

Experimental Characterization of the Vertical Position of the Trapped Charge in Si Nitride-Based Nonvolatile Memory Cells

Antonio Arreghini, Francesco Driussi, Elisa Vianello, David Esseni, *Senior Member, IEEE*,
Michiel J. van Duuren, Dusan S. Golubović, Nader Akil, and Rob van Schaijk

Abstract—We present a broad set of experiments on silicon nitride-based memories aimed at the investigation of the vertical position of the charge trapped in the nitride layer of silicon-oxide-nitride-oxide-semiconductor (SONOS) memories during program and erase in the tunneling regime. The results obtained for SONOS devices with conventional oxide-nitride-oxide and oxide-nitride-oxide-nitride-oxide gate stacks, as well as with high- κ top dielectric, have been validated by comparing different characterization techniques. It has been shown that, for SONOS cells, the charge centroid is located in the center of the silicon nitride layer, and its position is quite insensitive to the program or erase conditions and to the gate-stack composition.

Index Terms—Carrier separation, charge centroid, silicon nitride, silicon-oxide-nitride-oxide-semiconductor (SONOS).

I. INTRODUCTION

FURTHER scaling of modern floating-gate (FG) memory cells is hampered by the charge loss due to the stress-induced leakage current (SILC) in cycled cells [1]. In particular, SILC prevents the tunnel-oxide thinning below 8 nm [2] and, moreover, raises statistical-reliability issues for the nowadays very large cell arrays [3], [4]. Furthermore, a strong lateral scaling of nonvolatile-memory (NVM) arrays leads to reliability issues related to the parasitic capacitive coupling between adjacent FG cells [5].

Nitride-based memories, like the silicon-oxide-nitride-oxide-semiconductor (SONOS) devices, are good candidates to replace FG cells in the Flash memory chips of the future generations for high-density and embedded applications, owing to the SILC immunity achieved with localized-charge trapping, that guarantee very good endurance characteristics [6], [7]. Moreover, the relatively low program/erase (P/E) voltages and the excellent compatibility with the standard CMOS process of SONOS memories make them suitable for embedded applications [8]. Finally, the absence of erratic tail bits reduces the statistical-reliability issues in very large memory arrays, and the limited capacitive coupling between adjacent SONOS cells

is beneficial for the reliability issues related to the scaling of the spacing between cells.

However, the data retention of SONOS cells programmed and erased by direct tunneling is still a critical issue because of the thin tunnel oxide (≈ 2.5 nm) needed for the tunneling of substrate holes under negative gate voltages [9]. Therefore, in order to optimize the retention characteristics of SONOS memories, an accurate analysis of the trapping properties of the silicon nitride layer that is responsible for the charge storage in the cell is required. In this respect, the knowledge of the location of the charge trapped in the nitride and the type of carriers injected during the P/E operations are not yet complete [10]–[13] despite the importance of these pieces of information in order to understand the charge-loss rates from the nitride toward the gate and substrate and, therefore, to model accurately the retention characteristics.

In this framework, this paper presents an extensive analysis of the vertical position of the charge trapped in the gate dielectric for a wide set of SONOS devices with different oxide-nitride-oxide (ONO) stacks, including samples with nonconventional gate stacks and with a high- κ top dielectric. To accomplish this, we have extensively applied the experimental procedure developed in [14] to extract the trapped-charge centroid, thus fully validating it over several gate stacks for memory cells. Compared with other methods used to probe the charge trapped in the gate stack, our technique is based on the simplest MOSFET electrostatics, and no constraints are imposed to the P/E conditions during its application. In particular, whereas the low-frequency charge pumping (LFCP) [15] can extract the trap profile, but only within a few nanometers inside the ONO stack, our procedure can reach the deeper locations inside the gate stack, thus considering the effect of the whole trapped charge. However, the centroid alone does not provide the detailed knowledge of the charge profile inside the nitride, but it determines the effectiveness of a given distribution of charges in generating a corresponding V_T shift. Therefore, our technique can be seen as complementary to other methods used to probe the trapped-charge profile (e.g., the LFCP), overcoming some of their limitations in terms of accuracy and applicability.

A broad range of P/E conditions, i.e., different programming/erasing gate voltages (V_G), have been tested for each device, and the experimental results for the charge-centroid location have been confirmed by the comparison with those derived with the LFCP method, thus demonstrating the advantages and

Manuscript received November 26, 2007. This work was supported in part by the Italian MIUR under the PRIN 2004 and FIRB RBIP06YSJJ research programs. The review of this paper was arranged by Editor V. R. Rao.

A. Arreghini, F. Driussi, E. Vianello, and D. Esseni are with the Department of Electrical, Mechanical and Management Engineering, University of Udine, 33100 Udine, Italy (e-mail: francesco.driussi@uniud.it).

M. J. van Duuren, D. S. Golubović, N. Akil, and R. van Schaijk are with NXP Semiconductors, 3001 Leuven, Belgium.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TED.2008.919713

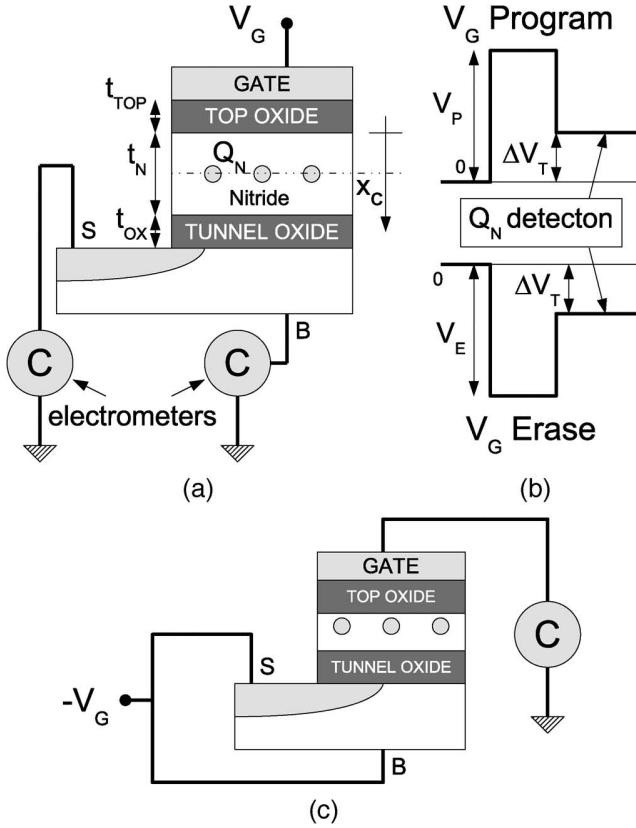


Fig. 1. (a) Measurement setup for carrier separation and charge-centroid extraction. Q_N is the trapped charge, x_C is the centroid position, and t_{TOP} , t_N , and t_{OX} are the thicknesses of the top oxide, nitride layer, and tunnel oxide, respectively. (b) Waveforms for the P/E pulses showing the three-level shape used to obtain a null contribution to the overall current integral of the displacement charge in the semiconductor. (c) Measurement setup for the charge-centroid extraction. By measuring Q_N at the gate while pulsing with a reversed waveform with respect to (b) at the source and bulk terminals, no compensation of possible leakage currents is needed.

limitations of each technique. In a limited number of cases, pMOS SONOS devices were fabricated, and carrier-separation experiments were performed to identify the type of carriers involved in the program/erase operation.

