×10^{-11}	0	63 Å
(-80)(+70)	23	2.5	23.7×10^{-11}	0.3×10^{-11}	65 Å
(-80)(+80)	28.7	3.75	30.1×10^{-11}	1.7×10^{-11}	67 Å

For this case Eq. (F5.2) becomes

$$\frac{\Delta Q_2}{C_2} + \frac{\Delta Q_1}{C_1} = f(v) \qquad \text{(F5.8)}$$

and Eq. (F5.5) becomes

$$\Delta Q_1 = C_I\left[f(v) + \Delta V_{FB}\right] \qquad \text{(F5.9)}$$

From Eq. (F5.9) ΔQ_1 may be determined. By using Eq. (F5.5). ΔQ_2 may be calculated as

$$\Delta V_{FB} = -\frac{\Delta Q_2}{C_2} + \Delta Q_1\left[\frac{1}{C_2} + \frac{1}{C_3}\right]$$

$$\Delta Q_2 = C_2\left[\Delta Q_1\left(\frac{1}{C_2} + \frac{1}{C_3}\right) - \Delta V_{FB}\right] \qquad \text{(F5.10)}$$

Here, C_2 and C_3 were determined from the material properties and the device geometry.

From Eq. (F5.8) x_1 may be determined:

$$\frac{1}{C_1} = \frac{1}{\Delta Q_1}\left[f(v) - \frac{\Delta Q_2}{C_2}\right]$$

$$\therefore \; x_1 = \frac{\epsilon_1 A}{\Delta Q_1}\left[f(v) - \frac{\Delta Q_2}{C_2}\right] \qquad \text{(F5.11)}$$

The results of this calculation are also shown in Table III. The dielectric constant, K_{ϵ_1} was chosen to be that of silicon dioxide.

The location of the center of charge at the I_1-I_2 interface is practically the same for the four pulse levels. There is some charge being stored at the I_2-I_1 interface at relatively large pulse levels.

CASE III - For these devices the I_3 layer was made conductive while the I_2 and I_1 layers were merged into one nonconductive layer of silicon dioxide. Therefore, the charge was transfered into the dielectric from the metal plate. Now, since $\Delta Q_1 = \Delta Q_2 = 0$, Eq. (F5.5) becomes

$$\Delta V_{FB} = -f(v) \qquad \text{(F5.12)}$$

and Eq. (F5.4) becomes

$$\Delta Q_I = \Delta Q_3 \qquad \text{(F5.13)}$$

The expressions for $f(v)$ and ΔV_{FB} are

$$\Delta V_{FB} = - \frac{\Delta Q_3}{C_3} \qquad (F.14)$$

$$f(v) = \frac{\Delta Q_3}{C_3} \qquad (F5.15)$$

The following table includes the results of pulsing one of these devices at three different levels.

TABLE IV

Results for a Case III Device

Pulse	ΔV_{FB}	f(v)
(-60)(+30)	0	0
(-60)(+40)	-2	2
(-60)(+50)	-6.7	6.75

As can be seen from Eq.(F5.12) all that is achieved from these two measurements is an identity relation. It is not possible to solve for the charge transferred, ΔQ_3 , without assuming a value for C_3 . This is equivalent to assuming a location for the charge center, which is what the measurements were, by themselves, attempting to achieve. However, it is possible to make several observations from the results of these measurements.

(1) Since ΔV_{FB} does, in fact, equal $-f(v)$, then the original assumption of no charge flow through the insulator adjacent to silicon is valid. If any charge were being transfered through this layer, it would be calculable from Eq. (F5.5) expressed in the form

$$\Delta Q_I = C_I\left[\Delta V_{FB} + f(v)\right] \qquad (F5.16)$$

(2) Because of the correspondence between $f(v)$ and $-\Delta V_{FB}$, the analog computer has proved to be a practical and accurate tool for the simulation of the charging processes occurring in this unique device.

6. Summary

The physical device model described in previous sections has been translated into its electrical circuit analog. The initial purpose for doing this was to provide an analog solution for the second order differential equation governing

the transient current conditions. In practice it was found that, while this was possible, the output was very noisy due to the final differentiation performed.

An analog circuit to study stored charge vs time (as against the previous current vs time) was therefore developed. The outcome of this work has been a sensitive method for ascertaining the location of stored charge in the dielectric. Indeed, it was found that charges stored at the two interfaces in a three-layer structure could be discriminated. Thus, this approach will be a useful tool in the future study of charge distributions in the dielectric.

G. MEMORY DEVICE FABRICATION

1. Introduction

In this section, an approach to the control of the fabrication of memory devices will be described. Historically, the first devices that exhibited good memory effects were MNS transistors that had been formed after the gate surfaces were cleaned particularly well and the silicon nitride deposited had been particularly well controlled in its oxygen content. As indicated before, early experiments pointed to the fact that a thin oxide layer was an important part of the structure. The absence of oxygen in the nitride near the silicon also had a strong effect (Fig. 22). Later experiments (Fig. 37) indicated that the thickness of the first layer need not be small, and that the materials of the first layer need not be oxide.

Nevertheless, it was decided to investigate the potentialities of the thin-oxide/oxide-free-nitride combination first. There were really three reasons for it. One was the fact that much operational experience had been accumulated with this structure, so that it was easier to proceed to refinements from there. Another was the fact that the sign of the accumulated charge indicated that the thin oxide layer was more conductive than the outside layer and that the conduction mechanism was by tunneling. Since the theoretical form of the J vs E curve for tunneling is more nonlinear than other conduction laws, it promised a larger write-in vs storage time ratio. Finally, storage of the charges nearer the silicon interface would result in a greater proportional effect of the charges on the induced threshold voltage. The major disadvantage, realized from the inception of this work, was the fact that oxide thicknesses in the range from 10 Å to 100 Å were hard to control. In addition, small absolute variations in thickness could be expected to have major effects on the conduction properties of the layer.

The series of experiments that will be described concentrates on the process to form that first layer. Five different approaches were taken to form the thin region adjacent to the silicon, followed by an oxygen-free nitride.

