Field and high-temperature dependence of the long term charge loss in erasable programmable read only memories: Measurements and modeling

M. Herrmann[a)]
Swiss Federal Institute of Technology, Reliability Laboratory, Gloriastrasse 35, CH-8092 Zürich, Switzerland

A. Schenk
Swiss Federal Institute of Technology, Integrated Systems Laboratory, Gloriastrasse 35, CH-8092 Zürich, Switzerland

(Received 14 July 1994; accepted for publication 6 January 1995)

High-temperature charge transport across an oxide-nitride-oxide sandwich of erasable programmable read only memories is mainly governed by the oxide conductivity as experimentally determined. It was verified in the examined devices that charge loss is not due to mobile ions. Since hole injection from the control gate into the nitride can be blocked by a 70-Å-thick top oxide we conjecture that charge loss is due to leakage of electrons; however, the observed leakage current is too large to be explained by pure electrode-limited charge transport (Richardson emission and direct tunneling). It was also verified that field gain on asperities and along edges cannot increase the charge loss current to the required range. Numerical evaluation of trap tunneling and resonant tunneling indicated that both mechanisms are weakly temperature dependent while charge loss has a typical activation energy of 1.2 eV in the range of 250–350 °C. Consequently, a multiphonon-assisted tunneling mechanism is proposed where electrons stored on the floating gate tunnel to oxide traps, then are emitted into the nitride. The coupling of the trap level to oxide phonons results in virtual energy levels in the oxide which allow for more effective transition paths. As a consequence of the electron-phonon coupling, the emission occurs close to the oxide conduction-band edge at temperatures between 250 and 350 °C, producing a strong temperature dependence for the mechanism.

I. INTRODUCTION

An erasable programmable read only memory (EPROM) cell consists of one floating gate transistor. During programming hot electrons are injected from the channel into the floating gate. They can be removed by UV light. Any change of the charge distribution between substrate and control gate leads to a change of the threshold voltage. Under normal operating conditions, a data retention time of 10 years is usually specified by the manufacturer and expected by the user. "Long term charge loss" labels the charge loss after steady state has been reached until the moment where the cell fails. This period of permanently decreasing threshold voltage determines the reliability of the memory device, besides the occurrence of conventional failures in integrated circuits. The floating gate is surrounded by the following dielectrics: the gate oxide between the floating gate and the substrate, the interpoly dielectric between floating gate and control gate, and the field oxide or the chemical-vapor-deposition (CVD) oxide between floating gates and adjacent cells. The interpoly dielectric is either an oxide film, or in recent technologies, an oxide-nitride-oxide (ONO) sandwich.

The floating gate charge can be determined either by a search algorithm for the threshold voltage of the floating gate transistor or by measuring the drain-source current at specified control gate-to-source and drain-to-source voltages. We used the second method which is a very accurate and quick way to measure many transistors. The actual threshold voltage can be easily determined from the measured drain-source current. High temperatures or high electric fields must be used to accelerate the charge loss. In our work, high temperatures have been chosen to measure the data retention characteristic. The electric field across the dielectrics is determined by the voltage applied to the control gate and by the floating gate charge, which is a function of stress time. During the bake no voltage was applied to the control gate; thus, the electric field across the dielectrics was the same as during normal standby operation.

The current density across the dielectric surrounding the floating gate is in the range of 10^{-14} to 10^{-11} A cm^{-2} for fields from 0.7 to 2 MV cm^{-1} at temperatures between 250 and 350 °C. This current is by far too large to be explained by pure electrode-limited charge transport, as shown in Fig. 1. There, the theoretical electrode-limited current is compared against the current density from the charge loss experiment at 300 °C and from high-field I-V measurements at 25 °C. Fowler–Nordheim tunneling[1,2] for high electric fields and Richardson emission[3] for low fields and high temperatures have been assumed. We have also verified numerically and with the help of conform transformation that field gain on asperities and along edges cannot increase the charge loss current to the mentioned range.

The following data loss mechanisms have been previously identified.

[a)]Present address: Philips Semiconductors, Binzstrasse 44, CH-8045 Zürich, Switzerland; Electronic mail: schenk@iis.ee.ethz.ch

 0021-8979/95/77(9)/4522/19/$6.00

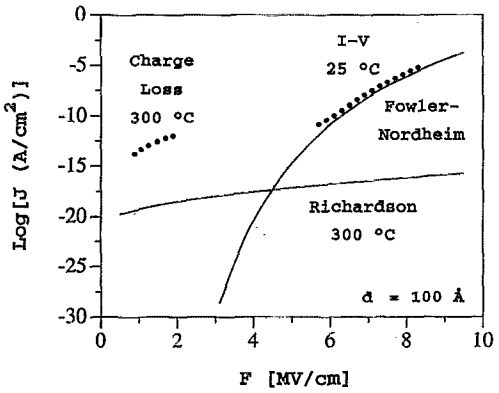


FIG. 1. Current density vs electric field: electrode-limited current (solid lines), charge loss, and *I-V* measurements (dots).

(1) Defects in the interpoly dielectric or gate oxide of EPROM cells leading to electronic leakage which result in charge loss or charge gain on the floating gate: The affected cells are generally randomly located in the memory array.[4,5]

(2) Ionic contamination can compensate the stored charge in EPROM cells. Positively charged ions move in the electric field that is generated by the negative charge on the floating gate. If the cells are erased, the ions can move due to their concentration gradient and cause charge gain. The failing cells form regions which generally increase in magnitude as function of the bake time.[4,6,7]

(3) During UV erasing of EPROM cells some of the electrons that were stored on the floating gate can be trapped in the interpoly dielectric or in the gate oxide. After the reprogramming of the cells and a second bake the release of electrons, previously trapped during UV erasing, results in a threshold voltage decrease just like charge loss due to electrons stored on the floating gate. The charge loss would stop when the traps were emptied. All memory cells are affected by this mechanism.[4]

(4) For stacked-gate avalanche-injection metal-oxide-semiconductor (SAMOS) structures and selective polysilicon oxidation (SEPOX) compatible EPROM (SEPROM) cells the thermionic electron emission model has been proposed to explain the temperature dependence of the charge loss.[8,9] The measured activation energy was interpreted as the energy barrier between the conduction-band edges of Si and SiO_2. This parameter had a value much lower than the energy barrier which results from optical or Fowler–Nordheim experiments. In optical experiments the irradiation energy at which the electrons surmount the barrier between the conduction-band edges in Si and SiO_2 is detected. The Fowler–Nordheim experiment uses this energy barrier as a fit parameter that is very sensitive to the slope of the *I-V* characteristic of MOS capacitors. By both types of experiments an energy barrier of about 3.1 eV was estimated.[10]

(5) The thermionic electron emission model has been modified by introducing a field-dependent barrier lowering due to the image force effect and a temperature-dependent charge accumulation around the floating gate.[11] At low temperatures more charges are assumed to accumulate as compared to high temperatures. This reduces the electric field in the region of electron injection to the oxide and, thus, prevents a charge loss from the floating gate. The model yields a good fit over a broad temperature range but still uses fit parameters that do not agree with other measurement methods.

(6) For EPROM cells with ONO interpoly dielectrics three distinct phases exist in the charge loss characteristic during bake time. An initial fast threshold voltage shift of less than 10 min at 250 °C can be explained by a rearrangement of carriers in the nitride or by a nitride polarization effect. The threshold voltage shift of the second phase is caused by the movement of trapped electrons which are mostly injected into the nitride during memory programming. This phase saturates after about 100 h at 250 °C. The third phase is the nonsaturating long-term threshold voltage shift. It is determined by electrons leaking through the bottom and top oxide. The first and second phases are a strong function of nitride thickness, while the third phase is dependent on the oxide thicknesses of the ONO.[12,13]

(7) For EPROMs with ONO interpoly dielectric it was found that trapped electrons at the nitride-oxide interface can directly tunnel through a thin (30 Å) top oxide.[14]

(8) In a subsequent article[15] the authors of Ref. 14 modeled the long-term charge loss above 300 °C with the Poole–Frenkel mechanism. From the temperature slope a barrier height for electrons trapped on the nitride-top oxide interface was calculated. The field dependence was assumed to be due to barrier lowering. The bottom oxide of the ONO structure was very thin and, therefore, not charge loss limiting.

(9) An analytical charge loss model for semiconductor-oxide-nitride-oxide-semiconductor (SONOS) devices was presented in Refs. 16 and 17. The model takes into account the following mechanisms: (1) electron back-tunneling from the nitride traps to the Si conduction band; (2) electron back-tunneling from the nitride traps to the Si/SiO_2 interface traps, and (3) hole injection from the Si valence band to the nitride traps. An amphoteric trap charge distribution was used in this model. Mechanism (1) determines the initial charge loss phase, while mechanisms (2) and (3) determine the long term charge loss phase.

(10) For electrically erasable programmable read only memories (E^2PROMs) with a thin tunnel oxide, high-field experiments at room temperature indicate that Fowler–Nordheim injection is responsible for the charge loss. At low fields, i.e., normal operating conditions, the mechanism must be replaced by direct tunneling.[18] Because of its strong dependence on insulator thickness, this mechanism cannot be applied to the thicker interpoly dielectric or gate oxide in EPROMs.

Many results from experiments performed on test capacitors have been published to explain charge transport in oxide (SiO_2), nitride (Si_3N_4), NO (Si_3N_4/SiO_2), and ONO ($SiO_2/Si_3N_4/SiO_2$), as follows.

(1) In oxide films, Fowler–Nordheim injection is dominant at high field strengths (higher than 5 MV/cm for a dielectric thickness of about 100 Å). This mechanism has a weak temperature dependence.[19] For high temperatures and low fields, Richardson injection becomes dominant. This mechanism is strongly temperature, but only weakly field,

dependent.[3] Direct tunneling becomes responsible for charge transport in thin oxides at low fields along with a weak temperature dependence.[20] Here the charge transport is due to electron conduction and is limited by injection from the electrode (electrode-limited conduction), if the influence of traps can be neglected.

(2) In nitride films the hole conduction is larger than the electron conduction.[21–24] The charge transport is limited by the bulk properties of the dielectric (bulk-limited conduction). According to Ref. 25, the current density is the sum of three contributions: At high fields and high temperatures the Poole–Frenkel effect dominates; at high fields and low temperatures the current is due to field ionization of trapped charge; and at low fields and moderate temperatures hopping from one trap to the other is dominant. The current density in silicon-nitride films is larger than in oxides at the same field.

(3) In NO stacked films, charge transport is due to the above-described mechanism for each film, respectively. Since the current-field characteristics are different for the two materials, the fields will adjust themselves by charge accumulation in the nitride close to the interface until current continuity is established.[26–32] The polarity of the trapped charge can be identified by measuring the flatband voltage shift.[33,34]

(4) In ONO stacked films, an anode-side oxide (positive electrode voltage with respect to the cathode-side) thicker than 30 Å can block hole injection into the nitride. The oxide on the cathode side (negative electrode voltage with respect to the anode side) determines the electron injection. Holes injected into the nitride are accumulated on the oxide-nitride interface on the cathode side, while electrons accumulate on the nitride-oxide interface on the anode side. Some of the charge can recombine in the nitride. The nitride thickness influences the amount of trapped charge.[35–40]

(5) The bulk-limited currents[20,41]—hopping conduction and space-charge-limited current—do not have an exponential field dependence. As is shown below, the charge loss data of our devices exhibit a field dependence that is similar to that of Poole–Frenkel conduction. Even the slope is quite similar, provided the oxide permittivity is used for the calculation. This corresponds to the results of Ref. 15. However, according to Ref. 20 the Poole–Frenkel mechanism is different in thin and thick dielectrics, respectively. In the case of thick dielectrics the activation energy has to be identified with the trap depth, while for thin dielectrics, which is the present case, it should be the height of the barrier between Si and SiO_2 lowered by the Poole–Frenkel effect. Thus, the activation energy should be about 3 eV which contradicts our measured activation energy of about 1.2 eV.

(6) Tunneling via traps that are located inside potential barriers had been recognized as a source of "leakage current" of *pn* junctions in the late 1950s already (so-called "excess current"), later on as a possible mechanism in metal-semiconductor contact barriers (Parker and Mead[42]), in metal-insulator-metal junctions (Gadzuk[43]), and in metal-nitride-oxide-silicon (MNOS) structures (Dorda and Pulver[44]). For thin insulators the probability is negligible to find more than one defect along the tunnel path. In a theoretical article, Svensson and Lundström[29] applied the trap-assisted tunneling mechanism to a MNOS structure. In the first step electrons tunnel by "modified Fowler–Nordheim injection" from the conduction band of the electrode into nitride traps across a very thin oxide layer. The second step, where the electrons are emitted to the nitride conduction band, was neglected because of the larger tunnel probability compared to that of the first step. Thus, the temperature dependence originated from the occupation probability in the conduction band of the electrode. Suzuki, Schroder, and Hayashi[45] considered high electric fields, and Fleischer, Lai, and Cheng[46] presented a closed formula for the two-step tunneling mechanism through a single insulating layer for intermediate electric fields neglecting the temperature dependence at all. Yasuda, Patel, and Toriumi[47] assumed an asymmetrical spatial trap distribution for that mechanism in order to explain an observed voltage-polarity dependence of stress-induced leakage currents across MOS capacitors.

(7) Other authors involved Poole–Frenkel conduction in combination with direct tunneling in a nonequilibrium condition in MNOS structures[48] or field emission from the cathode in combination with space-charge buildup in the body of a single-layer dielectric.[49] Reference 50 presents a model where electrons that have tunneled by the Fowler–Nordheim mechanism across the potential barrier, gain energy from the electric field in the oxide and lose energy by various scattering mechanisms. The hot electrons arriving at the anode lose their energy by emitting surface plasmons. The emitted surface plasmons decay via the excitation of electron-hole pairs and by the generation of both hot holes and hot electrons. The hot holes may be emitted over or tunnel through the potential barrier.

In this article we present a model wherein electrons stored on the floating gate are captured by traps in the bottom oxide and subsequently emitted into the nitride by a multiphonon-assisted tunneling process. After describing the experiments in Sec. II, a current-voltage characteristic of the charge loss leakage current is calculated from the measured drain-source current in Sec. III. Section IV presents the theory of multiphonon-assisted tunneling via oxide traps. The role of resonant tunneling as an alternative transport mechanism is studied in Sec. V and in the Appendix. In Sec. VI theoretical results are discussed in detail and compared with the measurements. Finally, conclusions are given in Sec. VII.

