Analytical model of the effects of a nonuniform distribution of stored charge on the electrical characteristics of discrete-trap nonvolatile memories

Luca Perniola
Dipartimento di Ingegneria dell’Informazione: Elettronica, Informatica, Telecomunicazioni, Università di Pisa and IMEP-ENSERG, Grenoble

Sandrine Bernardini
Provence Materials and Microelectronics Laboratory, Polytechnique–IMT, Marseille

Giuseppe Iannaccone
Dipartimento di Ingegneria dell’Informazione: Elettronica, Informatica, Telecomunicazioni, Università di Pisa

Pascal Masson
Provence Materials and Microelectronics Laboratory, Polytechnique–IMT, Marseille

Barbara De Salvo
Commissariat à l’Energie Atomique–Laboratory of Electronics, Technology, and Instrumentation, Grenoble

Gérard Ghibaudo
Institut de Microélectronique Electromagnétisme et Photonique–Centre National de la Recherche Scientifique/Institut National Polytechnique de Grenoble

Cosimo Gerardi
Central Research and Development, STMicroelectronics, Catania

Analytical Model of the Effects of a Nonuniform Distribution of Stored Charge on the Electrical Characteristics of Discrete-Trap Nonvolatile Memories

Luca Perniola, Sandrine Bernardini, Giuseppe Iannaccone, Pascal Masson, Barbara De Salvo, Gérard Ghibaudo, and Cosimo Gerardi

Abstract—We propose an analytical model of the effects of a nonuniform distribution of trapped charge on the electrical characteristics and on the perspectives of 2-bit operation of discrete-trap memories. To keep the model tractable, we consider an idealized nonuniform distribution, represented by a step function, so that the concentration of trapped charge can assume only two possible values in two different regions. Notwithstanding the simplicity of our assumptions, which limits the range of validity of our model to the subthreshold and weak inversion regions of the I–V characteristics, we can investigate a series of important aspects for 2-bit storage of nonvolatile memories. Our model is then validated through comparison with detailed numerical simulations performed with a commercial technology computer-aided design tool, and with the experimental electrical characteristics of nanocrystal flash memories under different bias conditions. Finally, we provide a method, based on our model, to extract an “effective” distribution of trapped charge, in which all charge is uniformly distributed in a localized region close to the drain.

Index Terms—Device modeling, discrete-trap memory, nanocrystal memory, nonvolatile memory.

I. INTRODUCTION

C

HANNEL HOT electron (CHE) injection is broadly used as a programming method for nonvolatile discrete-trap memories [1]–[3]. It provides the capability to inject the charge in a small localized region of the trapping medium, which allows to store two bits per cell [4]. The basic principle of 2-bit operation is shared by SONOS [3], NROM [1], and nanocrystal memories [5] since it essentially requires the presence of many discrete trapping sites in the storage medium.

With CHE programming, charges are trapped near one junction (drain or source). The memory is then read in the reverse mode, enhancing the electrostatic effect of such charges on channel conductivity (see Fig. 1).

It has been shown in the literature that the threshold voltage during the forward read \(V_{th,F}^{+} > 0\), \(V_{th,F}^{-}\), is lower than the threshold voltage \(V_{th,R}^{-}\) during the reverse read \(V_{th,R}^{-} < 0\) because of the strong 2-D effects near the charged junction [6]. If the effect of the charge trapped near the drain is suppressed by the high drain-to-source voltage \(V_{ds}\) applied in forward read (which induces a long pinchoff region), the transfer characteristics are very close to those of the fresh cell: we have a low \(V_{th,F}^{-}\). On the other hand, during the reverse read, even a high applied \(V_{ds}\) is not able to affect the potential profile at the source contact, and the flux of electrons is limited by the “bottleneck” near the low-voltage contact, which is controlled by the

Fig. 1. Fresh, forward, and reverse reading conditions after the charge via CHE injection. As it appears from the qualitative plot, the threshold voltage in the reverse reading \(V_{th,R}^{-}\) is higher than the threshold voltage in the forward reading \(V_{th,F}^{+}\). We highlight the fact that, in the literature, the forward-reverse read concentrates on the subthreshold regime of the \(I_{D} - V_{GS}\) behavior. Our approach models this working condition.

Manuscript received June 10, 2004; revised December 10, 2004. This work was supported in part by the European Union under Advanced Memories Based on Discrete-Traps Project IST-2001-34234, by the Network of Excellence Silicon-Based Nanodevices under IST-506844, and by the Italian National Research Council under the FIRB Project Sistemi miniaturizzati per elettronica e fotonica.

L. Perniola is with the Dipartimento di Ingegneria dell’Informazione, Università degli Studi di Pisa, 56122 Pisa, Italy and also with the Institut National Polytechnique de Grenoble—École Nationale Supérieure d’Électronique et de Radioélectricité de Grenoble, 38016 Grenoble, France.
B. De Salvo is with the Commissariat à l’Énergie Atomique–Laboratory of Electronics, Technology, and Instrumentation, 38054 Grenoble, France.
G. Iannaccone is with the Dipartimento di Ingegneria dell’Informazione, Università degli Studi di Pisa, 56122 Pisa, Italy and also with the Istituto di Elettronica e di Ingegneria dell’Informazione e delle Telecomunicazioni–Consiglio Nazionale delle Ricerche, 56122 Pisa, Italy.
S. Bernardini and P. Masson are with the Provence Materials and Microelectronics Laboratory, Polytechnique–IMT, 13451 Marseille, France.
G. Iannaccone is with the Dipartimento di Ingegneria dell’Informazione, Università degli Studi di Pisa, 56122 Pisa, Italy and also with the Istituto di Elettronica e di Ingegneria dell’Informazione e delle Telecomunicazioni–Consiglio Nazionale delle Ricerche, 56122 Pisa, Italy.
G. Ghibaudo is with the Institut de Microélectronique Electromagnétisme et Photonique–Centre National de la Recherche Scientifique/Institut National Polytechnique de Grenoble, 38016 Grenoble, France.
C. Gerardi is with Central Research and Development, STMicroelectronics, 95121 Catania, Italy.