Unfortunately, this oxygen-free nitride has an etch rate of about 10 Å/min under normal KPR processing conditions. In order to make such normal processing at all possible, only 200 Å of this nitride were deposited over the specially formed thin oxide layer. On top of this, 800 Å of oxynitride that had an etch rate of 200 Å/min were deposited in order to permit reasonably high operating voltages. Thus, the test device had really a three-layer structure. The conductivities

of the layers were ranked according to their distance from the silicon surface. Thus, the interface nearest the silicon would collect charges first, and these charges would be most effective. For longer applied pulses, some charge of equal sign to that at the first interface would be collected at the other interface. This defect in esthetics was felt to be unimportant compared to the ease of processing that the three-layer approach facilitated.

2. Test Procedure

Since it was considered too difficult to study the process variations by simple objective tests, it was decided to use the transistor as the test structure. The evaluation procedure of these transistors will now be described.

The transistors used had IGFET characteristics as described in Fig. 7. The experimental variations of the gate dielectric were not expected to affect such characteristics as junction breakdown voltage and transconductance to a material degree, so that these data were only taken for the record. It was the memory switching characteristics that were investigated. These were obtained with the help of a special measuring setup. It consisted of a pulse source that permitted single pulses of both polarities, with amplitudes from 0 to 100 V, to be placed across the gate of the transistor under test. A meter indicated the variable voltage applied in common to gate and drain when 10 μA flowed from the source to the drain of the IGFET. The switching characteristics of each transistor were determined by placing a 10 msec pulse across the transistor and measuring the threshold voltage of the transistor, and then placing the 10 msec pulse of opposite polarity across it and measuring the threshold voltage resulting from that pulse. The results were recorded directly on a plot of applied voltage vs induced threshold voltage, shown in Fig. 51. The applied voltages ranged generally from ±10 to +60 or +100 V. Since only the magnitude of the applied voltage was recorded regardless of polarity a Y-shaped plot was obtained. For a shorthand description of switching behavior, the applied voltage at the node of the Y was used as an index called V_{min}, and the voltages $+(V_{min}+10\ V)$ and $-(V_{min}+10\ V)$ were taken as the index of the amount of switching. In the following subsections, the averages of several fully mounted and bonded transistors will be reported as representative for each slice. The effect of processing steps on write-in time was left to a later stage in the investigation. This decision was taken because equipment capable of measuring this parameter simply and rapidly had not been designed. The transistors, however, are available for this measurement as soon as it is decided upon.

The storage behavior of the transistors was tested by setting their threshold voltages with either a positive or a negative pulse of 60 V and 10 msec, reading the threshold voltages immediately after the setting, and then recording it again after a week. Since more extended data (see Fig. 21) indicate a ΔV vs log t relationship, the results have been put on a comparative basis by calculating the time necessary to lose four of the last five induced volts above the "natural" threshold voltage. A linear extrapolation of these data is also quoted in the following experiments. Thus extreme upper and lower limits of storage can be deduced from these run-of-the-mill transistors.

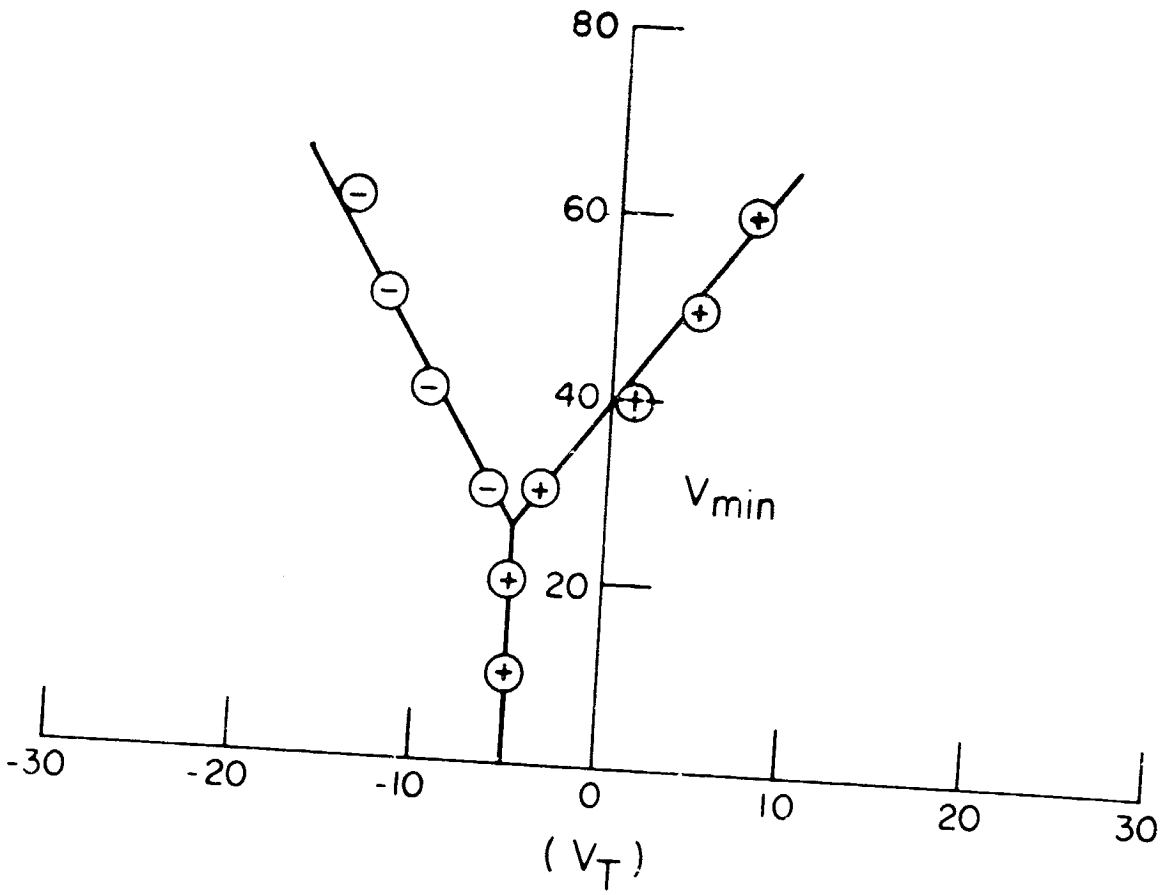


FIG. 51 Routine plot of switching characteristic.

