AN ABSTRACT OF THE THESIS OF

NORMAN LELAND JOHNSON (Name) for the DOCTOR OF PHILOSOPHY (Degree)

in Electrical and Computer Engineering (Major) presented on Nov. 29, 1973 (Date)

Title: AN ERASEABLE MOS READ-ONLY MEMORY

Abstract approved: Redacted for Privacy

James C. Looney

A non-volatile electrically alterable read-only memory using an MOS field effect transistor with a threshold voltage which depended upon the position of ionic charges for its memory was fabricated.

The conventional silicon dioxide layer was replaced by a composite layer consisting of thermal silicon dioxide, electron beam evaporated aluminum oxide, and pyrolytically deposited silicon dioxide which substantially reduced the migration rate of the ions.

The rate of drift of the ions was found to be dependent upon the polarity of the electric field. The ions had greater mobility under a negative applied bias. It was also shown that for the same electric fields the activation energy was higher for negative electric fields, and that the activation energy decreases with increasing electric fields. The migration rate was also found to be an exponential function of applied electric field.

AN ERASEABLE MOS READ-ONLY MEMORY

by

Norman Leland Johnson

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

June 1974

APPROVED:

Redacted for Privacy

Associate Professor of Electrical
and Computer Engineering
in charge of major

Redacted for Privacy

Head of Department of Electrical
and Computer Engineering

Redacted for Privacy

Dean of Graduate School

Date thesis is presented November 29, 1973

Typed by Brenda Fadness for Norman L. Johnson

ACKNOWLEDGEMENTS

The author would like to thank Professor James C. Looney of the Department of Electrical and Computer Engineering, Oregon State University. The basic idea of the MOS read-only memory was proposed by Professor Looney. It was through his helpful comments and suggestions by which the investigator was able to complete the investigation. The author would also like to acknowledge his helpful comments and criticisms in editing the final manuscript.

Financial support which was provided by the Department of Electrical and Computer Engineering in the form of a graduate assistantship made the investigation possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

AN ERASEABLE MOS READ-ONLY MEMORY

I. INTRODUCTION

The electrical characteristics of the first metal-oxide-semiconductor (MOS) transistors were not stable. The threshold voltage was not constant, but changed in the presence of an applied gate voltage. The phenomenon was studied extensively and the reason for such unstable behavior was discovered. The main cause of the instability was traced to ionic contamination, mainly sodium, which was present in the gate insulator (9, 10, 15, 16, 19). If a negative voltage was applied to the gate metal, the positive ions would migrate to the gate metal and have no effect on the threshold voltage. If a positive voltage was applied to the gate metal, the ions would migrate to the semiconductor-silicon dioxide interface, and change the threshold voltage. The change in threshold voltage is directly proportional to the charge of the ion and inversely proportional to the gate capacitance.

The experimenters finally solved this problem by using extremely clean fabrication techniques and reducing the mobility of ions in the gate oxide. This can be done by phosphorus doping. Other proposed solutions were also tried, using several different oxides or combinations of oxides and/or insulators. The MOS field effect transistor ROM that was fabricated and is described in this report used as the information storage mechanism the presence of the mobile positive

ions. The position of the ions in the gate oxide of the MOS field effect transistor determined the threshold voltage of the device or state of the memory. The conventional silicon dioxide gate insulator had to be replaced by a three-layer gate insulator made up of a layer of aluminum oxide sandwiched between two layers of silicon dioxide. This was necessary in order to gain a certain degree of stability in operation. The introduction of the oxide interfaces in the gate insulator slows the migration rate of ions in the gate insulator (1).

In the conventional MOS read-only memory the information is permanently established in the memory element during the fabrication of the device. An MOS field effect transistor is fabricated in such a manner as to have either a low threshold voltage or a high threshold voltage, which is done by controlling or establishing two different oxide gate thicknesses. A thick gate oxide in the order of 5000 Å would result in a transistor having a high threshold voltage. A thin gate oxide of normal thickness in the order of 1500 Å would result in a much lower threshold voltage. Thus, the mechanism for the information storage is the thickness of the oxide, and this can be detected by determining if the transistor will conduct at a gate voltage somewhere between the low and high threshold voltage values. The information in this type of ROM cannot be changed after fabrication.

Another type of ROM which is available is the avalanche-injection floating-gate transistor. In this device electrons are avalanche-

injected into a silicon layer in the gate insulator. This device is electrically programmable but has the additional feature of being eraseable by using radiation or ultraviolet lights.

Another method of establishing different threshold voltage values in an MOS field effect transistor for memory application is currently being investigated. The silicon dioxide present in the usual MOS field effect transistor is replaced by a silicon nitride-silicon dioxide layer. The resulting device is an MNOS memory. In this memory cell the threshold voltage is established by the presence of negatively charged particles. The memory transistors are indistinguishable after fabrication, but the threshold voltage can be established and changed by using suitable gate voltages (18).

The ROM which is described in this report has the same advantages of the MNOS ROM; the information can be electrically stored and erased in the storage cell. Instead of using the position of electrons to change the threshold voltage of the transistor as in the MNOS device, positive charge carriers are used.

II. THEORY OF OPERATION OF THE ERASEABLE ROM

Explanation of the Capacitance Versus Voltage Curve Shift Due to Ionic Contamination in the Oxide

One of the most useful techniques used to experimentally determine the properties of a metal-oxide-semiconductor (MOS) structure as drawn in Figure 1, is to determine its capacitance as a function of voltage (7, 20). This technique is a fast, accurate, and nondestructive method of determining many of the properties and characteristics of the MOS structure. By analyzing the resulting capacitance-voltage (C-V) curve, one can determine the conductivity type, carrier concentration near the surface of the silicon material, and surface state density of the semiconductor. The work function of the metal can also be determined because the C-V curve contains information of the work function difference between the metal and semiconductor.

The capacitance of the insulating layer can be determined by biasing the capacitor so that the surface of the semiconductor is accumulated. If the thickness of the oxide can be determined by some other method, then the dielectric constant of the insulating material or oxide can be calculated. The C-V curve will also furnish information on ionic impurities present in the oxide by a gate bias-temperature procedure.

A normalized metal-oxide-semiconductor capacitance versus

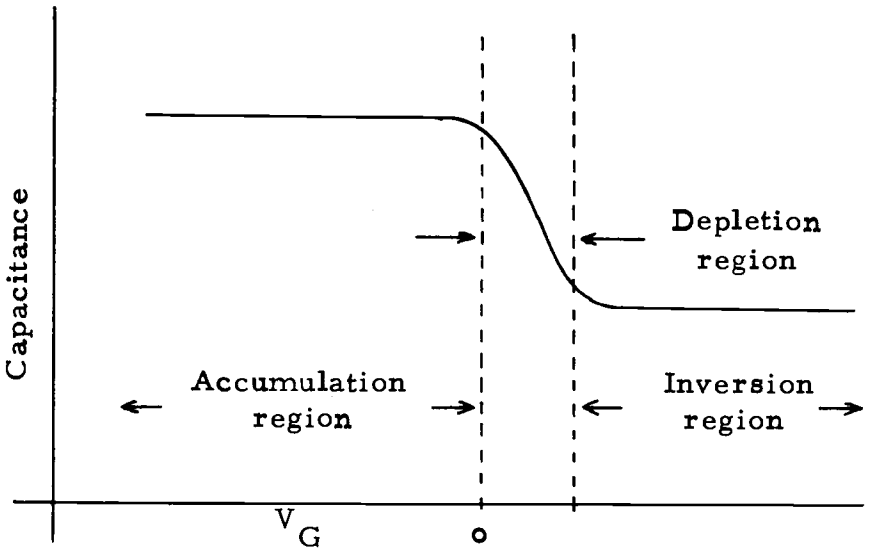

Figure 1. Typical capacitance-voltage (C-V) characteristics of an MOS capacitor constructed on a p-type substrate.

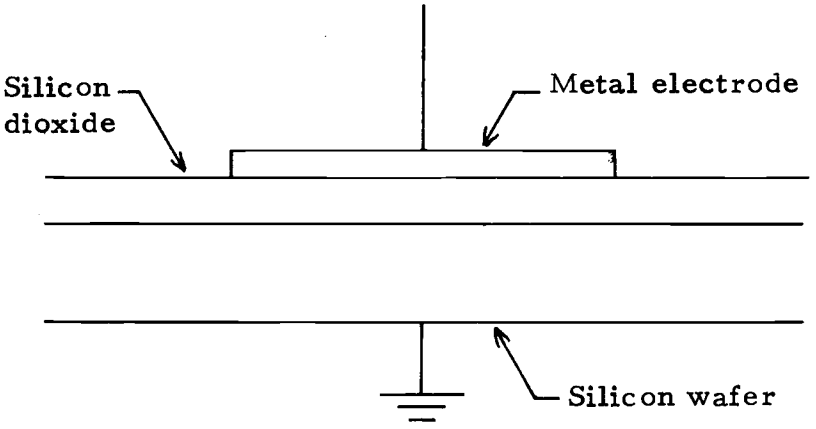

Figure 2. Schematic structure of an MOS capacitor (not to scale).

voltage is drawn in Figure 2. The figures shows a typical C-V curve which would be obtained when the semiconductor material is p-type, For an n-type material the polarity of the voltage would be reversed.

For an MOS structure the capacitance is given either by the series combination of the oxide capacitance and the capacitance of the surface space charge region associated with the p-n junction in the semiconductor, or the capacitance of the oxide alone, depending upon the bias voltage.

$$C = \frac{C_o C_s}{C_o + C_s} \tag{1}$$

where C_o = capacitance per unit area of the insulating material

C_s = capacitance per unit area of the surface space charge region

C = total MOS capacitance per unit area.

The capacitance per unit area of the insulating material is a function of dielectric constant of the oxide, permittivity of free space and oxide thickness.

$$C_o = \frac{K_o \epsilon_o}{x_o}$$

where ϵ_o = permittivity of free space

K_o = relative dielectric constant of the oxide

x_o = thickness of the insulating material.

If the insulating material consists of layers of more than one material, the capacitance per unit area of the oxide consists of the

series combination of the capacitance of each individual layer.

The MOS capacitance-voltage curve can be separated into three different regions as shown in Figure 1, depending upon the condition of the semiconductor surface which is controlled by the bias voltage.

For an MOS capacitor constructed on a p-type semiconductor, if the bias voltage is sufficiently negative, the surface will accumulate and there will be no space charge region. Consequently there will be no capacitance due to the space charge region and the total capacitance of the MOS structure will be only the capacitance due to the insulating layer.

If the gate is positive, but not sufficiently positive to invert the surface of the semiconductor, the capacitance versus voltage of the MOS structure would be described by the depletion region of the curve. This condition occurs when the surface of the semiconductor is depleted of charge carriers. The formula that relates the normalized capacitance per unit area to the bias voltage is (9, p. 272):

$$\frac{C}{C_o} = \frac{1}{\sqrt{1 - \frac{2K_o^2 \epsilon_o V_G}{qN_A K_s x_o^2}}} \qquad (3)$$

where K_s = relative dielectric constant of silicon

q = magnitude of the electric charge, 1.6×10^{-19} coulomb

N_A = concentration per unit volume of the acceptors (holes)

V_G = gate voltage.

Although x_d does not explicitly appear in Equation 3, the capacitance of the space charge region changes because x_d changes.

$$C_s = \frac{K_s \epsilon_o}{x_d} \tag{4}$$

where x_d = width of the space charge region

K_s = relative dielectric constant of the semiconductor.

As the bias voltage is continuously increased, the width of the space charge region will also increase until x_d reaches a maximum value. At this value of bias voltage the capacitance of the MOS structure will enter the inversion region of the C-V curve. In the inversion region the total capacitance is equal to capacitance of the insulator in series with the minimum capacitance of the space charge region that is associated with the p-n junction.

During the above discussion the metal semiconductor work function difference, the surface state charge, and the ionic charge density were not discussed. When these factors are taken into account they have the primary effect of shifting the C-V curve to the left or right according to the following equation (9, p. 281):

$$\Delta V = \Phi_{MS} - \frac{Q_{ss}}{C_o} - \frac{1}{C_o} \int_o^{x_o} \frac{x}{x_o} \rho(x)\, dx$$

where ΔV = shift of the C-V curve

Φ_{MS} = metal semiconductor work function difference

Q_{ss} = surface state charge density per unit area

$\rho(x)$ = ionic charge density distribution.

The metal-oxide interface is located at $x = 0$ with positive values of x in the direction of the semiconductor.

When the ionic charge density is due to ions that are mobile in the insulating gate oxide, then it is possible to shift the C-V curve with respect to the bias voltage axis depending upon the position of the ions in the insulating layer. The shift of the charged particles can be accomplished by the application of a negative or positive bias voltage, depending upon the position of the particles. The MOS capacitor may have to be heated, so that the ions will migrate in a reasonable amount of time. By inserting into the last term of Equation 5 the two possible ionic charge density distributions, the maximum threshold voltage shift due to the migrating ions can be determined.

When the ions are located near the metal-insulator interface, the ionic charge density distribution is equal to:

$$\rho(x) = Q_{ion}\,\delta(x) \tag{6}$$

where Q_{ion} = ionic charge density per unit area

$\delta(x)$ = delta function centered at $x = 0$.

Solving the equation yields $\Delta V = 0$.

When the ions are located at the insulator-semiconductor interface, the ionic charge density distribution is equal to:

$$\rho(x) = Q_{ion}\,\delta(x - x_o) . \tag{7}$$

Solving the equation yields:

$$\Delta V = \frac{Q_{ion}}{C_o} . \tag{8}$$

Consequently, when the ions are located at the metal-insulator interface they have no effect on the position of the C-V curve; when the ions are located at the insulator-semiconductor interface they have their maximum effect in a shift of the C-V curve equal to Q_{ion} divided by C_o. Equation 5 is only correct for a single-layer oxide, though it can be extended to a three-layer oxide. The resulting equation is:

$$\Delta V = \frac{1}{K_1 \epsilon_o} \int_o^{x_1} x\rho_1(x)\,dx + \frac{1}{K_2 \epsilon_o} \int_{x_1}^{x_1+x_2} \left[\frac{K_2}{K_1}\, x_1 + (x - x_1)\right] \rho_2(x)\,dx$$

$$+ \frac{1}{K_3 \epsilon_o} \int_{x_1+x_2}^{x_1+x_2+x_3} \left[\frac{K_3}{K_1}\, x_1 + \frac{K_3}{K_2} x_2 + (x - x_1 - x_2)\right] \rho_3(x)\,dx \tag{9}$$

where x_1 = thickness of the oxide nearest the metal

x_2 = thickness of the oxide in the middle

x_3 = thickness of the oxide nearest the semiconductor

$\rho_1(x)$ = ionic distribution density in the oxide nearest the metal

$\rho_2(x)$ = ionic distribution density of the middle oxide

$\rho_3(x)$ = ionic distribution density in the oxide nearest the metal

K_1 = relative dielectric strength of the oxide nearest the metal

K_2 = relative dielectric strength of the middle oxide

K_3 = relative dielectric strength of the oxide nearest the semiconductor.

If the ions are located at the metal-insulator interface: $\rho_1(x)=Q_{ion}\delta(x)$, $\rho_2(x) = 0$, and $\rho_3(x) = 0$. Equation 9 reduces to $\Delta V = 0$. If the ions are located at the insulator-semiconductor interface: $\rho_1(x) = 0$, $\rho_2(x) = 0$, and $\rho_3(x) = Q_{ion}\delta(x-x_1-x_2-x_3)$. Substitution yields:

$$\Delta V = \frac{1}{K_3\epsilon_o}\int_{x_1+x_2}^{x_1+x_2+x_3}\left[\frac{K_3}{K_1}x_1+\frac{K_3}{K_2}x_2+(x-x_1-x_2)\right]Q_{ion}\delta(x-x_1-x_2-x_3)dx \tag{10}$$

$$\Delta V = \frac{Q_{ion}}{K_3\epsilon_o}\left[\frac{K_3}{K_1}x_1+\frac{K_3}{K_2}x_2+x_3\right] \tag{11}$$

$$\Delta V = Q_{ion}\left[\frac{x_1}{K_1\epsilon_o}+\frac{x_2}{K_2\epsilon_o}+\frac{x_3}{K_3\epsilon_o}\right] . \tag{12}$$

But the total effective capacitance per unit area, C_T,

$$\frac{1}{C_T} = \left[\frac{x_1}{K_1\epsilon_o}+\frac{x_2}{K_2\epsilon_o}+\frac{x_3}{K_3\epsilon_o}\right]. \tag{13}$$

Therefore

$$\Delta V = \frac{Q_{ion}}{C_T} \quad . \tag{14}$$

When only the two extreme cases for the position of ions are examined, the maximum shift of the C-V curve depends only upon the concentration of the ion particles divided by the capacitance per unit area of the MOS capacitor.