Digital Object Identifier 10.1109/TNANO.2005.847033
trapped charge; a higher gate voltage is needed to have a transport condition comparable to the forward read. In this case, we have $V_{th-R} > V_{th-F}$ [2]–[6].

An additional effect highlighted in the literature is the sub-threshold slope degradation [7], which is a typical issue of programmed memories with localized storage nodes. It is apparent at low $|V_{ds}|$ both in the forward and reverse read or at high $|V_{ds}|$ only in the reverse read [8]. The presence of a very localized charge in the dielectric induces a potential barrier in the channel, which strongly suppresses the conductance. Indeed, the device region below the trapped charge behaves as an equivalent very short transistor with higher threshold voltage and relatively thick oxide, thus, with a very poor coupling between gate and channel. The conductance of the overall channel is essentially limited by such a small transistor whose performance is seriously degraded by SCEs and, therefore, the subthreshold slope is reduced.

Indeed, during the forward read, unless the equivalent charged transistor is pinched off completely by a high $V_{ds}$, the subthreshold slope will be worse than that of the fresh cell. In the reverse read, this detrimental effect of the equivalent charged transistor will remain for any reasonable value of $V_{ds}$.

Previous investigations on the subject are based on experiments [1], [4] or on two-dimensional (2-D) numerical approaches [2], [6], and models based on charge sharing [7] or on one-dimensional (1-D) gradual channel approximations [8].

The 2-D numerical approaches focus on the analysis of the surface potential, $\psi_S$. In our proposed approach, we are able to provide an analytical profile of $\psi_S$, taking into account 2-D effects along the channel, which are essential for device behavior.

In this paper, we propose a simple analytical model of the electrical properties in the subthreshold and weak inversion regimes of discrete-trap memories, where the distribution of charge is nonuniform. We have to make a strong assumption on the profile of the trapped-charge concentration in order to obtain an analytical expression: we assume a step function so that the concentration of trapped charge have only two values in the two regions in which we divide the device. Nevertheless, we will be able to investigate several properties of nonuniformly charged nonvolatile memories.

In addition, with the help of the proposed analytical approach, we are able to extract information on the trapped charge distribution in experimental devices in terms of the “effective” length of the charged region $L_2$ and of the concentration of trapped electrons per unit area $Q$. We need the two data available from electrical characterization: the total threshold voltage shift $\Delta V_{th-tot}$ (i.e., the difference between the threshold voltage in the reverse read $V_{th-R}$ and the threshold voltage of the fresh cell $V_{th-fresh}$) and the reverse-forward threshold voltage shift $\Delta V_{th-F} \equiv V_{th-R} - V_{th-F}$.

Let us stress the fact that we are aware that the distribution of trapped charge is actually a continuous function of the spatial coordinate. Here, we reduce it to our idealized “effective” distribution in the sense that from the point-of-view of the electrical properties of the device, it is equivalent to the actual one.

The viability of the model is tested with 2-D numerical simulations performed with a commercial technology computer-aided design (TCAD) code [9]. Finally, experimental results from nanocrystal memories will be compared with the analytical results under different bias conditions for $V_{ds}$ and for the bulk voltage $V_b$.

II. QUASI-2-D ANALYTICAL MODEL

Our analytical approach describes the behavior of the surface potential $\psi_S$ along the channel of a MOSFET with a nonuniform threshold voltage in subthreshold and weak inversion regions. In our study, the model developed by Liu et al. [10] has been used as a starting point, and it has been significantly extended to address our case. It is of utmost importance to properly describe the 2-D electrostatic effects in the region close to the junctions.

As it has been previously observed [6], any analytical model that is simply based on a 1-D approach (i.e., gradual channel approximation) is not able to reproduce the asymmetric read of the programmed memory in the subthreshold region. Indeed, it does not take into account the different impact of drain induced barrier lowering (DIBL) near the drain junction during forward or reverse read. The 1-D analytical approach shows an asymmetric read only in the linear region of the $I_D - V_{GS}$ characteristics, where the channel is inverted and the drift component of transport is dominant [8].

For the sake of simplicity, we assume that the programmed memory cell of channel length $L$ consists of two parts, as illustrated in Fig. 2: one written cell, equivalent to a MOSFET of channel length $L_2$ with high threshold voltage, in series to a fresh cell of channel length $L_1 = L - L_2$ with low threshold voltage. The difference between the threshold voltages of the two regions $\Delta V_{th}$ reads

$$\Delta V_{th} = q \frac{Q}{C_2} \quad (1)$$

where $q$ is the absolute value of the electron charge and $C_2$ is the control oxide capacitance per unit area, i.e., the capacitance between the charged region and top gate. When the charged length $L_2$ is small (a few tens of nanometers), the fringing field effects
must be taken into account. In this case, we assume that the potential in the small charged region can be considered constant and, therefore, use an approximated formula for the capacitance between a metal electrode with a rectangular cross section and plane. We retain only the terms corresponding to the contribution of the top face and of the two lateral faces of the electrode, and discarding the smaller third term providing the contribution of the bottom face of the electrode [11]

\[ C_2 = \frac{\varepsilon_{si}}{t_2} + \frac{2\varepsilon_{ox}}{L_2} \log \left( 1 + \frac{t_{ch}}{t_2} \right) \quad (2) \]

where \( \varepsilon_{si} \) (\( \varepsilon_{ox} \)) indicates the silicon (oxide) permittivity, \( t_2 \) is the control oxide thickness, and \( t_{ch} \) is the thickness of the negatively charged trapping layer. Let us also underline the fact that the second term of (2) is a strong approximation since the lateral coupling is also affected by the presence of the bottom channel.