This article is the full version of work presented in condensed form at the 1994 IEEE International Reliability Physics Symposium and at the 1994 International Conference on Solid State Devices and Materials.

II. EXPERIMENTS

The charge loss measurements have been performed on a commercial 4 Mbit EPROM. Due to test modes, it was possible to control the control gate-to-source and drain-to-source voltages of each floating gate transistor through the column and row decoder. The device was fabricated in 0.8 μm technology. The substrate was *p* doped while the floating gate and control gate were *n*-doped polysilicon. We measured two lots: one with oxide and the other with ONO as interpoly dielectric. Recent technologies use ONO dielectrics which

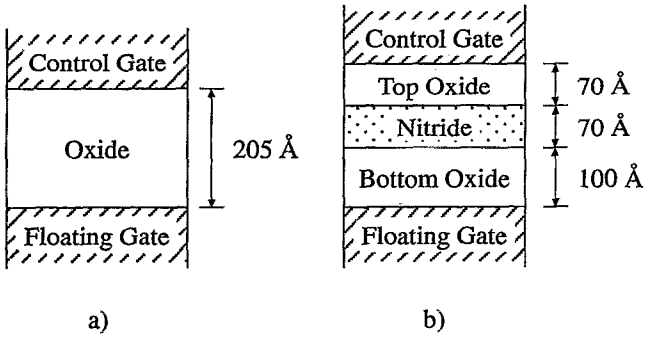


FIG. 2. Structure of the interpoly dielectric: (a) oxide, (b) ONO device. Long-term data loss can be due to defects in the isolation.

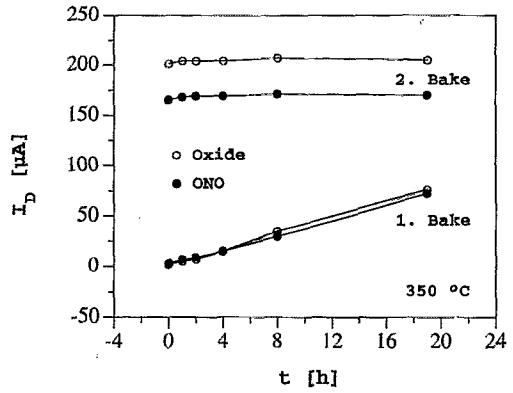


FIG. 3. Drain-source current vs bake time at 350 °C for oxide and ONO devices. The program-bake-erase-bake cycle proves that charge loss is not due to mobile ions.

have the following advantages: larger permittivity of nitride compared to that of SiO_2, smaller defect density of the stacked dielectric compared to that of a single oxide layer, and high breakdown voltage. The oxidation was performed at 1000 °C. The nitride was grown at 800 °C. The first poly was doped by implant. The nominal oxide thicknesses were 205 Å for the single oxide dielectric and 70, 70, and 100 Å from top to bottom for the ONO sandwich, respectively (see Fig. 2). The single oxide dielectric has the equivalent oxide thickness of the ONO dielectric, given by: $d_{G,o} = d_{b,o} + (\epsilon_o/\epsilon_n)d_n + d_{t,o}$, where $d_{G,o}$ is the equivalent oxide thickness, $d_{b,o}$, d_n, and $d_{t,o}$ are the thicknesses of bottom oxide, nitride and top oxide, respectively, ϵ_o and ϵ_n are the permittivities of oxide and nitride. This implies that the oxide field was the same for both lots at a given voltage drop between control gate and floating gate. The ratio ϵ_o/ϵ_n is about 1/2. The floating gate charge was determined by measuring the drain-source current at 8 V control gate-to-source voltage and 1 V drain-to-source voltage.

Long term data loss can be due to macroscopic defects in the isolation, mobile ions, or electronic leakage, respectively. Defects in the isolation[4,5] result in charge loss or charge gain on the floating gate. The affected cells are generally randomly located in the memory array. Defective devices can be screened out by high-gate voltage stress or high-temperature bake. Mobile ions[4,6,7] can compensate the stored charge. This is observed by a program-bake-erase-bake cycle. The ions are driven to the floating gate by the electric field from the stored negative charge during the first bake. Electric field and concentration gradient act concurrently upon the ions. UV erasing of the cell to the normal erased threshold voltage after the first bake leaves an excess of electrons which balance the attracted ionic charge, leaving the floating gate neutral. With the omission of the electric field, the concentration gradient causes the ions to diffuse out of the cell during the second bake. This appears as charge gain because the electrons have actually not been removed from the floating gate during the erase operation. Increasing temperature accelerates the process due to emission of mobile ions from traps. It has been reported that Na^+ ions can act as mobile ions leading to charge loss.[7] Figure 3 shows this program-bake-erase-bake cycle for both lots. The increasing drain-source current during the first bake indicates a decreasing threshold voltage. Since no decrease of the drain-source current occurs during the second bake, charge loss due to mobile ions can be excluded for these devices. It has been reported that for EPROMs with ONO interpoly dielectric a top oxide thicker than 30 Å can block hole injection from the control gate into the nitride.[37,38] Thus, we believe that leakage of electrons is responsible for the nondefective charge loss in the examined devices.

Figure 4 shows no significant difference in the charge loss characteristic between the devices with single oxide and ONO interpoly dielectric. The oxide field of both lots is the same for a given drain-source current, i.e., for a given floating gate-to-source voltage. This experiment leads to the conclusion that charge transport is limited by the oxide conductivity. In Refs. 13 and 12 charge loss characteristics of EPROM split lots with different nitride thicknesses of the ONO interpoly dielectric show two initial charge loss phases which saturate after about 100 h at 250 °C, and a long term charge loss phase. The initial charge loss phases are strong functions of nitride thickness, whereas the nitride has no influence on the long term charge loss phase. Similar activation energies[5,8,14,51] of both EPROM types from other manufacturers confirm this result. Values ranging from 1.0 to 1.9 eV have been measured for temperatures in the range of 200 to

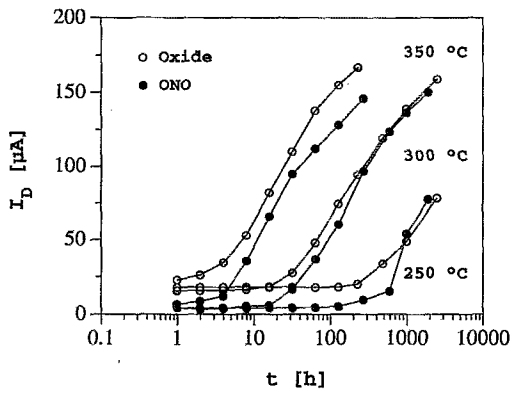


FIG. 4. Drain-source current of oxide and ONO devices vs time with bake temperature as parameter. The small difference between the two lots indicates that charge loss is limited by the oxide conductivity.

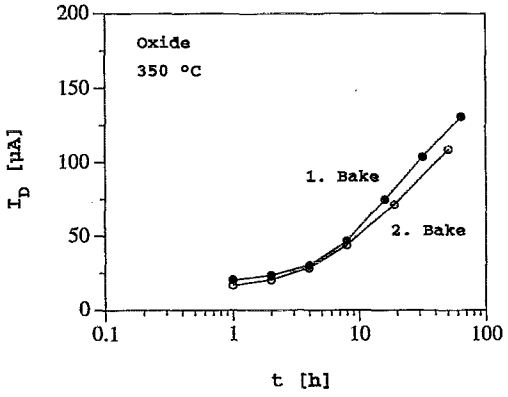


FIG. 5. Drain-source current of oxide devices vs time at 350 °C. The second bake proves that no charge accumulation in the isolation occurred during the first bake.

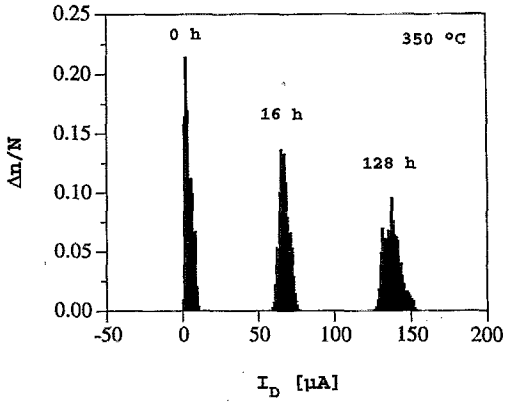


FIG. 6. Histogram of the drain-source current of an ONO device measured on $N=2000$ cells at 350 °C. The three intermediate measurements show an increasing deviation with increasing bake time.

425 °C. The activation energy in pure nitride capacitors[52] was reported to be less than in capacitors with oxide.

Reprogramming an already baked device and performing a second bake shows no difference of the drain-source currents compared to the first bake up to 10 h at 350 °C (see Fig. 5). After 10 h the charge loss during the second bake becomes slightly smaller compared to the first bake. A possible explanation for both observations is that there is no charge accumulation in the interpoly dielectric during the first bake; however, charge accumulation during the first bake might be removed due to the high electric field across the interpoly dielectric during reprogramming. Charge accumulation would cause a different initial condition compared to the first bake. On the other hand, charge loss through the interpoly dielectric during the first bake may create traps along the charge loss path that are filled during the second bake. The trapped negative charge would reduce the electric field on the floating gate side which would lead to a smaller charge loss during the second bake. For a reliable explanation, charge accumulation in equivalent test capacitors would have to be examined by measuring the flatband voltage shift during high-temperature storage with small electric field in the oxide or ONO dielectric.

In Refs. 13, 37, 39, and 40 extensive studies on the influence of the individual layer thicknesses in the ONO interpoly dielectric on the charge loss were published. The charge loss was measured during bakes at 250 and 300 °C. Reducing the top-oxide thickness of an ONO with a thick bottom oxide (150 Å) has no influence on the charge retention capability as long as the top oxide is thicker than 30 Å. For top oxides thinner than 30 Å the charge loss increases with decreasing top oxide thickness. To examine the influence of the nitride thickness two device groups were measured, one with top oxide thinner than 30 Å and one with thicker top oxide. When the top oxide is thin, a thicker nitride leads to enhanced charge loss if the equivalent oxide thickness is kept constant, i.e., the bottom oxide thickness has to be decreased with increasing nitride thickness. In the case where the top oxide is thicker than 30 Å the nitride thickness dependence on long-term charge loss was found to be small. The initial charge loss still increases with increasing nitride thickness as in the case of a thin top oxide. Scaling down the bottom oxide to about 100 Å does not lead to a degradation of the data retention capability. Missing bottom oxide (ON interpoly dielectric) leads to a large charge loss; however, if the top oxide thickness on the ON interpoly dielectric is increased, the charge loss decreases. All these results can be explained as follows: A top oxide thicker than 30 Å can block hole injection from the control gate into the nitride while the bottom oxide thickness has influence on electron injection from the floating gate into the nitride. The nitride conductivity is larger than the oxide conductivity. In the case of a thin top oxide, holes injected from the control gate can recombine with electrons leaking through the bottom oxide. Since the tunnel probability of electrons is larger compared to that of holes, the bottom oxide must be thicker than the top oxide. If the bottom oxide is missing (ON interpoly dielectric), charge loss must be prevented by a thicker top oxide.

III. CHARGE LOSS CHARACTERISTIC

After programming the cells of 15 devices with oxide and 15 devices with ONO as interpoly dielectric, a bake at 250, 300, and 350 °C was performed. On each device the drain-source current of 2000 cells was measured at 8 V control gate-to-source voltage and 1 V drain-to-source voltage. The following diagrams refer only to ONO devices. Figure 6 presents the histogram of the drain-source current measurements of one device. The three intermediate measurements show an increasing deviation with increasing bake time. The initial state of all measurements was about the same and no change occurred until the first intermediate measurement. The dots in Fig. 7 show the average of the drain-source current of all 10 000 cells belonging to one temperature. The error bars mark the minimum and the maximum. An activation energy of 1.2 eV has been estimated for this temperature range. The solid line fits are employed in the further calculations.

To calculate the floating gate charge from the measured drain-source current a model must be established for the cell. The floating gate transistor can be regarded as a customary

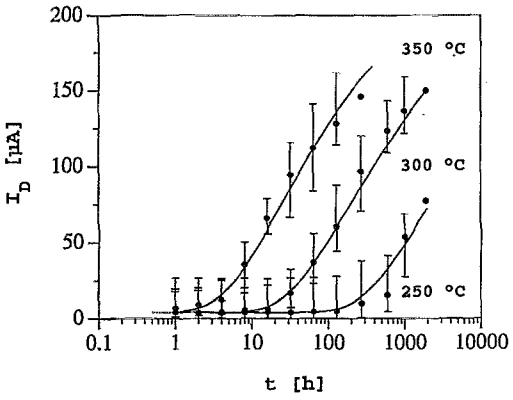


FIG. 7. Drain-source current of ONO devices vs time with bake temperature as parameter. Measured average (dots), minimum/maximum (error bars), and fit (solid lines) employed in further calculations.

field-effect transistor with the floating gate voltage on its gate.[53,54] The drain-source current can be calculated with Eqs. (1) and (2) valid for small drain-to-source voltages,[41]

$$I_D=\frac{\beta}{2b}(U_F-U_t)^2 \quad \text{for } U_t<U_F\leqslant U_t+bU_D, \tag{1}$$

$$I_D=\beta\left((U_F-U_t)U_D-\frac{b}{2}U_D^2\right) \quad \text{for } U_t+bU_D<U_F. \tag{2}$$

Here I_D is the drain-source current, U_F the floating gate-to-source voltage, U_D the drain-to-source voltage, and U_t the threshold voltage. U_t, β, and b are functions of electron mobility and doping concentration in the substrate, channel width and channel length, bulk potentials in substrate and floating gate, permittivity of oxide, and oxide thickness.