Equation 14 can be used to experimentally determine the concentration of the mobile ionic charges. The mobile ions migrate or diffuse through the oxide when the ions are under the influence of an electric field. The mobility of the ions increases as the temperature increases. Also the mobility of the ions is dependent on the magnitude of the electric field.

When the MOS capacitor is heated to approximately $200^{o}C$ with a negative bias on the capacitor, the positive ions will be attracted to the metal-insulator interface. When the MOS capacitor is cooled to room temperature, the positive ions are much less mobile; in this position the mobile ions will have a negligible effect on the C-V curve. If the MOS capacitor is again heated to approximately $200^{o}C$, this time with a positive bias on the capacitor, the mobile positive ions will be forced to the semiconductor-insulator interface. After the MOS capacitor is cooled to room temperature, the C-V curve can again be determined. When the ions are near the oxide-silicon interface they have the greatest effect on the position of the C-V curve.

The C-V curve will be shifted to the left or the negative direction. If the capacitance per unit area is known along with voltage shift, the total mobile charge can be determined by Equation 14.

Explanation of How the Ionic Contamination Affects MOS Field Effect Transistor Characteristics

As previously explained, an MOS capacitor that has been contaminated in some manner with mobile positive ions could be used as a memory device. The position of the C-V curve with respect to the bias voltage axis is dependent upon the position of the mobile positive ions. Furthermore, the position of the positive ions depends upon the direction of the electric field that has previously been applied to the capacitor. Therefore, the position of the C-V curve with respect to the bias voltage axis is dependent upon the previous history of the device. Thus the capacitor could be used as a memory element.

The memory capacitor has one serious undesirable characteristic. It would be necessary to plot the C-V curve for the capacitor, in order to "read out" the information, which would be very inefficient. Another undesirable characteristic is that with a memory capacitor it would be very difficult to establish a read-write technique for any type of matrix configuration. Another method of interrogating the memory would be highly desirable.

If two n-type regions are produced by diffusion techniques at

the ends or sides of the MOS capacitor extending slightly under the metal layer, an MOS field effect transistor would result. In the same manner that the position of the ionic contamination affected the position of the C-V curve of an MOS capacitor, the position of the ions would also affect the characteristics of the MOS field effect transistor. In order to explain how this is possible, a brief description of the operation of an MOS field effect transistor will be given.

Theory of an MOS Field Effect Transistor

A description of the electrical operating characteristics of the MOS field effect transistor has been given previously in detail by many authors (3, 14). Therefore, only the information that is directly applicable to the operation of an MOS ROM will be reported here.

The construction of an n-channel MOS field effect transistor is illustrated in Figure 3. N-type impurities are diffused into the p-type silicon wafer to form two n-type regions, named the source and the drain. The regions are placed so that one edge of a region is adjacent and parallel to one edge of the other region. The area between the two regions is called the channel. It is in the channel where the control of the conducting particles (electrons for an n-channel device) takes place.

Above the gate and overlapping the source and the drain is the gate oxide which is usually silicon dioxide, produced by the thermal

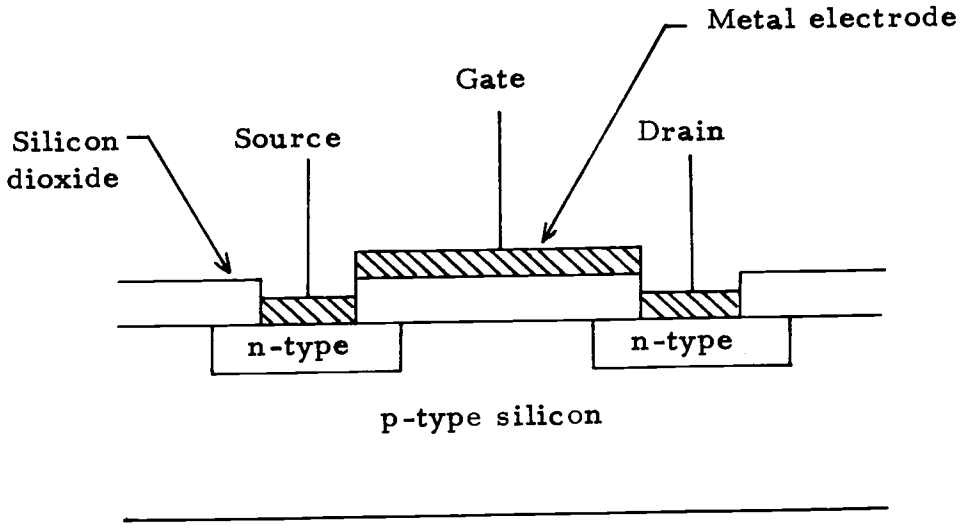

Figure 3. Schematic diagram of a cross-section of an n-channel MOS transistor (not to scale).

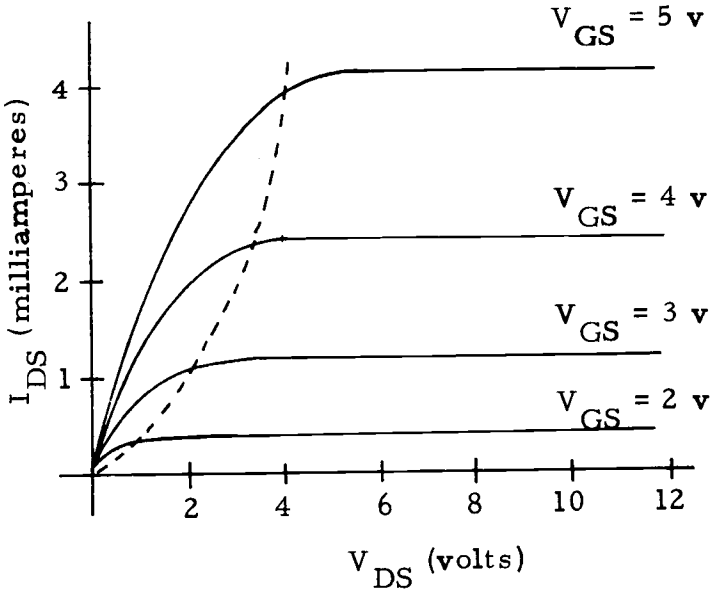

Figure 4. Typical current-voltage characteristics for an n-channel enhancement mode MOS field effect transistor.

oxidation of silicon. In the case of the ROM, the silicon dioxide layer was replaced by the three-layer insulator. Above this and slightly overlapping the source and the drain is the gate metal which is usually a layer of aluminum. Likewise, aluminum is usually used for the contacts to the source and drain regions.

Typical MOS field effect transistor characteristics are shown in Figure 4. The characteristic curves are given for an n-channel enhancement mode MOS field effect transistor. The dotted line delineates between two types of operation. If $|V_{DS}| \leq |V_{GS} - V_T|$, then the operation is to the left of the dotted line. This is the triode region of operation. If $|V_{DS}| \geq |V_{GS} - V_T|$, the field effect transistor is operating in the saturation region.

The current-voltage relationships are different for the different regions. For the triode region the relationship is given by (3, p. 29):

$$I_{DS} = \frac{W}{L} \mu_n C_o [(V_{GS} - V_T) V_{DS} - \frac{1}{2} V_{DS}^2] \tag{15}$$

where I_{DS} = drain current

W = channel width

μ_n = electron mobility

C_o = gate capacitance per unit area

V_{GS} = gate voltage

V_T = threshold voltage

V_{DS} = drain voltage.

For the saturation region the relationship is given by (3, p. 32):

$$I_{DS} = \frac{W}{2L} \mu_n C_o (V_{GS} - V_T)^2. \tag{16}$$

Another very useful method of describing the characteristics of an MOS field effect transistor is to plot the transfer characteristics. A typical plot for an n-channel enhancement mode MOS field effect transistor is given in Figure 5. The square root of the drain current is plotted as a function of the gate voltage for a particular value of V_{DS}.

By taking the square root of both sides of Equation 16 we have:

$$\sqrt{I_{DS}} = \sqrt{\frac{W}{2L} \mu_o C_o} \quad (V_{GS} - V_T). \tag{17}$$

Thus the slope of the line in Figure 5 is $\sqrt{\frac{W}{2L} \mu_n C_o}$ and the intercept is the threshold voltage. Thus, instead of plotting a C-V curve to determine the threshold voltage or state of the memory, it is possible to construct a transistor and have the threshold voltage determine the amount of current that flows between the source and drain, and therefore the state of the memory.

Description of a Memory Matrix

In order for any memory element to be useful there must be a method of arranging the elements into an array or matrix. A matrix organization permits the realization of high component density on a

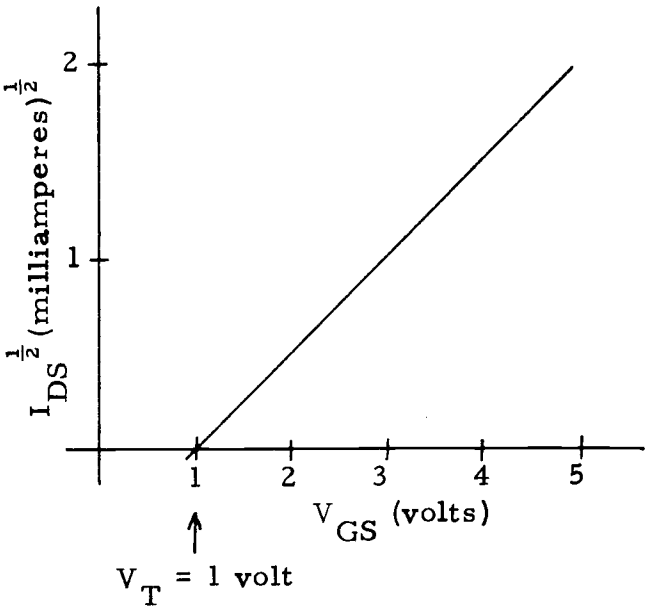

Figure 5. Transfer characteristics for a typical n-channel enhancement mode MOS field effect transistor.

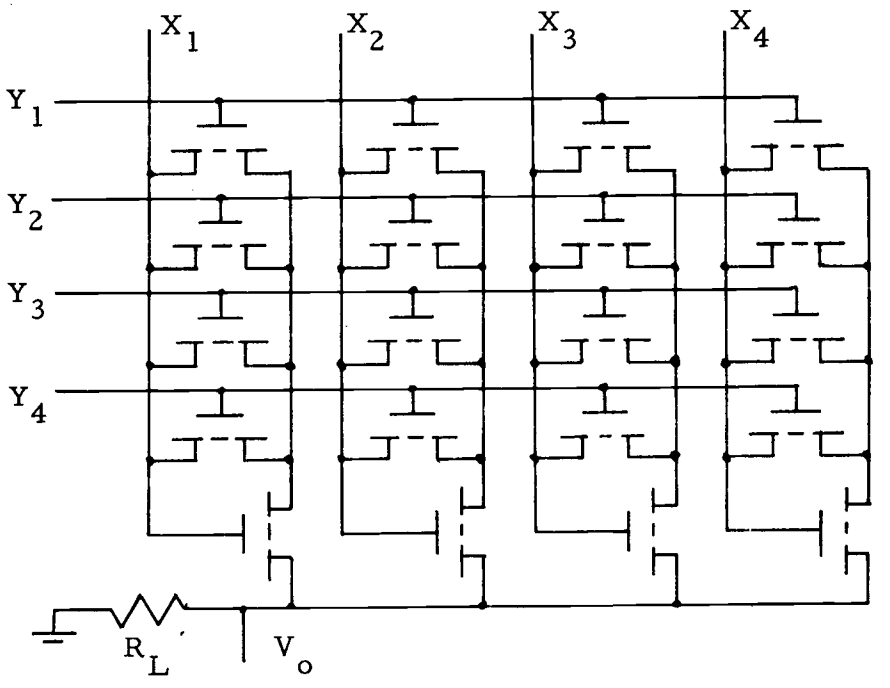

Figure 6. Schematic diagram of the 4 x 4 memory matrix and the associated readout transistors.

silicon wafer which can be maximized by utilizing the single field effect transistor per storage-bit feature of a three-layer MOS transistor. A matrix array also reduces the number of electrical input and output connections to the chip.

For information to be written into the memory field effect transistor, a voltage must be applied across the gate oxide between the gate and the silicon substrate. A method for the selection of any particular transistor in a matrix array as illustrated in Figure 6 must be available.

The voltage that appears across the dielectric of an MOS field effect transistor can be controlled by biasing the voltage at the drain terminal when there is an inversion layer connecting the source and drain areas. The source terminal must be kept floating and the polarity of the biasing voltage must not forward bias the p-n junction. Under these conditions the voltage across the gate will be determined by the voltage that is applied to the gate and the voltage that is applied to the drain.

For example, consider the MOS field effect transistor that appears in Figure 3. Assume a p-type substrate and a configuration such that the threshold voltage is 1 volt. If a positive 10 volts is applied to the gate terminal with the substrate biased at a negative 10 volts, an inversion layer or conductive channel will be induced. Now, depending upon whether one wants to move the ions from their

initial position at the silicon dioxide-aluminum interface, one can apply either a positive ten volts or a negative ten volts to the drain terminal. If a positive ten volts is applied to the drain terminal, there will be no voltage difference across the gate oxide. If a negative ten volts is applied to the drain terminal, there will be 20 volts across that gate oxide that will cause the ions to migrate and the threshold voltage of the transistor to change (13).

Method of Element Selection for Writing In and Reading Out Information in the Memory Matrix

A desirable feature for a field effect transistor which is to be used in a matrix scheme would be the existence of a critical voltage, a voltage higher than either the low, "0", or high, "1", value of threshold voltage but lower than the breakdown voltage of the p-n junctions in the transistor. Application of a voltage lower than the critical voltage would have no effect on the threshold voltage but application of a voltage higher than the critical would cause the threshold voltage of the transistor to change. The three-layer MOS field effect transistor that is proposed for application as a memory device does not have this feature. There does not exist a critical voltage which results in the threshold voltage shift of a selected element or memory transistory; the relationship between shift in threshold voltage and applied electric field is an exponential function. Therefore it is possible to shift only a finite number of memory transistors in a row or

a finite number of memory transistors in a column without significantly disturbing the threshold voltage of the other transistors.

Consider the memory matrix illustrated in Figure 6. At the start of write operation, assume that all of the memory transistors have a threshold voltage of one volt; this could correspond to "zero" in all of the storage bits. In order to store a "one" in a memory transistor we need to apply a higher electric field in the positive direction with respect to the substrate in that transistor than in any other memory transistor.

When a positive 20 volts is applied to a y-line or a row of gates, 0 volts is applied to the rest of the gate rows, and the substrate is lowered to -10 volts, an inversion layer will appear beneath all of the gates of the memory transistors when the output terminal is floating. If at the same time a negative ten volts is applied to an x-line or a column of drain connections and a positive ten volts is applied to the rest of the drain columns, the electric fields that appear across the gate oxides are as follows.

A positive 30 volts will appear across the gate oxide of a selected transistor (assuming the substrate is at ground potential). A positive ten volts will appear across the gate oxide of the transistors in the same row and in the same column as the selected transistor (again with reference to the substrate). The rest of the transistors have a negative ten volts appearing across the gate.

With the voltage selection scheme that was just proposed, the difference in gate voltage between the transistor that one desires to switch to a "one" or change its threshold voltage to a value between 0 and 1/2 volt, and those memory transitors that one does not want to switch, is 30 volts as compared to 10 volts. If the expotential relationship between the change in threshold voltage and gate voltage is expotential, one can see that it would be possible to switch a finite number of transistors of a column or row without significantly disturbing those transistors that one does not desire to switch.