In each of the two regions, the Liu model [10] has been properly modified. As it is sketched in Fig. 2, we start from Gauss’s law applied to a transversal slice of thickness \( dy \) along \( y \) (the longitudinal direction) and cross section equal to the average depletion width of the \( i \)th region \( X_{\text{dep}_i} \) \( (i = 1,2) \).

We can write the flux of the electric field through the faces perpendicular to \( y \) in terms of the surface potential \( \Psi_S \) as \( X_{\text{dep}_i}/\eta \cdot d\Psi_S/dy \). The fitting dimensionless parameter \( \eta \) represents the ratio of the \( y \)-component of the electric field at the surface to the average value over the depletion width \( X_{\text{dep}_i} \). In the Liu model, the charge consists only of fixed space charge and the mobile charge is neglected. The main differential equation reads

\[ \varepsilon_{si} \frac{d^2\Psi_S(y)}{dy^2} + \varepsilon_{ox} \frac{V_g - V_b - V_{fb} - \Psi_S(y)}{t_{ox}} = qN_{\text{sub}} X_{\text{dep}_i} \quad (3) \]

where \( V_g \) is the gate voltage, \( t_{ox} \) is the equivalent tunnel oxide thickness, and \( N_{\text{sub}} \) is the p-doping level of the bulk. Referring to Fig. 2, (3) is applied to the uncharged region \( 0 < y < L_1 \), and to the charged region \( 2L_1 < y < L \), and the continuity of \( \Psi_S(y) \) and \( d\Psi_S(y)/dy \) is enforced at \( y = L_1 \). The solution has the following expression:

\[ \Psi_S(y) = \left( \Psi^* - \Psi_{L_1} \right) \frac{\sinh \left( \frac{y}{\lambda_1} \right)}{\sinh \left( \frac{L_1}{\lambda_1} \right)} + \left( \Psi_{L_1} - \Psi_{L_2} \right) \frac{\sinh \left( \frac{L_1 - y}{\lambda_1} \right)}{\sinh \left( \frac{L_1}{\lambda_1} \right)} + \Psi_{L_1}, \]

for \( 0 \leq y \leq L_1 \)

\[ \Psi_S(y) = \left( \Psi^* - \Psi_{L_2} \right) \frac{\sinh \left( \frac{y - L_1}{\lambda_2} \right)}{\sinh \left( \frac{L_2}{\lambda_2} \right)} + \left( \Psi_{L_2} - \Psi_{L_3} \right) \frac{\sinh \left( \frac{L - y}{\lambda_2} \right)}{\sinh \left( \frac{L_2}{\lambda_2} \right)} + \Psi_{L_2}, \]

for \( L_1 < y \leq L \)

where \( \Psi^* = \Psi_S(L_1), \Psi_{L_1} = \Psi_S(0), \) and \( \Psi_{L_2} = \Psi_S(L) \). The potential is set to zero in the bulk. Therefore, if the bulk terminal is biased at voltage \( V_b \), the boundary conditions on the surface potential are \( \Psi_T = V_{th} + V_b - V_{fb} \) and \( \Psi_r = V_{th} + V_b - V_{fb} \) (\( V_s \) and \( V_d \) are the source and drain voltages, respectively). The bulk-source/bulk-drain pn junctions are considered “abrupt,” which explains the fact that the built-in voltage \( V_{th} \) drops in the active channel. Indeed, \( V_{th} \) is entirely included in the boundary conditions. The parameter \( \lambda_i \) is defined as follows:

\[ \lambda_i = \frac{\varepsilon_{si} t_{ox} X_{\text{dep}_i}}{\varepsilon_{ox} \eta}, \quad (6) \]

Clearly \( \lambda \) is a crucial parameter to describe short channel effects (SCEs), as it describes the influence of the pn junctions at the source and drain terminals on the behavior of \( \Psi_S \). The higher the value of \( \lambda \), the smoother the \( \Psi_S \) curve appears, and the weaker is the coupling between the gate and channel.

\( \Psi_{L_1} \) and \( \Psi_{L_2} \) are the long channel surface potentials, i.e., the surface potential in the case of \( L_1, L_2 \to \infty \).

A. Weak Inversion

The Liu model provides a solution for the surface potential in weak inversion, where mobile charges are completely neglected. In our model, if both regions of the memory cell are in weak inversion, the long channel surface potential \( \Psi_{L_i} (i = 1,2) \) in (4) and (5) is given by

\[ \Psi_{L_i} = \left( V_g - V_b - V_{fb} + \frac{\gamma^2}{2} \right) \times \left[ 1 - \sqrt{1 - \left( \frac{V_g - V_b - V_{fb}}{V_g - V_b - V_{fb} + \frac{\gamma^2}{2}} \right)^2} \right] \quad (7) \]

where \( \gamma = C_{ox}^{-1} \sqrt{2q\varepsilon_{ox} N_{\text{sub}}} \) is the body factor. The above expression comes from the solution of the 1-D Poisson equation along the transversal coordinate \( x \), where drain and source do not influence the electrostatic potential [12]. After this calculation, the depletion widths \( X_{\text{dep}_1} \) and \( X_{\text{dep}_2} \) are derived as follows:

\[ X_{\text{dep}_1} = \sqrt{\frac{2\varepsilon_{si} \Psi_{L_1}(V_g)}{q N_{\text{sub}}}}, \quad X_{\text{dep}_2} = \sqrt{\frac{2\varepsilon_{si} \Psi_{L_2}(V_g)}{q N_{\text{sub}}}}. \quad (8) \]

Equation (8) is then plugged in (6) to obtain \( \lambda_1 \) and \( \lambda_2 \) \( (\lambda_1 > \lambda_2) \). In order to keep the model simple, only one value of \( \eta \) is considered for both regions.