II. EXPERIMENTAL SETUP AND CHARGE-CENTROID-EXTRACTION PROCEDURE

Fig. 1(a) shows the vertical structure of a conventional SONOS device. The charge that is responsible for the threshold-voltage shift (ΔV_T) is supposed to be stored in the silicon nitride layer, which is sandwiched between the tunnel and the top oxides (this assumption will be relaxed later). If we let x_C denote the centroid of the trapped charge measured from the top-oxide/nitride interface [see Fig. 1(a)], then the threshold-voltage shift induced in the cell by the charge Q_N (uniformly injected over the device area A) can be calculated as [14], [16]–[18]

$$\begin{aligned} \Delta V_T &= -\frac{Q_N}{A} \cdot \frac{t_{TOP}}{\epsilon_{OX}} - \frac{1}{\epsilon_N} \int_0^{t_N} x \cdot q(x) dx \\ &= -\frac{Q_N}{A} \left(\frac{t_{TOP}}{\epsilon_{OX}} + \frac{x_C}{\epsilon_N} \right) \end{aligned} \quad (1)$$

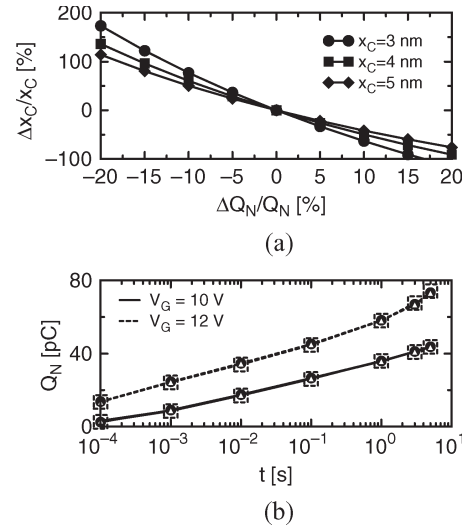


Fig. 2. (a) Sensitivity of the charge centroid on the possible error in the trapped-charge detection. Note that small errors in the measured Q_N lead to large errors in the estimation of x_C . (b) Measured Q_N as a function of the programming pulsewidth for three different devices (circles, squares, and triangles). Note the good repeatability of the measurements obtained with our setup.

where $q(x)$ is the charge density along the nitride depth x , ϵ_{OX} and ϵ_N are the permittivity of SiO_2 and Si_3N_4 , respectively, and t_{TOP} and t_N are the top oxide and nitride thicknesses, respectively. Therefore, if ΔV_T and Q_N are known, the centroid of the charge trapped in the nitride can be univocally determined as

$$x_C = -\epsilon_N \left(\frac{\Delta V_T \cdot A}{Q_N} + \frac{t_{TOP}}{\epsilon_{OX}} \right). \quad (2)$$

Hence, in order to determine x_C by using (2), we have developed a well-controlled experimental setup to directly measure both the programming (erasing) charge Q_N and the corresponding threshold-voltage shift $\Delta V_T = (V_{T \text{ final}} - V_{T \text{ initial}})$ produced by the program (erase) performed in tunneling conditions. For the sake of clarity, it is worth noting that we define program (erase) as the application of gate pulses with positive (negative) V_G , which increase (decrease) the threshold voltage V_T , regardless of the type of the device under test (nMOS or pMOS). In the experimental setup shown in Fig. 1(a), we connected two accurately zeroed electrometers at the source/drain and bulk terminals of the SONOS device. In this way, when an nMOS SONOS cell undergoes a P/E cycle, we can separately measure the contributions to the programming (erasing) charge Q_N of both the electrons detected at source/drain and the holes collected at the bulk, thus performing a carrier-separation experiment.

As for the x_C extraction, an analysis of (2) shows that, while the errors in the detection of ΔV_T are not so critical, x_C is largely sensitive to possible errors on the measured Q_N . With this regard, Fig. 2(a) shows, for a fixed actual centroid position (different symbols), the relative error in the extracted x_C induced by an error in the detection of the charge; we clearly see that even a relatively small error in Q_N can lead to an incorrect estimation of the charge centroid. Therefore, in order to minimize the charge errors, the leakage current through

the source-to-bulk junction has been compensated for, first, by measuring its spurious contribution to the charge during a pulse train with a reduced V_G value that prevents the injection into the gate stack and, then, by subtracting the obtained value from the Q_N measured during the P/E pulse with large V_G .

Alternatively, because the capacitor is a two-terminal device, the injected charge can be measured also by reversing the roles of the device contacts, hence, through an electrometer connected to the gate, while pulsing on the bulk and source terminals [Fig. 1(c); source and bulk are short-circuited]. In this configuration, a $-V_G$ value is applied during the pulse. In such a way, although carrier separation is not possible, no leakage compensation is needed on the gate terminal (where the charge is measured). Therefore, for the purpose of extracting the charge centroid, the measurement of Q_N by means of the experimental setup in Fig. 1(c) is highly preferable.

Moreover, in order to avoid the spurious contribution to Q_N due to a net, which is the nonnull-time integral of the displacement current, we forced the same ($V_G - V_T$), hence the same electric field E_{OX} at the Si/SiO₂ interface, which is before and after the program/erase pulse. This was obtained by applying to the gate three-level pulses, as shown in Fig. 1(b), where the final base level of the pulse is displaced by ΔV_T with respect to the initial base level. This approach requires the knowledge in advance of the ΔV_T induced by each program (erase) pulse, and therefore, the P/E characteristics have been preliminarily measured. Afterwards, the measured ΔV_T 's have been used to set the pulse shape and measure the Q_N , as shown in Fig. 1(c). Because the programming and the erasing are performed by applying a sequence of several P/E pulses, for each measurement, we monitored both the cumulative charge Q_N involved in the whole P/E sequence and the partial charge q_N injected during each individual P/E pulse.