3. Water Oxidation

The purpose of this experimental series was the formation of a reproducible thickness of oxide by boiling water. The method has been used before by Moll, Meyer and Bartelink (Phys. Rev. Letters 7, 87 (1961)). They found that they could reproducibly remove silicon at the rate of 34 Å per boil. This can be recalculated to result in a 71±4 Å oxide layer thickness, depending on the maximum and minimum density of silicon dioxide assumed. In effect, the thickness will be near 75 Å after boiling, and near 67 Å after the heating involved in the deposition of the silicon nitride layers.

The attractive feature of the approach is its self-lining nature: once the maximum oxide thickness has been formed, no further increase in thickness will occur. Second, it is a natural sequel to the final quench used after etching the silicon surface clean. Finally, the boiling condition imposes its own temperature control.

In the experimental tests, the p-diffused source and drain region were formed with the help of silicon nitride diffusion masks. Then the gate region was etched bare with buffered HF. After removal of the photoresist, the slice was briefly dipped in dilute buffered HF, quenched with deionized water, and then boiled for 15 minutes in a clean beaker containing fresh deionized water. The slice was then dried and transferred immediately to the RF heated cold wall reactor for the deposition of the two silicon nitride layers. The following table summarizes the memory data obtained:

Slice No.	V_{min}	$+(V_{min}+10\ V)$	$-(V_{min}+10\ V)$	Extrapolated Storage Time	
				Linear	Logarithmic
M13	45	-3	-13		
M14	43	-6	-18		
M16	30	-15	-23	130 days	10^2 years

These data exhibit a certain element of nonuniformity. In addition, it was found that this approach resulted in a high "natural" threshold voltage due to either some sodium contamination from the water or the beaker, or due to high surface state density, perhaps originating from the hydrogen atoms introduced by the water.

4. Thermal Oxidation at 900°C

The purpose of this approach was the formation of a reproducible thickness of oxide at 900°C. The oxidation chamber was the RF heated cold wall reactor used for silicon nitride deposition. Since 100% oxygen causes rapid oxidation at 900°C, a slower oxidant was needed to permit a limitation of the oxide thickness to a value of 50 Å. This reduction in oxidation rate was achieved by using 1% N_2O in argon as the oxidant. In preliminary experiments it was found that 5 minutes of exposure to 1% N_2O would result in the formation of roughly 50 Å of silicon dioxide at 900°C.

In order to eliminate the first thin oxide formed on the silicon by quenching in water and air oxidation, the bare gates of the transistors were briefly etched in halogen hydride before the oxidation was started. The following memory data are representative:

Slice No.	V_{min}	$+(V_{min}+10\ V)$	$-(V_{min}+10\ V)$	Extrapolated Storage Time	
				Linear	Logarithmic
J-113-2	53	-3	- 8.5		
D-125-B	50	-1	-13		
D-143-B	41	-3	-11	130 days	10^2 years

The results of these runs appear reasonably reproducible.

5. Thermal Oxidation at 750°C

The purpose of this experimental series was to form the oxide layer at a lower temperature. The reactor was a hot wall furnace that had been equipped for the deposition of silicon nitride. Preliminary experiments established that it required 30 minutes at 750°C for form 50 Å of oxide in 100% O_2. In this furnace halogen etching could not be performed, so that the oxide was grown in addition to the initial thin oxide layer formed after HF etching.

The results of this series are indicated in the following table.

Slice No.	V_{min}	$+(V_{min}+10\ V)$	$-(V_{min}+10\ V)$	Extrapolated Storage Time	
				Linear	Logarithmic
L16-2	26	-8	-14		
L18-1	45	+2	-13		
L18-2	33	-2	-10		
L21-1	45	-6	-13	120 days	10^2 years

The reproducibility of this series is poor. It is probably associated with the lack of reproducibility in the formation of the spontaneous oxide layer.

6. The Use of Pyrolytic Oxide

In this experimental series, the effect of a silicon oxide layer formed by the pyrolysis of silane and nitrous oxide on the memory characteristics was

investigated. Preliminary experiments showed that at 750°C, 1% N_2O in argon and 1% SiH_4 in argon would result in a growth rate of 3600 Å of oxide per hour. The 50 Å of oxide required therefore only 50 seconds. Again, as in the previous series, no surface treatments were made immediately before deposition. The results of this series are given in the following table.

Slice No.	V_{min}	$+(V_{min}+10\ V)$	$-(V_{min}+10\ V)$	Extrapolated Storage Time	
				Linear	Logarithmic
L6-1	45	-5	-7		
L14-1	45	-10	-9		
L16-1	40	-5	-9		
L12-1	45	+4	-3		
L12-2	45	-1.5	-5	30 days	30 days

It is clear that these data show the worst results of the slices so far. The amount of shift and even the polarity of the shift are unpredictable. On top of that, the storage life is significantly lower than that of preceding runs.

7. Silicon Oxynitride Interface

The experiments described in Fig. 22 had indicated that oxynitride in contact with a thin layer of oxide would not permit any shifts. It was surmised that perhaps both materials combined to act, at least electrically, as one layer under the conditions of the experiments. For this reason, a 50 Å layer of oxynitride was deposited over a freshly cleaned silicon surface in the hot wall furnace at 750°C. The deposition rates of all nitrides were 60 Å/min. The final structure had a 50 Å oxynitride, 200 Å pure nitride, and 800 Å oxynitride gate.

Slice No.	V_{min}	$+(V_{min}+10\ V)$	$-(V_{min}+10\ V)$	Extrapolated Storage Time	
				Linear	Logarithmic
L22-1	45	-7	-2		
L22-2	45	-7	-2		
L23-1	45	0	+5		
L23-2	45	0	+7	100 days	10^2 years

The main feature of this series is the reverse polarity shift observed, indicating that this oxynitride layer is acting as a bulk conductor rather than as a tunneling barrier. The results are the best obtained in the hot wall furnace. The storage life is comparable to that observed in all the other cases, except for the poor last one.

