Analysis and Modeling of Floating-Gate EEPROM Cells

AVINOAM KOLODNY, SIDNEY T. K. NIEH, BOAZ EITAN, MEMBER, IEEE, AND JOSEPH SHAPPIR, MEMBER, IEEE

Abstract—**Floating-gate MOS devices using thin tunnel oxide are becoming an acceptable standard in electrically erasable nonvolatile memory. Theoretical and experimental analysis of WRITE/ERASE characteristics for this type of memory cell are presented. A simplified device model is given based on the concept of coupling ratios. The WRITE operation is adequately represented by the simplified model. The ERASE operation is complicated due to formation of depletion layers in the transistor's channel and under the tunnel oxide. Experimental investigation of these effects is described, and they are included in a detailed cell model.**

In certain cell structures, a hole current can flow from the drain into the substrate during the ERASE oepration. This effect is shown to be associated with positive charge trapping in the tunnel oxide and threshold window opening. An experimental investigation of these phenomena is described, and a recommendation is made to avoid them by an appropriate cell design.

I. INTRODUCTION

ELECTRICALLY erasable nonvolatile memory (EEPROM) technology has emerged in recent years as a promising approach for implementing sophisticated VLSI systems [1]–[5]. Among the various devices that have been used to realize such a memory, floating-gate MOS transistors that employ a thin insulator for electron tunneling (FLOTOX) have been dominant [1]–[4]. This paper presents a theoretical analysis and experimental data of the programming and erasing of FLOTOX memory cells.

The general device structure is depicted schematically in Fig. 1. This is an n-channel double-poly transistor in which the first polysilicon is floating. A thin (~100 Å) dielectric layer between the floating gate and the drain enables the flow of electrons into and from the floating gate during WRITE/ERASE operations, by means of Fowler–Nordheim tunneling [6], [7].

In the WRITE operation the floating gate is charged negatively with electrons tunneling from the drain through the thin oxide. This is achieved by applying a positive voltage pulse to the top gate of the cell, while the source, drain, and substrate are grounded. The stored negative charge on the floating gate shifts the transistor's threshold voltage, as measured on the top gate, toward a more positive value. In a subsequent READ operation the transistor will not conduct channel current.

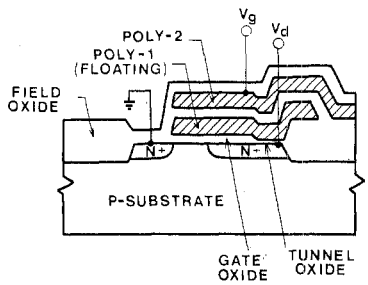


Fig. 1. Cross section of an EEPROM cell.

The ERASE operation removes electrons from the floating gate by applying a positive high-voltage pulse at the drain, while the source is floating and both the top gate and the substrate are grounded. The threshold voltage is shifted in the negative direction, and channel current would flow during subsequent READ operations.

During READ operations the voltages used are low enough such that tunnel current is negligible, and the floating gate is practically insulated. Charge retention in excess of 10 years can readily be obtained on the floating gate under normal operating conditions [9].

In memory circuits, a two-transistor cell is used [1]–[5]; the additional transistor is necessary in order to isolate the cell from adjacent cells during WRITE/ERASE operations.

This work focuses on analysis and modeling of the WRITE/ERASE operations, considering effects that occur during ERASE, which have not been discussed in the literature [9], [10]. An understanding of these effects is of major importance in cell design and optimization.

II. SIMPLIFIED DEVICE MODEL

A. Calculation of Tunnel Current

The tunneling current density through the tunnel oxide is approximated by the well-known Fowler–Nordheim equation [6], [7].

$$J_{tun} = \alpha E_{tun}^2 \cdot (\exp(-\beta/E_{tun})) \quad (1)$$

where E_{tun} is the electric field in the oxide, and α and β

Manuscript received March 13, 1985; revised December 31, 1985.

A. Kolodny was with Intel Corporation, Santa Clara, CA 95051. He is now at the Intel Israel Design Center, Haifa, Israel.

S. Nieh was with Intel Corporation, Santa Clara, CA 95051. He is now with Intel Corporation, Portland, OR.

B. Eitan is with WaferScale Integration, Inc., Fremont, CA 94538.

J. Shappir is with the Hebrew University of Jerusalem, Jerusalem, Israel.

IEEE Log Number 8608243.

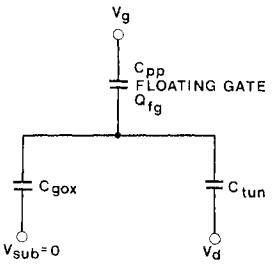


Fig. 2. A simplified capacitive equivalent circuit of the EEPROM cell.

are constants. The thin-oxide field E_{tun} is given by

$$E_{\text{tun}} = \frac{|V_{\text{tun}}|}{X_{\text{tun}}} \tag{2}$$

where V_{tun} is the voltage drop across the oxide and X_{tun} is its thickness. V_{tun} can be calculated from a capacitive equivalent circuit of the cell.

B. Calculation of V_{tun}

In order to gain insight into the basic device operation, a simplified equivalent circuit, shown in Fig. 2, is used. A more detailed analysis is given in Section V. In Fig. 2, C_{pp} is the interpoly capacitance, C_{tun} is the thin oxide capacitance, and C_{gox} is the capacitance of the gate oxide between the floating gate and the substrate. Q_{fg} is the stored charge on the floating gate. V_{tun} can be expressed for an electrically neutral floating gate in terms of simple coupling ratios

$$|V_{\text{tun}}|_{\text{WRITE}} = V_g \cdot K_w \tag{3}$$

where

$$K_w = \frac{C_{pp}}{C_{pp} + C_{\text{gox}} + C_{\text{tun}}} \tag{4}$$

and

$$|V_{\text{tun}}|_{\text{ERASE}} = V_d \cdot K_e \tag{5}$$

where

$$K_e = 1 - \frac{C_{\text{tun}}}{C_{pp} + C_{\text{gox}} + C_{\text{tun}}}. \tag{6}$$

The coupling ratios K_w and K_e denote the fraction of the applied voltage that appears across the tunnel oxide. Note that (3) and (5) are applicable only when $Q_{fg} = 0$. During WRITE operation buildup of negative stored charge of the floating gate will reduce the tunnel-oxide voltage according to

$$|V_{\text{tun}}|_{\text{WRITE}} = V_g \cdot K_w + \frac{Q_{fg}}{C_{pp} + C_{\text{gox}} + C_{\text{tun}}}. \tag{3'}$$

In the ERASE operation, the initial negative stored charge on the floating gate will increase the tunnel-oxide voltage according to

$$|V_{\text{tun}}|_{\text{ERASE}} = V_d \cdot K_e - \frac{Q_{fg}}{C_{pp} + C_{\text{gox}} + C_{\text{tun}}}. \tag{5'}$$

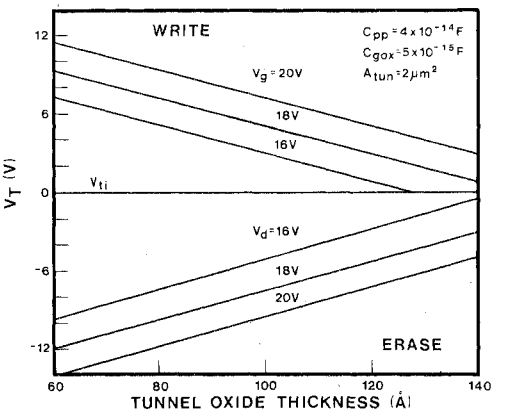


Fig. 3. WRITE/ERASE threshold window versus tunnel oxide thickness, calculated with the approximation of (8), (9), assuming that $V'_{\text{tun}} = 1 \times 10^7 \cdot X_{\text{tun}}$ at the end of the operation.

at the end of the ERASE operation when positive charge is built up on the floating gate, the last term in (5) will reduce the tunnel-oxide voltage.