Figure 8(a) aids in deriving the floating gate charge Q_F based on the measurement condition. The control gate and floating gate are highly doped and can thus be treated like a metal. By using Gauss's law the floating gate charge becomes

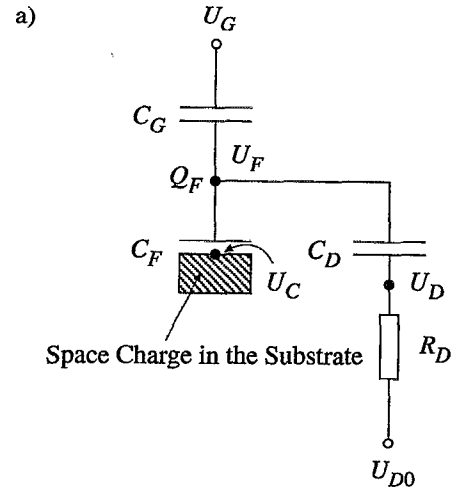
$$Q_F=C_G(U_F-U_G)+C_D(U_F-U_D)+\frac{C_F}{l_F}\int_0^{l_F}[U_F-U_C(l)]dl, \tag{3}$$

where U_G is the control gate-to-source voltage, U_C the voltage drop along the channel, l_F the channel length, and C_G, C_D, and C_F are the capacities indicated in Fig. 8. If the drain-to-source voltage is small and the transistor is not in saturation, the voltage along the channel is approximately linear, increasing from source to drain voltage. With this approximation and Eq. (3), the floating gate-to-source voltage becomes

$$U_F=\frac{Q_F}{C_T}+\alpha_G U_G+\alpha_D U_D+\alpha_F\frac{U_D}{2}, \tag{4}$$

with

$$\alpha_G=\frac{C_G}{C_T},\quad \alpha_D=\frac{C_D}{C_T},\quad \alpha_F=\frac{C_F}{C_T},\quad C_T=C_G+C_D+C_F, \tag{5}$$



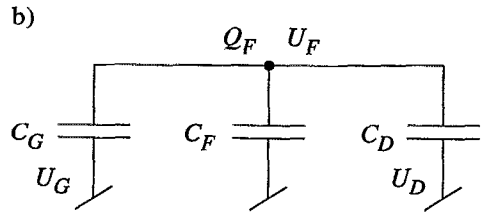


FIG. 8. Floating gate transistor model: (a) measurement condition; (b) bake condition.

where $U_D=U_{D0}-R_D I_D$ with R_D being the resistor from outside to the drain contact and U_{D0} the externally applied drain-to-source voltage.

Figure 9 shows the drain-source current versus gate-to-source voltage of one cell in different erase states. The dots represent the measurements while the solid lines were calculated. With the above described transistor model, the floating gate charge is calculated from the measured drain-source current. The derivative of this charge is the current leaking through the isolation: $I=dQ_F/dt$.

If the control gate-to-source and drain-to-source voltages are zero, there is still an electric field across the isolation due to the floating gate charge. This is the condition during the

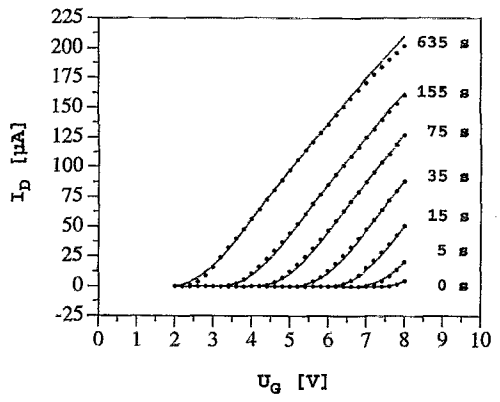


FIG. 9. Threshold voltage shift time with bake temperature as parameter.

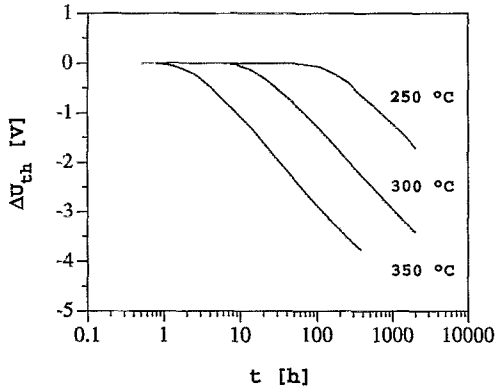


FIG. 10. Drain-source current vs gate-to-source voltage at 20 °C with UV erase time as parameter. Measurements (dots) and model (solid lines).

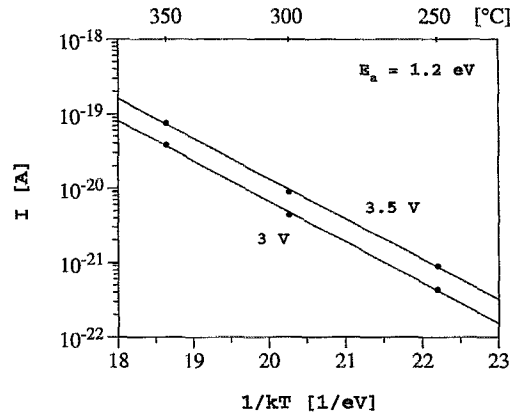


FIG. 12. Arrhenius plot of the leakage current vs temperature for 3 and 3.5 V floating gate-to-source voltage.

bake [see Fig. 8(b)]. The substrate is in accumulation; thus, the floating gate-to-source voltage is given by

$$U_F = \frac{Q_F}{C_T} + \alpha_F(\Psi_{b,S} - \Psi_{b,F}), \tag{6}$$

where $\Psi_{b,S}$ and $\Psi_{b,F}$ are the bulk potentials in the substrate and floating gate, respectively. $\Psi_{b,S}<0$, and $\Psi_{b,F}>0$, $\Psi_{b,S}-\Psi_{b,F}\approx -1$ V. Although the drain-source current of the cell transistor is used for the charge loss indicator, we plotted in Fig. 10 the total threshold voltage shift ΔU_{th} of the floating gate transistor to make comparison with previous work more easy. The total threshold voltage U_{th} becomes

$$U_{\text{th}} = \alpha_G\left(U_t - \frac{Q_F}{C_T} - \alpha_D U_D - \alpha_F \frac{U_D}{2}\right) \tag{7}$$

[see Eqs. (2) and (3)], and thus the total threshold shift is given by $\Delta U_{\text{th}} = \Delta Q_F / C_G$.

In Fig. 11 the long-term leakage current is plotted versus the floating gate-to-source voltage at corresponding bake times, and Fig. 12 shows the temperature dependence at 3 and 3.5 V floating gate-to-source voltage. The activation energy is calculated assuming that the temperature behavior of the leakage current is according to $\exp(-E_a/kT)$.

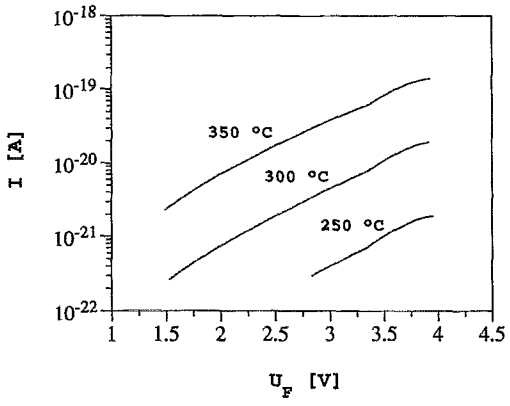


FIG. 11. Leakage current vs floating gate-to-source voltage calculated from the data retention experiment.

IV. MULTIPHONON-ASSISTED TUNNELING

It has been reported that the main path for nondefective charge loss in EPROMs with ONO interpoly dielectric is across the ONO.[13–15] This can be verified by applying high voltage on the control gate of programed and erased cells. The electric field of the programed cell is small across the gate oxide and high across the interpoly dielectric. The conditions for the erased cell are the reverse. If there is no charge gain on the erased cells but a charge loss of the programed cells during a bake, it can be concluded that the charge transport occurs through the interpoly dielectric. This experiment was performed within the scope of the present work with the control gate-to-source voltage between 4 and 7 V at 250 °C; but, since no difference was observed compared to the normal bake without bias, it has to be doubted that the voltage applied outside of the device actually reached the control gates. The devices were operating outside of the specified temperature range. The test was redone at 150 °C, but no change of the initial condition could be measured during 4000 h bake. Thus, we rely on the cited literature on this point.

We propose the following model for the leakage. Electrons stored on the floating gate can be captured by traps in the bottom oxide and subsequently be emitted into the nitride by a multiphonon-assisted tunneling process. In the nitride, the conductivity is larger than in the oxide. It will be assumed that electrons that have passed the bottom oxide can pass the top oxide likewise. Hole transport is prevented by the 70-Å-thick top oxide. Figure 13 illustrates the charge transport mechanism across the bottom oxide. Oxide and ONO devices have nearly the same charge loss characteristic (see Fig. 4) and, in particular, equal activation energies. Electrons stored on the floating gate have to pass the same energy barrier between floating gate and oxide in both device types. Thus, it can be expected that the charge loss is governed also by the same mechanism, i.e., that it is not limited by the bulk.

Tunnel-assisted release of electrons from traps with multiphonon participation was considered for the first time by Kriveris, Kudzmauskas, and Pipinys[55] in connection with thermoluminescence experiments. These authors convoluted

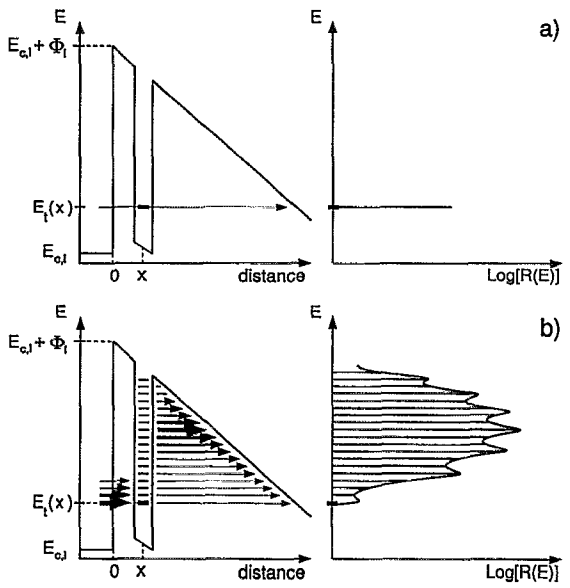


FIG. 13. Energy band diagram illustrating (a) resonant tunneling and (b) multiphonon-assisted tunneling. The left-hand-side parts indicate the probability for various transition paths, the right-hand-side parts show calculated transmission rates.

the probability of pure electronic tunneling with a Gaussian for the strong electron-phonon coupling, generalizing in this way the single-phonon-assisted band-to-band tunneling theory (Keldysh 1958[56,57]) to trap-to-band transitions. Dalidchik[58] coupled the WKB tunnel expression with a more refined multiphonon theory for the same problem.

Here we will follow the approach of Refs. 59 and 60, where Shockley–Read–Hall (SRH) lifetimes were calculated under the assumption that the capture of carriers is a multiphonon transition between a band-tail state, induced by the electric field, and the localized state of the center. In contrast to the case of a bulk semiconductor, for the present problem the initial state is localized within the floating gate, the final state in the nitride, and the tunnel barrier is formed by the fraction of the bottom-oxide gap between gate and trap. It is assumed that the electron traps in the bottom oxide are monoenergetic with a spatial density N_t and a thermal binding energy Φ_t. Figure 13(a) illustrates the two-step tunneling mechanism where the transition is restricted to a sharp energy given by the trap level. Coupling to oxide phonons [see Fig. 13(b)], represented by an effective phonon energy $\hbar\omega_0$, and two coupling constants (the lattice relaxation energy ϵ_R and the Huang–Rhys factor S), results in a thermal broadening of the trap level $E_t(x)$. A series of virtual states of the coupled electron-phonon system in the energy gap of the oxide occurs which allows for transition paths with an increased tunnel probability. The "ladder" of these states is indicated in Fig. 7. Each sublevel can serve as initial state for the multiphonon-assisted tunneling transition into the nitride after the electron has been captured by the trap. The occupation probability of the traps is given by

$$f_t=\frac{n_t}{N_t}, \tag{8}$$

where n_t is the density of filled traps. The rate of charge carriers per volume and time that occupy or leave the traps can be expressed by

$$R_l(x)=\tau_l^{-1}(x)N_t(1-f_t), \tag{9}$$

$$R_r(x)=\tau_r^{-1}(x)N_t f_t, \tag{10}$$

where τ_l and τ_r are the time constants for capture or emission of electrons, respectively. The difference of these two rates defines the variation of the trapped charge density with time,

$$\frac{dn_t(x)}{dt}=R_l(x)-R_r(x). \tag{11}$$

Under steady-state conditions $dn_t/dt=0$, or equivalently, $R=R_l=R_r$. Using the above equations and assuming steady state, the rate of charge carriers per volume and time passing the isolation via traps is given by

$$R(x)=N_t\frac{1}{\tau_l(x)+\tau_r(x)}, \tag{12}$$

and the current-density contribution from the traps in a small interval dx at a distance x is

$$dj(x)=qR(x)dx. \tag{13}$$

The time constants τ_l and τ_r contain characteristic quantities of the floating gate and the nitride, namely: the density of states in the conduction bands N_l and N_r, the occupation probabilities f_l and f_r, and the tunnel probabilities T_l and T_r from the floating gate to a trap at distance x and from this trap to the nitride,

$$\tau_l^{-1}(x)=\int_{E_t(x)}^{\infty}N_l(E)f_l(E)T_l(E,x)c_n(E,x)dE, \tag{14}$$

$$\tau_r^{-1}(x)=\int_{E_t(x)}^{\infty}N_r(E)[1-f_r](E)T_r(E,x)e_n(E,x)dE. \tag{15}$$

Here $c_n(E,x)$ denotes the capture rate,

$$c_n(E,x)=c_0\sum_{m=0}^{\infty}L_m(z)\,\delta[E-E_m(x)], \tag{16}$$

and $e_n(E,x)$ the emission rate,

$$e_n(E,x)=c_0\exp\left(-\frac{E-E_t(x)}{kT}\right)\sum_{m=0}^{\infty}L_m(z)\,\delta[E-E_m(x)]. \tag{17}$$

The prefactor c_0 is not used as a fitting parameter, but has been estimated using a 3D delta potential for the trap,[61]

$$V(\mathbf{x})=4\pi\Phi_t r_t^3\,\delta(\mathbf{x}-\mathbf{x}_t)(1+\mathbf{x}\cdot\nabla),$$

and by calculating the expectation value of the perturbation operator $qF\langle x\rangle_{ct}\approx\sqrt{(\hbar\Theta_o)^3/\Phi_{g,o}}$ neglecting Bloch factors,[62]

$$c_0=\frac{(4\pi)^2 r_t^3}{\hbar\Phi_{g,o}}(\hbar\Theta_o)^3. \tag{18}$$

Here r_t is the localization radius of the trapped electron, $\hbar\Theta_o$ denotes the electro-optical energy in the oxide,

$$\hbar\Theta_o=\left(\frac{q^2\hbar^2F^2}{2m_{c,o}}\right)^{1/3}, \tag{19}$$

q is the elementary charge, F the electric field in the oxide, $m_{c,o}$ and $\Phi_{g,o}$ are the effective electron mass and the band gap of the oxide, respectively, and E is the energy. The conduction-band edge on the left-hand side (floating gate) was chosen as energy zero. Only energy levels $E_m(x)=E_t(x)+m\hbar\omega_0$, where m is an integer, are allowed for the capture and emission process as a consequence of the one-mode approximation.