8. Summary

The first conclusion of this experimental series is that the treatment of the silicon-insulator interface is indeed an important aspect of the processing for the memory device. However, a tighter control must be achieved, since even within a series, a fairly large amount of variation existed. Surprisingly, in four out of five series the storage life was essentially the same and incidentally, quite high. A somewhat more tentative conclusion concerns the details of the treatments. It appears that better control was achieved when something was done deliberately to the "natural" or water-grown oxide layer. This is exemplified by boiling water treatment, where the water-grown layer was brought reproducibly to the same thickness. This is done also for the 900°C cold wall thermal oxidation, where the initial oxide was removed by halogen hydride etching. Finally, the last series points in the same direction, since here the effect of this layer was obliterated by its combination with the silicon oxynitride layer.

H. "STUDENT-001," AN ADAPTIVE CIRCUIT USING THE MNS-VTT

1. Introduction

An electronic device, "Student-001," has recently been constructed at SRRC to demonstrate the use of MNS-VTT's in circuits which appear to learn from experience. Upon being stimulated by the flipping of a switch, the device responds by lighting or not lighting a lamp. The device is then rewarded or punished (through switches) in order to encourage or discourage its most recent response.

Although no systematic experiments have been performed, several users of the devices have been convinced that Student-001 learns, statistically speaking, from its training. The object of this section is to provide a mathematical model for the device and its immediate logical extensions whereby their behavior may be predicted quantitatively and experiments may be designed and further extensions planned.

2. Theory of the Single-Light Student

A single-light Student's light is controlled by a variable threshold transistor whose input voltage, v_i, is randomly generated by a sampling circuit (activated by the "stimulate" switch) from a uniform distribution between v_{il} and v_{iu}. The light is turned on if the input voltage is exceeded by the threshold voltage v_t, that is, if $v_i < v_t$. The threshold voltage is altered by the "reward" and "punish" switches in such a direction as to raise or lower the probability of repeating the most recent response.

For instance, if the desired condition is "on", then an "on" response will lead to "reward", which raises v_t, say by an amount Δ^+v_t; if the response was "off", then the "punish" switch would likewise raise v_t by Δ^+v_t. Conversely, if the desired condition is "off", then responses of "on" and "off" will lead respectively to punishment and reward, both of which will lower v_t by an amount Δ^-v_t. The threshold voltage ranges between upper and lower limits, v_{tu} and t_{tl}. These may be adjusted to lie outside, on, or inside the corresponding input voltage limits; suppose for the present that they coincide, that is, $v_{tu} = v_{iu}$, and $v_{tl} = v_{il}$; suppose also that $\Delta^+v_t = \Delta^-v_t = \Delta v_t$.

Let the instantaneous state of the student be described by the probability that at the next stimulus the light will go on, let the goal variable g be defined by

$g = 0$ if desired response is "off"

$g = 1$ " " " " "on"

and let $m = \frac{v_{iu} - v_{il}}{\Delta v_t}$. Here v_t is being quantized to m discrete values. But since

$$p = \Pr\{v_i < v_t\}$$

$$= \Pr\{v_{il} \le v_i < v_t\} \qquad \text{(H2.1)}$$

and by the assumption of the uniform distribution of v_i over the interval $[v_{il}, v_{iu}]$,

$$p = \frac{v_t - v_{il}}{v_{iu} - v_{il}} \qquad \text{(H2.2)}$$

Thus p also is being quantized into m intervals, each defining a state of the Student.*

* A continuous analysis of student could certainly be carried out. The quantization, however, is convenient for the sake of exposition, especially in the multi-dimensional case

A state diagram of a one-dimensional Student is shown in Fig. 52(a). Figure 52(b) depicts the same Student, but with

$$v_{tu} - v_{iu} - \Delta v_t \tag{H2.3}$$

and

$$v_{tl} = v_{il} = -2\Delta v_t \tag{H2.9}$$

It is clear that the state-diagram structure remains essentially the same. Only the number of states and their labels are altered when the equality of voltage limits is relaxed.

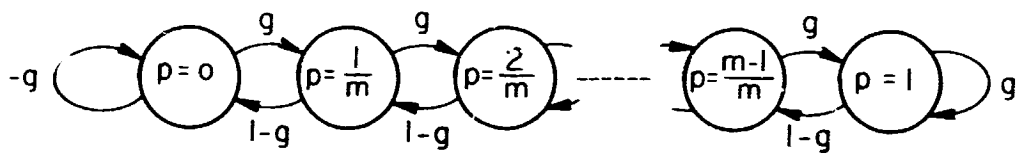
Another useful description of these processes is the transition matrix shown below:

$$M = \begin{vmatrix} 1-g & g & o & o & & o \\ 1-g & o & g & o & & o \\ o & 1-g & o & g & & \\ o & o & 1-g & o & & \\ & & & & & \\ & & & & o & g \\ o & o & & & 1-g & g \end{vmatrix}$$

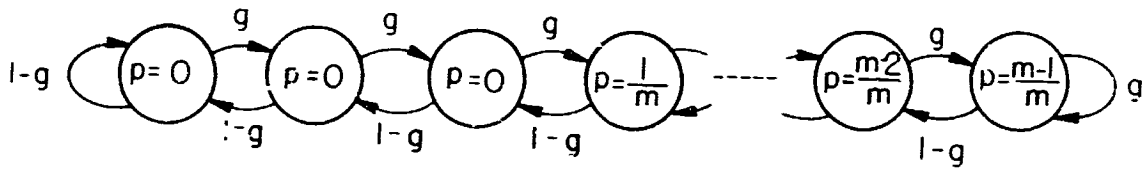
Since all entries are non-negative, and the row sums are unity, this matrix describes a stochastic (or Markov) process. For g either 0 or 1, the process has an absorbing state (either the first or last), and is connected to this state in the sense that it can be reached from any other. These conditions are sufficient to confirm mathematically the physical intuition that if g is held fixed for a sufficiently long time (at most m trials), the process will enter and remain in the intended stable or absorbing state. In other words, the Student will learn in m or fewer lessons. Obviously, the fastest learning occurs when $m=1$.