C. Calculation of Threshold Voltages

The initial threshold voltage of the cell, corresponding to $Q_{fg} = 0$ is denoted by V_{ti}. Stored charge shifts the threshold according to the relation

$$\Delta V_t = -\frac{Q_{fg}}{C_{pp}}. \tag{7}$$

Using (3') and (5') for Q_{fg} at the end of the WRITE/ERASE pulse, the cell's threshold voltages are

$$V_{tw} = V_{ti} - \frac{Q_{fg}}{C_{pp}} = V_{ti} + V_g\left[1 - \frac{V'_{\text{tun}}}{K_w \cdot V_g}\right] \tag{8}$$

$$V_{te} = V_{ti} - \frac{Q_{fg}}{C_{pp}} = V_{ti} - V_d\left[\frac{K_e}{K_w} - \frac{V'_{\text{tun}}}{K_w \cdot V_d}\right]. \tag{9}$$

Here V_{tw} is the threshold of a written cell, and V_{te} is the threshold of an erased cell. V_g and V_d are the WRITE/ERASE pulse amplitudes, respectively, and V'_{tun} is the tunnel-oxide voltage at the end of the pulse. Assuming that the WRITE/ERASE pulse is sufficiently long, the thin-oxide field will be reduced to below about 1×10^7 V/cm, when tunneling practically "stops." An approximation of V'_{tun} can be calculated from (2), and substituted in (8), (9) to give the approximate programming window of the cell and its dependence on cell parameters and programming voltage. Typical results are shown in Fig. 3.

In order to maximize the cell's window at a given tunnel-oxide thickness and WRITE/ERASE voltage, the coupling ratios should approach unity. Both coupling ratios can be increased by reducing C_{tun} and increasing C_{pp}. At a given tunnel-oxide thickness, this is usually achieved by minimizing the thin oxide area and adding extra poly-poly overlap area on the sides of the cell transistor. Typical coupling ratios are about 0.7 (K_e is always higher than K_w). Increasing the gate-oxide capacitance C_{gox} improved K_e but lowers K_w.

D. Dependence of Thresholds on WRITE/ERASE Time

An analytic expression for the cell's threshold versus programming time is obtained by solving the differential equation

$$\frac{dQ_{fg}}{dt} = A_{\text{tun}} \cdot J_{\text{tun}} \tag{10}$$

using the expressions in (1), (2), (3'), (5'), and (7). The resultant solutions are

$$V_{tw}(t) = V_{ti} + V_g - \frac{1}{K_w} \cdot \frac{B}{\ln (A \cdot B \cdot t + E_1)} \tag{11}$$

$$V_{te}(t) = V_{ti} - V_d \frac{K_e}{K_w} + \frac{1}{K_w} \cdot \frac{B}{\ln (A \cdot B \cdot t + E_2)}. \tag{12}$$

where

$$A = \frac{A_{\text{tun}} \cdot \alpha}{X_{\text{tun}} \cdot (C_{pp} + C_{\text{gox}} + C_{\text{tun}})}$$

$$B = \beta \cdot X_{\text{tun}}$$

$$E_1 = \exp\left[\frac{B}{K_w \cdot (V_g + V_{ti} - V_t(0))}\right]$$

$$E_2 = \exp\left[\frac{B}{V_d \cdot K_e + K_w \cdot V_t(0) + K_w \cdot V_{ti}}\right].$$

$V_t(0)$ is the cell's threshold at $t = 0$, which should not be confused with V_{ti}, the threshold of a neutral cell. A_{tun} is the tunnel-oxide area.

Note in (11) that the threshold voltage remains virtually constant at $V_t(0)$ if V_g is applied for a period that is less than a characteristic "time constant" τ defined by

$$\tau = \frac{1}{AB} \exp\left[\frac{B}{K_w(V_g + V_{ti} - V_t(0))}\right]. \tag{13}$$

For longer time t, the threshold asymptotically approaches the curve

$$V_{tw}(t) = V_{ti} + V_g - \frac{B}{K_w \cdot \ln (A \cdot B \cdot t)}. \tag{14}$$

Similar expressions are derived from (12) for the ERASE operation. These approximations are useful for cell design, and can be employed to evaluate the tradeoffs between programming time, retention time, threshold window, and operating voltages for any given set of cell parameters (A, B, V_{ti}, K_w, K_e). An example is given in Fig. 4, showing that a gate voltage of 5 V does not change the threshold even within several years. However, a pulse amplitude of 20 V can achieve a window of several volts within 1 ms.

III. Experimental Results and Comparison with the Simplified Model

In order to simulate the device performance, its physical parameters have to be measured. Some of the parameters (gate oxide thickness, poly-poly-oxide thickness) can be measured directly on large structures by well-known techniques (C–V or ellipsometry). However, most of the parameters are more difficult to measure and need special consideration for the floating-gate EEPROM device.

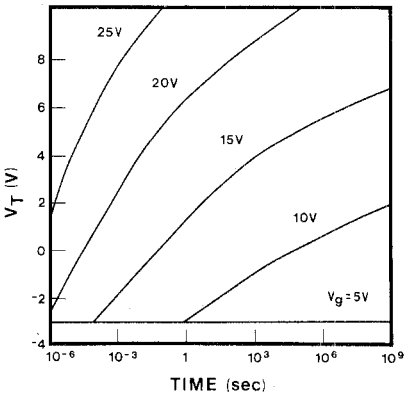


Fig. 4. Simulated threshold voltage versus time, for several WRITE pulse amplitudes (11). $V_t(0) = -3$ V, $V_{ti} = 0$ V, $C_{\text{total}} = 1 \times 10^{-13}$ F, $X_{\text{tun}} = 120$ Å, $A_{\text{tun}} = 2\ \mu\text{m}^2$, $K_w = 0.7$.