A method to read out the information from a selected bit must also be available in order to extract information from the memory array. A particular memory transistor can be selected by a suitable choice of an x-address line and y-address line which specifies a unique transistor. In order to determine if there is a one or a zero stored in the selected memory transistor, a five-volt pulse can be applied to the x-line and the y-line. The five-volt pulse applied to the x-address line on the gate will cause all of the transistors in the selected row to turn on. The five-volt pulse that is applied to the x-address line will cause the readout transistor of the selected column of memory transistors to turn on and there will be a path for current flow through the two transistors. The magnitude of this current will be dependent upon the threshold voltage of the selected memory transistors. A threshold voltage of one volt would cause appreciably less

current to flow (a zero), than if the memory transistor had a threshold voltage of zero volts (a one). The current can be sensed by putting a resistor between the output terminal and the ground. The voltage pulse that would appear across the resistor would be proportional to the magnitude of the current pulse.

The speed at which the mobile ions move in the oxide is an important point to consider, when considering the use of the position of the ion in the oxide to distinguish between stable states.

The ions must be mobile enough under certain conditions to be able to store information into the memory in a reasonable amount of time. The ions must be stable or not mobile when the memory is not in use. Also, the readout operation must be able to be performed a great many times without significantly disturbing the stored information.

Reading Operation of a Memory Transistor in a Matrix

In order to read out the information that is contained at a location in the matrix array, a voltage must be applied to the appropriate x- and y-lines. As the address voltage is applied to the x- and y-lines the selected transistor as well as the readout transistor associated with the appropriate column will turn on. Current will now be allowed to flow through both of the transistors which are turned on. This current can be sensed by the measuring of the voltage drop across a

resistor at the output. The value of the current will depend upon the value of the threshold voltage of the selected transistor. If the threshold voltage is one volt, a low value of current will be sensed which indicates a stored "0". If the threshold voltage is zero volts, a higher value of current will be sensed which will indicate a stored "1".

As previously stated, the value of the sensed current determines what state ("0" or "1") the memory transistor is in. It would seem desirable to give an indication of what values of current one might expect for the two states for a given memory matrix design. The following will be a description and calculation of the current levels for the 4 x 4 memory matrix that was fabricated in the laboratory.

Since in the readout of the ROM only one y-line or row of gates and only one x-line or column of common drain terminals will be activated to interrogate an individual memory cell or MOS FET, a matrix array can be simplified as shown in Figure 7. Transistor T_1 is the memory transistor that is to be interrogated. The voltage that is applied to the x-line is the drain voltage of the transistor. The voltage that is applied to the y-line is the gate voltage of the transistor. T_2 is the readout transistor for the column of transistors which includes the transistor that is to be sensed. Resistor R_L is an external resistor which is used to sense the current that is connected to the source terminal of the readout transistor.

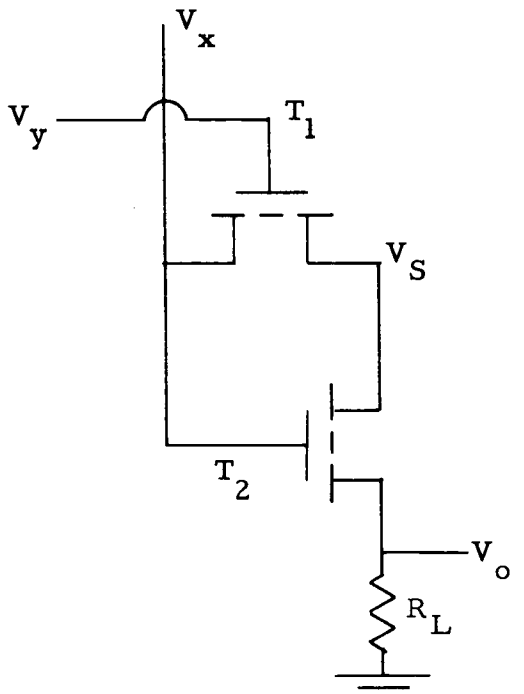

Figure 7. Schematic diagram of a typical element of the 4 x 4 memory matrix and its associated readout transistor.

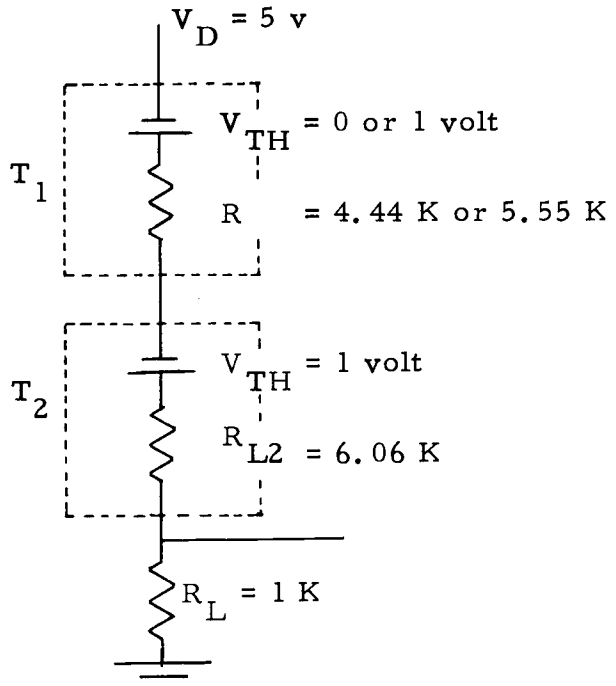

Figure 8. Equivalent circuit of the circuit shown in Figure 7.

Any MOS field effect transistor can be operated in two different regions. These are called the triode region and the saturation region. If $|V_{DS}| \geq |V_{GS} - V_T|$ then the MOS transistor is said to be operating in the saturation region. The transistor is operating in the non-saturation or triode region if $|V_{DS}| \leq |V_{GS} - V_T|$.

Returning to Figure 7, the two transistors are really in series. To determine what region the transistors are operating in, let us determine the relationship between V_{DS} and $V_{GS} - V_T$ for each of two transistors. Let us assume that $V_X = V_Y = 5$ volts, since there is no immediate advantage in using two different sense or address voltages for the x-line and y-line. The two values of threshold voltage for T_1 will be 0 and 1 volt. The voltage at the source of T_1 will be labelled V_S.

It is readily seen that $V_{DS} = 5 - V_S$ and $V_{GS} - V_T = 5 - V_S - V_T$ for T_1. Since V_T is either 0 or 1, the transistor is operating at a point near the interface of the two regions.

For T_2, $V_{DS} = V_S - V_O$ where V_O is the output voltage and $V_{GS} - V_T = 5 - V_O - V_T$. Since T_1 is operating in the saturation region the voltage drop across T_1 will be equal to 0 or 1 volts. V_T for T_2 is equal to 1 volt. Substituting yields for T_2: $V_{DS} = 5 - V_O$ or $4 - V_O$ and $V_{GS} - V_T = 4 - V_O$. Thus T_2 is also operating at a point near the interface of the two regions.

In order to determine V_O for the two possible states of T_1,

g_m must be calculated for T_1 and T_2 (3, p. 53):

$$g_m = \beta(V_{GS} - V_T) \tag{18}$$

and

$$\beta = \frac{W}{L} C_o \mu_n \tag{19}$$

where g_m = transconductance

W/L = channel width to channel length ratio

μ_n = mobility of the electrons in the channel

V_{GS} = gate voltage

V_T = threshold voltage

C_o = capacitance per unit area.

The channel width to length ratio for T_1 was 12 and it was 11 for T_2. The capacitance per unit area was the same for both transistors because they had identical dielectric structures. Since the thickness of each oxide leayer in the three-layer oxide can be controlled within experimental error along with the dielectric constant of the oxide, it is possible to calculate the value for the capacitance per unit area for each layer of the oxide and finally determine the total capacitance per unit area of the three-layer oxide. The dielectric constant of silicon dioxide has been previously reported to be 3.9 (9, p. 103). The dielectric constant of aluminum oxide that was produced under similar conditions has been reported to be 8.3 (6). The oxide thickness of the top and the bottom layers of the three-layer

oxide was 700 Å and the aluminum oxide thickness qas 300 Å. A capacitance per unit area of 2.24×10^4 picofarads per square centimeter is obtained if the previously mentioned values are used in the appropriate equations.

The mobility of the electrons in the channel cannot easily be predicted because its value is dependent upon the concentration of electrons in the channel and the condition of the silicon surface. A test MOS transistor was used to experimentally determine μ_n. The square root of the drain current was plotted as a function of the gate voltage in the saturation region (see Figure 25). According to Equation 17 the slope of the line equals $(W\mu_n C_o/2L)^{\frac{1}{2}}$ or 15.5 (microamperes)$^{\frac{1}{2}}$ per volt. The gate capacitance, C_o, was experimentally determined to be 2.42×10^4 picofarads per square centimeter. The W/L ratio equals 64. Solving for mobility yields 310 cm^2/volt-sec.

Now g_m can be calculated for the MOS transistors. Using Equation 19, $g_m = 450$ μmhos for T_1 assuming $V_{GS} = 5$ volts and $V_T = 0$ volts, and $g_m = 360$ μmhos for T_1 assuming $V_{GS} = 5$ volts and $V_T = 1$ volt. The threshold voltage for T_2 will be 1 volt which leads to a $g_m = 330$ μmhos.

The equivalent circuit for an MOS transistor that is operating in the saturation region consists of a battery equal to the threshold voltage and a resistor which is equal to two times the inverse of the transconductance, g_m (3, p. 90). Figure 8 is the resulting equivalent

circuit for the two transistors. This circuit can be used to calculate the two values of output voltage. If $V_T = 0$ for T_1, then $V_O = 0.35$ volts. If $V_T = 1$ for T_1, then $V_O = 0.24$ volts.

The readout time for the MOS three-layer memory transistor can be calculated in the same manner as an ordinary MOS transistor or an ROM. The readout time is restricted by the response time of the memory transistor, selection transistor, and by the capacitances between the gates and sources. Since in the matrix no MOS transistor is used to turn on another MOS transistor the speed of operation depends mainly upon how fast the address voltages can charge up the gates of their respective transistors and how fast the MOS transistor can charge up the load capacitance.

III. EXPERIMENTAL PROCEDURE

The MOS field effect memory transistors were fabricated using 0.3 - 0.5 Ω-cm p-type silicon wafers. The concentration of acceptors was 6×10^{16} cm^{-2}. The wafers were 3.175 centimeters in diameter. The crystal orientation was (100).

As with most MOS field effect transistors, extreme care had to be used in the preparation of the devices in order that they operate properly. This requires careful attention to cleaning procedures to prevent contamination of the device. It was determined that the devices with the smallest number of ionic charges in the gate insulator or the devices having the smallest threshold voltage shift had a voltage shift equal to approximately 1 volt.

Explanation of Masks Used

A necessary part which is used in the fabrication process for making MOS transistors is a set of photographic masks. Four photographic masks are used to mask areas on the silicon wafer during the various silicon dioxide and aluminum etching steps. The masks are named the source-drain diffusion, the gate oxide, the contact hole, and the metal contact masks.

Initially, the desired configuration or layout for the masks is cut out on 50.8 x 50.8 centimeter rubylith. A 40 to 1 reduction is

followed by a 10 to 1 reduction which was made by a step and repeat camera. Both of the reductions were made on 5.08 x 5.08 centimeter glass plates. The final mask had an effective reduction of 400 to 1 from the initial layout and contained 49 separate dice.

Figure 9 illustrates part of a composite layout of the 4 x 4 memory matrix made by overlapping the four separate masks. The layout also contained an enclosed drain geometry MOS field effect transistor, a diffused resistor, a thin oxide capacitor, and a set of squares which are used for alignment purposes.

The 4 x 4 memory matrix consists of 16 memory transistors, each row of which has a common source and a common drain. The drain of each column is connected to the gate of its readout transistors which all have a common source connection. There are four contact pads for the common drain of each column; the x-line. There are four contact pads for the y-line or gates of the memory transistor. There is also a substrate contact and finally a contact for the common source connections of the readout transistor.

These are some of the important dimensions of the mask which affect the operating characteristics. The channel widths are all 12.7 micrometers. The channel width divided by the channel length (W/L) ratio for the 16 memory transistors is 12. The W/L ratio for the readout transistor is 11.

The enclosed geometry MOS field effect transistor had a W/L

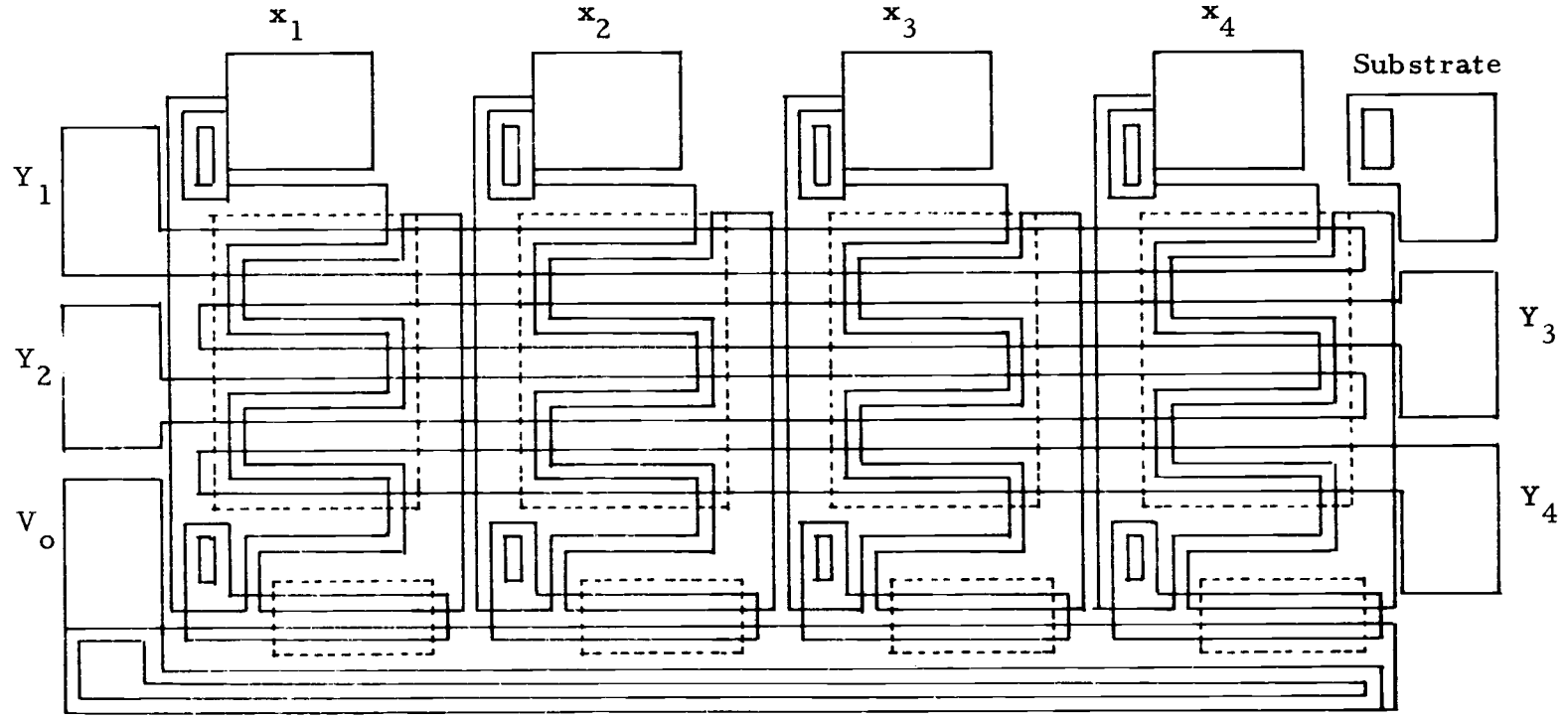

Figure 9. Composite layout for the 4 x 4 memory matrix.

ratio of 64. The purpose of this device was to obtain a value for the mobility of the electrons near the surface of the silicon. It was also used to obtain a value for the threshold voltage.

The thin oxide capacitor was used to plot the C-V curves which gave information on the mobility of the positive ions in the gate oxide.

Fabrication Procedure

The first step in the actual fabircation of the device was to properly clean the wafers prior to the initial thick oxide growth. Initially the silicon p-type wafers were cleaned using trichloroethylene and acetone which removed any contaminants such as greases and oils that may have been present due to fingerprints, etc. The wafer was placed for 30 minutes in a boiling solution of 50% nitric acid and 50% deionized water in order to grow a thin layer of silicon dioxide. The layer of oxide was then etched in a 4 to 1 buffered solution of hydroflouric acid which etched the oxide and exposed a fresh and clean silicon surface.