B. Strong Inversion

In our problem, during the gate voltage sweep, easily the uncharged region, near the source side, goes in strong inversion, while the charged region remains in weak inversion. Globally the device is in the subthreshold condition, therefore, it is still of interest for our model. The Liu model, as it starts from (3), does not include mobile charge, thus it is not able to take into account this new condition.
We can extend the analytical model to the strong inversion condition. During the gate sweep, once inversion is reached in the uncharged region, we can assume that $U_c$ is practically constant along the channel in region 1 (see Fig. 3) since the voltage $V_{th}$ drops completely over the charged region in subthreshold condition. Thus, in forward read, $U_c = 0$ (no body effect), and in reverse read, $U_c = [V_{th}]$. In this case, region 1, we can put in direct relation $U_c$ and $V_{th}$, the body effect of the equivalent fresh transistor, thus, the surface potential is approximately $2\phi_b + V_{th}$ [12].

When $\Psi_S \approx 2\phi_b + V_{th}$, a small positive variation of the surface potential causes a large buildup of charges at the interface semiconductor–oxide, thus, mobile charge cannot be neglected. Equation (3) must be modified as follows:

$$\frac{X_{dep}}{\eta} \cdot \frac{d^2\Psi_S(y)}{dy^2} + \frac{V_g - V_b - V_{th}}{\epsilon_{ox}} - V_{th} \cdot \frac{Q_d(2\phi_b - U_c - V_b)}{\epsilon_{ox}} = Q_d(2\phi_b - U_c - V_b) + Q_d(V_{gs}, V_{th}).$$  

(9)

$Q_d$ is the absolute value of the mobile charge, while $Q_d$ is the absolute value of the ionized doping charge at maximum depletion ($Q_d \equiv qN_{sub} \cdot X_{dep}$) ($2\phi_b - V_b$) in forward read; $Q_d \equiv qN_{sub} \cdot X_{dep}(2\phi_b - V_b - V_b)$ in the reverse read, considering the body effect). The inversion charge $Q_i$ can be computed as follows [13]:

$$Q_i(V_{gs}, V_{th}) = \frac{KT}{q} mC_{ox} \times \log \left(1 + \frac{C_{dep}}{mC_{ox}} \exp \left[\frac{q(V_{gs} - V_{th})}{mKT}\right]\right).$$  

(10)

where $KT$ is the thermal energy, $C_{dep} = \sqrt{\epsilon_{ox}qN_{sub}/2\phi_b}$ is the depletion layer capacitance, and $m$ is the subthreshold factor that can be extracted from the experimental transfer characteristics. In principle, a good approximation of this factor may be the theoretical $m = 1 + C_{dep}/C_{ox}$. The pedicles $s$ denotes quantities referred to the source of the fresh equivalent transistor. Let us note that, in (10), the quasi-Fermi energy level is not present, as it is considered constant along the channel of region 1. $V_{th}$ is the threshold voltage of the equivalent MOS capacitor; in forward read, we have

$$V_{th} = V_{th} + V_{fth} + 2\phi_b + C_{ox}^{-1} \sqrt{2q\epsilon_{ox}N_{sub}(2\phi_b - V_b)}$$  

(11)

while in reverse read, we have

$$V_{th} = V_{th} + V_{fth} + 2\phi_b + C_{ox}^{-1} \sqrt{2q\epsilon_{ox}N_{sub}(2\phi_b - V_b - V_b)}.$$  

(12)

The value of the inversion charge (10) is important because it modifies the definition of the long channel surface potential $\Psi_{L1}$ in (7). In forward read,

$$\Psi_{L1} = V_g - V_{fth} + Q_i \left(V_g + V_{th}, V_{th} \right) + Q_d(2\phi_b - V_b).$$  

(13)

and $\Psi_{L1}$ for the reverse read

$$\Psi_{L1} = V_g - V_{fth} + Q_i \left(V_g + V_{th}, V_{th} \right) + Q_d(2\phi_b - V_b).$$  

(14)

C. Drain Current

From the description of the surface potential, the drain current can be calculated. As in the subthreshold region, the current is a diffusion dominated process, the following expression can be used [14]:

$$I_D = \mu_{eff} \frac{W}{L} \sqrt{\frac{\epsilon_{ox}qN_{sub}}{2\Psi_{Smn} \cdot V_b}} \left(\frac{KT}{q}\right)^{2} \left(\frac{n_i}{N_{sub}}\right)^{2} \times \exp \left[\frac{q(\Psi_{Smn} + V_b)}{KT}\right] \left[1 - \exp \left(-\frac{qV_{th}}{KT}\right)\right].$$  

(15)

where $\mu_{eff}$ is the effective mobility of electrons in the channel, $n_i$ is the intrinsic electron concentration, and $\Psi_{Smn}$ is the minimum value of the surface potential. Even if the channel under the uncharged region is in strong inversion, the drain current is limited by the barrier $\Psi_{Smn}$ seen by electrons. Basically the drain current depends on the minimum value of the surface potential, and from (15), it is clear that once $[V_{th}] \gg KT/q$, $I_D$ does not depend anymore on the longitudinal component of the electric field.