As shown in Fig. 2(b), by using this measurement scheme on a given SONOS device, we could achieve a very good reproducibility of the measured Q_N and ΔV_T among different runs and dies, provided that all the P/E characteristics start from the same initial V_T value. An *ad hoc* reset procedure of the device consisting of several mild P/E cycles allowed us to enforce always the same starting V_T value within tens of millivolts. The possible degradation produced by the P/E cycling has been monitored by means of charge-pumping measurements, and it was found to be negligible in the range of relatively few P/E cycles used for our characterization experiments.

III. DEVICE DESCRIPTION

The devices used in this paper are nMOS and pMOS SONOS capacitors with a surrounding source diffusion and a gate area of $A = 3100 \mu\text{m}^2$. For both nMOS and pMOS devices, n+ and p+ polygates are available, and the substrate doping is approximately 10^{17}cm^{-3} . The vertical dimensions of the ONO gate stack have been extracted by means of transmission electron microscopy (TEM), ellipsometry, and capacitance-voltage (C-V) measurements. The nominal thickness of the thermal tunnel oxide is $t_{OX} = 2.2 \text{ nm}$ for all SONOS cells, whereas two thicknesses are available for both the dichlorosilane Si₃N₄ layer ($t_N = 3.3$ and 5.7 nm , respectively) and the high-temperature-

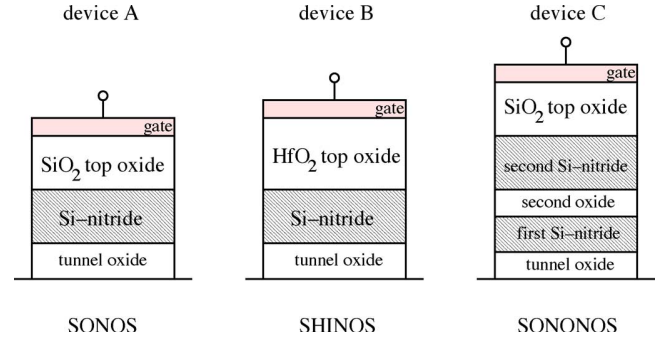


Fig. 3. Gate-stack composition of the devices used in this paper.

oxide (HTO) top oxide [$t_{TOP} = 8.9$ and 5.5 nm , respectively; Fig. 3, device A].

A gate stack featuring a 11.8-nm-thick HfO₂ layer as top dielectric, a 5.7-nm-thick Si nitride, and a 2.2-nm-thick tunnel oxide was also available (hereafter denoted as SHINOS; Fig. 3, device B).

Finally, the nonconventional gate-stack structure shown in Fig. 3 (device C, hereafter denoted as SONONOS) has been also used for the experiments related to the charge-centroid determination. This device is composed of two nitride layers sandwiched between the tunnel and top oxides and separated by an additional SiO₂ layer (second oxide) formed by the reoxidation of part of the first silicon nitride layer (the one on top of the tunnel oxide). In this device, the thicknesses extracted by means of TEM and C-V measurements are 1.3, 2.8, 1.5, 5.0, and 7.2 nm for the tunnel oxide, the first nitride, the second oxide, the second nitride, and the top oxide, respectively (Fig. 3, device C). All the measured devices are free of the passivation issue discussed in [19].

IV. CARRIER-SEPARATION EXPERIMENTS

Carrier-separation experiments are suitable to investigate quantitatively the contribution of electrons and holes during the program and erase operations. Previous experiments showed that, during program, the charge Q_N is almost entirely due to electrons [11], [13], [14], [17], whereas during erase, the contribution of holes can be significant [9], [14]. In this regard, we first successfully verified with the experimental setup in Fig. 1(a) that Q_N is due to electrons during program [14]. In analogy with the case of quantum-yield experiments, the hole recombination with electrons in the inversion layer is a very unlikely process [20], [21], and, thus, we conclude in agreement with Bachhofer *et al.* [11], Wang and White [13], and Libsch and White [17] that the hole detrapping is negligible during this phase.

In the case of erasing, because our nMOS SONOS test structures are rather large ($20 \times 155 \mu\text{m}$), the electrons that are possibly detrapped from the nitride must diffuse for several micrometers in the substrate before they can reach the source contact; thus, the recombination with substrate holes can take place, and the carrier-separation technique may underestimate the contribution to the erase of the detrapped electrons. The experiments on nMOS devices revealed that the measured Q_N during erase is mainly due to holes [14]; hence, either

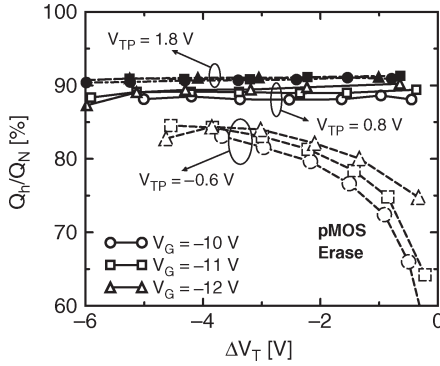


Fig. 4. Fraction of the charge due to holes measured during the pMOS SONOS erase as a function of ΔV_T for different values of V_G and of the initial threshold voltage V_{TP} . Note the decreasing contribution of holes for decreasing V_{TP} . $Q_h, Q_e > 0$ during erase.

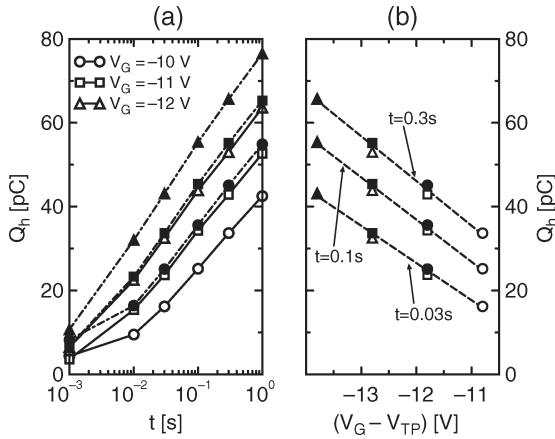


Fig. 5. Hole contribution to pMOS SONOS erase as a function of the (a) erasing time and (b) $(V_G - V_{TP})$. Open symbols for $V_{TP} = 0.8$ V and filled symbols for $V_{TP} = 1.8$ V.

the electron detrapping is very small or it is masked by the recombination in the p -substrate.

The accuracy of the carrier-separation experiments for determining the contribution of electrons and holes (Q_e and Q_h , respectively) to the erase is significantly improved if we use pMOS SONOS capacitors. In fact, under negative V_G , the holes are supplied by the inversion layer and are measured at the source, whereas the detrapped electrons go through the very thin hole inversion layer and, then, are collected at the bulk terminal of the n -type substrate [20], [21] so that the probability of recombination with holes is drastically reduced with respect to the experiment of the erase in nMOS SONOS.