3. Stochastic Training

The foregoing analysis has dealt with a perfect, or deterministic trainer. It is also instructive to consider an imperfect or stochastic trainer who sometimes errs in his reward or punishment of the Student. The simplest instance of such a trainer is one who errs with probability $1-q$ independent of his history (and of the Student's state). In terms of the



(a)



(b)

FIG. 52 (a) State diagram of one-dimensional student.
(b) Student of (a) with unequal input and threshold voltage limits.

above transition matrix M, in training the light to go on the trainer at each trial chooses $g=1$ with probability q.

It can be shown algebraically that for the purpose of this analysis this is equivalent to substituting $g=q$ in the transition matrix. Heuristically, this is explained by saying that instead of flipping a biased coin to determine which deterministic transition matrix is to apply to the Student, the trainer lets the Student flip the coin and make the appropriate transition.

With $g=q$, it can be confirmed that the stationary distribution vector $\underline{s}$ of M, that is, the unique state occupancy distribution $\underline{s}$ such that

$$\underline{s}M = \underline{s}$$

is given by

$$\underline{s} = \frac{1-r}{1-r^{m+1}}\left(r^m, r^{m-1}, \ldots, r, 1\right), \qquad r = \left(\frac{1-q}{q}\right) \tag{H2.5}$$

This means that after many trials the Student achieves a geometric state occupancy distribution, whose maximum occurs in the last state (the desired one, where the probability of lighting is $p_m = m/m = 1$), with a value there of

$$\frac{1-r}{1-r^{m+1}}, \qquad r = \left(\frac{1-q}{q}\right) \tag{H2.6}$$

The probability, after many trials, of the Student responding correctly is found by summing the correct response probabilities in all the states, weighted by the state occupancy probabilities:

$$\bar{p} = \sum_{i=0}^{m} p_i s_i = \sum_{i=0}^{m} \frac{i}{m} \frac{(1-r)r^{m-i}}{(1-r^{m+1})}$$

$$= \frac{m - (1+m)r + r^{m+1}}{m(1-r)(1-r^{m+1})} \tag{H2.7}$$

If $q > \frac{1}{2}$, then $r<1$, and the probability $\bar{p}$ of correct response approaches 1 as m grows large. This implies that by choosing the size of the threshold voltage increment sufficiently small, the Student will learn to give the correct response with probability arbitrarily close to 1 after many trials, despite the

imperfect trainer. (The trainer, however, may not be so imperfect as to train incorrectly more often than not.) The formula given above for p can be used to find the Student's ultimate probability of correct response for arbitrary voltage quantization and trainer error rate.

4. Electrical Description of "Student"

The one-light Student, Fig. 53, comprises three sections. These are the stimulus generator, the threshold storage and response unit, and the conditioning unit. The stimulus generator consists of a free-running multi-vibrator with a frequency of approximately 1 kHz whose output is clipped and used to drive a series RC integrating circuit. The capacitor in the integrating circuit is connected to the resistor through a double-throw switch marked "Stimulate". When this switch is thrown, the voltage appearing across the capacitor is held and transferred to the input of the threshold storage unit.

The threshold storage and response unit contains an MNS-VTT at its input. This element stores the present learning threshold in its gate insulator region. If the voltage transferred to its gate by the stimulus switch is more negative than the stored threshold, the MNS-VTT conducts and sets a flip-flop. The output of the flip-flop drives a lamp which indicates conduction of the MNS-VTT when it is illuminated. A second transfer contact on the "Stimulate" switch is used to prevent the flip-flop from being set when the voltage-transfer capacitor is not connected to the MNS-VTT gate. The flip-flop is reset by momentarily activating its resetting input by means of the "Reset" switch.

The conditioning unit contains logic elements which determine whether to apply a positive or a negative pulse to the gate of the MNS-VTT, based upon whether an impulse is received from the "Reward" or "Punish" switches and upon the state of the flip-flop. These pulses are shaped by single-shot multivibrators and used to drive the common-output, high-voltage pulse generators which connect to the MNS-VTT gate circuit.

Power is supplied to Student by internal batteries and is controlled by the two power switches on the front panel.

5. Summary

As an example of the adaptive use of the memory transistor, a "learning machine" was constructed. This circuit utilizes the controllable threshold voltage of the MNS-VTT to determine its electrical output. The setting of the threshold voltage is influenced by "reward" and "punish" stimuli. The theory of this circuit application has been developed. It is shown that learning can be accomplished even by an imperfect trainer. The only requirement of the trainer is that he make more correct decisions than incorrect decisions.

Two of these circuits have been built, of which one was delivered to the NASA Electronics Research Center. Both of these circuits have proven themselves in many successful tests.