It is of utmost importance to approximate $\Psi_{Smn}$ as accurately as possible to predict the experimental $I_D - V_{gs}$ characteristics. If the punchthrough near the drain region is not evident, $\Psi_{Smn}$ resides in the charged region, therefore, $d\Psi_{L2}(y)/dy = 0$ has to be evaluated from (5). The minimum is located at

$$\Psi_{min} = \frac{\chi}{2} \log \left[\frac{(\Psi_{s} - \Psi_{L2})e^{-\frac{\chi}{2}} - (\Psi_{r} - \Psi_{L2})e^{-\frac{\chi}{2}}}{(\Psi_{r} - \Psi_{L2}) - (\Psi_{s} - \Psi_{L2})e^{-\frac{\chi}{2}}}\right].$$  

(16)

If the minimum is under the uncharged region, it means that the electrostatic effect of the charged region is suppressed by pinchoff in the channel. Usually this is due either to the weak density of charge or to the very small charged region (both combined to the value of $V_d$ applied). In such a case, (16) still
remains valid, but must be applied to the uncharged region (pedices are \(i = 1\)). If (16) is plugged into (5) and the sinh functions are developed in exponentials, the minimum value of surface potential \(\Psi_{\text{Strain}}\) has the following expression:

\[
\Psi_{\text{Strain}} = \frac{1}{2} \left[ \frac{\Psi_r -\Psi_{L_2}}{\sinh \left( \frac{L_2}{\lambda_2} \right)} - \frac{\Psi^* -\Psi_{L_2}}{\sinh \left( \frac{L_2}{\lambda_2} \right)} e^{\frac{L_2}{\lambda_2}} \right] \\
\times \left[ \frac{\left(\Psi_r -\Psi_{L_2}\right) - \left(\Psi^* -\Psi_{L_2}\right)}{\sinh \left( \frac{L_2}{\lambda_2} \right)} e^{\frac{L_2}{\lambda_2}} \right]^{1/2} + \Psi_{L_2}. \tag{17}
\]

III. NUMERICAL SIMULATION

To test the viability of this analytical model with simulations and data, we have used nanocrystal memories from STMicroelectronics, Catania, Italy, and the commercial TCAD tool ATLAS [9]. Dimensions of the memory cell are \(L = 0.28\ \mu\text{m}, \ W = 0.16\ \mu\text{m}, \) tunneling oxide thickness \(t_{\text{ox}} = 5.5\) nm, and control oxide thickness \(L_2 = 8\) nm [5].

The charged nanocrystals have been modeled as a uniformly charged oxide region of thickness \(t_{\text{ch}} = 2\) nm between the tunneling oxide and control gate oxide. As specified before, we simply assumed that the charge distribution in the oxide is a step function where the charged length and density is to be determined.

The density of charge is chosen to fit different flat band voltages \(V_{FB}\)’s. We have simulated the presence of \(n\) charges in each dot, where the dot density is fixed at \(D_{\text{dot}} = 10^{12} \text{ cm}^{-2}\) (\(Q = n \cdot D_{\text{dot}}\)).

In the simulated structure, there are two overlapping regions of the gate on both drain and source junctions of 20 nm, in order to also investigate the effect of charge pockets localized over the drain/source junctions.

Results from numerical simulations are used to extract the value of the fitting parameter \(\eta\) used in (4) and (5). As \(\eta\) is present in the definition of \(\lambda\) [see (6)], it affects the modeling of SCEs. After comparison with fully 2-D simulations, \(\eta\) has been shown to be strongly dependent on the charged length \(L_2\), but not on the density of stored electrons \(Q\). In Figs. 3 and 4, we show two examples of plots, used to find the value of \(\eta\). In particular, Fig. 3 provides the behavior of the surface potential, and Fig. 4 provides a \(I_D = V_{\text{GS}}\) profile. In the inset of Fig. 4, a plot of the values \(\eta\) in the range \(L_2 = 17.5 - 140\) nm is provided.

A. Comparison of Results From the Analytical Model and Numerical Simulations

After inspection of analytical and numerical results, the following conclusions can be drawn.

1) 1-Bit Reading: It should be noticed that the values of \(\Psi_{L_1}\) and \(\Psi_{L_2}\) (7) stand for the asymptotes of the surface potential, which are actually reached if \(L_1 \gg \lambda_1\) and \(L_2 \gg \lambda_2\). In this limit, from (17), \(\Psi_{\text{Strain}} = \Psi_{L_2}\) so that the forward and reverse read characteristics of the programmed cell are identical. We must highlight that, for typical applied read bias, and the cell features described in Section III, \(\lambda_2\) does not exceed 30–35 nm; as it appears from Fig. 5, it is sufficient to have \(L_2 > 80 - 90\) nm to have identical forward and reverse read characteristics, which is in agreement with [6]. The main consequences of such observation are that if the trapping layer is only partially charged, we can have \(V_{\text{th, F}} = V_{\text{th, R}}\) (see Fig. 5), and that cells with the same density of injected charge \(Q\), but different charged length \(L_2\) (still larger than 80–90 nm), show similar threshold voltage shifts \(\Delta V_{\text{th, tot}}\). This property is evident in Fig. 6. \(\Delta V_{\text{th, tot}}\) becomes weakly dependent on the charged length \(L_2\).