To measure the gate coupling ratio K_w (see (4)), a floating-gate device is compared with the equivalent MOS device (an identical device with direct contact to the first polysilicon gate). The coupling ratio can be obtained by comparing thresholds, body coefficients, or transconductances. In the latter two techniques, the coupling is calculated as the ratio of the body coefficient or transconductance of the MOS device and the EEPROM cell. However, these parameters are sensitive to short-channel effects, mobility degradation, and threshold-adjust boron profile. Care should be taken to measure the two devices at the same operating point. In the first technique, the coupling ratio is calculated as

$$K_w = \frac{V_t(\text{MOS}) - V_{fb}(\text{MOS})}{V_t(\text{EEPROM}) - V_{fb}(\text{EEPROM})}. \tag{15}$$

For accurate results it is important to ensure that $Q_{fg} = 0$. This is achieved by UV light erasure of the EEPROM cell. The drain coupling ratio K_e (see (6)) is calculated from the slope of the V_t versus V_d characteristics. The threshold voltage as a function of the drain voltage can be measured by defining threshold as an arbitrary current level in the weak-inversion region. A significant fraction of the "coupling ratio" measured by this technique is actually due to drain-induced barrier lowering [11]. This short-channel effect is measured similarly on the equivalent MOS device, and subtracted from the slope of the EEPROM device

$$K_e = 1 - K_w[\text{slope}(V_t \text{ versus } V_d(\text{EEPROM})) - \text{slope}(V_t \text{ versus } V_d(\text{MOS}))]. \tag{16}$$

The large coupling between the drain and the floating gate has a substantial effect on the I–V characteristics as is shown in Fig. 5. The slope of the I–V characteristics in the saturation region is mainly due to the increase in the floating gate voltage with V_d.

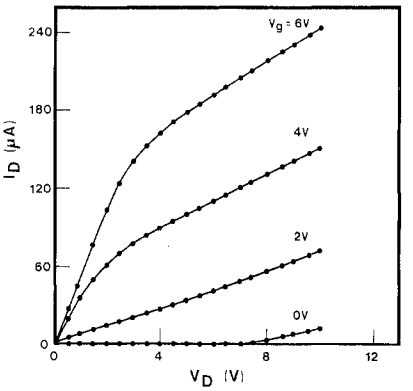


Fig. 5. I_d–V_d characteristics of the EEPROM cell, demonstrating the strong coupling from the drain to the floating gate. W/L = 5/3.5 μm, X_{tun} = 120 Å, K_w = 0.72, K_e = 0.88.

Extraction of the tunneling parameters α and β from the tunnel I–V characteristics is difficult and is a subject of controversy in the literature [7]. There are two major problems in determining the tunneling parameters; the first is the accuracy in which the tunnel-oxide thickness can be measured and the second is the change of the tunnel I–V characteristics with charge flow through the oxide due to positive and negative charge trapping [12], [13]. For simplicity, a compromise is made in which the oxide thickness is measured by the C–V technique assuming 3.9 for its relative dielectric constant, and tunnel coefficients are measured from an I–V taken after 0.1-C/cm^2 charge flowed through the oxide. This last decision is based on the fact that the positive charge trapping has saturated after this amount of charge has passed through the oxide, and the negative charge trapping is relatively small for the rest of the relevant device endurance [13]. Based on the above assumptions, it is found that

$$\alpha = 1.88 \times 10^{-6}\ \text{A/V}^2$$

$$\beta = 2.55 \times 10^{8}\ \text{V/cm}$$

for X_{tun} in the range 100–150 Å.

The calculated and measured results for the WRITE operation are compared in Fig. 6(a),(b). In Fig. 6(a) the threshold voltage as a function of WRITE time and in Fig. 6(b) the threshold voltage as a function of WRITE pulse amplitude are shown. For both examples the simulation results fit the measured data closely.

Trying to extend the same simplified model to the ERASE operation is shown in Fig. 7. The discrepancy between the simulated and measured results clearly demonstrates that the ERASE operation does not follow the simple model.

The main physical effects which cause the deviation of the ERASE characteristics from the simple model are: deep depletion under the gate, deep depletion under the tunnel oxide, and a current path for holes from under the tunnel oxide into the substrate. The detailed discussion of these mechanisms, along with the techniques to incorporate their effect in the device simulation, is the subject of the following sections.

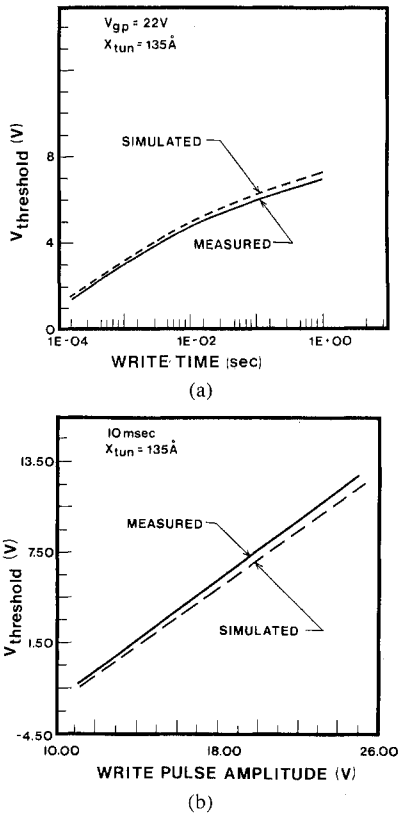


Fig. 6. (a) Measured and simulated threshold voltage as a function of WRITE time for a fixed pulse amplitude. (b) Measured and simulated threshold voltage as a function of WRITE pulse amplitude, for a fixed WRITE time.

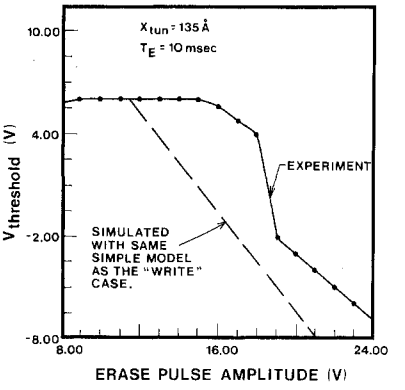


Fig. 7. Measured and simulated ERASE characteristics, using the simplified model. The discrepancy between the two curves is due to formation of depletion layers in the channel and under the tunnel oxide, and due to hole flow to the substrate.

IV. Mechanisms Affecting the Erase Operation

The "anomalous" behavior of the cell during ERASE is related to the formation of a depletion region in the channel and under the tunnel oxide, and to a hole current to the substrate. These three effects are shown schematically in Fig. 8 and will be analyzed below in more detail.

A. Deep Depletion in the Channel

During the ERASE operation the floating gate potential becomes positive due to the coupling of the positive drain voltage into the floating gate and the reduction in the neg-

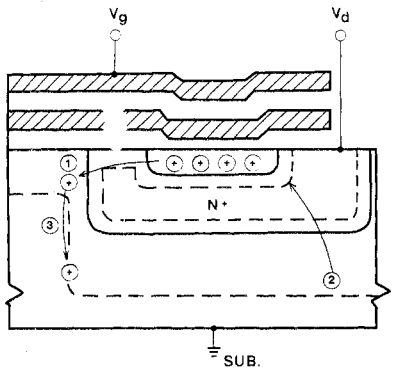


Fig. 8. Schematic illustration of three mechanisms affecting the ERASE operation: ① Deep depletion in the channel. ② Deep depletion under the tunnel oxide. ③ Current path for holes from under the tunnel oxide to the substrate.