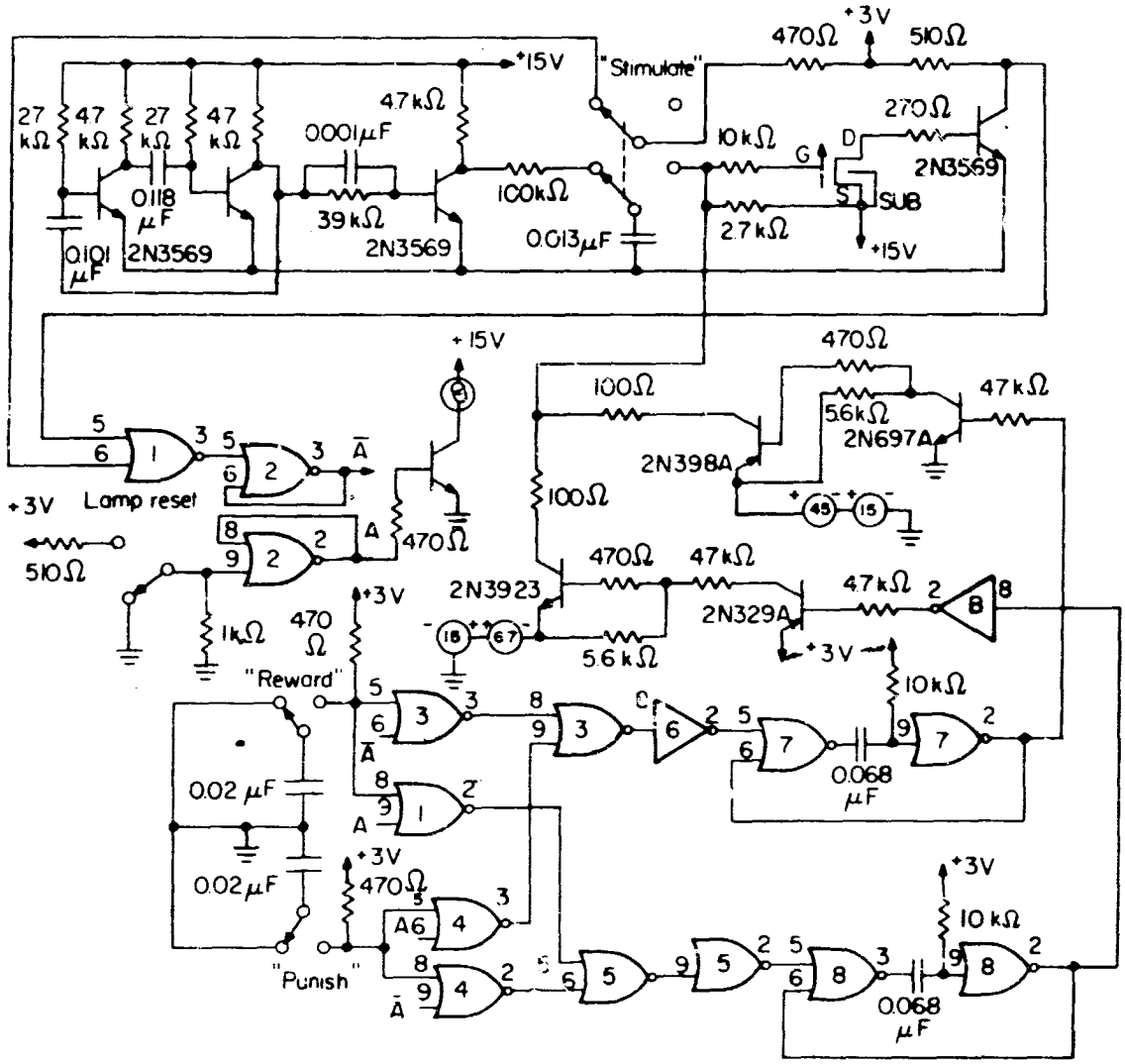


FIG. 53 Student 1-001 MNS-VTT learning system.

SECTION IV

PROGRAM SUMMARY

This report is a record of the progress made in the performance, fabrication, and understanding of the new nonvolatile semiconductor memory element - the MNS variable threshold transistor. The work had been sparked by the observation of a seemingly undesirable "instability". Measurements confirmed that this "instability" was in fact a reproducible effect. Capacitors and transistors incorporating this reproducible behavior promised potential new and unique application. It was at this point that the work reported here was begun.

We have now established the theory of this new device. We know how to fabricate it and have determined directions in which to refine the processing.

The equivalent circuit of the device will permit its more exact application in various design problems. A body of measurements has been established that delineates the range of capabilities of the new device. Concurrently, test techniques and equipment that permit a definitive description of the device have been acquired. In the course of this work, upwards of 100 memory transistors, embodying at least six different fabrication processes, have been delivered to the NASA Electronic Research Center. Finally, an adaptive circuit application, the "learning machine", was built and delivered to NASA.

It is clear, in retrospect, that we have developed an entirely new class of devices of which the memory transistor is only a particular embodiment. This new class depends on a thin film insulator as an electrically active part of the device operation. In response to the importance of the conductive properties of thin film insulators, increasingly exact work in the measurement and formation of these electrically active thin film insulators is required and is planned in the future.

SECTION V

RECOMMENDATIONS FOR FUTURE WORK

In the continuation of this program we propose to refine the results obtained from the previous work on the memory device. We will investigate in greater detail the physics of the device, with the purpose of utilizing our improved understanding in the design of devices that have better electrical performance and a wider range of applications. The study will be strongly supported by experimental work to substantiate this improvement. These experimental results will also lay the groundwork for the continuing betterment of device reproducibility. Finally, the feasibility of using the devices as a memory for signals other than electrical pulses will be evaluated.

Specifically, we will study the effect of the composition of silicon nitride on its conduction properties, and we will establish the process refinements that will result in the most reproducible properties. We will investigate theoretically and experimentally the effect of interfaces occurring in the double-layer memory structure on the operation of the device. We will study the spatial distribution of charges on the dielectric of the memory structure in order to minimize power dissipation and increase storage time.

Finally, we will study the effect of infrared, visible, and ultraviolet light on charge movement in the dielectric of the memory device.

APPENDIX A

TRANSIENT CONDITIONS

APPENDIX A

TRANSIENT CONDITIONS

We will start by recalling that the charge density accumulated at the insulator-insulator boundary is assumed to be distributed over such a small distance that it can be considered a delta function distribution. As the charge accumulates, then, it does not vary with respect to distance, but only with respect to time. We can therefore write

$$\sigma = \int j_I dt \tag{A1}$$

The most important condition during the time that such a current j_I flows is that of current continuity across both regions 1 and 2. This is expressed by

$$\frac{dD_1}{dt} + j_1 = \frac{dD_2}{dt} + j_2 = j \tag{A2}$$

where j is the total current flowing into and out of the MI_2I_1S capacitor electrodes; dD_1/dt and dD_2/dt are the displacement currents due to the charge build-up across regions 1 and 2, respectively; and j_1 and j_2 are the conduction currents flowing in regions 1 and 2.