Fig. 4 shows the carrier-separation measurements performed during the erase transient in pMOS SONOS devices, which, according to our definition, correspond to a reduction of the threshold voltage, i.e., to negative ΔV_T values. We see how the hole contribution Q_h to Q_N depends on the initial programmed state V_{TP} . In particular, for $V_{TP} = -0.6$ V, we observe a reduced contribution of holes, hence a large contribution of electrons detrapped from the nitride layer and a strong ΔV_T dependence of the (Q_h/Q_N) ratio. At higher V_{TP} , the relative contribution of holes remains constant versus ΔV_T , regardless of the erasing gate voltage V_G . Moreover, if we replot Q_h as in Fig. 5, we see how the hole component directly depends on the

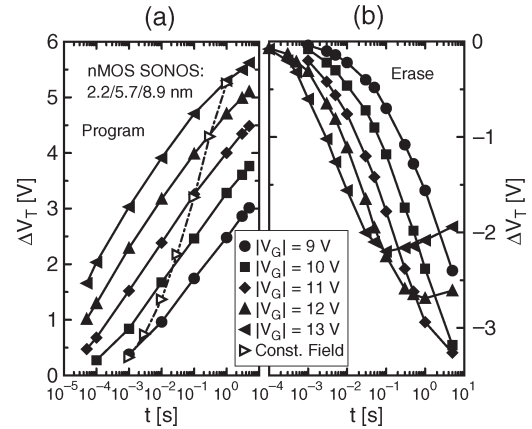


Fig. 6. (a) Program and (b) erase characteristics of an nMOS SONOS capacitor at constant gate voltage. Open triangles correspond to programming at constant tunnel-oxide field, which is obtained by compensating at each program pulse the ΔV_T produced by the previous pulse.

initial gate overdrive voltage ($V_G - V_{TP}$). This latter result can be justified by considering that both the tunneling rate and the hole-inversion-layer density depend on the $(V_G - V_{TP})$ value.

V. POSITION OF THE CHARGE CENTROID

After the identification of the different roles played by electrons and holes during P/E, we focused our attention on the extraction of the charge centroid x_C by using (2). In particular, to study the x_C dependence on the P/E conditions and the device structure, the experimental procedure described in Section II has been used to measure Q_N and the induced ΔV_T for several V_G values and for devices featuring different ONO gate stacks.

A. Experimental Results for SONOS and SHINOS Devices

Fig. 6 shows the P/E characteristics as functions of the total program (erase) time t of the nMOS SONOS with $t_{OX} = 2.2$ nm, $t_N = 5.7$ nm, and $t_{TOP} = 8.9$ nm (2.2/5.7/8.9). The ΔV_T 's have been evaluated by measuring the shifts of the $C-V$ curve after each P/E pulse. Then, the capacitor was reset to the initial state (by mild P/E cycling), and the ΔV_T data have been used to set the shape of the pulses [see Fig. 1(b)] during which we directly measured the injected Q_N with the experimental setup in Fig. 1(c). Finally, in Fig. 7(a), we have calculated the x_C of the cumulative charge Q_N . Each point of the curve is the centroid of the total charge Q_N injected from the beginning of the P/E characteristic to the time corresponding the ΔV_T value reported in the x -axis in Fig. 7(a).

We observe that the centroids for program and erase are very consistent with each other; at the beginning of the program (erase) phase, x_C is located near the middle of the nitride layer, which is in agreement with the results for HTO top oxides reported in [22]. Then, x_C slowly moves toward the top oxide. Moreover, the data in Fig. 7(a) do not show any dependence on the applied gate voltage V_G . In this respect, in order to further investigate a possible bias dependence of x_C , we have also performed measurements by keeping E_{OX}

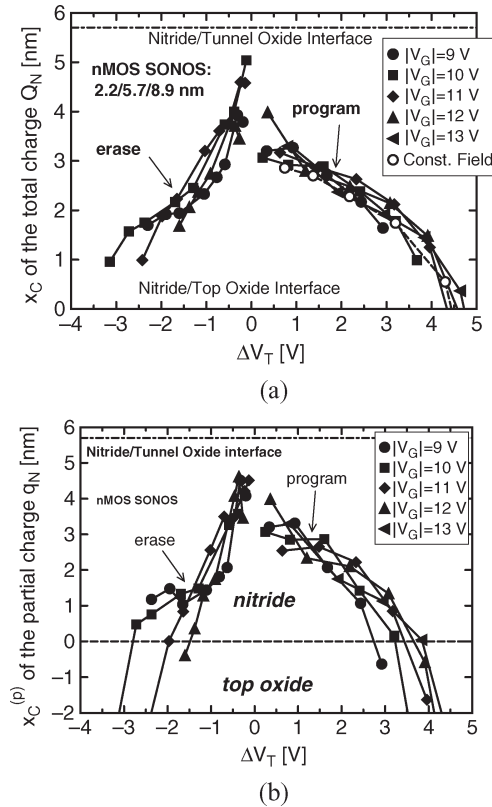


Fig. 7. (a) Centroid of the cumulative charge trapped in the nitride of the nMOS SONOS capacitor (2.2/5.7/8.9-nm ONO stack) extracted for the program and erase pulses at different V_G 's. Open circles correspond to the programming at constant electric field across the tunnel oxide (V_G is adjusted after each subsequent pulse). (b) Centroid of the partial charge injected into the nitride during each pulse of the programming/erasing of the nMOS SONOS (2.2/5.7/8.9-nm ONO stack). Note how during the last program/erase pulses (large $|\Delta V_T|$) the centroid of the partial charge is located in the top oxide, indicating that part of the injected charge has been lost through the gate.

at the Si/SiO₂ interface approximately constant during program [Figs. 6(a) and 7(a); open symbols] by increasing V_G after each pulse for an amount equal to the corresponding ΔV_T . This, of course, leads to a faster programming [Fig. 6(a); open symbols], but x_C still agrees well with the data obtained at constant V_G [Fig. 7(a)]. Therefore, we can conclude that there is no dependence of x_C on the injection conditions, at least in the bias range explored in this paper.