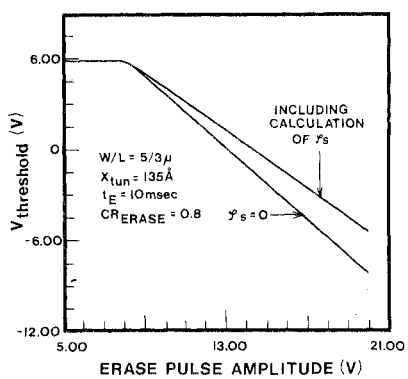


Fig. 9. Simulated ERASE characteristics with and without the effect of channel depletion. This effect reduces the efficiency of the ERASE operation.

ative stored charge. As a consequence, a depletion-layer is formed in the channel (region ① in Fig. 8), reducing the effective capacitance between the floating gate and the substrate, causing a reduction in the ERASE coupling ratio of the cell. In order to avoid channel current, which can overload the on-chip high-voltage generator supplying V_d, the source is floated. As a result the source is pulled up, tracking the floating-gate potential. This further reduces the ERASE coupling ratio.

Calculation of the channel depletion effect on ERASE characteristics is given in Section V. A typical example of simulated erased threshold as a function of ERASE drain voltage is shown in Fig. 9. The effect of the deep depletion in the channel is becoming more pronounced as the cell threshold is reduced, since the channel surface potential (ϕ_s) is becoming more and more positive and the ERASE coupling ratio is reduced.

B. Deep Depletion Under the Tunnel Oxide

1) The Origin of Deep Depletion: An electric field intensity above 1×10^7 V/cm is required for significant tunneling current. At this field, the n^+ region beneath the tunnel oxide is inverted or depleted, depending on its doping level and the availability of holes in this region. Hence, there is a voltage drop across the depletion layer in the n^+ region (denoted by ② in Fig. 8).

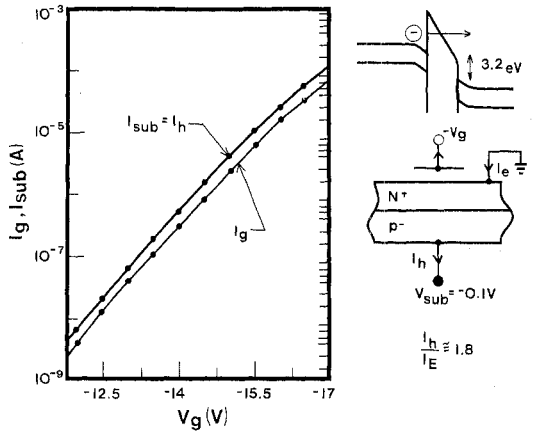


Fig. 10. Measurement of impact ionization by tunnel electrons, entering the silicon from the SiO_2 with the high energy (>3.2 eV, as shown in the insert). Electrons and holes are separated by the p – n^+ junction; from the ratio of the two currents, the ionization rate is 1.8 pairs per tunnel electron.

This voltage drop ϕ_{sn} is usually more than the equilibrium value of $2\phi_f$ since thermal equilibrium cannot be reached during the short ERASE operation. The holes in the depletion layer can be generated by any one of the following four mechanisms [14]–[16]:

1) thermal generation;
2) avalanche multiplication;
3) band-to-band tunneling in the n^+ Si; and
4) pair generation by tunnel electrons from the floating gate.

The first mechanism is very slow and for a typical ERASE operation (in the range of a few milliseconds) its effect is negligible.

The avalanche multiplication and band-to-band tunneling are fast generation mechanisms. Avalanche multiplication is dominant for doping levels below about 5×10^{17} cm^{-3}, and band-to-band tunneling for higher doping levels. The voltage drop across the depletion layer under the tunnel oxide is pinned to a value denoted by ϕ_{gen} that is determined by the onset of a fast hole generation by either one of the above mechanisms [14].

Yet another source of holes in the depletion layer under the tunnel oxide, are the pairs generated by the tunnel electrons entering the silicon [15], [16]. These electrons are very energetic (hot) in the Si conduction band due to the 3.2-eV energy difference between the Si and the SiO_2 conduction bands and their kinetic energy in the oxide conduction band (see insert in Fig. 10). A measurement of both the electron tunneling and the impact ionization hole current into the substrate is shown in Fig. 10. It is found that on the average every tunneling electron generates 1.8 hole electron pairs [17]. This measurement is done on a capacitor configuration in which the hole and electron currents in the substrate are separated [18]. The mechanism of hole generation accelerates the collapse of the deep depletion under the tunnel oxide into inversion.

2) The Effect of the Deep Depletion under the Tunnel Oxide on ERASE Characteristics: The deep depletion un-