2) Sensitivity of \(\Delta V_{\text{th R}}\) to Charge Density: Another important property that can be stated from our analytical approach and
from simulations is that of the reverse-forward threshold voltage shift $\Delta V_{\text{RF}}$ is only weakly sensitive to the charge density. Once a minimum density of electrons is stored in dots, it can be seen from Fig. 7 that different densities of electrons in the dots yield similar $\Delta V_{\text{RF}}$. This means that $\Delta V_{\text{RF}}$ becomes very weakly dependent on the density of charge $Q$. As is apparent from Fig. 7, in the case of a charged length of $L_2 = 70$ nm, the maximum value of $\Delta V_{\text{RF}}$ is reached with $Q = 2 \times 3 \times 10^{12} \text{ cm}^{-2}$, while for $L_2 = 47$ nm, the maximum is with $Q = 5 \times 10^{12} \text{ cm}^{-2}$. In both cases, these maxima are kept almost unaltered for higher values of the charge density. Let us note the fact that such maximum is more pronounced in the analytical model result than in numerical simulations.

3) Impact of Charge Over the Junction: The analytical model developed in Section II is not able to describe transport in the junctions, as they are doped with donors and they are polarized in accumulation while the rest of the active channel is

in inversion. The model works well if the region under analysis is either in weak inversion or in inversion. With this model, it is not analytically possible to consider the electrostatic impact of charges over the junction. On the other hand, TCAD simulations show that charges over the junction have a much smaller impact on the surface potential minimum than charges over the channel. Such an effect is clearly visible in Fig. 8. It is apparent that if we localize the overall region of charge on the channel or part on channel, part on the junction, we have different effects on $I_D - V_{\text{GS}}$ in reverse mode. In particular, the impact of the charged region of length $2L_2$ extending for a length $L_2$ on the channel and $L_2$ on the junction is similar to that of a charged region of length $L_2$ situated only on the channel. This result is in agreement with conclusions by [2] and [6]. In [15], the authors show some results consistent with ours on the effect of the charge over junction, but they arrive at different conclusions, claiming that most of the trapped charge resides on junction.

IV. EXPERIMENTS

Our analytical model has been tested with experiments on nanocrystal memory cells from STMicroelectronics. In particular, the behavior of the model has been measured when reading the cell under different bias conditions for $V_{\text{th}}$ and $V_{\text{ds}}$. In Fig. 9, the threshold voltage behavior with varying $V_{\text{ds}}$ is shown. After the CHE stress procedure, a reading is performed with $V_{\text{th}} = 0 \text{ V}$ and $0.5 \text{ V} < |V_{\text{ds}}| < 1.5 \text{ V}$.

As it is apparent, a very good agreement between the model and data is found. In order to fit these data, $n = 7.2$ electrons per dot ($D_{\text{dot}} = 10^{12} \text{ cm}^{-2}$) and $L_2 = 32.4$ nm. In particular, we note that the threshold voltage during the reverse read $V_{\text{th-R}}$ is almost independent of $V_{\text{ds}}$ as the electrostatic impact of the charged dots near the drain is essentially not affected by the voltage applied to the other contact.

On the other hand, the impact during forward read is significantly lowered by the voltage bias on the nearby contact. We can consider a complete punchthrough of the electrostatic barrier for $V_{\text{ds}} > 1.5 \text{ V}$ when $V_{\text{th-F}} = V_{\text{th-fresh}}$, but in this case, a
mental characteristics between devices, typically exhibit different threshold voltage shifts. Forw read disturb (discharge effect of the charged dots near the drain) is evident.

In Fig. 10, the dependence of the transfer characteristics on $V_b$ is shown. The measurements have been performed on a different sample, nominally identical with the previous one. It is worth noticing that, two nominally identical devices, programmed in the same way, typically exhibit different threshold voltage shifts and, therefore, different trapped charge distributions. As can be seen, analytical model reproduces the behavior of the experimental $I_D - V_G$ as a function of $V_b$. The shift of the experimental characteristics between $V_b = 0$ V and $V_b = -1$ V is well predicted. Good fit with the experimental is obtained with $n = 2.1 \text{cm}^2/\text{V}$, $L_2 = 90.1 \text{nm}$.

Fig. 9. Comparison of analytical model and experiments obtained from nanocrystal memories [5] as a function of drain-to-source $|V_{ds}|$ reading voltage. The fit is obtained with $L_2 = 32.4 \text{nm}$, $n = 7.2 \text{cm}^2/\text{V}$. (dfot $= 10^{12} \text{cm}^2$).

Fig. 10. Analytical model versus experiments on nanocrystal memories [5] as a function of bulk voltage $V_b$ for $|V_{ds}| = 1.5$ V. The fit is obtained with $L_2 = 90 \text{nm}$, $n = 2.1 \text{cm}^2/\text{V}$, $D_{\text{dot}} = 10^{12} \text{cm}^2$.

Fig. 11. Contour plots of $\Delta V_{\text{tot}}$ (dashed line) and $\Delta V_{\text{RF}}$ (solid line) as a function of the effective charged region $L_2$ and the density of charge $Q$. Parameters are $L = 0.25 \mu m$, $W = 0.1 \mu m$, $P_{\text{tot}} = 15.5 \text{nm}$, $|V_{ds}| = 1.5 \text{V}$, and $V_b = -1 \text{V}$. From the experimental data, we find the cross point between $\Delta V_{\text{tot}}$ and $\Delta V_{\text{RF}}$ values, and we are able to provide the description of charge distribution injected in the trapping medium through $L_2$ and $Q$.