The multiphonon transition probability $L_m(z)$ is given by

$$L_m(z)=\left(\frac{f_B+1}{f_B}\right)^{m/2}\exp[-S(2f_B+1)]I_m(z). \tag{20}$$

S denotes the Huang–Rhys factor,[63] which is a measure of the coupling strength of the electron-phonon interaction, $I_m(z)$ the modified Bessel function of order m with the argument $z=2S\sqrt{f_B(f_B+1)}$, and f_B the Bose function, which gives the phonon occupation number

$$f_B=\frac{1}{\exp(\hbar\omega_0/kT)-1}. \tag{21}$$

The lower integration limit $E_t(x)=E_{c,o}(x)-\Phi_t$ in Eqs. (14) and (15) indicates that all transitions via virtual states below the trap level were skipped because hole capture from the nitride and hole emission to the floating gate are neglected. $E_{c,o}(x)$ is the conduction-band edge energy in the oxide. The densities of state on the left- and right-hand sides of the oxide are taken in the form

$$N_l(E)=N_0\left(\frac{m_{c,l}}{m}\right)^{3/2}\left(\frac{E-E_{c,l}}{kT}\right)^{1/2}\Theta(E-E_{c,l}), \tag{22}$$

$$N_r(E)=N_0\left(\frac{m_{c,r}}{m}\right)^{3/2}\left(\frac{E-E_{c,r}}{kT}\right)^{1/2}\Theta(E-E_{c,r}), \tag{23}$$

with

$$N_0=\frac{\sqrt{kT}}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}. \tag{24}$$

These forms involve two approximations: First, parabolic bands are assumed, although the electron energy can be more than 1 eV above the bottom of the conduction bands; second, the density of states in the energy gap of the floating gate is assumed to be zero. This neglects localized states in the polysilicon conductivity gap and interface states at the poly-SiO_2 interface. $\Theta(E)$ denotes the step function, m is the electron rest mass, $m_{c,l}$ and $m_{c,r}$ are the effective electron masses, and $E_{c,l}$ and $E_{c,r}$ are the energies of the conduction band edges on the left- and right-hand sides of the oxide. The floating gate is highly n doped, thus, the Fermi level on the left-hand side can be approximated by $E_{f,l}\approx E_{c,l}$. On the right-hand side the following approximation is employed:

$$1-f_r(E)\approx 1, \tag{25}$$

i.e., every arriving electron is assumed to find an empty state. The tunnel probability from the left-hand side to a trap at distance x and from this trap to the right-hand side can be expressed using the WKB approach,

$$T_l(E,x)=\exp\left(-2\int_0^x\left|\kappa(\xi)\right|d\xi\right), \tag{26}$$

$$T_r(E,x)=\exp\left(-2\int_x^d\left|\kappa(\xi)\right|d\xi\right), \tag{27}$$

with

$$\kappa^2(\xi)=\frac{2m_{c,o}}{\hbar^2}[E_{c,o}(\xi)-E]. \tag{28}$$

Neglecting the image force effect, assuming a constant electric field in the oxide and an abrupt potential well introduced by the trap, the tunnel probabilities become

$$T_l(E,x)=\exp\left[-\frac{4}{3}\left(\frac{\varphi(E,0)}{\hbar\Theta_o}\right)^{3/2}+\frac{4}{3}\left(\frac{\varphi(E,x)}{\hbar\Theta_o}\right)^{3/2}\right], \tag{29}$$

$$T_r(E,x)=\exp\left[-\frac{4}{3}\left(\frac{\varphi(E,x)}{\hbar\Theta_o}\right)^{3/2}+\frac{4}{3}\left(\frac{\varphi(E,d)}{\hbar\Theta_o}\right)^{3/2}\Theta[E_{c,o}(d)-E]\right], \tag{30}$$

with

$$\varphi(E,x)=\Phi_l-qFx\doteq E=E_{c,o}(x)-E. \tag{31}$$

Equations (16)–(25) and (29)–(31) define all the quantities in (14) and (15). Due to the delta function in (16) and (17), the energy integrals (14) and (15) turn into sums over discrete energies $E_m(x)$,

$$\tau_l^{-1}(x)=\sum_{m=0}^{\infty}\tau_{l,m}^{-1}(x) \quad\text{and}\quad \tau_r^{-1}(x)=\sum_{m=0}^{\infty}\tau_{r,m}^{-1}(x), \tag{32}$$

with

$$\tau_{l,m}^{-1}(x)=\tau_0^{-1}\frac{N_l(E_m)}{N_0}f_l(E_m)T_l(E_m,x)L_m(z), \tag{33}$$

$$\tau_{r,m}^{-1}(x)=\tau_0^{-1}\frac{N_r(E_m)}{N_0}T_r(E_m,x)\exp\left(-\frac{m\hbar\omega_0}{kT}\right)L_m(z), \tag{34}$$

and the time constant τ_0 given by

$$\tau_0=\frac{1}{N_0c_0}. \tag{35}$$

Using Eq. (13), the total current density becomes

$$j=qN_t\int_0^d\frac{1}{\tau_l(x)+\tau_r(x)}\,dx, \tag{36}$$

where the integration can be done numerically.

V. RESONANT TUNNELING

In this section we study the role of resonant tunneling as an alternative mechanism of charge loss. The potential wells

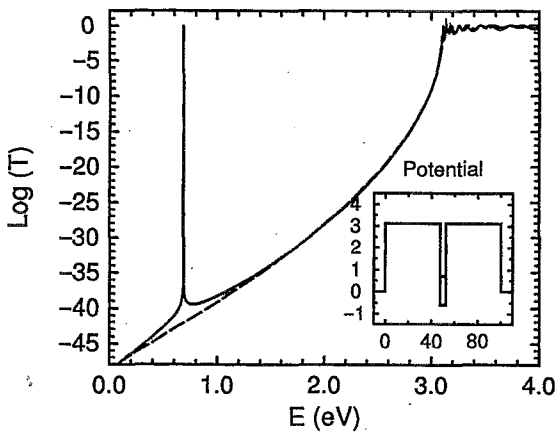


FIG. 14. Calculated transmission probability for a 100-Å-thick oxide containing a centered square well with a single bound state $\Phi_t=2.4$ eV.

of oxide traps offer the possibility for gate electrons to tunnel with an increased probability along a "path" containing the trap. It is not at first place the reduction of the tunneling length by the diameter of the potential well which gives rise to the larger probability, but the existence of resonance levels associated with the trap potential. At these energies the transmission probability peaks sharply, and the tunnel current is solely determined by the width of the resonance. Figure 14 illustrates the single resonance of a (properly chosen) square-well potential, which is located exactly in the middle of the oxide barrier between floating gate and substrate (nitride not considered for simplicity). Such an idealized symmetrical potential makes the oxide completely transparent at the resonance energy. Here, the exponentially rising solution of the Schrödinger equation exactly compensates for the exponentially decaying.

Any disturbance of the symmetrical shape, as by an electric field $F\neq 0$, a shift of the trap position toward one interface, or by the coupling of the resonance level to charges, phonons, etc., will damp the transmission peak. This is demonstrated for the case of a repulsive trap in Fig. 15. The

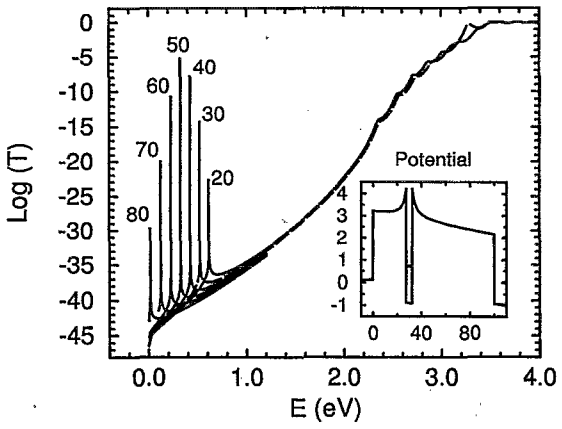


FIG. 15. Dependence of the resonance peak of a repulsive trap on the position within a 100-Å-thick oxide for $F=1$ MV/cm. The numbers are the distance of the trap from the left-hand-side oxide boundary. At $x\approx 80$ Å the resonance disappears because the trap level passes the silicon band edge.

position of the trap was varied in steps of 10 Å starting at a distance of 20 Å from the gate-oxide interface and shifting the center toward the opposite interface. The applied field was 1 MV/cm, and the calculation was performed using the transfer-matrix method (see, e.g., Refs. 64–66) with 80 partial rectangular barriers. Now the maximum transmission remains below 1 and the trap position where the transmission probability becomes maximum is less than half the barrier width, which can be seen from the difference by three orders of magnitude between the peaks of 40 and 60 Å distance. The resonance disappears from the energy scale at a trap position of about 80 Å.

It may be concluded that depending on the potential shape, i.e., the applied field, only traps in a narrow sheet at a certain distance from the gate-oxide interface contribute essentially to the resonant tunneling current. The latter will be calculated analytically under the same simplifying assumptions as in the previous section (no oxide charges, no image forces, abrupt trap potential). Similar problems were solved by Duke and Alferieff[67] in connection with field emission through atoms adsorbed on a metal surface, and by Gadzuk[43] in connection with resonant trap tunneling across metal-insulator-metal structures. Ricco, Azbel, and Brodsky[68] considered resonant tunneling via defect-related states in thin SiO_2 films under high electric field in a more qualitative way. Dalidchik[69] calculated the effect of an ordered distribution of scattering centers on the resonant tunnel current.

In the present article the treatment will be generalized with respect to the tunnel probability, and the current will be given in a form suitable for comparison with the two-step capture-emission mechanism described in the previous section. We also turn the square-well potential into a delta function (see, e.g., Ref. 70), which results in an even simpler form for the resonant tunneling current.

The transition rate due to all traps situated in a plane of distance x from the gate-oxide interface is obtained by multiplying with the ratio of the total cross section of all traps and the contact area $\pi r_t^2 \nu_t(x)$ (ν_t—2D trap density at x, traps are assumed to be distributed homogeneously within that plane)

$$R_{\rm res}=\pi r_t^2 N_t(x)\,\frac{1}{4\pi^3\hbar}\int_{-\infty}^{\infty} dE\; T_{\rm res}(E,x)\int d^2\mathbf{k}_\perp f_l(E,\mathbf{k}_\perp) \tag{37}$$

$$=\frac{r_t^2 N_t(x) m_{c,l} kT}{2\pi\hbar^3}\int_{-\infty}^{\infty} dE\; T_{\rm res}(E,x)$$

$$\times\ln(1+e^{-(E-E_{f,l}/kT)}). \tag{38}$$

As indicated by the last expression, the so-called supply function (number of electrons per energy range that are available for tunneling through the barrier) is given by the density of states and the thermal occupation probability in the floating gate. Penley[71] pointed out that the resonance of the wave corresponds to the electron being trapped for some time in the potential well which limits the number of electrons that can pass through the well at any given time. Penley showed that the delay time in the trap is given by $\tau_{\rm del}=\hbar\partial\Phi/\partial E$, where Φ is the total phase shift across the

barrier. Therefore, the effective supply function would result from the conductivity limited by traps in series with the conductivity limited by the supply of electrons from the gate. We will assume here that the delay time τ_{del} is always much smaller than the time constant of supply from the gate. This is supported by the assumption of a strong electron-phonon coupling of the defect state which limits τ_{del}.

The transmission probability T_{res} can be evaluated explicitly for the simplified trap potential model. The major steps of the derivation are presented in the Appendix. First, T_{res} is obtained in Lorentzian form taking special care of the damping term. Since the resonance level is extremely sharp, the Lorentzian is then transformed into a delta function of energy. The information about potential parameters which is not available, as depth and width of the well, becomes dispensable, if the square well is turned into a 1D delta potential. Then, a simple and symmetrical expression follows for T_{res},

$$T_{res}(E,x)=32\pi\frac{\Phi_t}{\Phi_l\Phi_r}\left(\frac{m_{c,r}}{m_{c,l}}\right)^{1/2}\sqrt{E_lE_r}\times\sqrt{(\Phi_l-E_l)(\Phi_r-E_r)}\,\frac{T_lT_r}{\sqrt{T_l^2+T_r^2}}\,\delta[E_t(x)-E], \tag{39}$$

where $E_l\equiv E_t(x)-E_{c,l}$, $E_r\equiv E_t(x)-E_{c,r}$, and Φ_l,Φ_r denote the energy barriers on the left- and right-hand sides, respectively. Now we are able to compare directly the rate of resonant tunneling with that of zero-phonon trap tunneling. Inserting Eq. (39) into Eq. (38) we obtain

$$R_{res}(x)=R_{0,res}(x)\,\frac{\sqrt{m_{c,l}m_{c,r}}}{m}\,\frac{\sqrt{E_lE_r}\sqrt{(\Phi_l-E_l)(\Phi_r-E_r)}}{\Phi_l\Phi_r}\times\Theta(E_l)\Theta(E_r)\ln(1+e^{-(E_t(x)-E_{f,l}/kT)})\,\frac{T_lT_r}{\sqrt{T_l^2+T_r^2}}, \tag{40}$$

with

$$R_{0,res}(x)=\frac{8N_t(x)}{\hbar}\,\Phi_t kT\left(\frac{2mr_t^2}{\hbar^2}\right). \tag{41}$$

Switching off the electron-phonon coupling (Huang–Rhys factor $S\to 0$, $L_m(z)\to\delta_{m,0}$) in the multiphonon-assisted trap tunneling mechanism [Eqs. (16) and (17)], the two-step process is restricted to the resonance energy $E_t(x)$, and the resulting zero-phonon rate becomes

$$R_{0-ph}(x)=R_{0,0-ph}(x)\,\frac{m_{c,l}^{3/2}m_{c,r}^{3/2}}{m^3}\,\frac{\sqrt{E_lE_r}}{kT}\,\Theta(E_l)\Theta(E_r)f_l[E_t(x)]\,\frac{T_lT_r}{f_l[E_t(x)]\sqrt{E_l/kT}(m_{c,l}/m)^{3/2}T_l+\sqrt{E_r/kT}(m_{c,r}/m)^{3/2}T_r}, \tag{42}$$

with

$$R_{0,0-ph}(x)=\frac{N_t(x)}{\tau_0}=\frac{8N_t(x)}{\hbar}\,\sqrt{kT}\,\frac{(\hbar\Theta_o)^3}{\Phi_{g,o}}\left(\frac{2mr_t^2}{\hbar^2}\right)^{3/2}. \tag{43}$$

It is interesting that the structure of Eq. (40) is quite similar to that of Eq. (42) despite the different physical starting points. Figure 16 shows that both mechanisms yield the same order of magnitude. The main differences arise from the assumption that the delay time τ_{del} is much smaller than the time of supply of electrons from the floating gate. This assumption makes resonant tunneling actually a one-step process. Hence, there is no additivity for reciprocal time constants proportional to $1/T_l$ and $1/T_r$, and the current always depends on the occupation in the floating gate. In contrast, the two-step process becomes independent of the latter, if $T_l \gg T_r$ holds, i.e., if the emission from the trap site is the bottleneck for the charge loss. The above calculation shows that the two-step mechanism without phonon participation as a special case of multiphonon-assisted trap tunneling contains the same physics as resonant tunneling, however, requires a much simpler algebra only. Resonant tunneling including phonon coupling of the trap state, as demonstrated by Fu and Willander[72] for a double-barrier resonant tunneling structure, would result in the multiphonon-assisted tunneling mechanism described in Sec. IV.