The next step was to oxidize the silicon wafer in order to grow a silicon dioxide layer of approximately 8500 Å which is necessary in order to properly mask the phosphorus impurities from the silicon wafer. The wafer was oxidized in a specially constructed double-wall oxidation furnace tube. Initially, wet oxygen (oxygen bubbled through 95°C deionized water at 11.3 liters per hour) was passed through the

central core and the jacket simultaneously for 30 minutes. This cleaning procedure was required to remove contaminants that may have been present on the surfaces of the furnace tubes. The wafer was oxidized in wet oxygen for two hours at 1100°C. The wet oxygen flow was removed from the middle tube and nitrogen at 28.3 liters per hour was allowed to flow into the furnace tube over the surface of the wafer. The nitrogen provided an inert atmosphere in which no further oxidation could take place. The high temperature annealed the wafer, reducing the surface state density.

The wafer was now ready for the first photographic and silicon dioxide etching step. It was coated with a thin layer of AZ-1350 positive photoresist. The wafer was then masked, using the source drain photographic plate, and was then exposed to light. The photoresist was developed and rinsed with deionized water. A 4 to 1 buffered solution of hydroflouric acid was used to etch the source and drain windows. After etching, acetone was used with the aid of an ultrasonic cleaner to remove any remaining photoresist. This left the appropriate source and drain areas of the silicon wafer clear of any silicon dioxide, and masked those areas where the p-type diffusion was not desired.

Phosphorofilm was used to dope the wafer in the deposition process. The wafer was coated with phosphorofilm, diluted to a ratio of 10 to 1 with deionized water, and then baked for three minutes at

150°C in air which evaporated the solvent material from the wafer. The wafer was then placed in the diffusion furnace for 15 minutes at 1000°C. After the diffusion process, the wafer was placed in boiling deionized water for 30 minutes as part of the cleaning procedure. A 10 to 1 buffered solution of hydroflouric acid was used to remove any remaining phosphorofilm and the silicon dioxide that had been grown during the deposition process.

The next step was to remove the silicon oxide that was covering the gate areas of the wafer. The photoresist masking steps were repeated using the gate mask. The silicon oxide was etched with a 4 to 1 buffered solution of hydroflouric acid and acetone was used to remove any remaining photoresist material.

Now the wafer was ready for the three-layer oxide. The first layer was silicon dioxide which was thermally grown. The double-wall oxidation furnace was cleaned with wet oxygen. After 30 minutes the wafer was placed in the furnace. Dry oxygen at a rate of 11.3 liters per hour was passed over the wafer for 15 minutes. The oxidation process resulted in a silicon dioxide thickness of 700 Å, brown in color, as determined by the color reference method.

The second layer of the three-layer oxide was aluminum oxide. The aluminum oxide film was produced by evaporating sapphire. Sapphire was used instead of aluminum oxide powder because the crystal form stays in the crucible better during the evaporation and being

a solid it can be heated more uniformly with the electron beam gun than the powder form. The wafer was placed in a bell jar along with the sapphire crystal. An electron gun was used to heat the crystal under a vacuum to a temperature where it evaporated to form a film on the wafer. The starting pressure and final pressure in the bell jar were typically 1×10^{-5} Torr and 5×10^{-5} Torr, respectively. An electron beam current of 40 ma resulted in a deposition rate of approximately 100 Å per minute. A layer of aluminum oxide having a thickness of 300 Å was produced.

The last layer of the three-layer oxide was silicon dioxide. Since thermal oxidation of the silicon substrate could not be used to produce this layer, a pyrolytic technique was used (2). A layer of silicon dioxide would result if a heated wafer was placed in a gaseous atmosphere consisting of silane, oxygen, and nitrogen in appropriate amounts. The wafer was heated to 475°C, bringing about a deposition rate of 750 Å per minute. A layer of silicon dioxide having a thickness of 700 Å was grown by this method.

In order to have electrical contact to the n-type regions of the wafer, contact holes had to be etched through the three-layer oxide. The photoresist masking steps were again repeated, this time using the source-drain contact mask. A 4 to 1 buffered solution of hydroflouric acid was used to etch the silicon dioxide layers while an aluminum etch was used for the aluminum oxide layer (12, p. 24).

The aluminum etch was a solution consisting of 80 parts concentrated phosphoric acid (85% by volume), 16 parts deionized water, and 5 parts concentrated nitric acid (70% by volume).

The wafer was now prepared for the aluminum film by removing the remaining photoresist with acetone and then rinsing in deionized water. The silicon dioxide was then densified at 1100°C for 10 minutes in nitrogen. An aluminum layer of about 5000 Å was produced by using an electron gun to evaporate aluminum (99.9% pure) in a vacuum. The pressure in the bell jar during evaporation was 5×10^{-5} Torr or below. An electron beam of 300 ma was used to evaporate the aluminum. The desired metallic electrode pattern was then produced by using the standard photoresist masking technique with the metal pattern photographic mask. The metal pattern was etched with the phosphoric acid aluminum etch. The last processing step was to anneal the wafer at 400°C in nitrogen for two hours.

By sectioning, angle lapping, and staining the water, the p- and n-type areas were delineated and a junction depth of 1.35 micrometers was measured by counting the interference fringes from the light of a sodium lamp.

The sheet resistance of the n-type diffusions was 25 Ω/square after processing.

IV. EXPERIMENTAL RESULTS

Experimentally Determined Metal-Semiconductor Work Function Differences

Since the characteristics of the memory MOS field effect transistor depends upon the properties of an insulating layer and these properties are affected by the processing steps in the fabrication of the device, it would be convenient to use metals other than aluminum. When aluminum is evaporated onto a silicon dioxide-silicon system, the silicon is damaged by the radiation that is given off by the electron gun. This silicon wafer must be annealed for the MOS transistor or capacitor to operate properly. During the aluminum evaporation and subsequent annealing, sodium and other contaminants can be introduced into the system.

MOS capacitors can be made using gold, gallium, or mercury for the electrode. These capacitors can be fabricated without the radiation damage of e-gun vacuum deposition. The main requirement is that the metal must be in intimate contact with the insulator. Gold, gallium, and mercury are either fluid or ductile enough to form a close contact with their own weight alone or a certain amount of pressure.

One ohm-centimeter, n-type silicon wafers were used in this study. A layer of 1500 Å of silicon dioxide was thermally grown using

dry oxygen at 1100°C. The aluminum capacitors were formed by evaporating the metal through a metal mask of circular dots. The gallium and mercury capacitors were made by placing a small ball of the respective metal on the silicon dioxide. Contact to the aluminum, gallium, and mercury metals was made by means of a tungsten probe. A gold field plate was made by melting the tip of a gold wire into a ball. A certain amount of force had to be used to force the gold ball to deform and make contact to the silicon dioxide.

The flat-band capacitance of n-type, one ohm-centimeter silicon is 0.9 times the oxide capacitance for an oxide thickness of 1500 Å (7). This occurs at zero volts for the ideal MOS capacitor. If the metal-semiconductor work function difference and the surface state density are taken into account, the flat-band capacitance will be displaced from zero by an amount described by the following equation:

$$V_{FB} = \frac{Q_{ss}}{C_o} - \Phi_{MS} \tag{20}$$

where V_{FB} = flat-band voltage.

The flat-band voltage was experimentally determined for the capacitors formed by the four metals. The values are tabulated in the Appendix, Table I.

Deal and Snow reported that the metal-semiconductor work function difference for aluminum on n-type, one ohm-centimeter silicon is -0.3 volts (4). Sze reported that the metal work function

for aluminum is 4.1 volts (16). Substitution of these values into Equation 20 along with the oxide capacitance of 0.022 microfarads per centimeter squared, yields a Q_{ss} value of 1.51×10^{11} cm^{-2} (5).

Knowing the work function for aluminum and the aluminum-semiconductor work function difference, the work function for the other metals can be calculated. These values are also listed in the Appendix, Table I.

Evaluation of the MOS Capacitor and Migration Rates of Ions Through the Three-Layer Oxide

Since the gate insulator plays an active part in the operation of an MOS transistor and the characteristics of an MOS transistor are very sensitive to the properties of the insulator, much information can be gained by evaluating the MOS capacitor and extending this information to the MOS field effect transistor.

The MOS capacitor was fabricated so that its dielectric was identical to the gate oxide of the MOS field effect transistor. The metal that was used for the electrodes of both devices was also identical.

The most important measurement to determine for an MOS capacitor is obviously its capacitance per unit area. The capacitance per unit area of the MOS three-layer oxide was determined by measuring the absolute capacitance of the MOS capacitance and dividing by

the cross sectional area. The electrodes of the fabricated capacitors were square with each side being 0.249 centimeters. Since the thickness of each oxide layer in the three-layer oxide is known within experimental error along with the dielectric constant of the oxide, it is possible to predict a value for the capacitance per unit area for each layer of oxide and finally the total capacitance per unit area of the three-layer oxide. A capacitance value of 2.24×10^4 picofarads per square centimeter is obtained as previously mentioned in the theory section of this report.

The capacitance per unit area was determined over the surface of a wafer that was used to fabricate the various MOS devices. To ensure that the capacitance that was measured was the oxide capacitance and not the series capacitance of the oxide, a sufficient high value of negative bias was applied to the capacitor to ensure no depletion region was present at the time of measurement. An average experimental value of 2.42×10^4 picofarads per square centimeter was found, with a standard deviation of 8.42×10^2 picofarads per square centimeter. Thus the actual value of capacitance per unit area was close to the expected value. The variation in the experimental value was probably the result of an uneven thickness of the oxide layers and variation in capacitor area.

A series of measurements were taken in order to determine the mobility of the ions in the three-layer oxide. This was necessary

in order to determine the operating characteristics of the memory, the "read" time and the "write" time. In order to determine the mobility of the ions, a C-V curve had to be plotted at various time intervals during the migration of the ions. The migration rate or mobility of ions had to be determined for different conditions of sample temperature and applied electric field. Since many C-V curves had to be taken, it was necessary to use an automatic measuring technique.

A schematic diagram of the instruments used is shown in Figure 10. The equipment consisted of a Tektronix type 564 storage oscilloscope. A type 67 time-base plug-in and a type 3C66 carrier amplifier plug-in were used. By the use of a specially constructed instrument, various positive and negative bias voltages could be applied to the capacitor and also to the horizontal scan of the oscilloscope. The vertical scan was the capacitance value. The measurement frequency was 25 KHz.

An environmental chamber was used to establish and maintain the temperature of the MOS capacitor during the measurement.

Normally a sweep speed of 1 msec per division was used to sweep the bias voltage of the C-V curve. This was usually fast enough for the drift measurement. A fast sweep was necessary in order that the impurity distribution of the ions was not significantly disturbed by the C-V curve plotting. By the use of this technique it

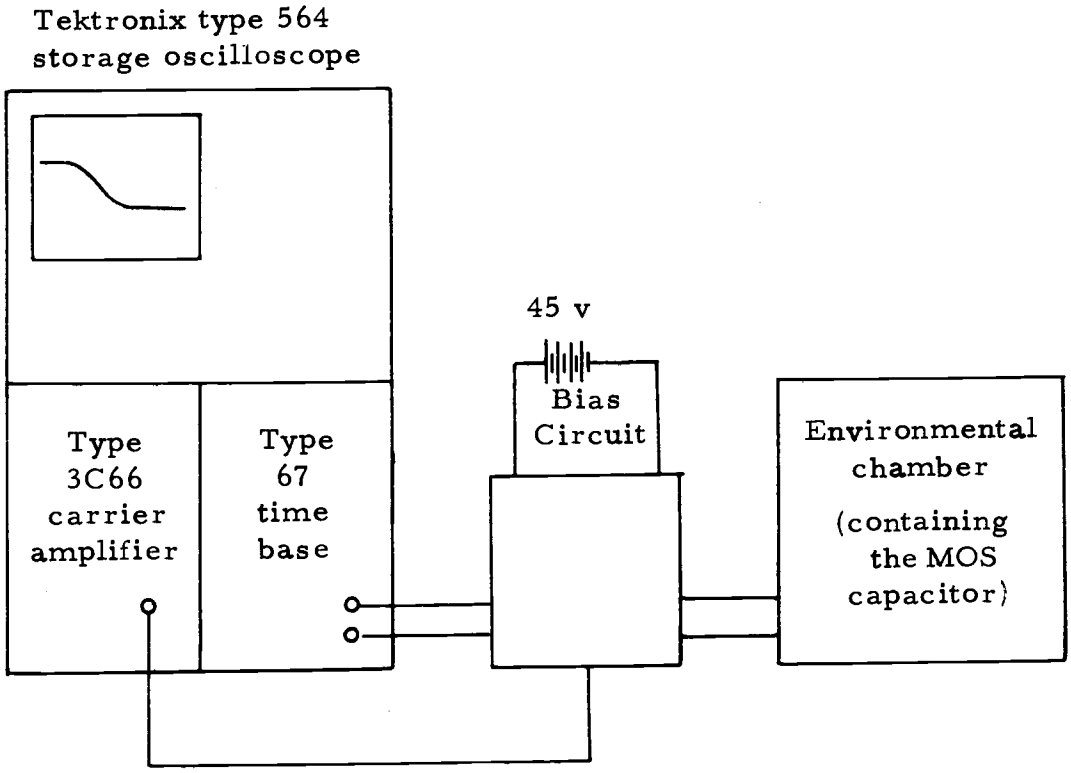

Figure 10. Diagram of the automatic record and display of the capacitance versus voltage for the MOS capacitors.

was possible to observe the shifting of the C-V curve on the voltage axis without bringing the MOS capacitor down to room temperature to plot the C-V curve.

The rate at which the ions drifted in the three-layer oxide was measured at various temperatures between 40°C and 100°C. At temperatures higher than 100°C the ions drifted too fast to make accurate measurements; the act of applying the voltage sweep to the MOS capacitor had a significant effect on the positions of the ions in the oxide layers. At temperatures lower than 50°C the drift was so slow that a great deal of time was necessary to make the measurements. Hours or even days would have been required for the C-V curve of the MOS capacitor to approach a saturation value.