Fig. 12. Proof of the viability of the analytical model. Experiments refer to nanocrystal memories made by STMicroelectronics: the cell is stressed with $V_b = 4 \text{V}$, $V_d = 3.5 \text{V}$, and $t = 100 \text{ms}$. Firstly, experiments are fitted with parameters taken from the contour plots of Fig. 11. $Q = 6.65 \times 10^{12} \text{cm}^{-2}$ and $L_2 = 40 \text{nm}$ are deduced, these values are used in the ATLAS input deck and numerical $I_D - V_{GS}$ characteristics are plotted here. Good agreement between data and numerical simulation is found, which means that $L_2$ and $Q$ from the analytical model are reliable parameters.

V. ANALYTICAL MAP FOR EXTRACTING THE INFORMATION ON THE CHARGE DISTRIBUTION

The analytical model can be used as a tool for extracting information on the distribution of charged charge in programmed devices. Indeed, from experiments, we have at our disposal the transfer characteristics and values of $V_{b0} = \text{fresh}$, $V_{b1} = \text{F}$, and $V_{b1} = \text{R}$. We can then compute the two quantities defined as $\Delta V_{\text{tot}}$ and $\Delta V_{\text{RF}}$. With the help of the presented analytical model, it is possible to create a map with the contour plots of $\Delta V_{\text{tot}}$ and of $\Delta V_{\text{RF}}$ as a function of the density of electrons $Q$ and the charged length $L_2$. The experimental $\Delta V_{\text{tot}}$ and $\Delta V_{\text{RF}}$ identify one point on the map, from which we can find the corresponding values of $Q$ and $L_2$.

In Fig. 11, an example of such maps is shown for the case of nanocrystal memories with a density of dots of $10^{12} \text{cm}^{-2}$. In order to test the validity of such a map, we compare results from
our model with experiments. We extract from the map $Q$ and $I_2$, for a pair of experimental $\Delta V_{tn,tot}$ and $\Delta V_{RF}$ and use them in the 2-D TCAD simulator to see if simulated electrical characteristics are in agreement with the experiments. The comparison between experiments and numerical simulations is shown in Fig. 12 for a case of $\Delta V_{tn,tot} = 1.5$ V and $\Delta V_{RF} = 0.49$ V: the reasonable agreement obtained shows that the proposed maps are useful to extract an “effective” distribution of the trapped charge.

In general, a good agreement is found, except for very small $L_2$ (below $\approx 20$ nm) and very high $Q$ (order of $10^{13}$ cm$^{-2}$). For small values of $L_2$, we have to consider the fact that the effect of a small portion (order of few nanometers) of the charged dots next to the drain is suppressed by the high voltage $V_{ds}$ during the forward reading. An analytical approach to quantify such effects should depend on the reading voltage $V_{ds}$ and on geometrical parameters (i.e., distance of trapping layer from the junction). If $L_2$ is of the order of tens of nanometers, the percentage of dots whose effect is suppressed becomes high, and the model is not accurate. On the other hand, if $Q$ is very high, the depletion region along the longitudinal direction of the channel presents an abrupt discontinuity at the junction between the charged and uncharged region: the depletion width of the uncharged region $X_{trp}$ is drastically larger than the depletion width of the charged region $X_{trp}$ [see (8)]. This is not what happens in the real device, as from 2-D numerical simulation the depletion width does not present strong discontinuities. In such a case, the model is not able to follow the correct behavior of the surface potential and the minimum $\psi_{cri}$ is not well interpreted for (15). Due to this problems, this analytical model is accurate up to a value of $\Delta V_{RF} < 0.8 - 1$ V, and $\Delta V_{tot} < 3.5 - 4$ V.

In this two cases ($L_2$ small and $Q$ large), an extension of the presented model would be required to obtain an accurate analytical solution.

VI. CONCLUSION

In this paper, we have provided an original analytical model of the effect of an “idealized” nonuniform distribution of trapped charge on the electrical characteristics and on the perspective of 2-bit operation of discrete-trap memories. Our model also provides a method to extract an “effective” distribution of trapped charge in which all charge is uniformly distributed with density per unit area $Q$ in a limited region of length $L_2$ close to the drain region. The only fitting parameter we have used, i.e., $\eta$, has been carefully extracted with numerical simulations.

We have validated our model through comparison with detailed numerical TCAD simulations and with experiments on nanocrystal flash memories in different bias conditions.

We have quantitatively addressed important aspects of dual-bit operation of discrete-trap memories such as the extension of the charged region to achieve dual-bit behavior, and the sensitivity of the programmed threshold voltage on the amount of trapped charge. In addition, via the help of the numerical simulator, it has been shown that the trapped charge over the junction has a small effect on electrical characteristics.

We believe that our model is a powerful tool for the analysis of the behavior of discrete-trap nonvolatile memories such as SONOS, NROMs, and nanocrystal memories, charged with nonuniform injection techniques such as CHE. However, the model has one main weakness since it loses accuracy if the pockets of trapped charge are very small (in the cases considered, smaller than 20 nm). An extension of the model to such a condition would be very much needed.

REFERENCES


Luca Perniola

Luca Perniola was born in Florence, Italy, in 1978. He received the B.S. and M.S. degrees in nuclear engineering from the Politecnico di Milano, Milan, Italy, in 2002, and is currently working toward the Ph.D. degree in microelectronics at the Institut National Polytechnique de Grenoble (INPG), Grenoble, France, and the Università degli Studi di Pisa (UniPi), Pisa, Italy. His master’s thesis concerned multinanocrystal memory devices. He spent one academic year with the Commissariat à l’Energie Atomique (CEA)—Laboratory of Electronics, Technology, and Instrumentation (LETI), Grenoble, France, where he was involved with the electrical characterization of multinanocrystal memories. His research interests are in simulation of physics of silicon dot devices and electrical properties of multinanocrystal memories.
Sandrine Bernardini was born in Marseille, France, in 1977. She received the B.S. and M.S. degrees in fundamental physics from Paul Cizanne University, Marseille, France, in 1999, the Engineering degree in materials and microelectronics from the National Institute of Applied Sciences (INSA), Lyon, France, in 2001, and the Ph.D. degree in microelectronics from Provence University, Marseille, France, in 2004.