For $\Delta V_T > 4$ V, the x_C values move toward the top oxide and eventually become negative. The negative x_C , however, is most likely an artifact due to an electron leakage through the top oxide at the largest programming times, hence largest ΔV_T 's. In fact, when the cell is strongly programmed and the nitride is almost completely charged, the electric field across the top oxide is maximum, and the electrons injected in the nitride can tunnel out of the storage layer during the last part of the programming phase. This charge leakage offsets our estimate of Q_N , which hence appears less effective with respect to the measured threshold-voltage shift and results in a smaller value of x_C [see (2)]. In order to more precisely assess the ΔV_T range where the determination of x_C is reliable, in Fig. 7(b), we calculated the centroid $x_C^{(p)}$ of the partial charge packet q_N injected during the last pulse of the P/E sequence used to reach

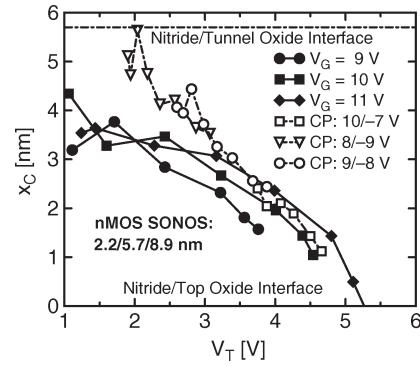


Fig. 8. Centroid x_C for the nMOS SONOS capacitor (2.2/5.7/8.9-nm ONO stack) extracted during programming. The data obtained with the (filled symbols) direct-measurement technique summarized by (2) are in good agreement with those provided by the (open symbols) LFCP experiments.

a given ΔV_T value, hence injected between two adjacent points of the curves in Fig. 7(a). The centroid $x_C^{(p)}$ is calculated by inserting in (2) the charge q_N and the threshold-voltage shift corresponding to the single (individual) P/E pulse. It is worth to note how, for $\Delta V_T > 3$ V, $x_C^{(p)}$ is negative, thus indicating that we are measuring a charge packet q_N that is larger than the one needed to justify the partial threshold-voltage shift obtained during the last P/E pulse. Hence, part of the charge q_N must have been lost through the gate. For $\Delta V_T < 3$ V, on the other hand, the $x_C^{(p)}$ of each partial charge packet is located well inside the nitride, thus indicating that we are not losing charge through the gate and validating the x_C values extracted for the cumulative charge Q_N in this range of ΔV_T values.

Moreover, the analysis of the $x_C^{(p)}$ values shows that the centroid position is independent of the starting V_T , because each point in Fig. 7(b) corresponds to the charge centroid obtained, starting from a different V_T level.

During erase, $x_C^{(p)}$ becomes negative when the electron injection from the gate takes place. In this situation, a steady-state current flows from the substrate to the gate, thus yielding a saturation of the erase characteristics that limit the observable ΔV_T range [Fig. 6(b)]. The charge measured in this condition is largely exceeding the charge responsible for the threshold-voltage shift, and therefore, the extracted x_C becomes unreliable while approaching this region. On the basis of such observations, we estimated that, for the measured device, the x_C values are reliable for $-2 < \Delta V_T < 3$ V and, in this ΔV_T range, the charge centroid stays close to the center of the nitride layer and it only slightly moves toward the top oxide for both program and erase [14], [18], [22].

As stated in Section IV, the carrier-separation experiments showed that, in our nMOS SONOS capacitors, the measured Q_N during program is essentially due to electrons whereas, during erase, it is mainly due to holes. Thus, the net result of a complete P/E cycle is a recombination charge, and we can also use an LFCP technique to detect such a recombined charge Q_N [15]. Fig. 8 shows the x_C values calculated by using Q_N detected by means of LFCP with different base and top pulse levels. We see that the results agree well with the x_C obtained with the direct measure of Q_N on the same device, except at small ΔV_T , where the LFCP method underestimates Q_N , as

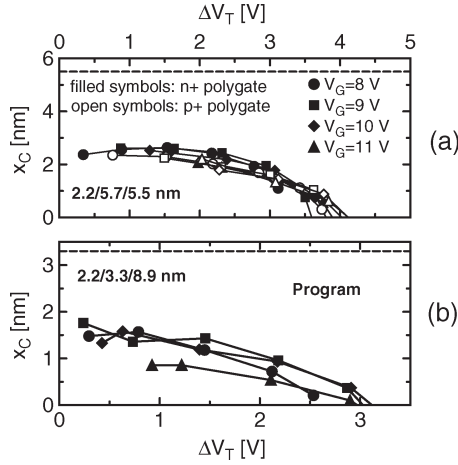


Fig. 9. Centroid x_C extracted during the programming of SONOS devices, featuring two different ONO stacks: (a) 2.2/5.7/5.5 and (b) 2.2/3.3/8.9. Filled and open symbols are the data for SONOS with n+ polygate and p+ polygate, respectively. The dashed line indicates the tunnel-oxide/silicon-nitride interface.

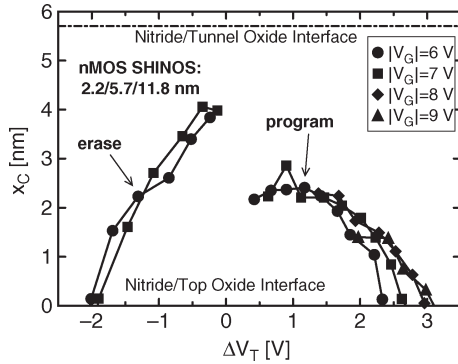


Fig. 10. Charge centroid extracted during the program and the erase at different V_G 's for the nMOS SHINOS cells, featuring a HfO_2 top oxide of 11.8 nm.

shown in [14], thus resulting in a deviation of x_C toward the tunnel oxide [see (2)].

Fig. 9 shows the centroid x_C extracted during the programming of SONOS capacitors with different top oxide [$t_{\text{TOP}} = 5.5$ nm; (a)] and silicon nitride thicknesses [$t_N = 3.3$ nm; (b)], whereas all other stack dimensions are the same, as shown in Fig. 7. Also, in this case, we find that x_C is located near the middle of the nitride (slowly shifting toward the top oxide as programming continues) and no gate bias dependence is observed. Moreover, there is no significant dependence of the measured x_C on the polygate type (open and filled symbols). Fig. 10 shows the x_C values extracted from the SHINOS cells featuring the HfO_2 top dielectric (Fig. 3, device B). Also, in this case, x_C is located inside the nitride, no dependence on V_G is observed, and the data for program and erase are consistent with each other. The results obtained for the pMOS SONOS devices are very similar [14].

As a summary, Fig. 11 shows the centroid position normalized to the nitride thickness t_N for all the measured SONOS devices. We see how, for moderate ΔV_T values, x_C is always located near the center of the nitride irrespective of the gate-stack composition and the substrate doping type.