Positive and negative values of the bias voltage were used to establish the electric field in the MOS capacitor. Since the capacitor is not symmetrical, there is no reason to assume that the rate of migration will be the same for positive or negative electric fields. Although both layers of silicon dioxide were the same thickness, they probably do not have the same structure because they were produced by different means at different temperatures. Also, one layer of silicon dioxide is next to a layer of aluminum. Figure 11 shows two typical C-V curves. The one at the right results when the ions are located at the silicon dioxide-aluminum interface. The curve at the left is obtained when the ions are located at the silicon dioxide-silicon

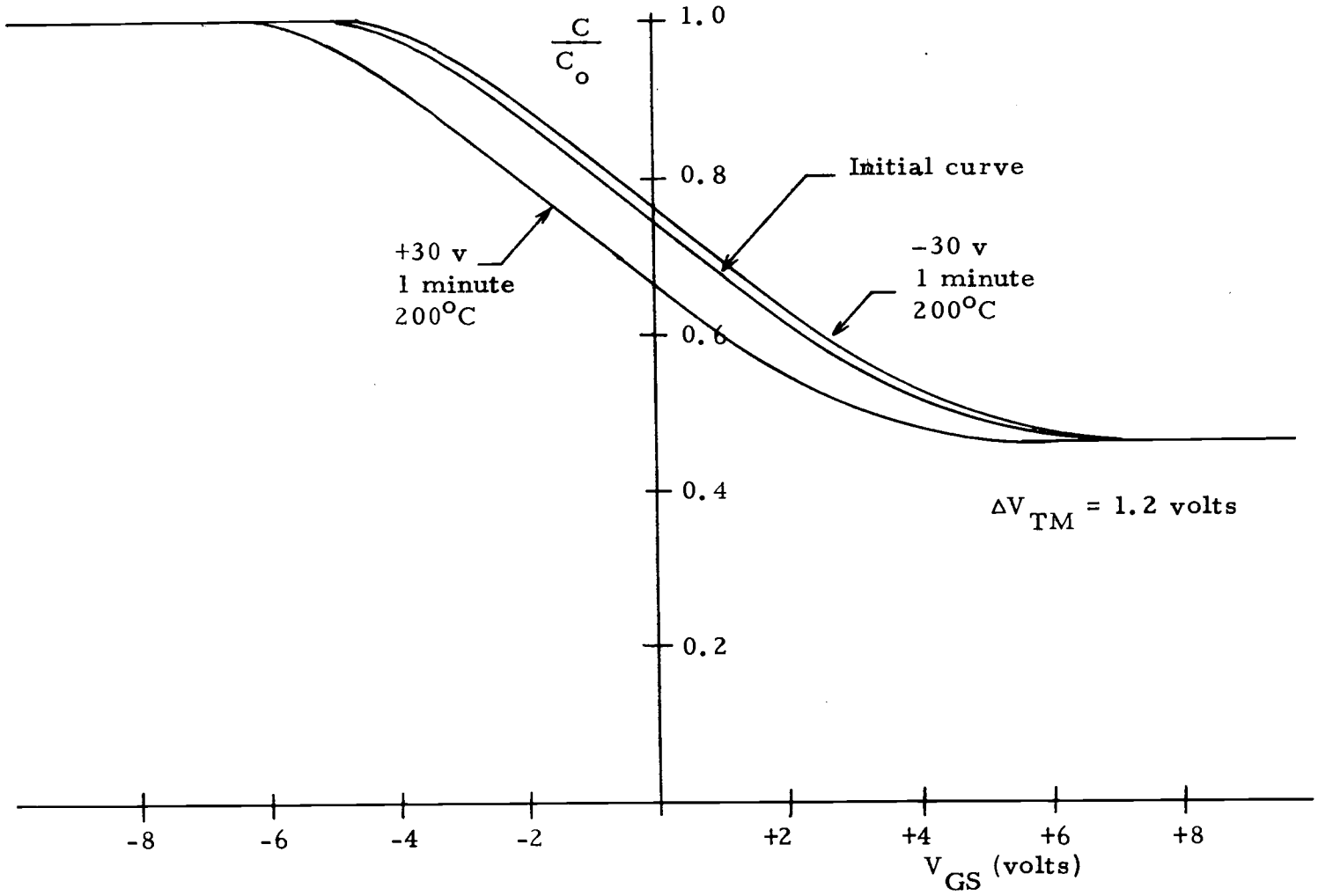

Figure 11. Bias-temperature capacitance versus voltage for an MOS capacitor having a maximum ΔV_{TM} = 1.2 volts.

interface. By measuring the capacitance per unit area and assuming one electronic charge per ion, the concentration of the ions can be determined. Instead of plotting the entire C-V curve for each time or time interval, only the translation of the threshold voltage was noted. Therefore one can plot the change in threshold voltage versus time for a given voltage and temperature.

Figure 12 shows a typical set of data that was just described. ΔV_{TM} is the maximum change in threshold voltage, a constant depending only on the mobile ion density and capacitance. ΔV_T is the amount that the threshold voltage has changed during the application of a pulse of a given magnitude and duration. The vertical axis has been normalized. The data was taken at temperatures between $40^{\circ}C$ and $100^{\circ}C$. The applied bias voltage was a positive 22.2 volts. Thus the graph shows the time that it takes for the positive ions to migrate from the aluminum-silicon dioxide interface. Figures 13 and 14 show similar data for bias voltages of +18.3 volts and +11.0 volts respectively. By comparing the three graphs, one can see that the higher bias or applied electric fields resulted in a larger threshold voltage shift for the same time interval. Also, the ions are more mobile as the temperature increases.

A time constant can be obtained from each curve. The time constant, τ_F, will be defined as the time at which the change in

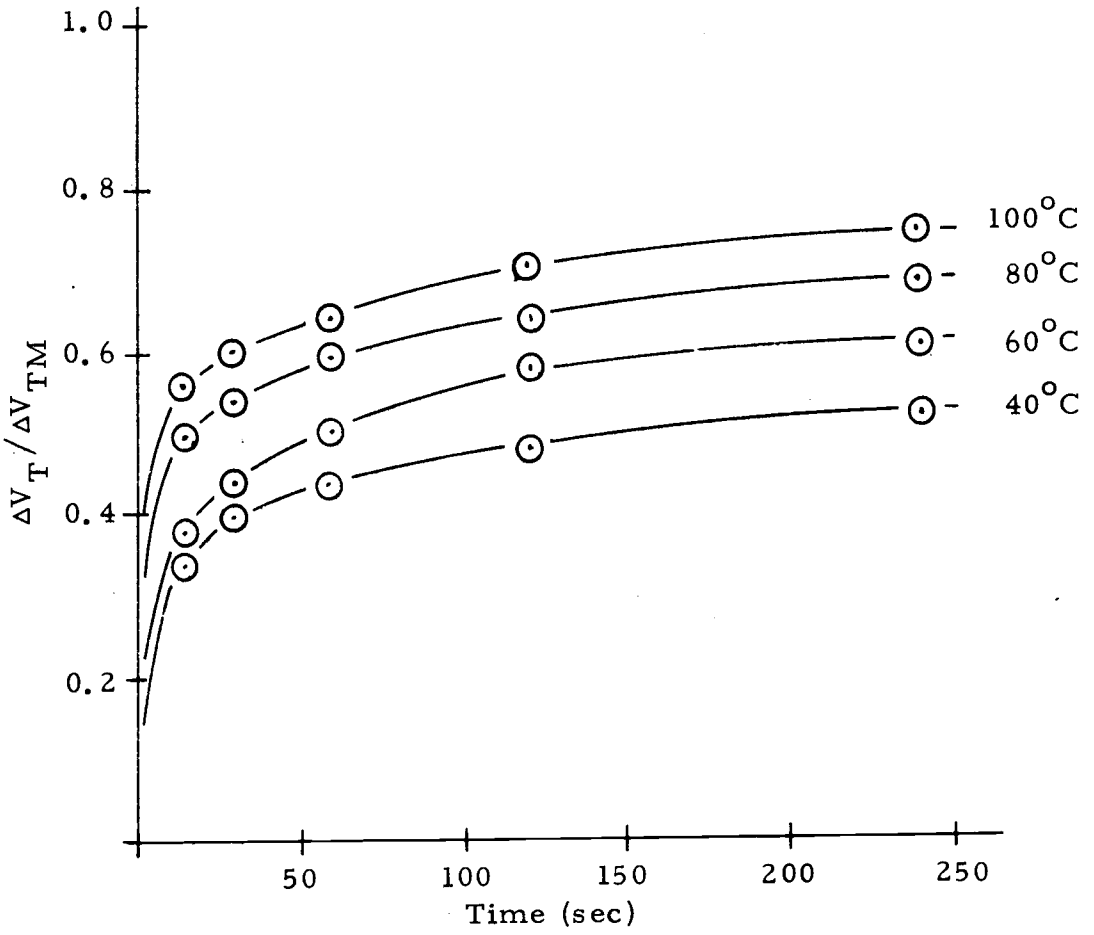

Figure 12. Normalized threshold voltage shift as a function of time (bias voltage = +22.2 volts and ΔV_{TM} = 5.0 volts).

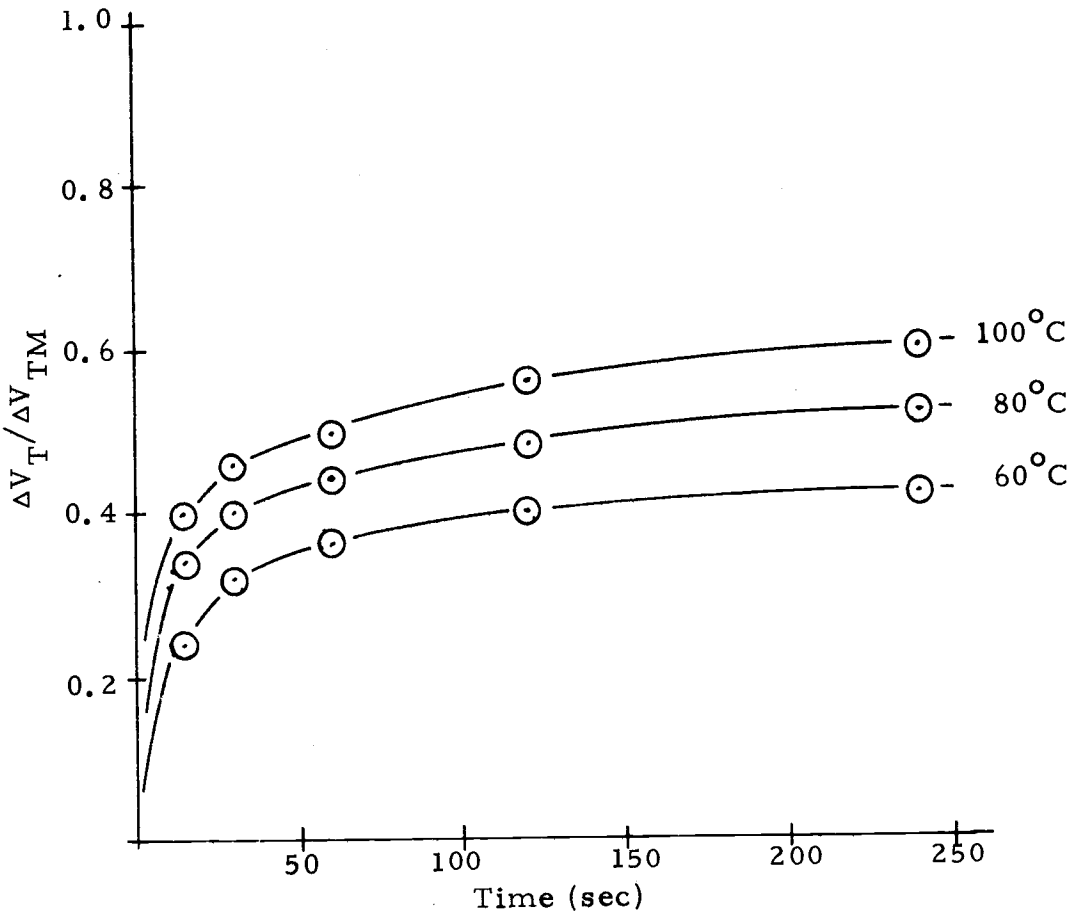

Figure 13. Normalized threshold voltage shift as a function of time (bias voltage = +18.3 volts and ΔV_{TM} = 5.0 volts).

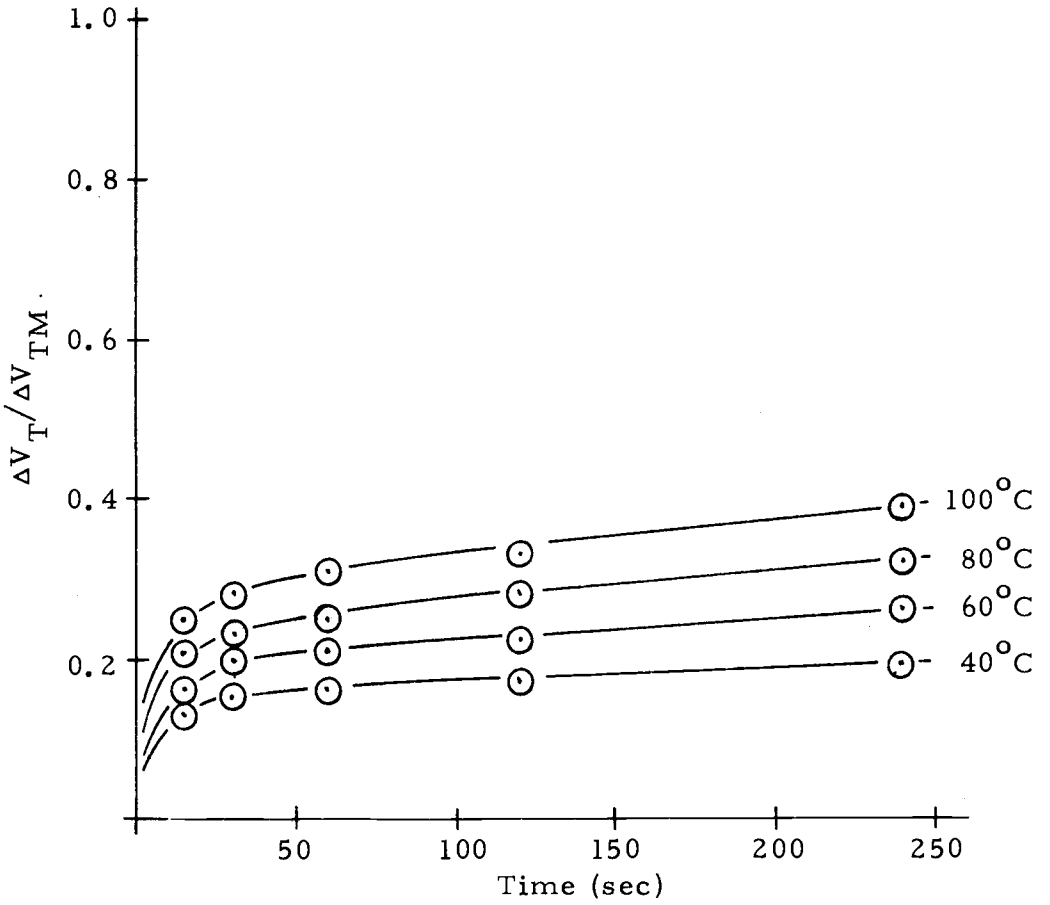

Figure 14. Normalized threshold voltage shift as a function of time (bias voltage = +11.0 volts and ΔV_{TM} = 5.0 volts).

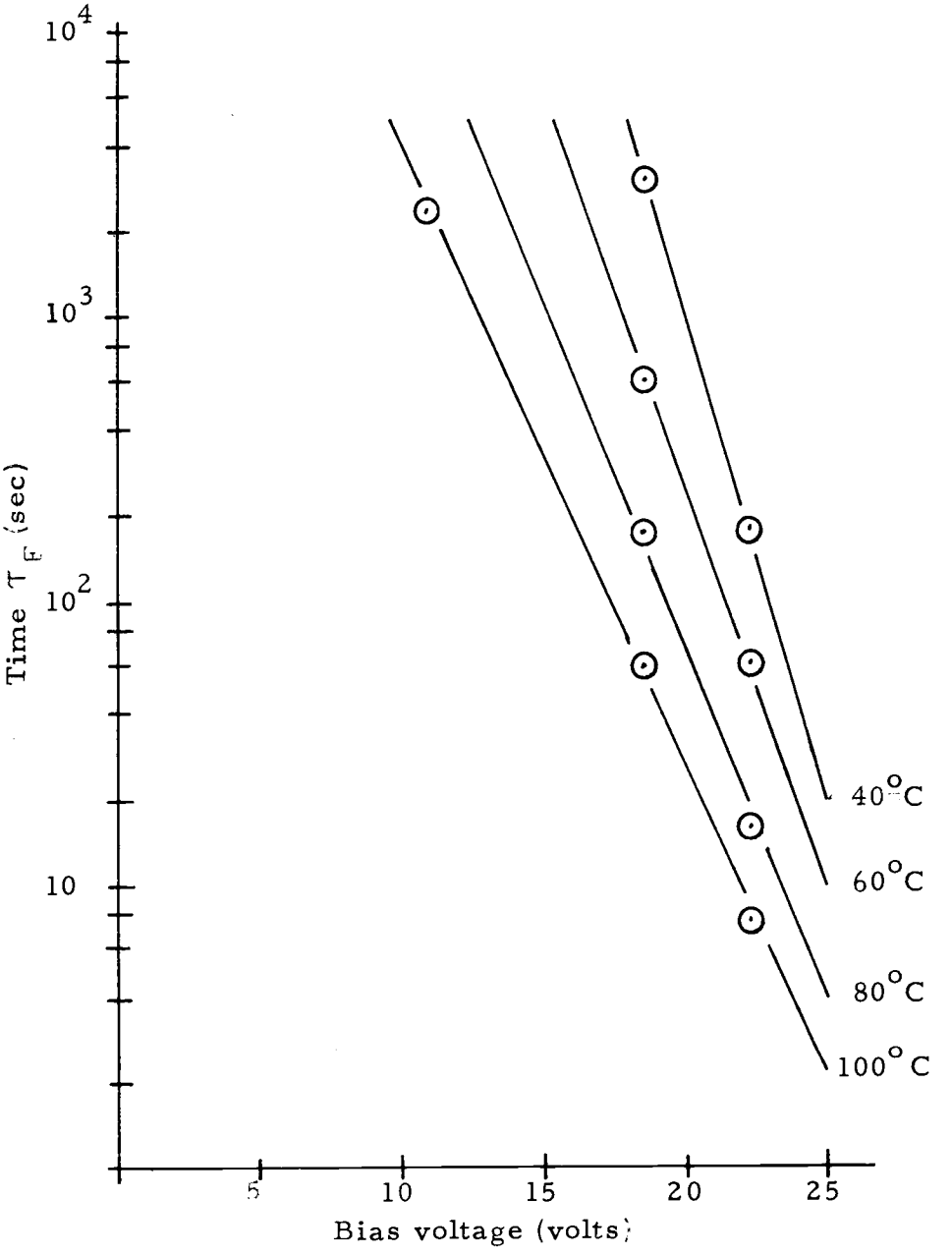

Figure 15. Bias voltage versus forward drift time constant for various temperatures.