In the exploration of the current vs time relationships that follow, the functional dependence j_1 vs E_1 will not be specified; the correct formulas can be inserted when they become known. The structure we will consider is an MI_2I_1S capacitor with the silicon having n-type conductivity. A resistor will be placed in series with it. We will, for the present, assume that the initial flatband voltage of the capacitor is zero. When we apply a negative bias to this capacitor we have the added complication that a space charge region forms in the reverse-biased n-type substrate. We will start by allocating potential drops across the various parts of the circuit indicated in Fig. A-1. We can write

$$V_{appl} = V_2 + V_1 + V_{SC} + V_R \tag{A3}$$

where $V_1 = E_1X_1$, $V_2 = E_2(x_o - x_1)$, and $V_R = IR = jAR$, where A is the area of the MI_2I_1S capacitor.

We will now derive V_{SC} in terms of E_1. D-vector continuity at the silicon-insulator interface demands that:

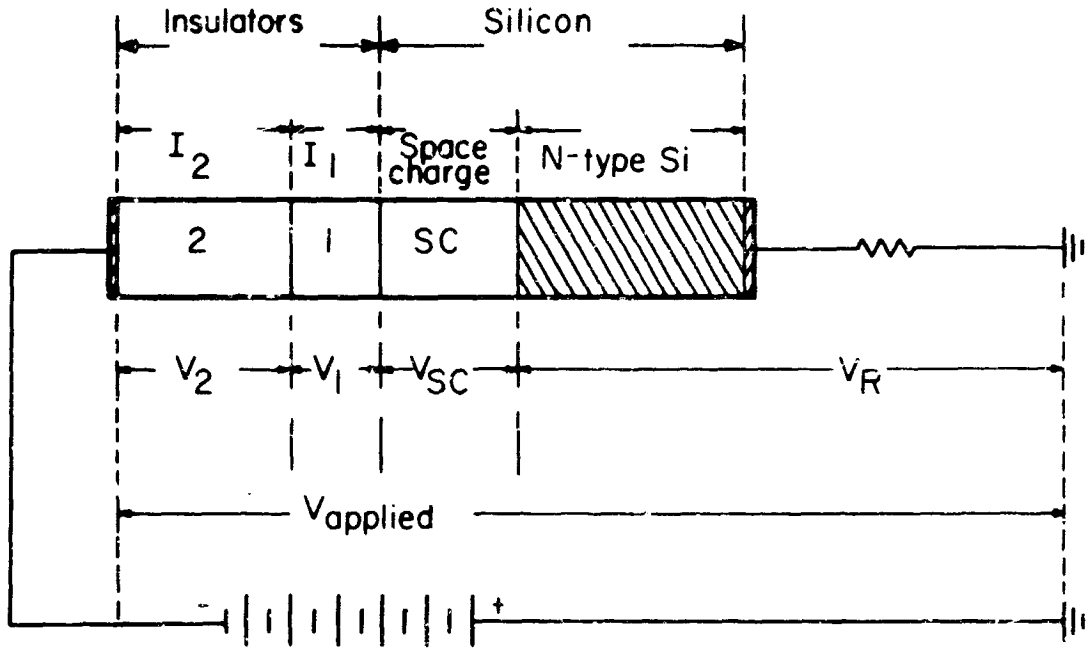


FIG. A-1 Potential drops across $RC_{MI_2I_1S}$ circuit.

(a) $D_1 = D_{SC}$ and $E_{SC} = E_1 \frac{\epsilon_1}{\epsilon_{Si}}$ (A4)

(b) The space charge region in the silicon obey Poisson's equation

$$\frac{dE_{SC}}{dx} = \frac{eN_D}{\epsilon_{Si}}$$

where N_D is the excess donor density in the silicon substrate.

(c) Integrating, we find

$$E_{SC} = \frac{e}{\epsilon_{Si}} \int_o^w N_D \, dx = \frac{eN_D}{\epsilon_{Si}} w$$

where w is the width of the space charge region.

(d) Integrating again to obtain V_{SC}, we obtain

$$V_{SC} = \int_o^w \int_o^x \frac{eN_D}{\epsilon_{Si}} dx = \frac{eN_D}{2\epsilon_{Si}} w^2$$

(e) Substituting (c) into (d) to eliminate w, we find

$$V_{SC} = \frac{eN_D}{2\epsilon_{Si}} \left(\frac{E_{SC}\epsilon_{Si}}{eN_D}\right)^2 = \frac{E_{SC}^2 \epsilon_{Si}}{2eN_D}$$

(f) From (a), $V_{SC} = \frac{E_1^2 \epsilon_1^2}{2\epsilon_{Si} eN_D}$.

To express j in terms of E_1 we will make use of the continuity condition, Eq. (A2). To express E_2 in terms of E_1 we also make use of the continuity equation

$$\frac{dE_2}{dt} = \frac{j_1 - j_2}{\epsilon_2} + \frac{\epsilon_1}{\epsilon_2} \frac{dE_1}{dt} \quad (A5)$$

$$E_2 = \int \frac{j_1 - j_2}{\epsilon_2}\, dt + \frac{\epsilon_1}{\epsilon_2} E_1$$

Now we can express everything in terms of E_1, except that the dependence of $j_2 = f(E_2)$ results in the fairly cumbersome

$$j_2 = f\left[\frac{\epsilon_1}{\epsilon_2} E_1 + \int\left(\frac{j_1 - j_2}{\epsilon_2}\right) dt\right].$$

Going back to Eq. (A3) and expressing everything in terms of E_1, we find

$$V_{appl} = E_1 x_1 + \int \frac{j_1 - j_2}{\epsilon_2}\, dt - \frac{\epsilon_1}{\epsilon_2} E_1 (x_o - x_1)$$

$$+ \frac{1}{2}(\epsilon_1^2 / 2\epsilon_{Si} e N_D) + (j_1 + \epsilon_1 \frac{dE_1}{dt}) AR \qquad (A6)$$

In order to eliminate the integral sign, this equation can be differentiated with respect to time. This results in a nonlinear, inhomogenous equation of second order in E_1, which up to now could not be solved.

The term containing E_1^2 is important only if the total transient takes less than about 10 msec in capacitors; in transistors even shorter transients make it unimportant. Any longer time will permit the formation of an inversion layer at the semiconductor-insulator interface, with the attendant partial disappearance of the carrier-free space charge region.