VI. DISCUSSION

For the numerical calculations we used the following parameters: The electron effective masses in Si on the left-hand side in SiO_2, and in Si_3N_4 on the right-hand side, were set to $m_{c,l}=0.33m$, $m_{c,o}=0.42m$, and $m_{c,r}=0.42m$, respectively. The energy barriers at the Si-SiO_2 interface and at the SiO_2-Si_3N_4 interface were given the values $\Phi_l=3.1$ eV and $\Phi_r=1.1$ eV. Furthermore, we used for the energy gap of SiO_2 $\Phi_{g,o}=8.9$ eV, for the energy of the defect electron $\hbar^2/(2mr_t^2)=3.81r_t^2$ eV with r_t in Å (e.g., 0.61 eV for $r_t=2.5$ Å), and for the electro-optical energy

$$\hbar\Theta_o=7.25\times10^{-6}(F^2m/m_{c,o})^{1/3}\ \text{eV}$$

with F in V/cm (e.g., $\hbar\Theta_o=0.15$ eV for $F=2$ MV/cm). The time constant τ_0 is 5.9×10^{-13} s for $F=2$ MV/cm. This is in the range from 10^{-12} to 10^{-14} s, as given in Ref. 29.

In Secs. IV and V the multiphonon-assisted tunneling and resonant tunneling mechanisms were presented. In order

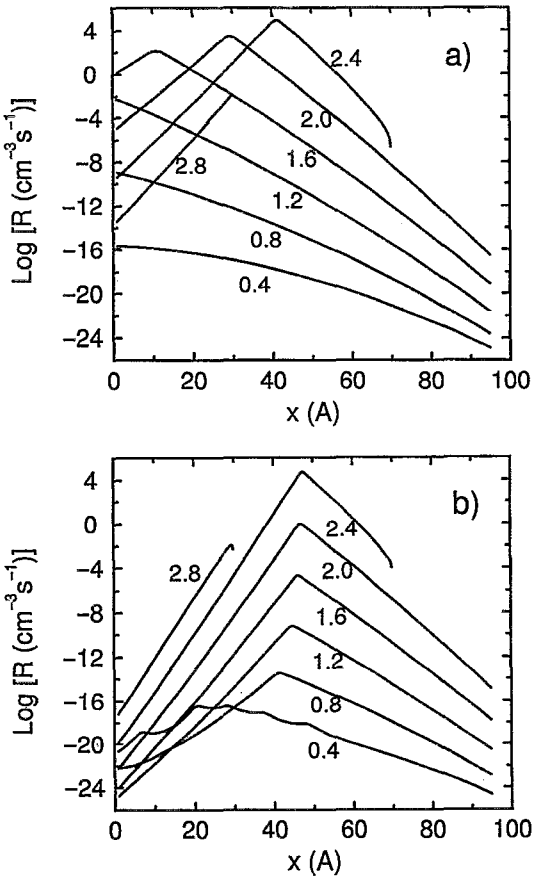


FIG. 16. Transition rate R as function of trap position for $T=300$ K, $F=1$ MV/cm and different trap depths Φ_t measured from the oxide conduction-band edge. Comparison of (a) two-step (zero phonon) tunneling and (b) resonant tunneling.

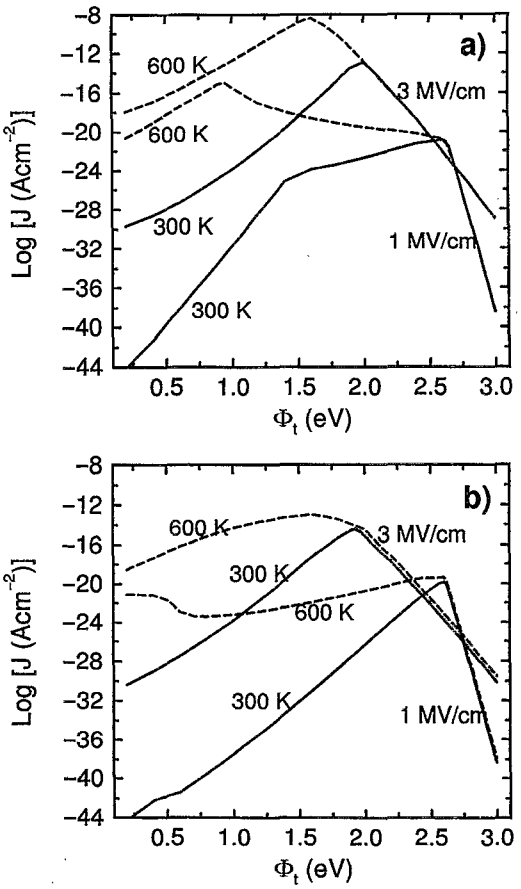


FIG. 17. Current density as function of trap depth Φ_t for different temperatures and field strengths. Comparison of (a) two-step (zero phonon) tunneling and (b) resonant tunneling.

to compare both mechanisms, we switch off the electron-phonon coupling. Then, the multiphonon-assisted mechanism turns into the two-step tunneling mechanism, discussed, e.g., in Ref. 47.

Figure 16 compares the transition rates of two-step tunneling $R_{0-\mathrm{ph}}$ against resonant tunneling R_{res}, as defined in Eqs. (42) and (40), respectively. The transition rates are plotted as a function of trap position with the trap depth measured from the oxide conduction-band edge as family parameter. The oxide field strength is 1 MV/cm. For such a field the trap depth of 2.6 eV gives the largest rate in both cases. Surprisingly, even the absolute value is the same, despite the different physical bases. The essential difference is that in the case of two-step tunneling those traps located in the left-hand-side half of the oxide (closer to the cathode) yield much higher contributions than in the case of resonant tunneling. The shallower traps ($\Phi_t<1.4$ eV) are even most effective, if they are located directly at the interface. This is because the capture process restricts the whole transition rate due to the very small thermal occupation of the corresponding energy levels in the conduction band on the left-hand side of the oxide. Both mechanisms yield an abrupt drop of the transition rate for trap levels $E_t(x)$ lower than $E_{c,l}$ ($\Phi_t>2.4$ eV and $x>70$ Å in the figure), since we did not consider possible tunneling transitions out of tail and interface states of the floating gate.

Figure 17 shows the comparison of the "spectral" current, i.e., the current density as function of the trap depth, for different field strengths and temperatures. The trap depth of maximum current density decreases with increasing field. In the case of two-step tunneling at 1 MV/cm, the peak at $\Phi_t=2.6$ eV is exceeded by a new peak at $\Phi_t=0.9$ eV, if the temperature rises from 300 to 600 K. Increasing temperature for both mechanisms enhances the contribution of shallower

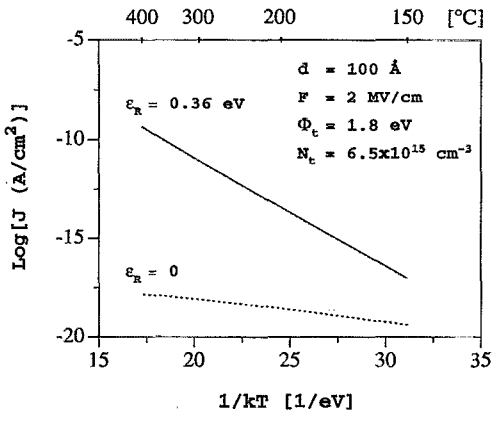


FIG. 18. Arrhenius plot of the current density vs temperature. Multiphonon process (solid line) and no electron-phonon coupling (dashed line).

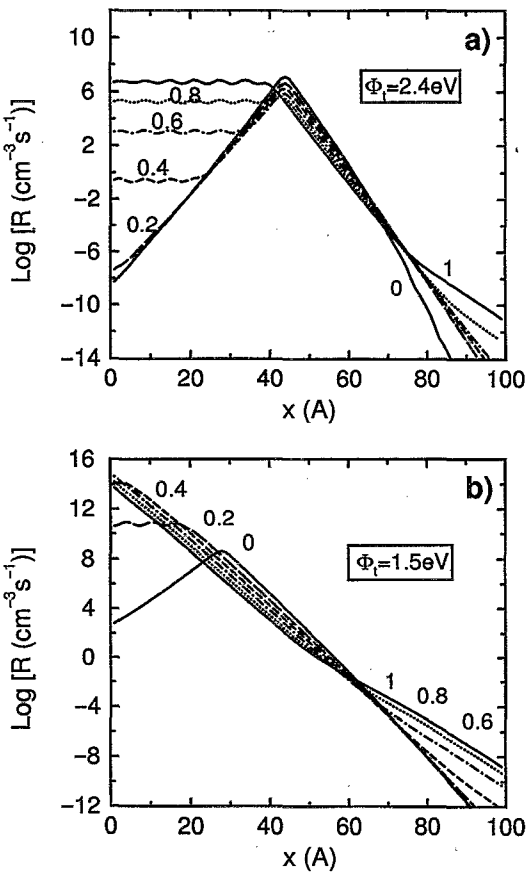


FIG. 19. Transition rate R as function of trap position for T=573 K, F=1 MV/cm, $\hbar\omega_0$=0.063 eV; and different lattice relaxation energies ϵ_R in eV: (a) Φ_t=2.4 eV, (b) Φ_t=1.5 eV.

traps. Two-step tunneling shows a larger increase of the current density with increasing temperature compared to resonant tunneling.

It has been verified experimentally in Sec. III that charge loss has a strong temperature dependence as shown in Fig. 12. The theoretical temperature dependence of both two-step tunneling and resonant tunneling is too weak to explain the charge loss measurements. Coupling of the trap state to oxide phonons is essential for the activation behavior. Figure 18 shows the influence of the electron-phonon coupling on the temperature dependence. The solid line was calculated with S=6 (ϵ_R=0.36 eV), but the dashed line with the coupling switched off (S=0). A multiphonon coupling resulting in a lattice relaxation energy of 0.36 eV can explain the measured activation energy of 1.2 eV.

Figure 19 illustrates the effect of electron-phonon coupling on the transition rate for two different trap depths at a temperature of 573 K and a field strength of 1 MV/cm. The family parameter is the lattice relaxation energy ϵ_R in eV. For weak electron-phonon coupling deep traps ($\Phi_t \approx 2.4$ eV) contribute most, if they are located in the center of the oxide layer. With increasing electron-phonon coupling traps located in the left-hand-side half of the oxide (closer to the cathode) become more important for the total transition rate. This is because the electron-phonon coupling enhances the emission process more than the capture process. The emission occurs closer to the conduction-band edge of the oxide where the tunnel probability is large [compare Fig. 13(b)]. Shallower traps (Φ_t=1.5 eV) are most effective, if they are located directly at the interface on the left-hand side of the oxide.

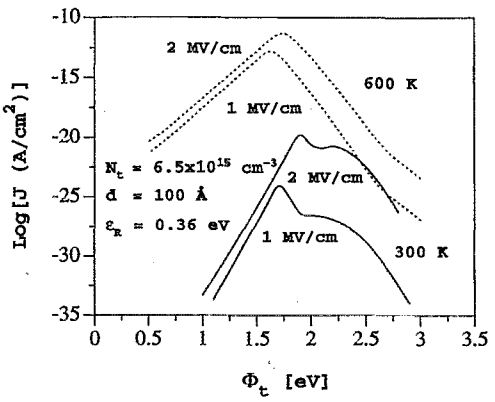


FIG. 20. Current density vs trap energy measured from the oxide conduction-band edge.

In Fig. 20 the current density is plotted versus trap depth Φ_t for two values of the electric field and two different temperatures. At room temperature maxima appear at 1.6 and 1.7 eV, respectively. The current density changes by many orders of magnitude as the trap depth varies. Various trap levels have been reported for SiO_2.[73] Since only a very low trap density is sufficient to explain the measured current density, we make the assumption that all energy levels are available in the range between 1.5 and 1.7 eV. Figure 21 shows the dependence of the trap depth resulting in the largest current density on electric field and temperature.

Instead of integrating the current density contributions over all trap energies, only the maximum has been used. The half-width of the current density distribution at 600 K is about 0.2 eV (see Fig. 20). For the density of trap states we assume that the value of N_t refers to an energy interval of about 0.2 eV now, i.e., $N_t(E)=5N_t$ (eV)$^{-1}$. The deep-seated traps can also be used by leaking carriers but with a much lower probability. It is assumed that all traps have similar

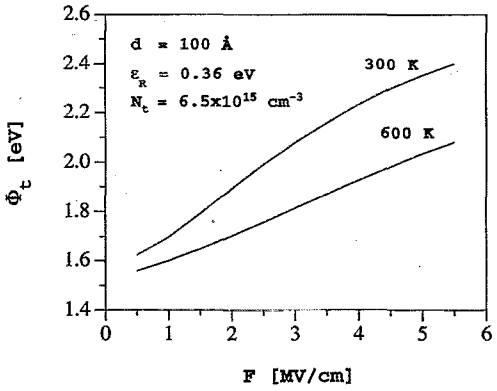


FIG. 21. Trap level (measured from the oxide conduction-band edge) resulting in maximum current density vs electric field with temperature as parameter.