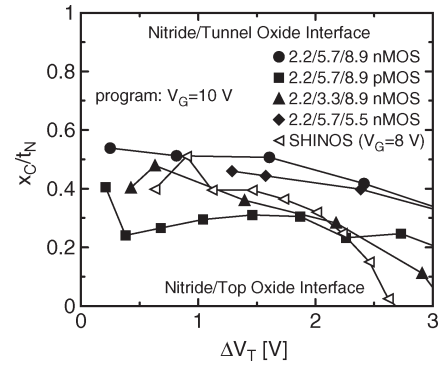


Fig. 11. Normalized position along the nitride layer of the charge centroid extracted from different devices. Note that the relative position of x_C is near the center of nitride for all the samples, as long as ΔV_T is not very large.

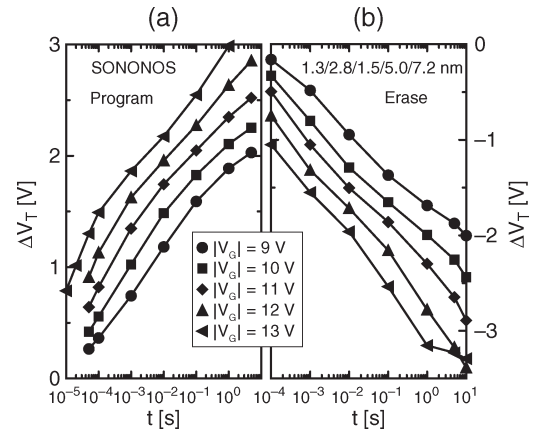


Fig. 12. (a) Program and (b) erase characteristics of the SONONOS cell for different V_G values.

B. Experimental Results for SONONOS Devices

We applied the charge-centroid-extraction procedure also to the SONONOS devices featuring the gate-stack structure shown in Fig. 3 (device C). Fig. 12 shows the P/E characteristics of this SONONOS capacitor, which show a faster programming and erasing compared with SONOS cells (see Fig. 6). In order to perform the x_C extraction accurately, particular care has been taken in the ΔV_T measurement. In fact, the application of even small V_G values produces appreciable threshold-voltage shifts, due to the very thin tunnel oxide (1.3 nm). This results in a considerable hysteresis in the $C-V$ characteristics obtained, spanning a relatively large V_G range [Fig. 13(a)]. In our capacitors, the ΔV_T detection is performed by measuring the shift of the $C-V$ characteristic, and this can be accomplished by detecting the capacitance at different points, hence V_G values. However, in the SONONOS device, we had to choose an appropriate evaluation point for the ΔV_T [Fig. 13(a), point B] in order to avoid the application of too large ($V_G - V_T$) values that induce a spurious programming during the ΔV_T extraction, as in the case of the choice of the point A in the $C-V$ curve in Fig. 13(a). In the latter case, the gate disturbance leads to a parasitic threshold-voltage shift that depends on the measurement time [Fig. 13(b), in squares]. The experimental results presented in the following have been extracted by detecting the ΔV_T at

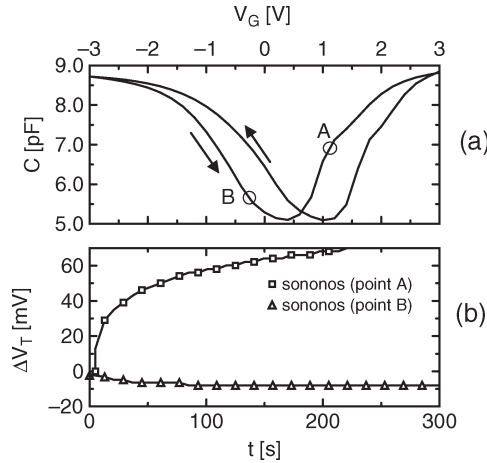


Fig. 13. (a) C - V curves of the SONONOS, showing a large hysteresis during the measurement. The arrows indicate the direction of the V_G sweep. (b) The ΔV_T value is measured as a shift of the C - V curve. The selection of the right evaluation point in the C - V curve avoids a spurious V_T shift induced by a gate disturb during the detection of the ΔV_T .

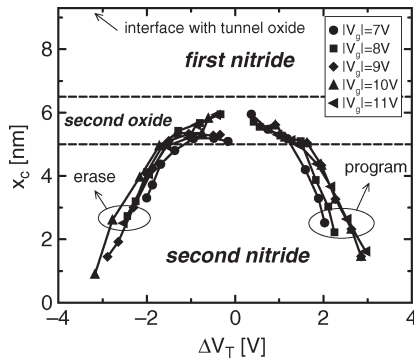


Fig. 14. Charge centroid extracted from the SONONOS devices. Initially, the charge is trapped near the interface between the first nitride and the second SiO_2 layer, and then, it fills the traps in the center of the second nitride. See Fig. 3 (device C) for the gate-stack definition.

(triangles) point B, where the gate disturbance is essentially negligible.

On the other hand, it is worth noting that, in the SONOS samples, the discussed issue is not present due to the fact that the C - V curve is not hysteretic because of the thicker tunnel oxide. In this case, the ΔV_T detection can be done at any point of the C - V curve, because the application of the V_G bias that is necessary for the capacitance measurement does not also produce spurious threshold-voltage shifts if this V_G value is applied to the device for long times.

The x_C values extracted from the SONONOS cells are shown in Fig. 14. Also, in this case, there is no dependence on the applied P/E conditions, and the data for program and erase are consistent. However, unlike standard SONOS devices, SONONOS structures exhibit a centroid x_C that is initially located very close to the second oxide layer (Fig. 3, device C). This is probably due to the fact that the second oxide has been formed by the reoxidation of part of the first nitride layer. This oxide-fabrication method is known to produce a large number of traps at the interface between the first silicon nitride layer and the second oxide layer [22]–[24]. The density of these interface traps can be larger than the concentration in the bulk

of the nitride layer, and the results in Fig. 14 are consistent with the assumption that the carriers are mainly trapped at the $\text{Si}_3\text{N}_4/\text{SiO}_2$ interface at the beginning of the program/erase operation, namely, for small $|\Delta V_T|$ values. Because in this case, x_C is located in the SiO_2 layer that has a different permittivity ϵ_{OX} with respect to nitride, for the evaluation of the x_C values, (2) has been modified as follows:

$$x_C = -\epsilon_{\text{OX}} \left(\frac{\Delta V_T A}{Q_N} + \frac{t_N}{\epsilon_N} \right) + t_N - t_{\text{TOP}}. \quad (3)$$

By increasing the program (erase) level, electrons (holes) start filling the bulk traps located in the second nitride film, and x_C moves to the middle of the nitride thickness as in the case of SONOS structures. It seems, however, that no significant amount of charge is trapped in the first nitride layer, although a comparable concentration of free traps should be available with respect to the second nitride layer.

These results also indicate that our experimental technique is indeed sensitive; in fact, it can detect changes of the x_C value due to different trapping properties in the gate stack.