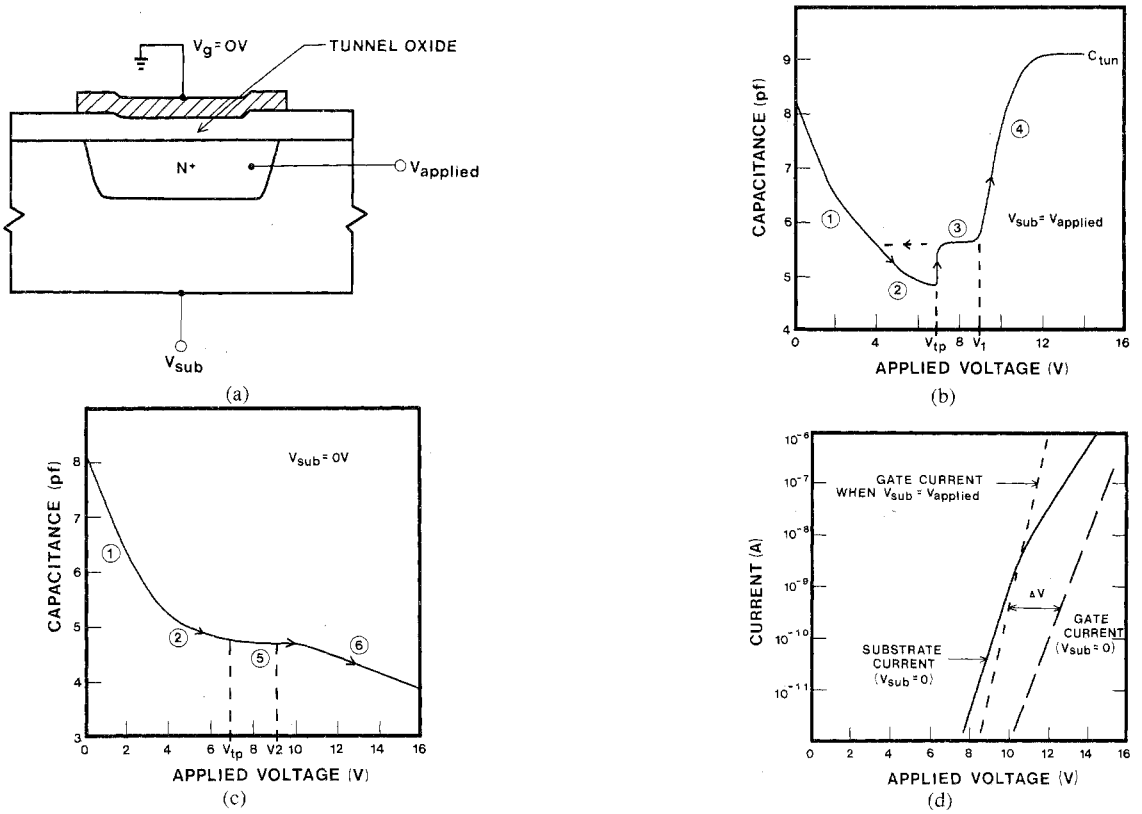


Fig. 11. (a) A cross section of the tunnel oxide capacitor. (b) A C–V curve measured with the substrate connected to the n^+ region. (c) A C–V curve measured with the substrate connected to the gate (grounded). (d) Gate and substrate currents measured on the same device.

der the tunnel oxide is associated with a potential drop ϕ_{sn} in the n^+ region. The effective ERASE voltage at the Si-tunnel oxide interface is $V_d - \phi_{sn}$. This is equivalent to a shift of the horizontal axis by ϕ_{sn} in the V_t versus V_d characteristic.

For practical modeling purposes, ϕ_{sn} is assumed to be constant throughout the ERASE operation. In a typical EEPROM cell where the n^+ concentration is higher than 1×10^{18} cm^{-3}, the initial ϕ_{sn} is equal to ϕ_{gen}, corresponding to hole generation by band-to-band tunneling. More holes are generated by impact ionization of the tunnel electrons (mechanism 4) until ϕ_{sn} is reduced from ϕ_{gen} to $2\phi_f$. An average value of ϕ_{sn} is used in the model. This value is actually a fitting parameter, typically between 1 and 1.5 V.

C. Hole Flow into the Substrate

A surface channel from the inversion layer in the drain region beneath the tunnel oxide to the substrate might be turned on allowing the flow of holes into the substrate. This is the case in the cell structure shown in Fig. 8, where the path of hole flow is denoted by ③. Continuous removal of holes from the inversion layer forces a deep depletion condition, in which holes are continuously generated and accelerated toward the surface. It is experimentally shown that this effect enhances positive charge trapping in the tunnel oxide and alters the Fowler–Nordheim characteristic. The discussion of hole flow into the substrate starts with its manifestation in the C–V and the I–V curves of a tunnel capacitor, followed by its effect on positive charge trapping in the oxide, and concluded by its effect on the ERASE characteristics.

1) Investigation of a Tunnel Oxide Capacitor: The effects of deep depletion under the tunnel oxide and hole flow to the substrate have been investigated by C–V and I–V measurements on the test structure shown in Fig. 11(a). This device is a 20 μm $\times$ 180 μm thin-oxide capacitor fabricated on an ion-implanted n^+ region in a p-type substrate, which is equivalent to the tunnel-oxide capacitor in the EEPROM cell.

A typical C–V measurement with a grounded gate is shown in Fig. 11(b). Bias is applied to the n^+ and the substrate terminals, which are connected together. This C–V curve reveals the existence of a surface channel for holes under the thicker oxide at the edge of the tunnel capacitor. The various segments of the C–V curve are:

1) Depletion: The n^+ surface under the tunnel oxide is depleted of electrons.

2) Deep depletion: The voltage drop across the depletion layer exceeds $2\phi_f$ due to the shortage of holes and the relatively fast ramping (10 V/s).

3) Collapse of the deep depletion to inversion at $V = V_{tp}$ the depletion voltage under the thicker oxide in the edges of the capacitor reaches $2\phi_f$, and a surface channel

is formed for holes. The p-type substrate acts as a source, and the inversion layer under the tunnel oxide acts as a virtual drain of a p-channel transistor. The measurement corresponds to equilibrium inversion high-frequency capacitance.

4) Strong conduction in the surface channel: At about $V = V1$ the conductivity of the surface channel is high enough to short-out the depletion capacitance, so that the full thin-oxide capacitance is observed (i.e., the transit time for holes to and from the inversion layer becomes shorter than the period of the measurement ac signal). This is similar to low-frequency behavior.

The surface channel described above has an important effect on the ERASE characteristics of the EEPROM cell. To simulate the ERASE conditions, a C–V measurement is performed on the same device with the substate grounded together with the gate, and bias applied to the n^+ region. The C–V curve is shown in Fig. 11(c). It coincides with the previous result up to $V = V_{tp}$. At this point, when the surface p-channel is turned on, the substrate acts as a drain for the p-channel transistor, and the inversion layer of holes beneath the tunnel oxide acts as a virtual source. In the range denoted by (5), the capacitance is clamped to a value C_{gen}, corresponding to a potential drop ϕ_{gen} in the depletion region. The generated holes are continuously removed into the substrate through the surface channel so that equilibrium cannot be reached. Above $V = V2$ (range (6)), the potential barrier at the virtual source is reduced beyond ϕ_{gen}, and the capacitor is forced into deeper depletion with increasing bias voltage.

The interpretation of C–V curves is confirmed by I–V measurements on the same device, shown in Fig. 11(d). With a grounded substrate, a large substrate current appears at voltage above V_{tp} (as more and more holes are generated under the tunnel oxide and collected by the substrate).

By comparing the I–V curves of the gate current, one with grounded substrate and the second with substrate connected to the n^+ region, a voltage difference ΔV is observed. This voltage drop is the additional drop across the deep-depletion layer in the n^+ region when the substrate is grounded. Surface potentials ϕ_{sn} extracted from capacitances in Fig. 11(c) agree very well with $\Delta V + 2\phi_f$ as measured on Fig. 11(d).

In summary, the edge transistor can be understood analyzing the C–V/I–V characteristics of the tunnel capacitor. The main consequences of this edge transistor are: deep depletion under the tunnel oxide which reduces the tunneling current, and hole flow into the substrate.

2) Positive Charge Trapping in the Tunnel Oxide: Enhanced positive charge trapping in the tunnel oxide, correlated with the above hole current flow to the substrate, has been observed in C–V measurements on capacitor structures. Positive charge trapping in thin oxide has been previously discussed in the literature [12], [13] as a side effect of tunneling; it is exhibited experimentally by hysteresis in the I–V and C–V curves.