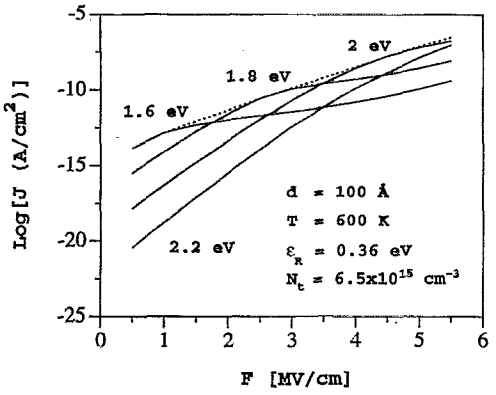


FIG. 22. Current density vs electric field in the oxide. Single trap level (solid lines) and varying trap depth resulting in the largest current density at each field strength (dashed line).

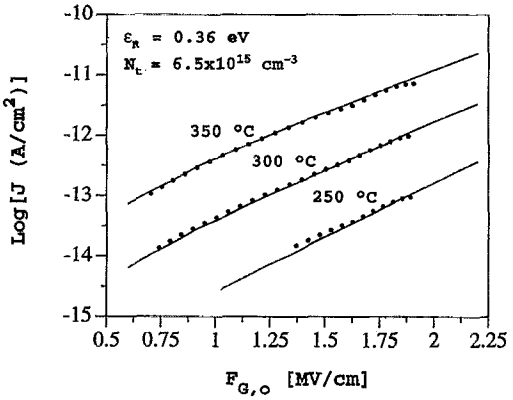


FIG. 24. Current density vs oxide field: measurement (dots) and calculation (solid lines).

microscopic coupling constants. Thus, the transition rates differ due to Φ_t only. More permanent occupation of deeper traps has to be expected. That contributes to a steady-state charge density in the oxide. Such charges could behave like "fixed" charges, i.e., they are present before and after repeated write operations.

The dependence of the current density on the electric field is plotted in Fig. 22. The solid lines were computed for different single trap levels, the dashed line corresponds to a variable trap depth according to Fig. 21. In Fig. 23 the current density versus temperature is plotted in an Arrhenius plot. The activation energy is increasing with temperature while the trap depth giving the largest contribution to the current density is slightly decreasing. A large activation energy indicates a strong temperature dependence. At high temperatures the electron-phonon coupling has a dominant influence on the current density. With decreasing temperature tunneling becomes more important for the capture and emission process. Previous models used a constant activation energy to describe the temperature dependence, which may not be realistic over such a large temperature range as from 300 °C to room temperature.

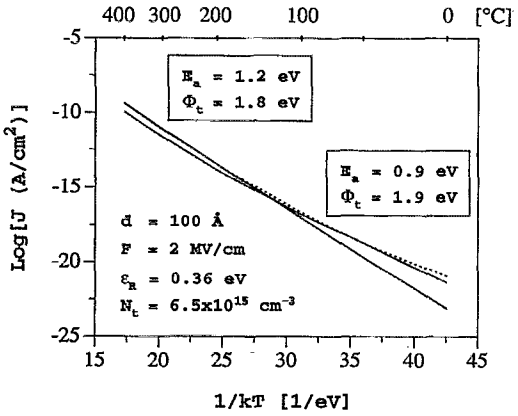


FIG. 23. Arrhenius plot of the current density vs temperature. Single trap level (solid lines) and varying trap depth resulting in the largest current density (dashed line).

The bake experiment has been performed between 250 and 350 °C. The electric field in the oxides of the ONO, given by

$$F_{G,o}=U_F/(d_{b,o}+\epsilon_o d_n/\epsilon_n+d_{t,o}),$$

has been calculated to be below 2 MV/cm ($d_{b,o}$, d_n, and $d_{t,o}$ are the thicknesses of bottom oxide, nitride, and top oxide, respectively, ϵ_o and ϵ_n are the permittivities of oxide and nitride). Figure 24 compares the measurement with the calculation. A trap density $N_t=6.5\times10^{15}$ cm^{-3} and a Huang–Rhys factor $S=6$ have been used as fit parameters. The third fit parameter—the lattice relaxation energy ϵ_R—contains the effective phonon mode $\hbar\omega_0$ in the amorphous oxide. Varying S and $\hbar\omega_0$ but keeping $\epsilon_R=S\hbar\omega_0$ constant has only little influence on the fit. Therefore, the only relevant fit parameter in the model is ϵ_R if we assume a quasicontinuous distribution of trap levels. Even with a variation of ϵ_R in the range from 0.2 to 0.6 eV the fit only slightly deteriorates. Experimental evidence for lattice relaxation was found by Hwang, Or, and Forbes[74] who measured a difference between thermal activation energy and photoemission energy of electrons bound to traps in 20-nm-thick oxides. The reemission time constant after the oxide had been stressed was strongly temperature dependent indicating a multiphonon process; however, the quantitative determination of the lattice relaxation energy from their data is inhibited, because it remains unclear to which extent the "thermal" barrier was lowered by tunneling. In order to estimate this effect one would have to know the depth of the created traps in the oxide, which depends on the stress conditions.

To verify whether the multiphonon-assisted tunneling mechanism is dominant even at room temperature, charge loss measurements would have to be extended to a time period of more than 10 years. Using the same parameter set as for the high-temperature range, we can calculate the expected current density at room temperature. The results for multiphonon-assisted tunneling, resonant tunneling (without phonon participation), and direct tunneling are compared in Fig. 25, where the current density across a single oxide layer of 10 nm width was calculated assuming a homogeneous distribution of the traps in current direction (N_t=const) but only one discrete energy level (Φ_t=1.6 eV). It can be seen

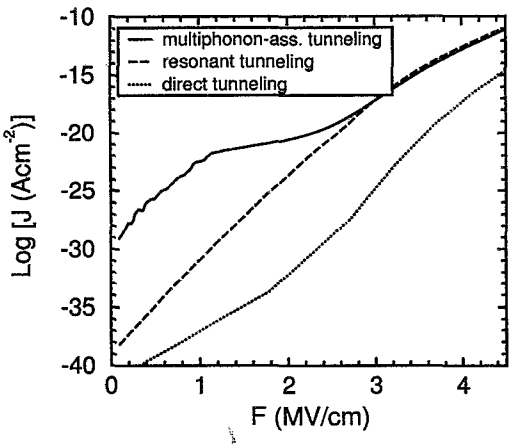


FIG. 25. Current density vs oxide field in the low-field range for a single oxide layer of 10 nm width: Comparison between multiphonon-assisted tunneling, resonant tunneling, and direct tunneling. Parameters: Φ_t=1.6 eV, ϵ_R=0.36 eV.

that thermal emission plays a dominant role up to fields of 2 MV/cm. Between 2 and 3 MV/cm the tunneling component of the second step becomes superior, and multiphonon-assisted tunneling approaches resonant tunneling. Direct tunneling, which turns into Fowler–Nordheim tunneling for $F>3$ MV/cm, remains unimportant up to 5 MV/cm.

Multiplying the trap density with the volume of the bottom and the top oxide results in 210 traps per EPROM cell in an energy interval of about 0.2 eV around the trap level that is giving the largest current density. Assuming this to be an average value over the entire energy gap of SiO_2 a total of 9500 traps per EPROM cell can be estimated. Only a fraction of these traps will be actually occupied. Thus, the total trapped charge is comparable to the 4000 electrons that have left the floating gate during the initial charge loss phase. Quantitative conclusions cannot be drawn about the trapped charge in deep traps from this experiment.

VII. CONCLUSIONS

In this article the nondefective charge loss mechanism was identified and modeled by a trap-assisted tunneling process considering the coupling of trap level to oxide phonons. From a program-bake-erase-bake cycle it was found that long term charge loss is not due to mobile ions. An eventually positive ion accumulation in the floating gate region during the first bake resulting in a charge gain during the second bake has not been observed. A charge loss experiment with two lots of EPROMs—one with ONO, the other with oxide for the interpoly dielectric—verified that long-term charge loss is limited by the oxide conductivity. The oxide field of both lots was the same for a given floating gate-to-source voltage. Since there was no significant difference in the charge loss characteristic of the two device types, it can be concluded that the nitride has no influence on the long-term charge loss. Therefore, it can be summarized that the long-term charge loss of the examined devices is limited by electronic leakage through the bottom oxide of the ONO interpoly dielectric. Electrons that have passed the bottom oxide can pass the top oxide likewise. In the nitride the conductivity is larger than in the oxide. Hole injection from the control gate into the nitride is prevented by the 70-Å-thick top oxide.

Since there is not much charge accumulation in the isolation surrounding the floating gate, a simple model can be derived to calculate the floating gate charge and voltage from the measured drain-source current. The normal one-dimensional long-channel transistor model with capacitive coupling from the control gate and drain to the floating gate was used. It was found that the leakage current is too large to be explained by pure electrode-limited charge transport. The bulk-limited currents (hopping conduction and space-charge-limited current) do not have an exponential field dependence, while the Poole–Frenkel conduction is not appropriate for thin dielectrics. Field ionization of trapped electrons is not temperature dependent. Field gain on asperities and along edges can not increase the charge loss current to the required range.

A long-term charge loss mechanism for EPROM cells with ONO interpoly dielectric has been presented wherein electrons from the floating gate tunnel to oxide traps and are then emitted to the nitride. The coupling of the trap level to oxide phonons results in virtual energy levels in the oxide which allows for transition paths with an increased tunnel probability. As a consequence, the second step is thermally activated and the total transition rate is enhanced. A simple two-step tunneling mechanism could not explain the strong temperature dependence of the measured charge loss data. The same mechanism allows the electrons to escape through the top oxide into the polysilicon of the control gate. Since the trapped charge in the interpoly dielectric is small, this step is not rate limiting. A quasicontinuous energetic distribution of trap energies was assumed. Trap levels between 1.5 and 1.7 eV measured from the conduction-band edge give the dominant contribution to the current density at high temperatures and low electric fields. Furthermore, only traps near the floating gate-bottom oxide interface are active in the steady-state trapping and detrapping process under these conditions. Consequently, the mechanism is only weakly dependent on the oxide thickness. This may explain why the two examined lots have the same charge loss characteristic, although they are of different absolute oxide thicknesses in the interpoly dielectric. A trap density of $N_t=6.5\times10^{15}$ cm^{-3} and a lattice relaxation energy ϵ_R=0.36 eV result in excellent agreement with the measured field and temperature dependence of the steady-state leakage current. A variation of ϵ_R in the range from 0.2 to 0.6 eV has only little influence on the fit, which shows the robustness of the model. The estimated trap density results in 210 traps per EPROM cell in an energy interval of about 0.2 eV around the trap level, which is giving the largest current density.

Previous charge loss models postulated a reduced energy barrier for the polycrystalline Si-SiO_2 interface to explain the measured temperature dependence, although photocurrent measurements proved that polycrystalline Si-SiO_2-Al and crystalline Si-SiO_2-Al MOS structures have identical interface barrier heights. The present model, based on multiphonon-assisted tunneling, yields an activation energy of 1.2 eV, even using the measured barrier height of 3.1 eV.

ACKNOWLEDGMENTS

The authors wish to thank the SGS-Thomson R&D Technology Group for the access to test modes and process-specific information. Thanks are also to Professor A. Birolini and Professor W. Fichtner for supporting the work, as well as for technical discussions.

APPENDIX

The transmission probability $T_{\rm res}$ is determined by the component M_{22} of the transfer matrix M (see e.g., Ref. 65),

$$T_{\rm res}(E,x)=\frac{m_{c,r}(E)}{m_{c,l}(E)}\frac{k_l(E)}{k_r(E)}\frac{1}{|M_{22}(E,x)|^2}, \tag{A1}$$

where M is composed as

$$M(E,x)=M_r(E)[M_t(E,x)][M_l(E)], \tag{A2}$$

with the product matrix

$$M_t(E,x)=M_{t,r}(E,x)[M_{t,l}(E,x)]$$

containing the matching conditions at the trap potential walls at $x\pm r_t$, and the matrices M_l and M_r describing the matching at the gate-oxide and oxide-substrate interfaces, respectively. The component M_{22} can easily be evaluated from Eq. (A2),

$$M_{22}=m_{12}^l m_{21}^r\left(m_{11}^t+\frac{m_{22}^r}{m_{21}^r}m_{21}^t\right)+m_{22}^l m_{22}^r\left(m_{22}^t+\frac{m_{21}^r}{m_{22}^r}m_{12}^t\right), \tag{A3}$$

where the matrix elements follow from

$$M_r=\begin{pmatrix} -\frac{1}{2}e^{-ik_r d}\left(\mathrm{Ai}(\xi_r)+\frac{i}{k_r r_o}\mathrm{Ai}'(\xi_r)\right) & \frac{1}{2}e^{-ik_r d}\left(\mathrm{Bi}(\xi_r)+\frac{i}{k_r r_o}\mathrm{Bi}'(\xi_r)\right) \\ -\frac{1}{2}e^{-ik_r d}\left(\mathrm{Ai}(\xi_r)-\frac{i}{k_r r_o}\mathrm{Ai}'(\xi_r)\right) & \frac{1}{2}e^{-ik_r d}\left(\mathrm{Bi}(\xi_r)-\frac{i}{k_r r_o}\mathrm{Bi}'(\xi_r)\right)\end{pmatrix}, \tag{A4}$$

$$M_l=\begin{pmatrix} \pi[ik_l r_o\,\mathrm{Bi}(\xi_l)+\mathrm{Bi}'(\xi_l)] & -\pi[ik_l r_o\,\mathrm{Bi}(\xi_l)-\mathrm{Bi}'(\xi_l)] \\ \pi[ik_l r_o\,\mathrm{Ai}(\xi_l)+\mathrm{Ai}'(\xi_l)] & -\pi[ik_l r_o\,\mathrm{Ai}(\xi_l)-\mathrm{Ai}'(\xi_l)]\end{pmatrix}, \tag{A5}$$

$$M_{t,s}=\begin{pmatrix} \pi[\mathrm{Ai}'(\xi_{t,s})\mathrm{Bi}(\xi_{t,s+})-\mathrm{Ai}(\xi_{t,s})\mathrm{Bi}'(\xi_{t,s+})] & \pi[\mathrm{Bi}'(\xi_{t,s})\mathrm{Bi}(\xi_{t,s+})-\mathrm{Bi}(\xi_{t,s})\mathrm{Bi}'(\xi_{t,s+})] \\ -\pi[\mathrm{Ai}'(\xi_{t,s})\mathrm{Ai}(\xi_{t,s+})-\mathrm{Ai}(\xi_{t,s})\mathrm{Ai}'(\xi_{t,s+})] & -\pi[\mathrm{Bi}'(\xi_{t,s})\mathrm{Ai}(\xi_{t,s+})-\mathrm{Bi}(\xi_{t,s})\mathrm{Ai}'(\xi_{t,s+})]\end{pmatrix}. \tag{A6}$$