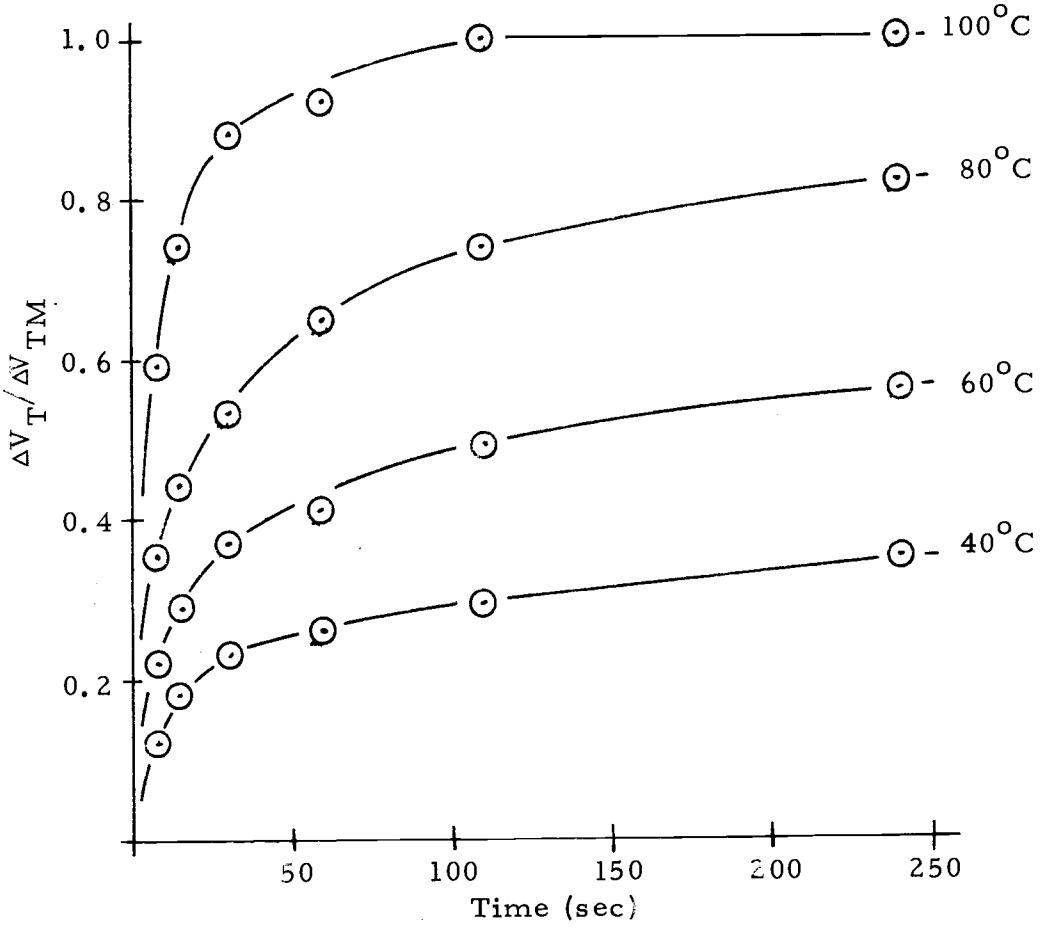

Figure 16. Normalized threshold voltage shift as a function of time (bias voltage = -11.4 volts and ΔV_{TM} = 6.8 volts).

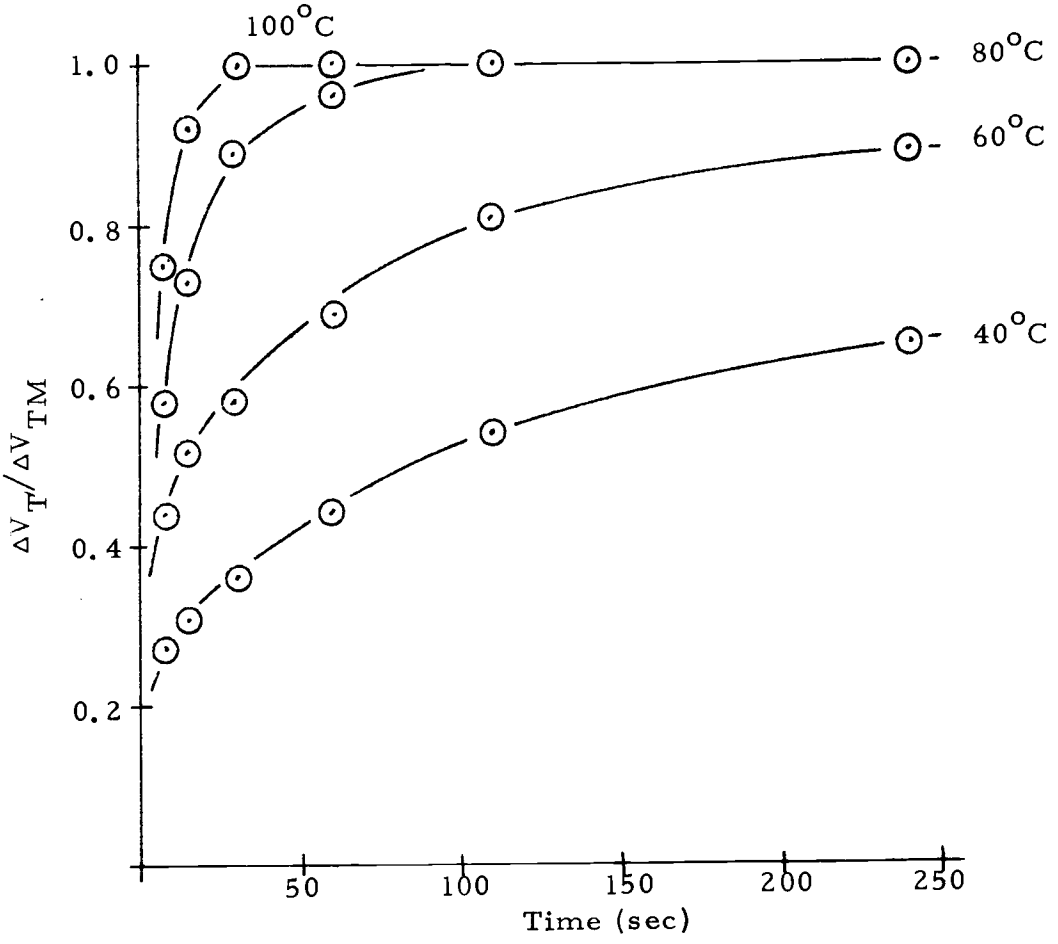

Figure 17. Normalized threshold voltage shift as a function of time (bias voltage = -22.6 volts and ΔV_{TM} = 5.2 volts).

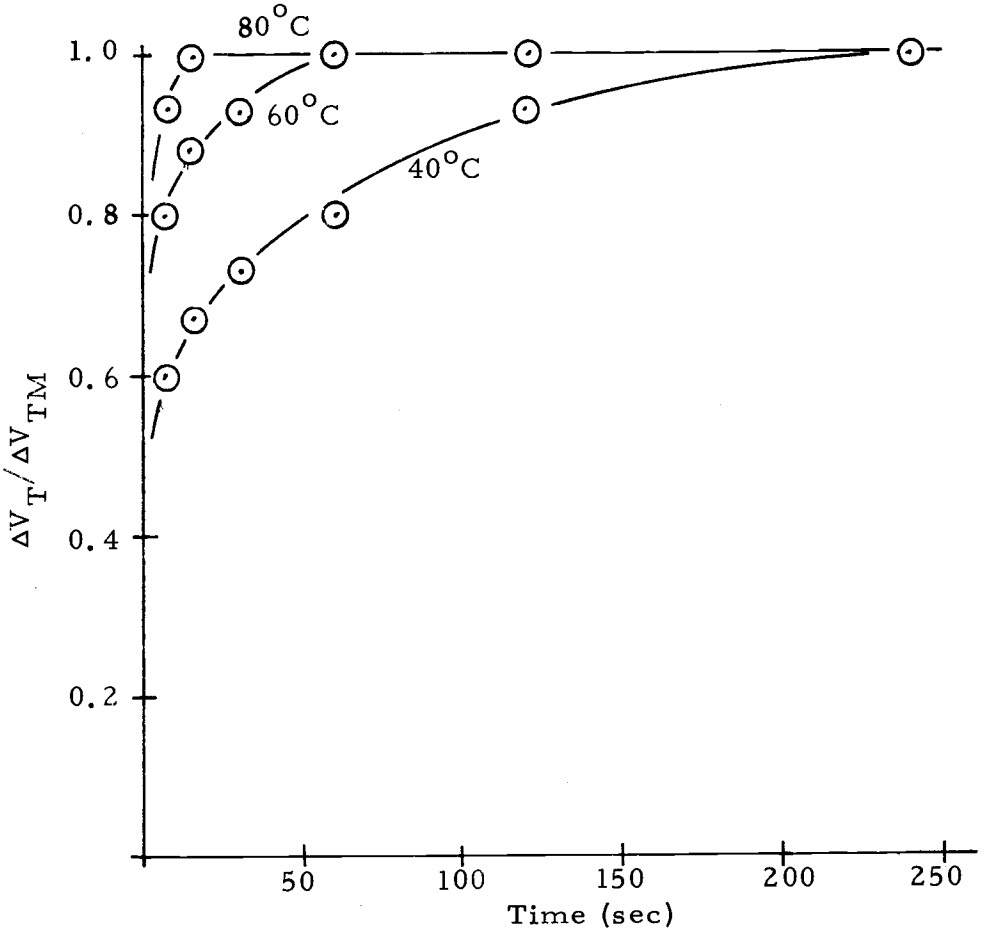

Figure 18. Normalized threshold voltage shift as a function of time (bias voltage = -35.2 volts and ΔV_{TM} = 6.0 volts).

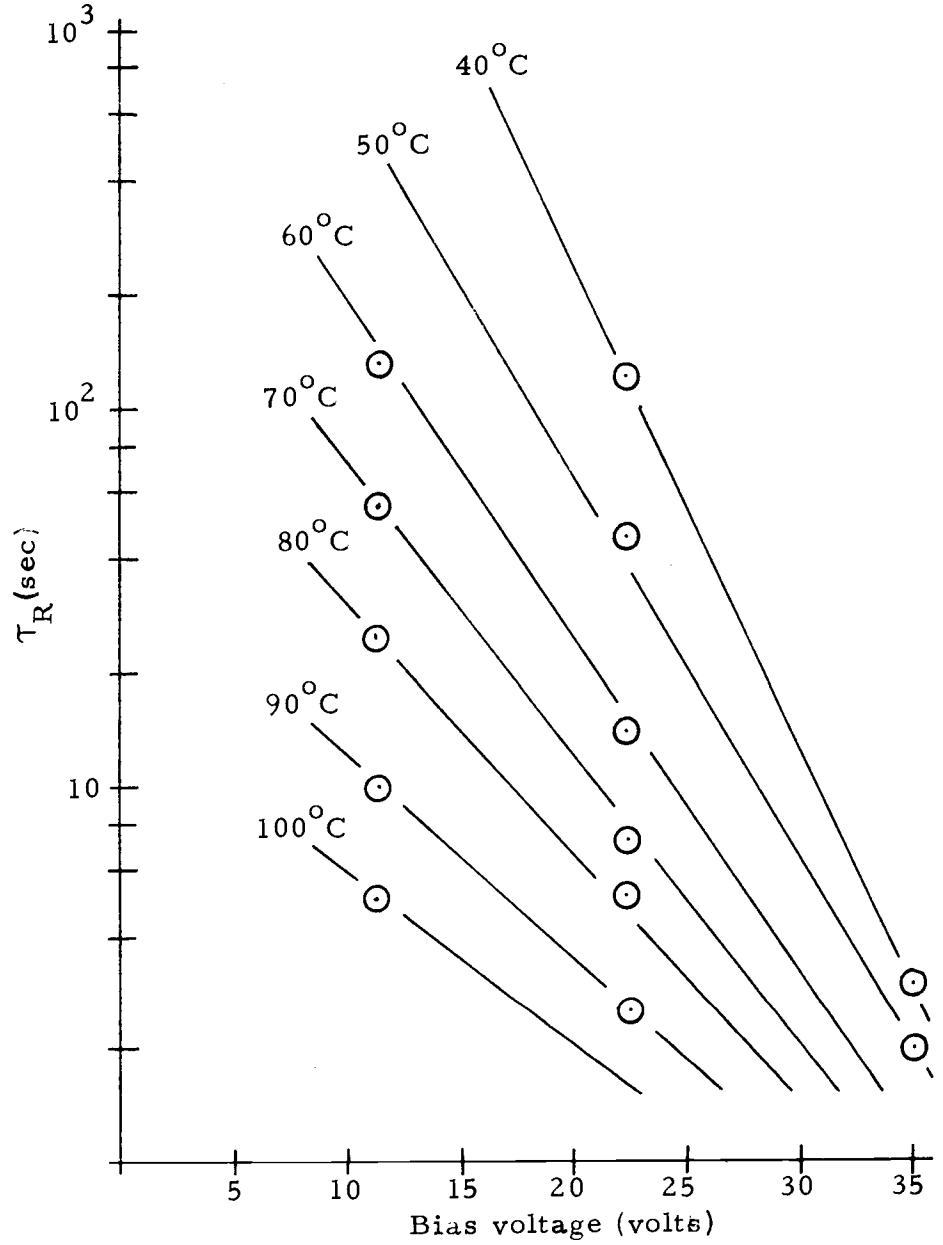

Figure 19. Bias voltage versus reverse drift time constant for various temperatures.

threshold voltage is equal to one half of the maximum change. The subscript is used to denote the forward time constant or the time constant obtained with a positive bias on the aluminum electrode.

Figure 15 shows the forward time constant versus bias voltage or applied electric field. The forward time constant has an exponential dependence on the applied electric field.

Similar bias-temperature measurements were also performed on some of the MOS capacitors using a negative applied electric field. In this case at the start the ions were located at the silicon-silicon dioxide interface. When a negative bias was applied to the MOS capacitor the ions drifted to the silicon dioxide-aluminum interface. Figures 16, 17, and 18 show the threshold voltage shift in the reverse direction for bias voltages of -11.4 volts, -22.6 volts, and -35.2 volts. A reverse time constant, τ_R, can also be defined as the time at which the change in threshold voltage is equal to one half of the maximum change. It was discovered that the reverse time constant was shorter than the forward time constant.

Figure 19 shows the reverse time constant for the MOS capcitors plotted for several different temperatures as a function of applied electric field.