VI. CONCLUSION

We have extensively used carrier-separation experiments with direct measurement of the charge injected during the program and erase operations of SONOS memory cells to study the type of carrier that is mainly involved in each operation as well as the location of the trapped charge Q_N . Preliminary experiments showed that, whereas Q_N is entirely due to electrons during program, during erase, there is a significant contribution of holes, which follows a $(V_G - V_{\text{TP}})$ dependence.

Concerning the extraction of the charge centroid, a wide set of experiments in several P/E conditions on SONOS cells with different gate-stack structures has consistently shown that the centroid of the cumulative charge (x_C) is always located close to the center of the trapping layer and it slowly shifts toward the top oxide irrespective of the applied P/E voltages. Moreover, the x_C values do not significantly depend on the top oxide and nitride thicknesses, the composition of the top dielectric, or the substrate and gate doping type.

The centroid results for the SONONOS gate stack instead show that the carriers are initially trapped near the intermediate oxide layer likely because the reoxidation of the first nitride layer has generated a large number of traps at the $\text{Si}_3\text{N}_4/\text{SiO}_2$ interface [22]–[24]. For larger P/E levels, the carriers start filling the bulk traps of the second Si_3N_4 layer, so that the centroid moves toward the middle of the second nitride layer, which is consistent with the results obtained for SONOS devices. Also, for these devices, the behavior of x_C is essentially independent of the P/E conditions.

This large set of experimental data demonstrates the sensitivity of the characterization procedure proposed in this paper, which is thus useful to study the trapping properties of different gate-stack structures and support the design of the nitride-based NVM cells. The data also represent a valid reference for the calibration of program/erase and retention models of Si nitride-based NVM [25].

ACKNOWLEDGMENT

The authors would like to thank Prof. L. Selmi for the fruitful discussions and the constant encouragement. Parts of the samples were fabricated in the framework of the MEDEA+ project NEMeSyS.

REFERENCES

- [1] J. Blauwe, J. Van Houdt, D. Wellekens, R. Degraeve, P. Roussel, L. Haspeslagh, L. Deferm, G. Groeseneken, and H. E. Maes, "A new quantitative model to predict SILC-related disturb characteristics in flash E²PROM devices," in *IEDM Tech. Dig.*, 1996, p. 343.
- [2] K. Yoshikawa, S. Mori, E. Sakagami, N. Arai, Y. Kaneko, and Y. Ohshima, "A flash EEPROM cell scaling including tunnel oxide limitations," in *Proc. Eur. Solid State Device Res. Conf.*, 1990, pp. 169–172.
- [3] F. Driussi, F. Widdershoven, D. Esseni, L. Selmi, and M. van Duuren, "Experimental characterization of statistically independent defects in gate dielectrics—Part I: Description and validation of the model," *IEEE Trans. Electron Devices*, vol. 52, no. 5, pp. 942–948, May 2005.
- [4] D. Ielmini, A. S. Spinelli, A. L. Lacaita, and M. Duuren, "Correlated defect generation in thin oxides and its impact on Flash reliability," in *IEDM Tech. Dig.*, 2002, pp. 143–146.
- [5] A. Ghetti, L. Bortesi, and L. Vendrame, "3D Simulation study of gate coupling and gate cross-interference in advanced floating gate non-volatile memories," *Solid State Electron.*, vol. 49, no. 11, pp. 1805–1812, Nov. 2005.
- [6] M. H. White, D. A. Adams, and J. Bu, "On the go with SONOS," *IEEE Circuits Devices Mag.*, vol. 16, no. 4, pp. 22–31, Jul. 2000.
- [7] R. van Schaijk, M. van Duuren, P. Goarin, W. Y. Mei, and K. van der Jeugd, "Reliability of embedded SONOS memories," in *Proc. ESSDERC*, 2004, pp. 159–162.
- [8] Y. Kamigaki and S. Minami, "MNOS nonvolatile semiconductor memory technology: Present and future," *IEICE Trans. Electron.*, vol. E84-C, no. 6, pp. 713–723, Jun. 2001.
- [9] E. Suzuki, K. Miura, Y. R. Tsay, and D. Schroder, "Hole and electron current transport in metal-oxide-nitride-oxide-silicon memory structures," *IEEE Trans. Electron Devices*, vol. 36, no. 6, pp. 1145–1149, Jun. 1989.
- [10] Y. Yang, A. Purwar, and M. White, "Reliability considerations in scaled SONOS nonvolatile memory devices," *Solid State Electron.*, vol. 43, no. 11, pp. 2025–2032, Nov. 1999.
- [11] H. Bachhofer, H. Reisinger, E. Bertagnolli, and H. Philipsborn, "Transient conduction in multilayered silicon-oxide-nitride-oxide semiconductor structures," *J. Appl. Phys.*, vol. 89, no. 5, pp. 2791–2800, Mar. 2001.
- [12] S. Chung, P.-Y. Chiang, G. Chou, C.-T. Huang, P. Chen, C.-H. Chu, and C.-H. Hsu, "A novel leakage current separation technique in a direct tunneling regime gate oxide SONOS memory cell," in *IEDM Tech. Dig.*, 2003, p. 617.
- [13] Y. Wang and M. White, "An analytical retention model for SONOS nonvolatile memory devices in the excess electron state," *Solid State Electron.*, vol. 49, no. 1, pp. 97–107, Jan. 2005.
- [14] A. Arreghini, F. Driussi, D. Esseni, L. Selmi, M. van Duuren, and R. van Schaijk, "Experimental extraction of the charge centroid and of the charge type in the P/E operation of SONOS memory cells," in *IEDM Tech. Dig.*, 2006, pp. 499–502.
- [15] A. Arreghini, F. Driussi, D. Esseni, L. Selmi, M. van Duuren, and R. van Schaijk, "New charge pumping model for the analysis of the spatial trap distribution in the nitride layer of SONOS devices," *Microelectron. Eng.*, vol. 80, pp. 333–336, 2005.
- [16] S. M. Sze, *Physics of Semiconductor Devices*. Hoboken, NJ: Wiley, 1981.
- [17] F. R. Libsch and M. White, "Charge transport and storage of low programming voltage SONOS/MONOS memory devices," *Solid State Electron.*, vol. 33, no. 1, pp. 105–126, Jan. 1990.
- [18] P.-Y. Du, H.-T. Lue, S.-Y. Wang, E.-K. Lai, T.-Y. Huang, K.-Y. Hsieh, R. Liu, and C.-Y. Lu, "Study of the gate-sensing and channel-sensing transient analysis method for monitoring the charge vertical location of SONOS-type devices," *IEEE Trans. Device Mater. Rel.*, vol. 7, no. 3, pp. 407–419, Sep. 2007.
- [19] F. Driussi, L. Selmi, N. Akil, M. van Duuren, and R. van Schaijk, "On the passivation of interface states in SONOS test structures: Impact of device layout and annealing process," in *Proc. IEEE ICMTS*, 2006, pp. 51–55.
- [20] C. Chang, C. Hu, and R. Brodersen, "Quantum yield of electron impact ionization in silicon," *J. Appl. Phys.*, vol. 57, no. 2, pp. 302–309, Jan. 1985.
- [21] A. Ghetti, M. Alam, J. Bude, D. Monroe, E. Sangiorgi, and H. Vaidya, "Stress induced leakage current analysis via quantum yield experiments," *IEEE Trans. Electron Devices*, vol. 47, no. 7, pp. 1341–1348, Jul. 2000.
- [22] H.-T. Lue, Y.-H. Shih, K.-Y. Hsieh, R. Liu, and C.-Y. Lu, "A transient analysis method to characterize the trap vertical location in nitride-trapping devices," *IEEE Electron Device Lett.*, vol. 25, no. 12, pp. 816–818, Dec. 2004.
- [23] E. Suzuki, H. Hiraishi, K. Ishii, and Y. Hayashi, "A low-voltage alterable EEPROM with metal-oxide-nitride-oxide-semiconductor (MONOS) structures," *IEEE Trans. Electron Devices*, vol. ED-30, no. 2, pp. 122–127, Feb. 1983.
- [24] M. Naich, G. Rosenman, and Y. Roizin, "Profiling of deep traps in silicon oxide-nitride-oxide structures," *Thin Solid Films*, vol. 417, no. 1/2, pp. 166–169, Jan. 2005.
- [25] A. Arreghini, N. Akil, F. Driussi, D. Esseni, L. Selmi, and M. van Duuren, "Characterization and modeling of long term retention in SONOS non volatile memories," in *Proc. ESSDERC*, 2007, pp. 406–409.