The index s is either l or r, and $r_o=\hbar\Theta_o/(qF)$. The arguments of the Airy functions have the following explicit form:

$$\begin{aligned}
\mathrm{Ai}(\xi_l)&=\mathrm{Ai}\left(\frac{\varphi(E,0)}{\hbar\Theta_o}\right),\\
\mathrm{Ai}(\xi_r)&=\mathrm{Ai}\left(\frac{\varphi(E,d)}{\hbar\Theta_o}\right),\\
\mathrm{Ai}(\xi_{t,l})&=\mathrm{Ai}\left(\frac{\varphi(E,x-r_t)}{\hbar\Theta_o}\right),\\
\mathrm{Ai}(\xi_{t,r+})&=\mathrm{Ai}\left(\frac{\varphi(E,x+r_t)}{\hbar\Theta_o}\right),\\
\mathrm{Ai}(\xi_{t,l+})&=\mathrm{Ai}\left(\frac{\varphi(E+V_t,x-r_t)}{\hbar\Theta_o}\right),\\
\mathrm{Ai}(\xi_{t,r})&=\mathrm{Ai}\left(\frac{\varphi(E+V_t,x+r_t)}{\hbar\Theta_o}\right).
\end{aligned} \tag{A7}$$

The arguments of the functions Ai′, Bi, and Bi′ were labeled in the same way. The quantity $\varphi(E,x)$ is given by Eq. (31) and V_t denotes the depth of the trap potential measured from the oxide conduction-band edge. For the trap levels and field strengths considered here we have $\Phi_t\gg\hbar\Theta_o$ and also $|\Phi_t-V_t|\gg\hbar\Theta_o$. Therefore, it follows for the arguments of the Airy functions at the resonance level $E_t(x)=\Phi_t-qFx-\Phi_t$ that

$$\begin{aligned}
&\varphi[E_t(x),0]\gg\hbar\Theta_o,\quad \varphi[E_t(x),x\pm r_t]\gg\hbar\Theta_o,\\
&\varphi[E_t(x)+V_t,x\pm r_t)\ll-\hbar\Theta_o.
\end{aligned} \tag{A8}$$

That allows to use the respective asymptotic forms[75] at the gate-oxide interface and at the trap potential walls. Only at the oxide-substrate interface do the full Airy functions have to be applied. For the matrix elements of M_l we obtain

$$m_{12}^l=-\sqrt{\pi}\left(\frac{ik_l r_o}{|\xi_l|^{1/4}}-\left|\xi_l\right|^{1/4}\right)e^{S_l}, \tag{A9}$$

$$m_{22}^l=-\frac{\sqrt{\pi}}{2}\left(\frac{ik_l r_o}{|\xi_l|^{1/4}}+\left|\xi_l\right|^{1/4}\right)e^{-S_l}, \tag{A10}$$

where S_ν denotes the action $S_\nu=\frac{2}{3}|\xi_\nu|^{3/2}$. The position and width of the resonance are determined by M_t, the elements of which read

$$m^t_{\substack{11\\22}}=\frac{1}{2}\,e^{\pm(S_{t,r+}-S_{t,l})}\Bigg[\cos(S_{t,r}-S_{t,l+})\times\left(\left|\frac{\xi_{t,l}\xi_{t,r}}{\xi_{t,l+}\xi_{t,r+}}\right|^{1/4}+\left|\frac{\xi_{t,l+}\xi_{t,r+}}{\xi_{t,l}\xi_{t,r}}\right|^{1/4}\right)\pm\sin(S_{t,r}-S_{t,l+})\times\left(\left|\frac{\xi_{t,l}\xi_{t,r+}}{\xi_{t,l+}\xi_{t,r}}\right|^{1/4}-\left|\frac{\xi_{t,l+}\xi_{t,r}}{\xi_{t,l}\xi_{t,r+}}\right|^{1/4}\right)\Bigg],\tag{A11}$$

$$m^t_{\substack{12\\21}}=\begin{Bmatrix}-1\\-\frac{1}{4}\end{Bmatrix}e^{\pm(S_{t,r+}+S_{t,l})}\Bigg[\cos(S_{t,r}-S_{t,l+})\times\left(\left|\frac{\xi_{t,l}\xi_{t,r}}{\xi_{t,l+}\xi_{t,r+}}\right|^{1/4}-\left|\frac{\xi_{t,l+}\xi_{t,r+}}{\xi_{t,l}\xi_{t,r}}\right|^{1/4}\right)\pm\sin(S_{t,r}-S_{t,l+})\times\left(\left|\frac{\xi_{t,l}\xi_{t,r+}}{\xi_{t,l+}\xi_{t,r}}\right|^{1/4}+\left|\frac{\xi_{t,l+}\xi_{t,r}}{\xi_{t,l}\xi_{t,r+}}\right|^{1/4}\right)\Bigg].\tag{A12}$$

The actions S in Eqs. (A11) and (A12) can be developed with respect to the small potential drop qFr_t across the trap radius

$$S_{t,r}-S_{t,l+}\approx 2\,\frac{r_t}{r_o}\left(\frac{-\varphi(E+V_t,x)}{\hbar\Theta_o}\right)^{1/2}=2\,\frac{r_t}{r_o}\sqrt{-\xi_{t,\mathrm{in}}},\tag{A13}$$

with $\xi_{t,\mathrm{in}}=(\xi_{t,l+}+\xi_{t,r})/2$. Accordingly,

$$S_{t,r+}-S_{t,l}\approx -2\,\frac{r_t}{r_o}\left(\frac{\varphi(E,x)}{\hbar\Theta_o}\right)^{1/2}=-2\,\frac{r_t}{r_o}\sqrt{\xi_{t,\mathrm{out}}},\tag{A14}$$

with $\xi_{t,\mathrm{out}}=(\xi_{t,l}+\xi_{t,r+})/2$. Developing the algebraic factors in Eqs. (A11) and (A12) as well and neglecting the quadratic Stark effect gives

$$\left|\frac{\xi_{t,l}\xi_{t,r}}{\xi_{t,l+}\xi_{t,r+}}\right|^{1/4}\approx 1+\frac{1}{2\xi_{t,\mathrm{out}}}\,\frac{r_t}{r_o},\tag{A15}$$

$$\left|\frac{\xi_{t,l+}\xi_{t,r+}}{\xi_{t,l}\xi_{t,r}}\right|^{1/4}\approx 1-\frac{1}{2\xi_{t,\mathrm{out}}}\,\frac{r_t}{r_o},\tag{A16}$$

$$\left|\frac{\xi_{t,l}\xi_{t,r+}}{\xi_{t,l+}\xi_{t,r}}\right|^{1/4}\approx\left(\frac{\xi_{t,\mathrm{out}}}{-\xi_{t,\mathrm{in}}}\right)^{1/2}.\tag{A17}$$

The diagonal elements of M_t become

$$m^t_{\substack{11\\22}}=e^{\mp 2\sqrt{\xi_{t,\mathrm{out}}}(r_t/r_o)}\cos^2\alpha\left[\left(\frac{\xi_{t,\mathrm{out}}}{-\xi_{t,\mathrm{in}}}\right)^{1/2}\mp\tan\alpha\right]\times\left[\left(\frac{-\xi_{t,\mathrm{in}}}{\xi_{t,\mathrm{out}}}\right)^{1/2}\pm\tan\alpha\right),\tag{A18}$$

with $\alpha=\sqrt{-\xi_{t,\mathrm{in}}}\,r_t/r_o$. The bound state of the square-well potential is reproduced by the resonance condition

$$\left(\frac{\xi_{t,\mathrm{out}}}{-\xi_{t,\mathrm{in}_k}}\right)^{1/2}=\tan\alpha,\tag{A19}$$

and the symmetry relation $m^t_{22}(-F)=m^t_{11}(F)$ ensures that the same level occurs if the polarity of the field is changed. The off-diagonal elements that determine the damping of the transmission probability turn into

$$m^t_{\substack{12\\21}}=\begin{Bmatrix}2\\1\\\frac{1}{2}\end{Bmatrix}e^{\pm 2S_t}\cos^2\alpha\left(\frac{1}{2\xi_{t,\mathrm{out}}}\,\frac{r_t}{r_o}\,\frac{(\xi_{t,\mathrm{out}}+\xi_{t,\mathrm{in}})}{\xi_{t,\mathrm{in}}}\pm\frac{(\xi_{t,\mathrm{out}}-\xi_{t,\mathrm{in}})}{\xi_{t,\mathrm{in}}}\right),\tag{A20}$$

which holds in the vicinity of the resonance energy. Inserting Eqs. (A4), (A9), and (A10) into Eq. (A3) we obtain

$$|M_{22}|^2=\frac{\pi}{4}\,e^{2S_l}[m^{t2}_{11}(r^{-2}_{\mathrm{Ai}}+j^{+2}_{\mathrm{Ai}})+m^{t2}_{21}(r^{-2}_{\mathrm{Bi}}+j^{+2}_{\mathrm{Bi}})-2m^t_{11}m^t_{21}(r^-_{\mathrm{Ai}}r^-_{\mathrm{Bi}}+j^+_{\mathrm{Ai}}j^+_{\mathrm{Bi}})]+\frac{\pi}{16}\,e^{-2S_l}[m^{t2}_{22}(r^{+2}_{\mathrm{Bi}}+j^{-2}_{\mathrm{Bi}})+m^{t2}_{12}(r^{+2}_{\mathrm{Ai}}+j^{-2}_{\mathrm{Ai}})-2m^t_{22}m^t_{12}(r^+_{\mathrm{Bi}}r^+_{\mathrm{Ai}}+j^-_{\mathrm{Bi}}j^-_{\mathrm{Ai}})]+\frac{\pi}{4}[(r^-_{\mathrm{Ai}}m^t_{11}-r^-_{\mathrm{Bi}}m^t_{21})(r^+_{\mathrm{Bi}}m^t_{22}-r^+_{\mathrm{Ai}}m^t_{12})+(j^+_{\mathrm{Bi}}m^t_{21}-j^+_{\mathrm{Ai}}m^t_{11})(j^-_{\mathrm{Bi}}m^t_{22}-j^-_{\mathrm{Ai}}m^t_{12})],\tag{A21}$$

with the abbreviations

$$r^{\pm}_{\mathrm{Ai}}=\frac{k_l}{k_r|\xi_l|^{1/4}}\,Ai'(\xi_r)\pm|\xi_l|^{1/4}\,\mathrm{Ai}(\xi_r),\tag{A22}$$

$$j^{\pm}_{\mathrm{Ai}}=\frac{k_l r_o}{|\xi_l|^{1/4}}\,Ai(\xi_r)\pm\frac{|\xi_l|^{1/4}}{k_r r_o}\,\mathrm{Ai}'(\xi_r),\tag{A23}$$

and the corresponding definitions for $r^{\pm}_{\mathrm{Bi}}$ and $j^{\pm}_{\mathrm{Bi}}$. For not too small field strengths, i.e., as long as $S_l\gg S_r$ can still be assumed, various terms in Eq. (A22) are negligible. The remaining are

$$|M_{22}|^2\approx\frac{\pi}{4}\,e^{2S_l}(r^{-2}_{\mathrm{Ai}}+j^{+2}_{\mathrm{Ai}})(m^t_{11})^2+\frac{\pi}{4}\,e^{2S_l}(r^{-2}_{\mathrm{Bi}}+j^{+2}_{\mathrm{Bi}})(m^t_{21})^2+\frac{\pi}{16}\,e^{-2S_l}(r^{+2}_{\mathrm{Ai}}+j^{-2}_{\mathrm{Ai}})(m^t_{12})^2+\frac{\pi}{4}\,(j^+_{\mathrm{Ai}}j^-_{\mathrm{Ai}}-r^+_{\mathrm{Ai}}r^-_{\mathrm{Ai}})m^t_{11}m^t_{12}-\frac{\pi}{2}\,e^{2S_l}(r^-_{\mathrm{Ai}}r^-_{\mathrm{Bi}}+j^+_{\mathrm{Ai}}j^+_{\mathrm{Bi}})m^t_{11}m^t_{21}.\tag{A24}$$

The last but one term accounts for the shift of the resonance level, if the trap is located very close to the gate-oxide interface. The last term describes the respective shift for a trap situated very close to the oxide-substrate interface. These shifts are due

to the delocalization of the wave function as one potential barrier becomes very thin. At the same time, the damping term for those traps strongly increases [second line in Eq. (A24)], which reduces the total transmission probability. Therefore, we skip the last two terms of Eq. (A24).