Figures 20 and 21 show plots of τ_F and τ_R vs $1/T$ for various applied electric fields. From the two graphs it is obvious that τ_F and τ_R are an exponential function of $1/T$. Also the

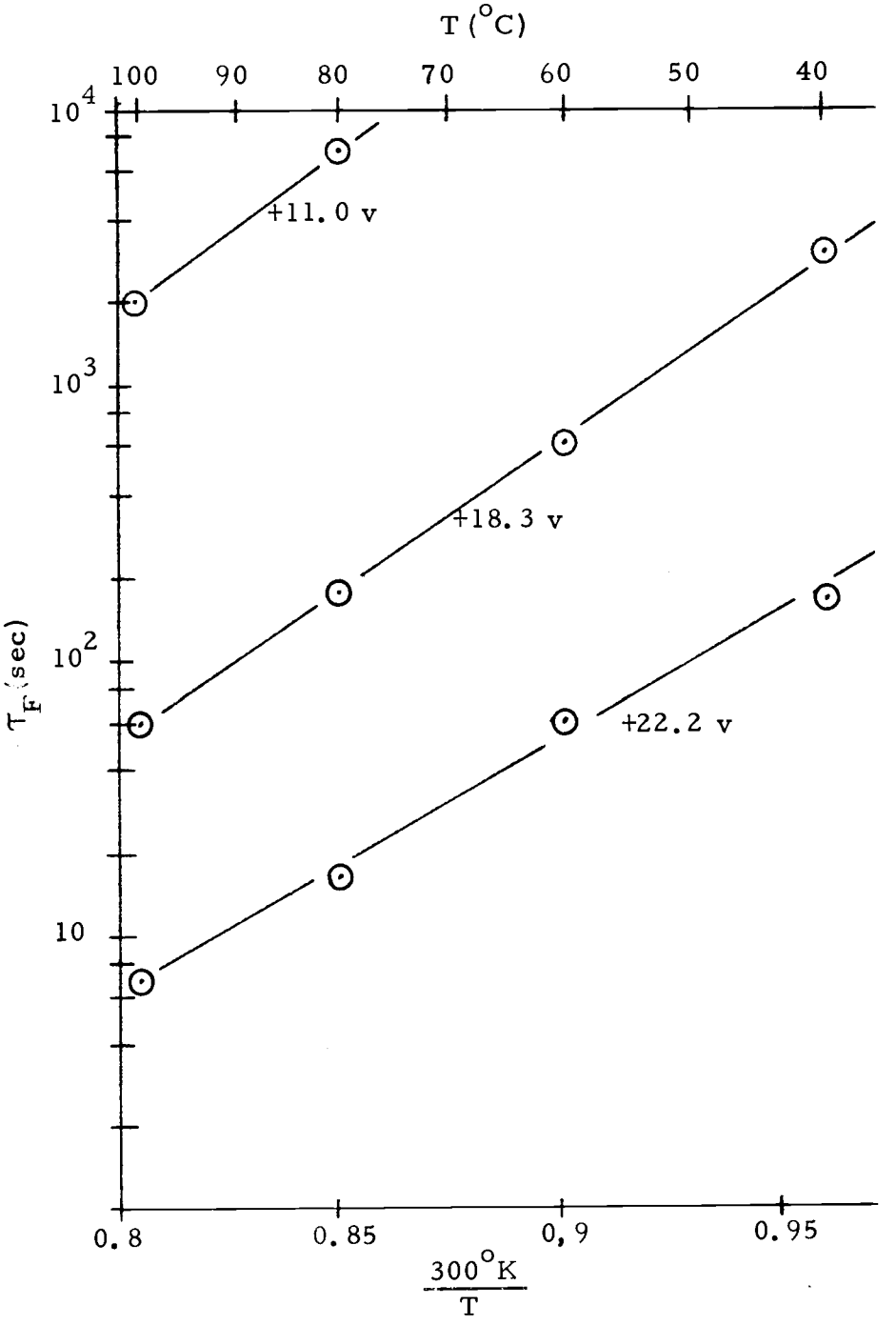

Figure 20. Temperature dependence of forward drift time constant for various bias voltages.

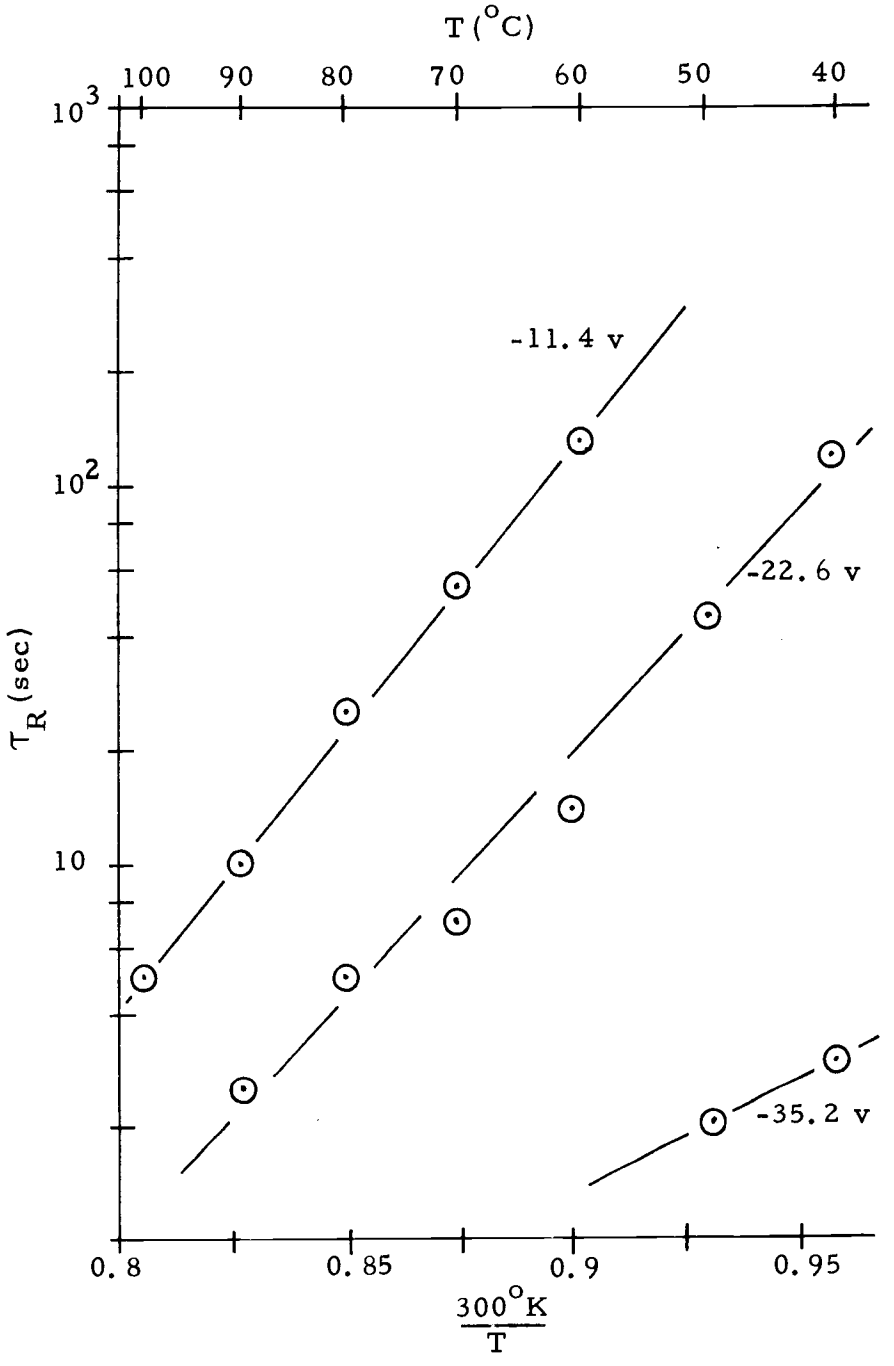

Figure 21. Temperature dependence of the reverse drift time constant for various bias voltages.

relationship depends upon the magnitude and direction of the applied electric field. This suggests that the time constant is expressed by the equation:

$$\tau_F = A\, e^{-\epsilon_A / K_B T}$$

where K_B = Boltzman constant

T = temperature in degrees Kelvin

A = a constant

ϵ_A = activation energy which is dependent upon the applied electric field.

Activation energies can be calculated from the slopes of the curves on Figures 20 and 21. The results appear in the Appendix, Table II.

These results are consistent with a trapping model proposed by Abbott and Kamins (1). As the applied electric field increases, the activation energy decreases. The ion needs to acquire a certain amount of energy to overcome the potential barrier and migrate through the oxide. This energy is called the activation energy. The activation energy depends upon the result of an interaction of a coulomb potential and an applied electric field. The electric field increases the slope of the potential by which the ion is influenced. The net result is that the activation energy decreases.

From the data contained in Figure 15 it is possible to estimate how many times a memory can be read. If a five-volt pulse is

applied to the gate in order to read the state of the transistor, this will cause the ions in the oxide to start to migrate through the oxide and charge the state of the transistor. The forward drift constant can be estimated by extrapolation of the data in Figure 15. The extrapolated time constant is 10^6 seconds. If a microsecond long pulse is used to read the memory, after 10^{12} pulses of 5 volts have been applied to the gate, the threshold voltage will have changed one half of its maximum change.

Experimentally Determined Migration Rates of Ions Through Silicon Dioxide

The three-layer gate insulator was proposed because of the necessity to reduce the mobility of the mobile ions in the gate dielectric. Therefore it is necessary to present some experimental evidence to support this idea.

An MOS capacitor having a gate dielectric consisting of 1500 Å of thermally grown silicon dioxide was fabricated. An aluminum electrode was vacuum deposited by an electron gun.

Figure 22 shows a change in threshold voltage normalized by the total change in threshold voltage versus time for a positive 5 volts on the gate. Figure 23 shows the same data except that a negative 5 volts is applied to the gate.

From Figures 22 and 23 it is possible to determine the forward

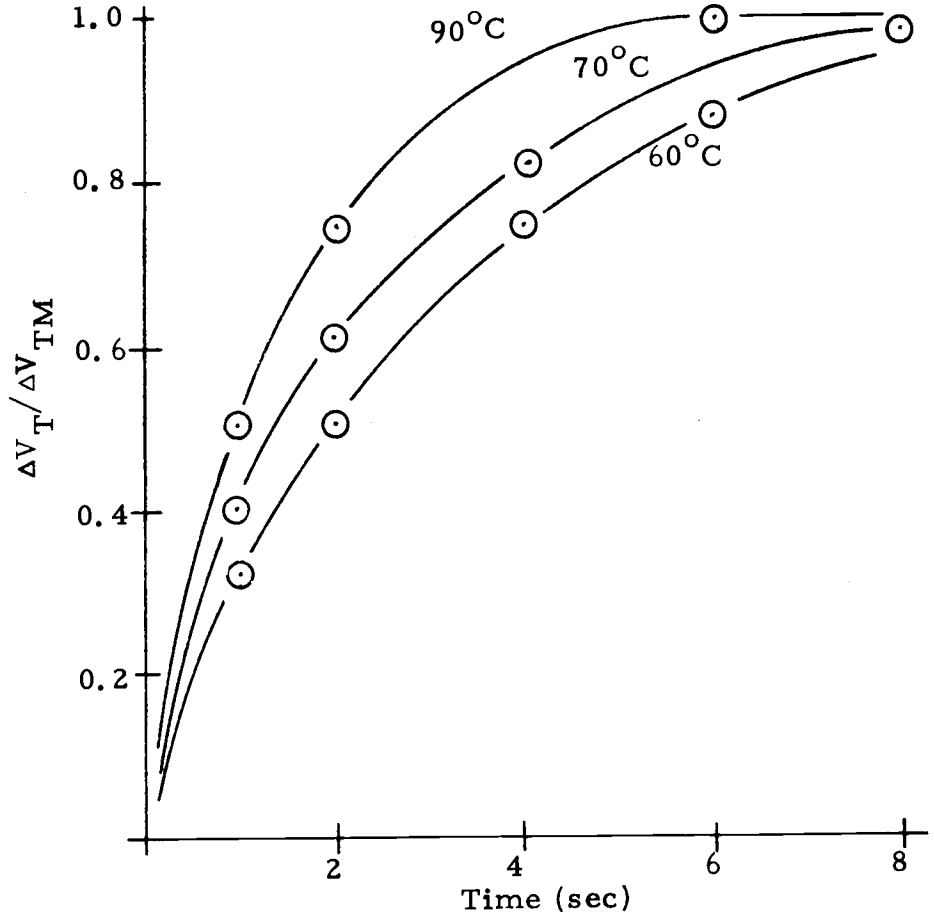

Figure 22. Normalized threshold voltage shift as a function of time (bias voltage = +5 volts and ΔV_{TM} = 6.0 volts).

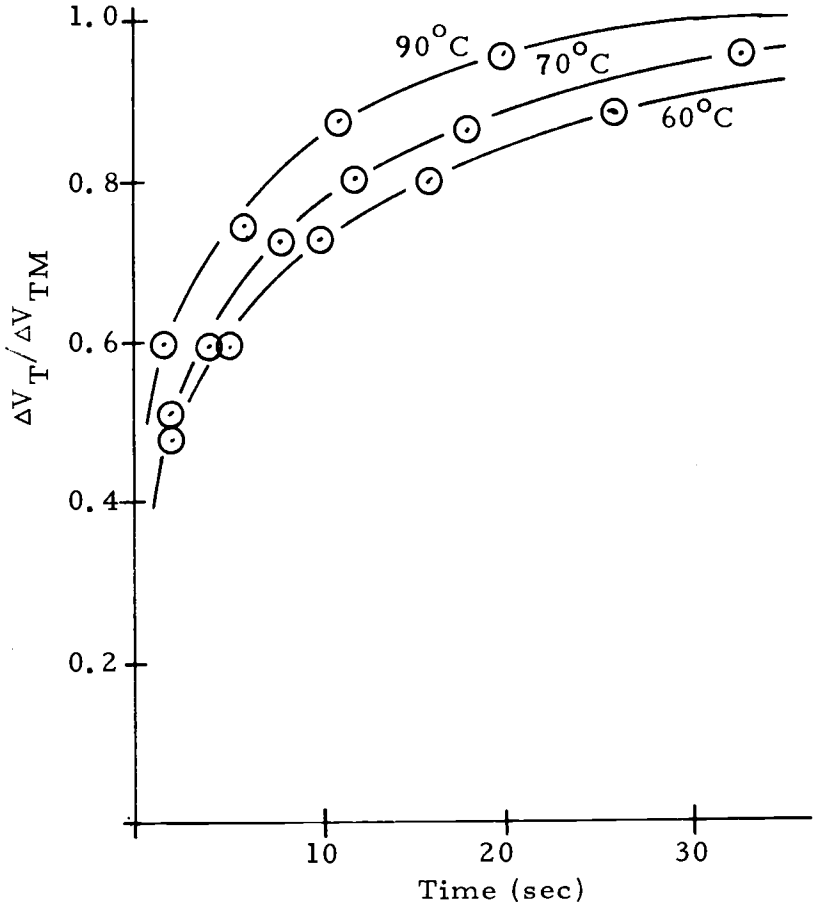

Figure 23. Normalized threshold voltage shift as a function of time (bias voltage = -5 volts and ΔV_{TM} = 6.0 volts).

time constant and reverse time constant for the temperatures that are given. From the forward and reverse time constants for the three temperatures chosen, it is possible to calculate the activation energy of the ionic motion as shown in Figure 24. The activation energies for the electric field of 3.3×10^5 v/cm are shown in the Appendix, Table III.