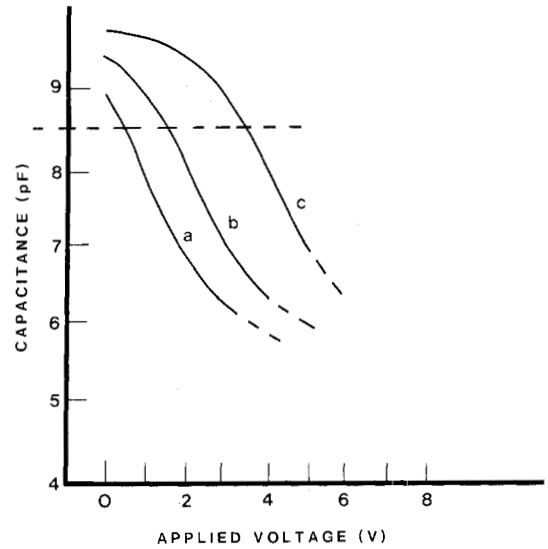


Fig. 12. C–V curves measured around flat band on a tunnel oxide capacitor: (a) Before stress, (b) After stress without hole flow to the substrate ($t = 20$ s, $J = 3.1 \times 10^{-2}$ A/cm^2, $V_{sub} = V_n^+$). (c) After stress with hole flow to the substrate ($t = 20$ s, $J = 3.1 \times 10^{-2}$ A/cm^2, $V_{sub} = V_g = 0$). Enhanced positive charge trapping is indicated by the flatband shift in curve c.

The C–V curves of a tunnel capacitor around flat band are shown in Fig. 12. Curve a is the initial C–V before any stress. The second (curve b) is the C–V after tunneling only (substrate connected to the n^+ region). There is a flat-band shift which indicates positive charge trapping. However, repeating the C–V measurement after tunneling with the substrate connected to the gate, results in a large flat-band shift due to enhanced positive charge trapping (curve c). The enhanced trapping is the result of hot-hole injection in the deep depletion layer under the tunnel oxide, which is also manifested by hole current to the substrate (see previous section). Effective interface charge densities up to 1×10^{13} cm^{-2} have been observed, in correlation with the depth of the depletion region as measured from the capacitance during the stress. The observation of positive charge trapping in oxides due to hot-hole injection is well established in the literature [7], [19]. In this work, however, the hole injection is not over the oxide potential barrier since the total voltage drop in the silicon is less than the potential barrier for holes. A possible explanation for the injection involves a two-step mechanism: energetic holes are injected toward the interface and tunnel into trapping levels in the oxide. The positive charge in the oxide is very unstable and can be annihilated by electron injection or annealed by a high-temperature cycle [20].

3) The Effect of Hole Flow on the ERASE Characteristics: The result of an experiment done on an EEPROM cell with a structure permitting hole flow into the substrate is shown in Fig. 13. The cell is written and erased repetitively, and the threshold voltage is measured after each operation. For writing the cell, fixed conditions are used ($V_g = 20$ V, $t = 10$ ms). For erasing the cell, the pulse

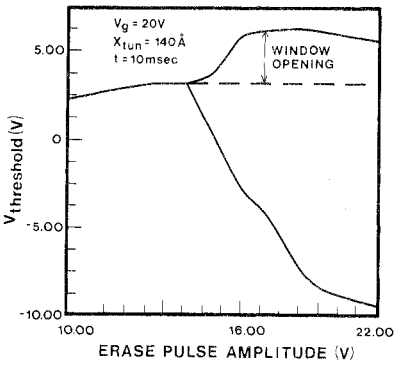


Fig. 13. Cell's threshold window measured versus ERASE pulse amplitude, with a constant WRITE pulse amplitude. The data exhibits window opening due to positive charge trapping, associated with hole flow into the substrate.

amplitude applied to the drain is gradually increased at each cycle. The first WRITE operation shifted the cell's threshold from ϕ V to about 2.5 V. The first few ERASE pulses did not cause any change in threshold, because the amplitude of the ERASE pulse was still too small, and the repetitive WRITE operations slightly moved the threshold voltage upward. With ERASE pulse amplitude above 14 V, a measurable threshold window could be observed. The important feature in this experiment is the window opening on the WRITE side. The written cell threshold was increased by about 3 V when the ERASE pulse amplitude reached approximately 16 V, instead of following the dashed line, as would be expected with fixed amplitude of the WRITE pulses. This window opening is a result of positive charge trapping in the tunnel oxide, which enhances the tunnel current by increasing the oxide electric field at the injecting electrode [8]. The window opening was found to be correlated with the appearance of a current spike in the substrate during the ERASE operation, verifying the relation between hole injection and the positive charge trapping. Note also the nonlinear shape of the V_{te} curves in Figs. 13 and 7. This is an indication of the voltage drop across the deep-depletion layer in the n^+ region, as well as the increase in tunneling field after the enhanced positive-charge trapping begins. The positive charges associated with window opening can be removed either by high-temperature anneal or by cycling the cell at low ERASE pulse amplitude.

The hole flow to the substrate in the ERASE operation is associated with an unpredictable and unstable threshold window. Furthermore, the current flow from the high drain voltage to the substrate may prevent the use of an on-chip charge pump for "5 V only" application. Therefore, it is important to avoid hole flow into the substrate in a good cell design. This is accomplished by isolating the tunnel area from the substrate by a thick gate oxide overlapping the n^+ region, so that the virtual p-channel transistor cannot be turned on when the cell is erased. In the case that a thin oxide is intentionally used over a n^+-p junction [2], the undesirable consequences of the hole flow to the substrate cannot be avoided.

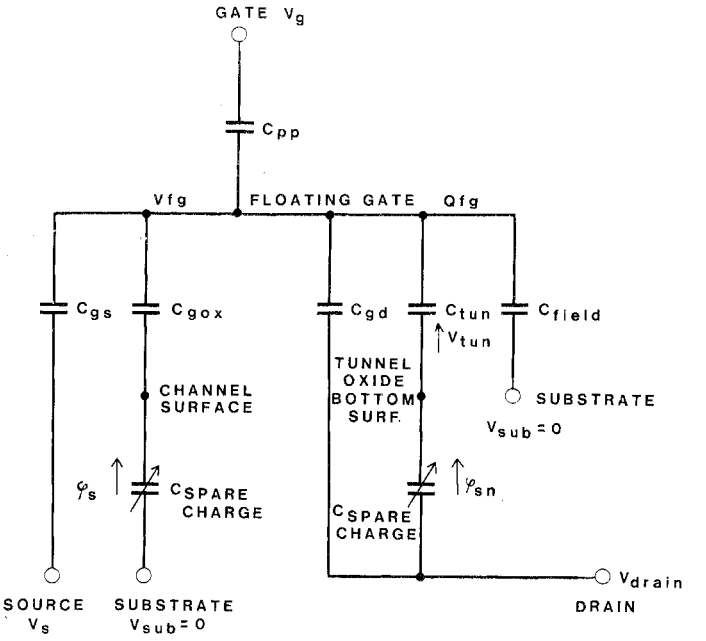


Fig. 14. Detailed capacitive equivalent circuit of the EEPROM cell, including space charge and parasitic capacitance.