In order to obtain a Lorentzian for $T_{\rm res}$ we linearize m_{11}^t in the energy and evaluate m_{12}^t and m_{21}^t at the resonance level. That gives

$$m_{11}^t(E)\rightarrow\frac{1}{2}\,e^{-2\sqrt{\Phi_t/\hbar\Theta_o}(r_t/r_o)}\,\frac{V_t}{\Phi_t(V_t-\Phi_t)}\left[1+\left(\frac{\Phi_t}{\hbar\Theta_o}\right)^{1/2}\frac{r_t}{r_o}\right][E_t(x)-E], \tag{A25}$$

$$m_{\substack{12\\21}}^t\rightarrow\left\{\begin{matrix}1\\ \frac{1}{4}\end{matrix}\right\}e^{\pm 2S_t[E_t(x)]}\left(\frac{\hbar\Theta_o}{\Phi_t}\,\frac{r_t}{r_o}\,\frac{(V_t-2\Phi_t)}{V_t}\mp 2\right). \tag{A26}$$

Inserting Eqs. (A25) and (A26) into Eq. (A24), $1/|M_{22}|^2$ takes Lorentzian form and can be transformed into a delta function, since the resonance is extremely sharp (see Fig. 15),

$$\frac{1}{|M_{22}|^2}=\frac{1}{A[E_t(x)-E]^2+B}\rightarrow\frac{\pi}{\sqrt{AB}}\,\delta[E_t(x)-E]. \tag{A27}$$

The prefactor follows from comparison with Eq. (A24) and inserting Eqs. (A22) and (A23),

$$AB=\left(\frac{\pi}{16}\right)^2 e^{-4\sqrt{\Phi_t/\hbar\Theta_o}(r_t/r_o)}\,\frac{V_t^2}{\Phi_t^2(V_t-\Phi_t)^2|\xi_l|}\left(1+\frac{k_l^2r_o^2}{|\xi_l|}\right)^2\left[1+\left(\frac{\Phi_t}{\hbar\Theta_o}\right)^{1/2}\frac{r_t}{r_o}\right]^2\left[\frac{1}{4}\,e^{4(S_l-S_t)}\left(\mathrm{Ai}^2(\xi_r)+\frac{\mathrm{Ai}'^2(\xi_r)}{k_r^2r_o^2}\right)\right.$$
$$\left.\times\left(\mathrm{Bi}^2(\xi_r)+\frac{\mathrm{Bi}'^2(\xi_r)}{k_r^2r_o^2}\right)\left(\frac{\hbar\Theta_o}{\Phi_t}\,\frac{r_t}{r_o}\,\frac{(V_t-2\Phi_t)}{V_t}+2\right)^2+e^{4S_t}\left(\mathrm{Ai}^2(\xi_r)+\frac{\mathrm{Ai}'^2(\xi_r)}{k_r^2r_o^2}\right)^2\left(\frac{\hbar\Theta_0}{\Phi_t}\,\frac{r_t}{r_o}\,\frac{(V_t-2\Phi_t)}{V_t}-2\right)^2\right]. \tag{A28}$$

We now turn the square-well potential into a delta potential $V(x)=-\sqrt{2\hbar^2/m_{c,o}(V_t-\Phi_t)}\,\delta(x-x_t)$. In this limit the transition rate does not depend on the potential parameters and can be directly compared to the capture/emission process, where a 3D delta potential was used. With $V_t\rightarrow\infty$ and $r_t\rightarrow 0$ in Eq. (A28) the transmission probability for resonant tunneling takes the form

$$T_{\rm res}(E,x)=\frac{16\Phi_t\,\dfrac{m_{c,r}k_l}{m_{c,l}k_r}\left(1+\dfrac{m_{c,l}}{m_{c,o}}\,\dfrac{E_l}{\Phi_t+qFx}\right)^{-1}\delta[E_t(x)-E]}{\sqrt{\xi_l}\left[T_l^{-2}\left(\mathrm{Ai}_r^2+\dfrac{\hbar\Theta_o}{E_r}\,\mathrm{Ai}_r'^2\right)\left(\mathrm{Bi}_r^2+\dfrac{\hbar\Theta_o}{E_r}\,\mathrm{Bi}_r'^2\right)+T_r^{-2}4e^{4S_r}\left(\mathrm{Ai}_r^2+\dfrac{\hbar\Theta_o}{E_r}\,\mathrm{Ai}_r'^2\right)^2\right]^{1/2}}, \tag{A29}$$

where $\mathrm{Ai}_r=\mathrm{Ai}(\xi_r)$, etc., $E_l\equiv E_t(x)-E_{c,l}$ and $E_r\equiv E_t(x)-E_{c,r}$. Furthermore, we have introduced the WKB probabilities T_l and T_r [Eqs. (29) and (30)] for tunneling into and out of the trap well, respectively. If the field strength is such that $\xi_r\gg 1$ holds, we can apply the WKB approximation also at the oxide-substrate interface,

$$\mathrm{Ai}_r^2+\frac{\hbar\Theta_o}{E_r}\,\mathrm{Ai}_r'^2\rightarrow\frac{1}{4\pi}\,\frac{1}{\sqrt{\xi_r}}\left(1+\frac{m_{c,o}}{m_{c,r}}\,\frac{[-qF(d-x)+\Phi_t]}{E_r}\right)e^{-2S_r}, \tag{A30}$$

$$\mathrm{Bi}_r^2+\frac{\hbar\Theta_o}{E_r}\,\mathrm{Bi}_r'^2\rightarrow\frac{1}{\pi}\,\frac{1}{\sqrt{\xi_r}}\left(1+\frac{m_{c,o}}{m_{c,r}}\,\frac{[-qF(d-x)+\Phi_t]}{E_r}\right)e^{2S_r}. \tag{A31}$$

Inserting into Eq. (A29) we obtain the final form of $T_{\rm res}$,

$$T_{\rm res}(E,x)=32\pi\,\frac{\Phi_t}{\Phi_l\Phi_r}\left(\frac{m_{c,r}}{m_{c,l}}\right)^{1/2}\sqrt{E_lE_r}\times\sqrt{(\Phi_l-E_l)(\Phi_r-E_r)}\,\frac{T_lT_r}{\sqrt{T_l^2+T_r^2}}\,\delta[E_t(x)-E]. \tag{A32}$$

[1] M. Lenzlinger and E. H. Snow, J. Appl. Phys. **40**, 278 (1969).
[2] S. Nagano, M. Tsukiji, K. Ando, E. Hasegawa, and A. Ishitani, J. Appl. Phys. **75**, 3530 (1994).
[3] J. G. Simmons, Phys. Rev. Lett. **15**, 967 (1965).
[4] N. R. Mielke, in International Reliability Physics Symposium, 1983, pp. 106–113.
[5] R. E. Shiner, J. M. Caywood, and B. L. Euzent, in International Reliability Physics Symposium, 1980, pp. 238–243.
[6] G. Crisenza, G. Ghidini, S. Manzini, A. Modelli, and M. Tosi, in IEDM Technical Digest, 1990, pp. 107–110.
[7] E. Sakagami, N. Arai, H. Tsunoda, H. Egawa, Y. Yamaguchi, E. Kamiya, M. Takebuchi, K. Yamada, K. Yoshikawa, and S. Mori, in International Reliability Physics Symposium, 1994, pp. 359–367.
[8] H. Nozawa and S. Kohoyama, in Jpn. J. Appl. Phys. **21**, 111 (1982).
[9] H. Nozawa, Y. Niitsu, N. Mątsukawa, J. Matsunaga, and S. Kohoyama, IEEE Trans. Electron Devices **ED-31**, 1413 (1984).
[10] D. J. DiMaria and D. R. Kerr, Appl. Phys. Lett. **27**, 505 (1975).

[11] H. H. Heimeier, Ph.D. thesis, Eidgenössische Technische Hochschule Zürich, 1992.
[12] K. J. Wu, C. S. Pan, J. J. Shaw, P. P. Freiberger, and G. Sery, in International Reliability Physics Symposium, 1990, pp. 145–149.
[13] C. S. Pan, K. J. Wu, P. P. Freiberger, and A. Chatterjee, IEEE Trans. Electron Devices **ED-37**, 1439 (1990).
[14] C. S. Pan, K. Wu, D. Chin, G. Sery, and J. Kiely, IEEE Electron Device Lett. **EDL-12**, 506 (1991).
[15] C. S. Pan, K. Wu, and G. Sery, IEEE Electron Device Lett. **EDL-12**, 51 (1991).
[16] A. Roy and M. H. White, Solid-State Electron. **34**, 1083 (1991).
[17] Y. Hu and M. H. White, Solid-State Electron. **36**, 1402 (1993).
[18] M. Lanzoni, R. Menozzi, C. Riva, P. Olivo, and B. Ricco, QRE Int. **7**, 293 (1991).
[19] R. H. Fowler and L. Nordheim, Proc. R. Soc. London, Ser. A **119**, 173 (1928).
[20] P. Hesto, in *Instabilities in Silicon Devices*, edited by G. Barbottin and A. Vapaille (Elsevier, North-Holland, 1986), Chap. 5, pp. 265–314.
[21] Z. A. Weinberg and R. A. Pollak, Appl. Phys. Lett. **27**, 254 (1975).
[22] Z. A. Weinberg, Appl. Phys. Lett. **29**, 254 (1976).
[23] P. C. Arnett and Z. A. Weinberg, IEEE Trans. Electron Devices **ED-25**, 1014 (1978).
[24] F. T. Liou and S. O. Chen, IEEE Trans. Electron Devices **ED-31**, 1736 (1984).
[25] S. M. Sze, J. Appl. Phys. **38**, 2951 (1967).
[26] D. Frohman-Bentchkowsky and M. Lenzlinger, J. Appl. Phys. **40**, 3307 (1969).
[27] I. K. Lundström and C. M. Svensson, IEEE Trans. Electron Devices **ED-19**, 826 (1972).
[28] C. M. Svensson and I. K. Lundström, Electron. Lett. **6**, 645 (1970).
[29] C. M. Svensson and I. K. Lundström, J. Appl. Phys. **44**, 4657 (1973).
[30] C. M. Svensson, J. Appl. Phys. **48**, 329 (1977).
[31] K. K. Young, C. Hu, and W. G. Oldham, IEEE Electron Device Lett. **EDL-9**, 616 (1988).
[32] M. Aminzadeh, S. Nozaki, and R. V. Giridhar, IEEE Trans. Electron Devices **ED-35**, 459 (1988).
[33] S. Manzini and F. Volont, J. Appl. Phys. **58**, 4300 (1985).
[34] S. Nozaki and R. V. Giridhar, IEEE Electron Device Lett. **EDL-7**, 486 (1986).
[35] T. Watanabe, A. Menjoh, M. Ishikawa, and J. Kumagai, in IEDM Technical Digest, 1984, pp. 173–176.
[36] T. Watanabe, A. Menjoh, T. Mochizuki, S. Shinozaki, and O. Ozawa, in International Reliability Physics Symposium, 1985, p. 1823.
[37] S. Mori, E. Sakagami, H. Araki, Y. Kaneko, K. Narita, N. Arai, Y. Ohshima, and K. Yoshikawa, in International Reliability Physics Symposium, 1990, pp. 132–144.
[38] S. Mori, E. Sakagami, H. Araki, Y. Kaneko, K. Narita, Y. Ohshima, N. Arai, and K. Yoshikawa, IEEE Trans. Electron Devices **ED-38**, 386 (1991).
[39] S. Mori, E. Sakagami, Y. Kaneko, Y. Ohshima, N. Arai, and K. Yoshikawa, in International Reliability Physics Symposium, 1991, pp. 175–182.
[40] S. Mori, E. Sakagami, Y. Kaneko, Y. Ohshima, N. Arai, and K. Yoshikawa, IEEE Trans. Electron Devices **ED-39**, 283 (1992).
[41] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley, New York, 1981).
[42] G. H. Parker and C. A. Mead, Appl. Phys. Lett. **14**, 21 (1969).
[43] J. W. Gadzuk, J. Appl. Phys. **41**, 286 (1970).
[44] G. Dorda and M. Pulver, Phys. Status Solidi A **1**, 71 (1970).
[45] E. Suzuki, D. K. Schroder, and Y. Hayashi, J. Appl. Phys. **60**, 3616 (1986).
[46] S. Fleischer, P. T. Lai, and Y. C. Cheng, J. Appl. Phys. **72**, 5711 (1992).
[47] N. Yasuda, N. Patel, and A. Toriumi, in Extended Abstracts on Solid State Devices and Materials SSDM, Chiba, Japan, 1993, pp. 847–849.
[48] R. A. Williams and M. M. Beguwala, IEEE Trans. Electron Devices **ED-25**, 1019 (1978).
[49] J. J. O'Dwyer, J. Appl. Phys. **37**, 599 (1966).
[50] G. Hemink, T. Endoh, and R. Shirota, in Extended Abstracts on Solid State Devices and Materials SSDM, Chiba, Japan, 1993, pp. 464–466.
[51] R. M. Alexander, in International Reliability Physics Symposium, 1978, pp. 229–232.
[52] A. Spitzer and R. Baunach, in *Applied Surface Science* (Elsevier, Amsterdam, 1989), Vol. 39, pp. 192–199.
[53] S. T. Wang, IEEE Trans. Electron Devices **ED-26**, 1292 (1979).
[54] L. C. Liong and P. C. Liu, IEEE Trans. Electron Devices **ED-40**, 146 (1993).
[55] A. Kriveris, S. Kudzmauskas, and P. Pipinys, Phys. Status Solidi A **37**, 321 (1976).
[56] L. V. Keldysh, Sov. Phys. JETP **6**, 763 (1958).
[57] L. V. Keldysh, Sov. Phys. JETP **7**, 665 (1958).
[58] F. I. Dalidchik, Eks. Theo. Fiz. **74**, 472 (1978).
[59] A. Schenk, J. Appl. Phys. **71**, 3339 (1992).
[60] A. Schenk, Solid-State Electron. **35**, 1585 (1992).
[61] G. Lucovski, Solid State Commun. **3**, 299 (1965).
[62] A. Schenk, Ph.D. thesis, Humboldt University, Berlin, 1986.
[63] K. Huang and A. Rhys, Proc. R. Soc. London Ser. A **204**, (1950).
[64] D. Hsu, M. Dsu, C. Tan, and Y. Y. Wang, J. Appl. Phys. **72**, 4972 (1992).
[65] Y. Ando and T. Itoh, J. Appl. Phys. **61**, 1497 (1987).
[66] W. W. Lui and M. Fukuma, J. Appl. Phys. **60**, 1555 (1986).
[67] C. B. Duke and M. E. Alferieff, J. Chem. Phys. **46**, 923 (1967).
[68] B. Ricco, M. Ya. Azbel, and M. H. Brodsky, Phys. Rev. Lett. **51**, 1795 (1983).
[69] F. I. Dalidchik, Sov. Phys. Solid State **25**, 1289 (1983).
[70] B. Gu, M. Mangiantini, and C. Coluzza, J. Appl. Phys. **64**, 6867 (1988).
[71] J. C. Penley, Phys. Rev. **128**, 596 (1962).
[72] Y. Fu and M. Willander, J. Appl. Phys. **73**, 1848 (1993).
[73] B. Balland, in *Instabilities in Silicon Devices*, edited by G. Barbottin and A. Vapaille (Elsevier, Amsterdam, 1986), Chap. 2, pp. 101–153.
[74] N. Hwang, B. S. S. Or, and L. Forbes, IEEE Trans. Electron Devices **ED-40**, 1100 (1993).
[75] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1972).