Evaluation of the Single MOS Field Effect Transistor

The bias-temperature test on the MOS capacitors was used to determine the maximum shift of the capacitance versus voltage of the MOS capacitors. Theory predicts that the voltage shift of the C-V curve of an MOS capacitor is the same as the shift in the threshold voltage of an MOS field effect transistor. In order to verify this, the threshold voltage was determined after various bias voltages were applied to the gate electrode for the single MOS field effect transistor. The threshold voltage of an enhancement type MOS field effect transistor is the value of gate voltage that initiates conduction between the source and the drain. Therefore it would be reasonable to plot the drain current as a function of the gate voltage for a particular value of drain voltage, assuming that the source is grounded. The intersection of line with the gate voltage axis would then be the threshold voltage. But the line never crosses the gate voltage axis because this would require current flow in the opposite direction. The line

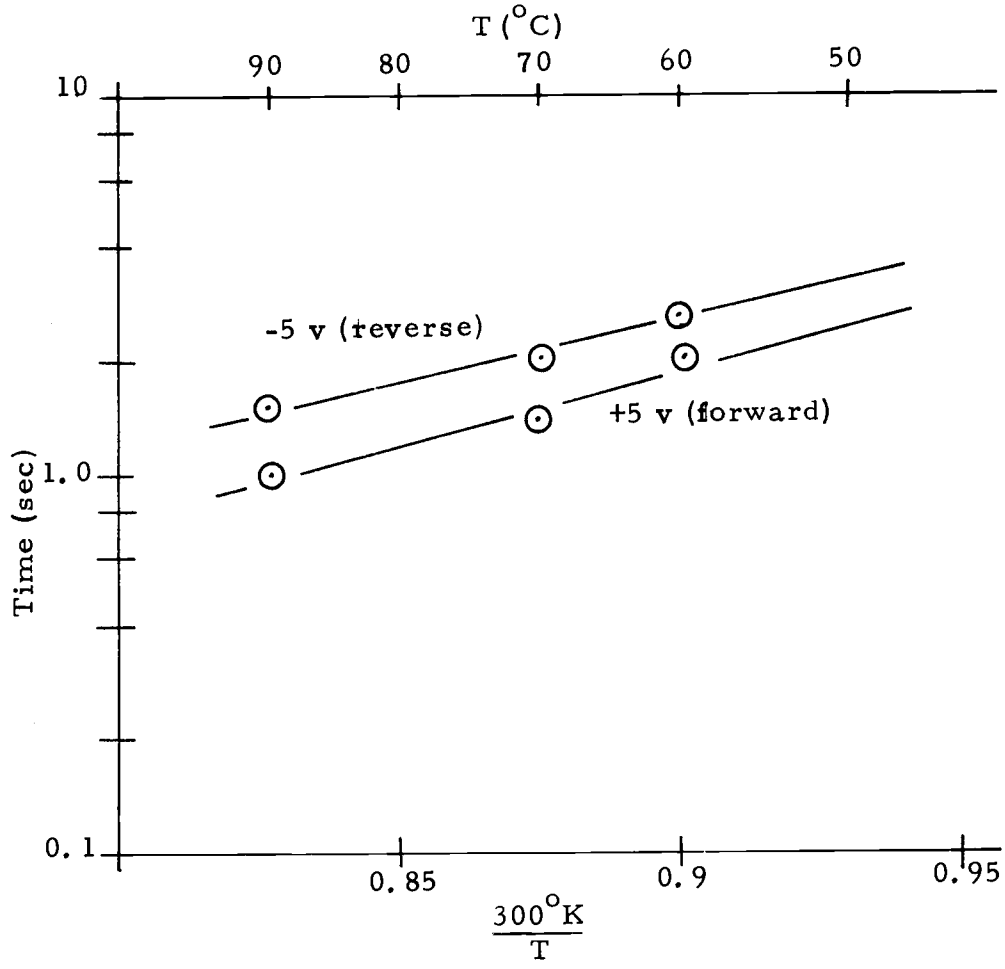

Figure 24. Temperature dependence of the forward and the reverse drift time constant for a negative and a positive field of 3.3×10^5 v/cm for silicon dioxide.

only approaches the gate voltage asymptotically. One way of getting this problem is to arbitrarily define the threshold voltage as the gate voltage at which there is a small amount of drain current. Manufacturers of transistors typically use 20 microamperes.

A better and less arbitrary method but more time consuming is to plot the square root of the drain current as a function of the gate voltage. Then the resulting plot is not a curve but a straight line which can be extended to cross the gate voltage axis.

Figure 25 contains two straight line plots for the square root of drain current obtained after -34 volts had been applied to the gate to ensure that all of the mobile ions were near the aluminum-silicon oxide interface. This would give the highest value of threshold voltage. The left line was obtained after +34 volts had been applied to the gate to ensure that all of the mobile ions were near the silicon-silicon dioxide interface. This would give the lowest value of threshold voltage. The difference between the two values for threshold voltage was 0.9 volts which compares reasonably close to the value obtained from the voltage shift of the C-V curve.

Another feature of plotting the characteristics of an MOS field effect transistor in the preceding manner is that if the oxide capacitance and W/L ratio for the transistors is known, it is possible to calculate the mobility of the electrons near the surface of the wafer from the slope of the line. Equation 17 shows the relationship.

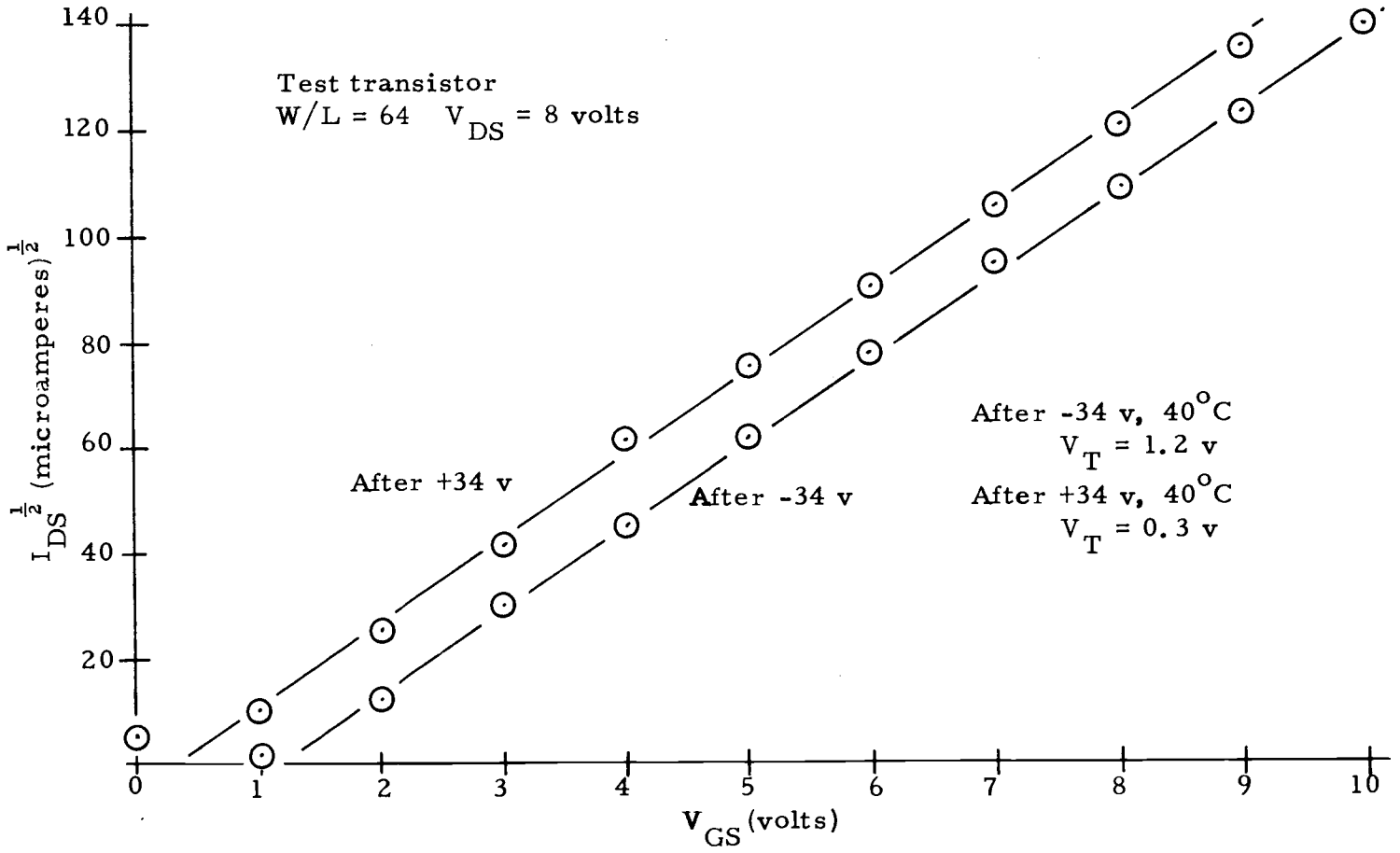

Figure 25. Transfer characteristics for the single MOS field effect transistor after positive and negative gate bias.

Evaluation of the MOS Field Effect Transistor as an Element in a Matrix

The operation of an MOS read-only memory has been previously described from a theoretical point of view. The writing operation and the reading operation have been explained. Several devices including MOS capacitors and an MOS field effect transistor have been fabricated and studied to give an understanding of just how an MOS transistor employing a three-layer dielectric may be used as a read-only memory.

The memory matrix that contained the 16 MOS field effect transistors were tested using an Adar pulse generator. The pulse generator had the capability of supplying 16 separate outputs of pulses, either positive or negative. The sequence of the train of output pulses could be programmed. The electrical pulses that were applied to the matrix were a positive five volts for a duration of 10 msecs.

Essentially, the matrix of MOS field effect transistors was tested to determine the level of the output voltage or current for the two states of the individual transistors. The current was sensed by measuring the voltage developed across a 1 K resistor. Five volts were applied simultaneously to a gate row and a drain column to specify an individual transistor. The two voltages caused a voltage at the output to rise to 0.29 mv. This is the "0" state or the state

determined by the one-volt threshold voltage.

In order to determine the value of current for the "1" state, the individual transistor in the matrix had to be switched. This was accomplished by applying +20 volts to the row of gates which contained the transistor that was to be switched, 0 volts to the rest of the gates, -10 volts to the drain row which contained the transistor that was to be switched, +10 volts to the other drain connections, and -10 volts to the substrate. These voltages were held for 10 seconds.

The readout operation was then repeated using five volts for the x-line, five volts for the y-line, and a 1 K resistor to sense the current. The average current for the MOS transistor that was desired to be switched was now 0.36 ma. The higher current was due to the fact that the mobile ions were forced to the silicon dioxide-silicon interface which lowered the threshold voltage. The five-volt pulses were also applied to the MOS transistors that were adjacent to the switched transistor. The average value of the current was now 0.30 ma, indicating that a significant change in their threshold voltages had not taken place.

The values of current for the "0" state and "1" state compared favorably to those predicted by theory as prescribed in a previous section.

V. CONCLUSIONS

The electrically alterable solid state semiconductor memories which are currently being manufactured lack one desirable feature. The memories are volatile; power has to be applied to the circuit in order to have the circuit retain the information.

The MOS device that was investigated was an electrically alterable memory. The state of the memory was determined by the threshold voltage of an MOS transistor.

The threshold voltage of an MOS transistor is very sensitive to the quantity and the position of charged particles in the gate insulator. Some of these charged particles, mainly sodium, were responsible for the instabilities that plagued the first MOS devices. Since these ions can be controlled by the application of electric fields between the metal gate and the silicon substrate, one has the possibility of constructing an electrically controlled and alterable non-volatile semiconductor memory.

For the variable threshold memory field effect transistor the single layer of silicon dioxide was replaced by a three-layer dielectric. The first layer or layer next to the silicon was 700 Å of silicon dioxide. The next layer was 300 Å of aluminum oxide which was deposited in a vacuum. The last layer was a pyrolytically deposited 700 Å layer of silicon dioxide. The composite structure was

necessary in order to reduce the mobility of the mobile ions in the gate insulator. For an electric field of 3.3×10^5 volts per centimeter and a temperature of $60^{\circ}C$, the time which it takes the threshold voltage to change one half of the total change in the forward direction is increased from 2.0 seconds to approximately 10^6 seconds, and the reverse time constant is increased from 2.5 seconds to approximately 5×10^2 seconds. The forward direction is defined by a positive voltage on the gate. The reverse direction is defined by a negative voltage on the gate.

For the composite three-layer dielectric, the rate of drift of the ions is an exponential function of applied electric field for either polarity. The forward time constants are larger than the reverse time constants. This indicates a stronger binding force of the ions near the silicon interface than the aluminum interface. Also as the applied electric field increases the activation energy decreases, indicating a lowering of the barrier which traps the ionic charges.

The usefulness of an MOS memory which employs the position of mobile ions to determine the state of the memory can be questioned. The speed at which the memory can be made to change states is relatively slow. Also since there is no true switching threshold voltage, a voltage at which no change in state takes place no matter how the voltage is applied, it would be difficult to operate large arrays of memory transistors.

BIBLIOGRAPHY

1. Abbott, R.A. and T.I. Kamins. Sodium migration through electron-gun evaporated Al_2O_3 and double layer Al_2O_3-SiO_2 structures. Solid-State Electronics 13:565-576. 1970.

2. Chu, T.L., J.R. Szedon, and G.A. Gruber. Silica films by the oxidation of silane. Transactions of the Metallurgical Society of AIME 242:532-538. 1968.

3. Crawford, R.H. MOSFET in circuit design. New York, McGraw-Hill, 1967. 136 p.

4. Deal, B.E., E.H. Snow, and C.A. Mead. Barrier energies in metal-silicon dioxide-silicon structures. The Journal of Physics and Chemistry of Solids 27:1873-1879. 1966.

5. Deal, B.E., M. Sklar, A.S. Grove, and E.H. Snow. Characteristics of the surface-state charge (Q_{ss}) of thermally grown oxidized silicon. Journal of the Electrochemical Society 114: 266-274. 1967.

6. Delzer, D.R. Vacuum evaporated dielectrics in MOS structures. Doctoral thesis. Corvallis, Oregon State University, 1969. 153 numb. leaves.

7. Goetzberger, A. Ideal MOS curves for silicon. Bell System Technical Journal 45:1097-1122. 1966.

8. Grove, A.S., B.E. Deal, E.H. Snow, and C.T. Sah. Investigation of thermally oxidized silicon surfaces using metal-oxide-semiconductor structures. Solid-State Electronics 8:145-163. 1965.

9. Grove, A.S. Physics and technology of semiconductor devices. New York, Wiley, 1967. 366 p.

10. Hofstein, S.R. An investigation of instability and charge motion in metal-silicon oxide-silicon structures. Institute of Electrical and Electronics Engineers Transactions on Electron Devices ED-13:222-237. 1966.

11. Hofstein, S. R. Proton and sodium transport in SiO_2 films. Institute of Electrical and Electronics Engineers Transactions on Electron Devices ED-14:749-759. 1967.

12. Hwang, Yun-Sheng. Vacuum-deposited aluminum-oxide MOS transistors. Master's thesis. Corvallis, Oregon State University, 1970. 66 numb. leaves.

13. Nakanuma, S., T. Tsujide, R. Igarashi, K. Onoda, T. Wada, and M. Nakagiri. Read-only memory using MAS transistors. Institute of Electrical and Electronics Engineers Journal of Solid-State Circuits SC-5:203-207. 1970.

14. Sah, C. T. Characteristics of the metal-oxide-semiconductor transistors. Institute of Electrical and Electronics Engineers Transactions on Electron Devices ED-11:324-345. 1964.

15. Snow, E. H., A. S. Grove, B. E. Deal, and C. T. Sah. Ion transport phenomena in insulating films. Journal of Applied Physics 36:1664-1673. 1965.

16. Snow, E. H., and B. E. Deal. Polarization effects in insulating films on silicon-a review. Transactions of the Metallurgical Society of AIME 242:512-522. 1968.

17. Sze, S. M. Physics of semiconductor devices. New York, Wiley, 1969. 812 p.

18. Wallmark, J. T. and J. H. Scott. Switching and storage characteristics of MIS memory transistors. RCA Review 30:335-365. 1969.

19. Yon, E., W. H. Ko, and A. B. Kuper. Sodium distribution in thermal oxide on silicon by radiochemical and MOS analysis. Institute of Electrical and Electronics Engineers Transactions on Electron Devices ED-13:276-280. 1966.

20. Zaininger, K. H. and F. P. Heiman. The C-V technique as an analytical tool. Solid State Technology 13:49-56. 14:46-55. 1970.

APPENDIX

Table I. Flat-band voltages, metal-semiconductor work functions, and metal work functions for various gate electrode materials.

	V_{FB}	ϕ_{MS}	ϕ_M
Gold	0.7 eV	+0.4 eV	+4.8 eV
Gallium	1.6 eV	-0.5 eV	+3.9 eV
Mercury	1.3 eV	-0.2 eV	+4.2 eV
Aluminum	1.4 eV	-0.3 eV	+4.1 eV

Table II. Activation energies for the positive ions in a three-layer oxide for various electric fields.

V_{GS}	Electric Field	Activation Energy
+11.0 v	6.5×10^5 v/cm	0.72 eV
+18.3 v	10.7×10^5 v/cm	0.62 eV
+22.2 v	13.0×10^5 v/cm	0.56 eV
-11.4 v	6.7×10^5 v/cm	0.85 eV
-22.6 v	13.3×10^5 v/cm	0.72 eV
-35.2 v	20.7×10^5 v/cm	0.37 eV

Table III. Activation energies for the positive ions in silicon dioxide.

V_{GS}	Electric Field	Activation Energy
+5.0 v	3.3×10^5 v/cm	0.20 eV
-5.0 v	3.3×10^5 v/cm	0.22 eV