V. Detailed Device Model

A. Calculation of Floating Gate and Channel Potentials

An equivalent circuit for the EEPROM cell, including parasitic capacitances and depletion-layer capacitances, is shown in Fig. 14. The effect of hole flow into the substrate is excluded from the model, assuming that the cell is appropriately designed (see Section IV-C). C_{gs} and C_{gd} are overlap gate-oxide capacitances, C_{fld} is a field-oxide capacitance from the floating gate to the substrate. The voltage drop on the depletion-layer capacitances is ϕ_s and ϕ_{sn} for the channel and the n^+ region, respectively. The stored charge on the floating gate Q_{fg} is the sum of all capacitor charges

$$Q_{fg} = C_{pp}(V_{fg} - V_g) + C_{gd}(V_{fg} - V_d) + C_{\text{fld}}(V_{fg} - V_{\text{sub}}) + C_{\text{tun}}(V_{fg} - (V_D - |\phi_{sn}|)) + C_{gs}(V_{fg} - V_s) + C_{\text{gox}}(V_{fg} - (V_{\text{sub}} + |\phi_s|)). \quad (17)$$

During the WRITE operation, the n^+ region is accumulated and ϕ_{sn} is assumed to be zero. The channel is formed so that the channel surface and the floating source assume the voltage of the drain $V_d = 0$. Thus, V_{fg} can be solved explicitly from (17).

During the ERASE operation, ϕ_{sn} is assumed to be constant, as discussed in Section IV-B above. The condition of the channel surface is determined in the following manner for any given Q_{fg}: First, depletion is assumed, and the last term in (17) is replaced by

$$Q_{\text{dep}} = A_{ch} \cdot \sqrt{2q \cdot \epsilon_{si} \cdot \epsilon_0 \cdot N_b \cdot \phi_s}. \quad (18)$$

For this assumed condition, V_{fg} is related to ϕ_s by

$$V_{fg} = V_{fb} + \phi_s + \frac{A_{ch}}{C_{\text{gox}}} \sqrt{2q \cdot \epsilon_{si} \cdot \epsilon_0 \cdot \phi_s}. \quad (19)$$

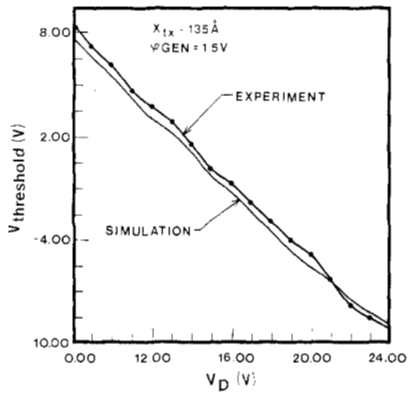


Fig. 15. A comparison between measured and simulated ERASE characteristics for a cell without hole flow path to the substrate using the detailed model.

This expression is inserted into (17), and the resultant quadratic equation is solved for $\sqrt{\phi_s}$. If there is no positive solution, the channel surface is accumulated and ϕ_s is taken as zero. The voltage V_s at the floating source is equal to ϕ_s. Equation (17) is then solved for V_{fg} with the appropriate value of ϕ_s.

B. Calculation of WRITE and ERASE Characteristics

Once all the internal voltages are determined, the tunnel current density can be calculated from (1). Starting from an initial stored charge $Q_{fg}(0)$, the differential equation

$$dt = \frac{dQ_{fg}}{A_{tun} \cdot J_{tun}} \tag{20}$$

is integrated numerically, to obtain the floating-gate charge as a function of time. Each step in the integration involves the calculation of voltages and tunnel current as outlined above.

The cell's threshold voltage is related to Q_{fg} by

$$V_t = V_{ti} - \frac{Q_{fg}}{C_{pp}} \tag{21}$$

where V_{ti} is the neutral cell threshold. V_{fb} is adjusted in (17) to yield $Q_{fg} = 0$ when $V_g = V_{ti}$ accounts for any fixed surface charges in the double-poly structure.

Using (21) and the solution of (20), the dependence of threshold on programming time is simulated for any set of device parameters and programming waveforms. An example is shown in Fig. 15 for ERASE characteristics of a cell without hole flow into the substrate, exhibiting a reasonable fit to measured data.

VI. Conclusions

A simplified model for FLOTOX EEPROM cells based on the concept of coupling ratios has been presented. The principal considerations in cell design can be derived from the simplified model. However, this model cannot be used for accurate simulation of the ERASE operation. Physical effects which complicate the ERASE operation have been presented and analyzed. Depletion in the transistor's channel reduces the coupling ratio during erasure. Deep depletion in the n^+ region under the tunnel oxide causes a voltage drop that leads to a further reduction in the threshold shift. The role of these depletion layers has been included in a detailed cell model.

A flow path for holes from under the tunnel oxide into the substrate during erasure leads to further detrimental consequences. On-chip generation of the drain voltage by charge pumps may not be possible due to the large substrate current. Enhanced positive charge trapping in the oxide, associated with the hole flow, has been observed. The role of hot-hole injection in this process has been discussed. Unstable opening of the cell's threshold window has been shown to occur as a result of the enhanced positive charge trapping. It is concluded that hole flow into the substrate can be avoided by an appropriate cell design.

Acknowledgment

The authors are indebted to J. Lee, N. Mielke, S. Lai, G. Gongwer, and E. Hellman for their contribution to this work.

References

[1] W. Johnson, G. Perlegos, A. Renninger, G. Kuhn, and T. Ranganath, "A 16K bit electrically erasable non-volatile memory," in *Dig. Tech. Papers IEEE Int. Solid-State Circuits Conf.*, p. 152, Feb. 1980.

[2] C. Kuo, J. R. Yeargain, W. J. Downey, K. A. Ilgenstein, J. R. Jorvig, S. L. Smith, and A. R. Bormann, "An 80-nS 32-K EEPROM using the FETMOS cell," *IEEE J. Solid-State Circuits*, vol. SC-17, p. 821, 1982.

[3] G. Yaron, S. J. Prasad, M. S. Ebel, and B. M. K. Leong, "A 16K EEPROM employing new array architecture and designed-in reliability features," *IEEE J. Solid-State Circuits*, vol. SC-17, p. 833, 1982.

[4] A. Gupta, T. L. Chiu, M. S. Chang, A. Renninger, and G. Perlegos, "A 5V-only 16K EEPROM utilizing oxynitride dielectrics and EPROM redundancy," in *Dig. Tech. Papers IEEE Int. Solid-State Circuits Conf.*, p. 184, Feb. 1982.

[5] T. Hogiwara, Y. Yatsuda, R. Kendo, S. I. Minami, T. Aoto, and Y. Itoh, "A 16 Kbit electrically erasable PROM using n-channel Si-gate MNOS technology," *IEEE J. Solid-State Circuits*, vol. SC-15, p. 346, 1980.

[6] M. Lenzlinger and E. H. Snow, "Fowler–Nordheim tunneling into thermally grown SiO," *J. Appl. Phys.*, vol. 40, p. 278, 1969.