Since 2001, she has been with the Laboratory of Materials and Microelectronics of Provence (L2MP), Marseille, France, where she has been involved with simulations and characterizations of MOS devices. Her research interests include now MOS device physics, modeling and simulation of MOSFETs electrical properties, and memory device physics, in particular, discrete-trap nonvolatile memories.

Giuseppe Iannaccone was born on April 28, 1968. He received the Laurea degree (cum laude) in electrical engineering and Ph.D. degree from the University of Catania, Catania, Italy, in 1992 and 1996, respectively. His Ph.D. thesis concerned transport and noise phenomena in ultra-small structures.

Since January 2001, he has been an Associate Professor with the Dipartimento di Ingegneria dell’Informazione, Università degli Studi di Pisa. He is also the principal investigator in other national and international projects. His interests include transport and noise modeling in nanoscale devices, devices and architectures for nonelectronics, the design of passive RF identification (RFID) transponders, and the exploitation of quantum effects in conventional electron devices. He has participated in a series of European, National, and research projects as consortium coordinator or principal investigator. He has authored or coauthored over 80 papers in peer-reviewed journals.

Barbara De Salvo received the B.S. and M.S. degrees in electronics engineering from the University of Parma, Parma, Italy, in 1996, and the Ph.D. degree in microelectronics from the Polytechnics Institute of Grenoble, Grenoble, France, in 1999.

From 1995 to 1996, she was with the Department of Information Technology, University of Parma, where she was involved with electrical characterization of A/D converters. From 1996 to 1999, she was with the Laboratory of Physics of Semiconductor Devices (LPSC), Centre National de la Recherche Scientifique/Institut National Polytechnique de Grenoble (CNRS/INPG), Grenoble, France, where she was involved with research on reliability of nonvolatile memories, gate oxide transport mechanisms, and degradation phenomena. In 1999, she joined the Laboratory of Electronics, Technology and Instrumentation (LETI), Commissariat à l’Energie Atomique (CEA), Grenoble, France, as a Research Engineer. Her current interests and activities cover the engineering and physics of advanced memory devices (in particular, silicon dots-based) for ULSI technologies. From March 2000 to February 2002, she was the coordinator of the ADAMANT Project (IST-2001-342343) funded by the European Commission within the Fifth Framework Programme. She currently manages the Advanced Memories Group, LETI, covering several projects with industrial partners and advanced topics such as 3-D memory devices and molecular memories. She has authored or coauthored over 100 papers appearing in international refereed journals and conferences. She has supervised several Master and Ph.D. students.

Pascal Masson was born in France, in 1969. He received the Eng. degree from the Institut National Polytechnique de Grenoble (INPG), Grenoble, France, in 1996, and the Ph.D. degree in microelectronics from the National Institute of Applied Sciences (INSA), Lyon, France, in 1999.

From 2000 to 2001, he was an Assistant Professor with the INSA, Lyon, France, where his main research activities concerned traps study in gate dielectric of deep-sub-0.1-μm MOS transistors. He is currently a Professor with the Provence Materials and Microelectronics Laboratory (L2MP), Provence University, Marseille, France. He is a member of the Memories team and heads the New Memories on Silicon (MISI) Group. He is currently involved with the modeling and characterization of advanced memories such as 1TDRAM, nanocrystal Flash, single-poly Flash, etc.

Gérard Ghіbauдо was born in France, in 1954. He graduated from the Polytechnics Institute of Grenoble, Grenoble, France, in 1979. He received the Ph.D. degree in electronics and State Thesis degree in physics from the Polytechnics Institute of Grenoble in 1981 and 1984, respectively.

In 1981, he became an Associate Researcher with the Centre National de la Recherche Scientifique (CNRS), Grenoble, France, where he is currently Director of Research of the Laboratories of Physics of Semiconductor Devices (LPSC/ENSERG) [now the Laboratories of Microelectronics, Electromagnetism, and Photonics (IMEP/ENSERG)]. During the 1987–1988 academic year, he spent a sabbatical year with the Naval Research Laboratory, Washington, DC, where he was involved with the characterization of MOSFETs. He has authored or coauthored over 210 papers appearing in international refereed journals and conferences. He has been involved in the organization of eight work-shops/summer schools over the last ten years.

Cosimo Gerardi received the Laurea degree (cum laude) in physics from the University of Lecce, Lecce, Italy, in 1989. His research thesis concerned electronic transport in thin III–V semiconductors.

From 1990 to 1992, he was involved in the field of semiconductor surface characterization techniques with the University of Padova, Padova, Italy. From 1992 to 1997, he was a Researcher and Senior Researcher with the Center for National Research (CNRS), Brindisi, Italy, where he was involved with surface, interface, and thin-film analysis by using SIMS, XPS, TEM, and XRD. In 1997, he joined Central Research and Development, STMicroelectronics, Catania, Italy, where he is involved with thin dielectrics characterization, Flash memories technology, and nanocrystal memories development. He has been coordinator for STMicroelectronics for several international and national research and industrial projects. He is Manager of the Device Engineering Section, Central Research and Development, Catania, Italy. He has authored or coauthored over 100 publications. He holds two patents.