For positive applied voltage, the same steps leading to Eq. (A6) are applicable, except that an accumulation region forms at the silicon-insulator interface instead of a space charge region. This means that in (A6) the term containing E_1^2 is again zero. Therefore, for a positive applied voltage

$$V_{appl} = E_1 x_1 + \left(\int \frac{j_1 - j_2}{\epsilon_2}\, dt + \frac{\epsilon_1}{\epsilon_2} E_1\right) (x_o - x_1)$$

$$+ (j_1 + \epsilon_1 \frac{dE_1}{dt}) AR \qquad (A7)$$

This equation presents nearly the same difficulties as (A6), and has not been solved to date.

We can, however, obtain information about the time dependence of the total transient current j by realizing that it consists of a displacement current that simply charges up the capacitor and a conduction current that transfers charge into the insulator.

Let us assume first that the conduction current into the insulator is very high for any given E_1, so that as the voltage builds up across the MI_2I_1S capacitor the conduction current into the insulator keeps pace, so that, instantaneously, the temporary voltage across the capacitor V_{appl} fulfills the condition

$$V'_{appl} = V'_{FB} + V'_{min}$$

The final V_{FB} will then be reached at once when the MI_2I_1S capacitor has been completely charged. Therefore, the shortest time to build up V_{FB} is that determined by the $RC_{MI_2I_1}$ time constant.

For a positive bias, with no space charge region formed in the silicon, the transient current simply follows the trivial

$$i = \frac{V_{appl}}{R} \exp\left(-t/RC\right) \tag{A8}$$

For a negative bias the space charge region complicates matters. We can use Eq. (A6), setting currents j_1 and j_2 equal to zero. Then

$$V_{appl} = E_1\left[x_1 + \frac{\epsilon_1}{\epsilon_2}\left(x_o - x_1\right)\right] + E_1^2\left(\epsilon_1^2/2\epsilon_{Si}eN_D\right) + \epsilon_1 AR\frac{dE_1}{dt} \tag{A9}$$

Clearly, the variables can be separated to form

$$\int \frac{dE_1}{V_{appl} - E_1x'_o - E_1^2\beta^2} = \frac{t}{\epsilon_1 AR} + \text{constant} \tag{A10}$$

where

$$x_o' = x_1 + \frac{\epsilon_1}{\epsilon_2}\left(x_o - x_1\right)$$

$$\varepsilon^2 = \varepsilon_1^2/2\epsilon_{Si}\ eN_D$$

which can be converted into

$$\frac{dE_1}{E_1 - B^2E_1^2 - E_A} = -\frac{t}{RC} + \text{constant} \qquad \text{(A11)}$$

where

$$B^2 = \varepsilon^2/x_o'$$

$$C = \frac{\epsilon_1 A}{x_o'}$$

$$E_A = \frac{V_{appl}}{x_o'}$$

This can be solved with a standard solution from a table of integrals and from the condition that $E_1 = 0$ when $t = 0$. The solution is

$$E_1 = \frac{1}{2B}\left[\frac{K(1 + M\exp(-Kt/RC)}{1 - M\exp(-Kt/RC)} - 1\right] \qquad \text{(A12)}$$

where

$$K = (4E_A B + 1)^{1/2}$$

$$M = (1/K) \ln [(1 - K)/(1 + K)]$$

From this equation, the build-up of other fields or voltages can be derived. The transient current is of course $i = \epsilon_1 A \frac{dE_1}{dt}$, which results in

$$i = \frac{x_o'}{R}\left(E_A - \frac{1}{4B}\left\{\left[\frac{K(1 + M \exp(-Kt/RC)}{1 - M \exp(-Kt/RC)}\right]^2 - 1\right\}\right) \qquad (A13)$$

Let us assume now that the current density due to conduction in the dielectric is very low, so essentially all charge transfer occurred only when the MI_2I_1S capacitor was charged up long ago. In that case, the iR drop due to the series resistor can be neglected, and we can write

$$V_{appl} = E_1 x_1 - \left(\frac{j_1 - j_2}{\epsilon_2} dt + \frac{\epsilon_1}{\epsilon_2} E_1\right)(x_o - x_1) \qquad (A14)$$

which can be differentiated with respect to d/dt and rearranged to form

$$\int_{E_1 f}^{E_A} dE_1 \Big/ \left[j_1(E_1) - j_2(E_1)\right] = t \Big/ \left(\frac{\epsilon_2 x_1}{x_o - x_1} + \epsilon_1\right) \qquad (A15)$$

for positive bias. For negative bias, similar manipulations result in

$$\int_{E_1 f}^{E_a} \left[1 + E_1 \epsilon_1^2 / \epsilon_{Si} e N_D x_o\right] dE_1 \Big/ \left[j_1(E_1) - j_2(E_1)\right] = t \Big/ \left(\frac{\epsilon_2 x_1}{x_o - x_1} + \epsilon_1\right)$$

by setting the last term in Eq. (A6) equal to zero. (A16)

When we apply these equations to a real MI_2I_1S structure we have to consider the presence of a charge already present, which can be expressed as a flatband voltage V_{FB}. If V_{FB} is negative and V_{appl} is negative, the capacitor will behave as a simple MI_2I_1S capacitor until E_1 is zero; that is, for a short time t' defined by

$$V_{FB} = V_{appl}[1 - \exp(t'/RC)]$$

Then depletion sets in, and Eqs. (A6), (A9), or (A12) apply, with V_{appl} in those equations replaced by $(V_{appl}-V_{FB})$.

If V_{FB} is negative and V_{appl} is positive, both fields add to E_1 and we can replace V_{appl} in Eqs. (A7), (A8), and (A11) by $(V_{appl}+V_{FB})$. If V_{FB} is positive and V_{appl} is negative, both fields again add to form E_1. But since we started from a static condition, an inversion layer has already formed that neutralized the field due to V_{FB}. Therefore, Eqs. (A6), (A8), and (A12) can be used without modifying V_{appl}. Finally, if V_{FB} is positive and V_{appl} is positive, Eqs. (A7), (A8) and (A11) can be used in replacing V_{appl} with $(V_{appl}-V_{FB})$.