[7] Z. A. Weinberg, "On tunneling in metal-oxide-silicon structures," *J. Appl. Phys.*, vol. 53, p. 5052, 1982.

[8] B. Euzent, N. Boruta, J. Lee, and C. Jenq, "Reliability aspects of a floating gate EEPROM," in *Proc. Int. Reliability Physics Symp.*, 1981.

[9] P. I. Suciu, B. P. Cox, D. D. Rinerson, and S. F. Cagnina, "Cell model for EEPROM floating-gate memories," in *IEDM Tech. Dig.* (San Francisco), p. 737, 1982.

[10] S. T. Wang, "Charge retention of floating-gate transistors under applied bias conditions," *IEEE Trans. Electron Devices*, vol. ED-27, p. 297, 1980.

[11] L. D. Yau, "A simple theory to predict the threshold voltage of short channel IGFET's," *Solid-State Electron.*, vol. 17, pp. 1059–1063, 1974.

[12] Y. Nissan-Cohen, D. Frohman-Bentchkowsky, and J. Shappir, "Characterization of simultaneous bulk and interface high-field trapping effects in SiO_2," in *IEDM Tech. Dig.* (Washington, DC), paper 8.2, 1983.

[13] M. Itsumi, "Positive and negative charging of thermally grown SiO_2 induced by Fowler–Nordheim emission," *J. Appl. Phys.*, vol. 52, pp. 3491–3497, 1981.

[14] E. H. Nicollian and J. R. Brewz, *MOS Physics and Technology*. New York: Wiley, 1982, ch. 9.

[15] E. Suzuki and Y. Hayashi, "Transport processes of electrons in MNOS structures," *Appl. Phys.*, vol. 50, pp. 7001–7006, 1979.

[16] M. S. Liang, C. Chang, Y. T. Yeow, C. Hu, and R. W. Brodersen,

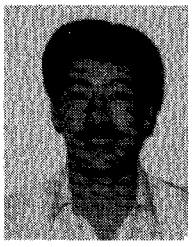
"Creation and termination of substrate deep depletion in thin oxide MOS capacitors by charge tunneling," *IEEE Electron Device Lett.*, vol. EDL-4, pp. 350–352, 1983.
[17] E. O. Kane, "Electron scattering by pair production in silicon," *Phys Rev.*, vol. 159, pp. 624–631, 1967.
[18] B. Eitan and A. Kolodny, "Two components of tunneling current in metal-oxide-semiconductor structures," *Appl. Phys. Lett.*, vol. 43, pp. 106–108, 1983.
[19] J. M. Aitken and D. R. Young, "Avalanche injection of holes into SiO_2," *IEEE Trans. Nucl. Sci.*, vol. NS-24, pp. 2128–2134, 1977.
[20] C. Jenq *et al.*, "High-field generation of electron traps and charge trapping in ultra-thin SiO_2," in *IEDM Tech. Dig.*, pp. 388–391, 1981.
[21] Y. Tarui, Y. Hayashi, and K. Nagai, "Electrically programmable nonvolatile semiconductor memory," in *Proc. 5th Conf. Solid State Devices*, supplement to *J. Japan Soc. Appl. Phy.*, vol. 43, 1974.
[22] Y. N. Hsieh, R. A. Wood, and P. O. Wang, "Electrically alterable programmable logic array," in *IEDM Tech. Dig.*, pp. 598–601, 1980.
[23] H. Schaver *et al.*, "A high-density, high performance EEPROM cell," *IEEE Trans. Electron Devices*, vol. ED-29, p. 1178, 1982.

*

Avinoam Kolodny received the B.Sc., M.Sc., and D.Sc. degrees in electrical engineering from the Technion-Israel Institute of Technology.

From 1980 to 1981, he was a Lecturer in the Electrical Engineering Department at the Technion. He was engaged in research of narrow-bandgap semiconductor devices and infrared detectors. In 1981, he joined Intel Corporation in Santa Clara, CA, where he worked on research and development of nonvolatile memory devices in silicon. Since 1984, he is working on the development of computer-aided design software for VLSI, at the Intel Israel Design Center in Haifa.

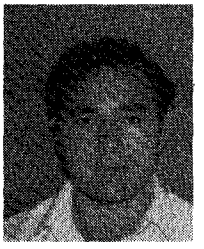
*



Sidney T. K. Nieh was born in Hong Kong in 1942. He received the B.S.E.E. degree from Southern Methodist University, Dallas, TX, in 1966, and the M.S.E.E. and Ph.D. (E.E.) degrees from Stanford University, Stanford, CA, specializing in quantum electronics, in 1968 and 1971, respectively.

He has worked for Hewlett-Packard, Addington Laboratories, the University of Southern California, and Technion from 1966 through 1978 on microwave semiconductor devices and infrared detectors. In 1978, he joined Intel Corporation, Santa Clara, CA, first working in process reliability and subsequently in nonvolatile memory technology development. He is currently engaged in SRAM and logic process development in Portland, OR.

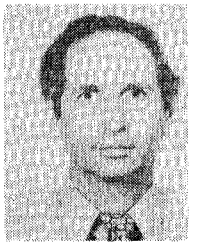
*



Boaz Eitan (M'82) was born in Israel on October 18, 1948. He received the B.Sc. degree in mathematics and physics, the M.Sc. degree in applied physics, and the Ph.D. degree, all from the School of Applied Science and Technology, Hebrew University of Jerusalem, Jerusalem, Israel, in 1976, 1978, and 1981, respectively. His doctoral work concerned short-channel MOS devices.

In 1981, he joined Intel Corporation where he was involved in technology development. He joined WaferScale Integration, Inc., Fremont, CA, in 1983, where he is currently involved in CMOS and nonvolatile memory development for high-speed and ASIC applications.

*



Joseph Shappir (M'75) was born in Hadera, Israel, on May 29, 1938. He received the M.Sc. and Ph.D. degrees in physics from the Hebrew University of Jerusalem, Jerusalem, Israel, in 1961 and 1969, respectively.

He served in the Israeli Defense Forces from 1961 to 1964 as a technical officer. In 1961, he joined the N.V. Philips Gloielampenfabrieken, Nijmegean, The Netherlands, working on technology development of CMOS integrated circuits. In 1972, he joined the Electrical Engineering Department, Techion-Israel Institute of Technology, Haifa, Israel, as Senior Lecturer. He worked on semiconductor devices for logic and imaging applications. On his sabbatical leave in 1977, he came to Intel Corporation, Santa Clara, CA, where he worked on technology development for RAM and nonvolatile memories. In 1979, he joined the School and Applied Science and Technology, Hebrew University of Jerusalem, as Associate Professor of Applied Physics. He continues his connections with Intel Corporation, Santa Clara, as consultant for advanced technology and device development in RAM and nonvolatile memories. His main interests are in the area of semiconductor device physics and technology for logic, memory, and imaging applications.

Dr. Shappir has served as Chairman of the Israeli IEEE Chapter on Electron Devices and Hybrids and Packaging since 1981.