Antonio Arreghini was born in Udine, Italy, in 1979. He received the Master Laurea degree in electronic engineering from the University of Udine, Udine, Italy, in 2004, where he is currently working toward the Ph.D. degree in industrial and information engineering in the Department of Electrical, Mechanical and Management Engineering.

His research topics include the modeling and the experimental characterization of silicon nitride-based nonvolatile memory cells. From July 2006 to July 2007, he was a Visiting Scientist with NXP Semiconductor, Leuven, Belgium, where he worked on modeling and simulations of long-retention program and erase transients of silicon nitride-based nonvolatile memories.



Francesco Driussi received the Laurea and Ph.D. degrees in electronic engineering from the University of Udine, Udine, Italy, in 2000 and 2004, respectively, working on the reliability of Flash memory devices.

From October 2002 to September 2003, he held a student position with Philips Research Leuven, Leuven, Belgium. Since 2005, he became a Research Associate with the University of Udine, where he is currently with the Department of Electrical, Mechanical and Management Engineering. His research activities are mainly in the field of nonvolatile memory cell reliability and, in particular, the characterization of device degradation and gate dielectric reliability. In particular, he has worked on substrate-enhanced hot-electron phenomena, and he has investigated on their implications for Flash electrically erasable programmable read-only memory devices. More recently, he worked on the development of physical and statistical models for the study of stress-induced leakage current in large Flash memory arrays and of the oxide trap generation and distribution. He is also working on the characterization and modeling of SONOS memory cells and the trapping properties of silicon nitride. Part of his current activities are also in the field of the experimental characterization and modeling of the carrier mobility in MOSFET devices, featuring strained silicon and high- κ gate stacks.



Elisa Vianello was born in Rovereto, Italy, in 1981. She received the Laurea degree in electronic engineering from the University of Udine, Udine, Italy, in 2006, where she is currently working toward the Ph.D. degree in electronic engineering in the Department of Electrical, Mechanical and Management Engineering.

Her research interests are in the field of the characterization and modeling of nonvolatile memory cells and, in particular, SONOS memory cells.



David Esseni (S'98–M'00–SM'06) received the Laurea and Ph.D. degrees in electronic engineering from the University of Bologna, Bologna, Italy, in 1994 and 1998, respectively.

In 1999, he became an Assistant Professor and, later on, an Associate Professor with the University of Udine, Udine, Italy, where he is currently in the Department of Electrical, Mechanical and Management Engineering. In 2000, he was a Visiting Scientist with Bell Laboratories, Lucent Technologies, Murray Hill, NJ. His research activities include characterization techniques for electrical parameters in MOS devices and several aspects related to the hot-electron phenomena in MOSFETs and Flash electrically erasable programmable read-only memories (EEPROMs). In particular, he has studied the reliability of thin oxide transistors, the statistical distribution of stress-induced leakage current in nonvolatile-memory arrays, and the low-voltage and substrate-enhanced hot-electron injection for applications to Flash EEPROM devices. More recently, he has been involved in the experimental characterization and modeling of low-field mobility in ultrathin silicon-on-insulator MOSFETs and in MOSFETs with ultrathin gate oxides. His modeling interests also include the quantization in MOS devices beyond the effective mass approximation and the semiclassical Monte Carlo approach for decananometric MOSFETs.

Dr. Esseni is a member of the Technology Computer Aided Design Committee of the IEEE Electron Devices Society. He was member of the Technical Committee of the IEEE International Electron Devices Meeting, from 2003 to 2004. He is currently in the Technical Committee of the IEEE European Solid State Device Research Conference and the IEEE International Reliability Physics Symposium.



Michiel J. van Duuren was born in Groenlo, The Netherlands, in 1970. He received the M.S. degree in physics and the Ph.D. degree from the University of Twente, Enschede, The Netherlands, in 1993 and 1998, respectively.

After he graduated, he was with Philips Research Laboratories, Eindhoven, The Netherlands, where he started to work on reliability aspects, characterization methods, and physical modeling of floating gate and—since some years—SONOS memories. Since 2001, he has been with Philips Research Leuven (now NXP Semiconductors), Leuven, Belgium, where he is currently leading the “embedded memories” project.



Dusan S. Golubović received the M.S.E.E. degree from the University of Belgrade, Belgrade, Serbia, in 2000 and the Ph.D. degree in applied physics from Katholieke Universiteit Leuven, Leuven, Belgium, in 2005.

Since 2005, he has been a Senior Scientist with Philips Research Leuven (now NXP semiconductors), Leuven. His research interests are in the field of process integration and characterization of advanced embedded nonvolatile memories.



Nader Akil received the M.S. degree in physics from University of Reims Champagne-Ardenne, Reims, France, in 1995 and the Ph.D. degree from Metz University, Metz, France, in 1998. He spent the last year of his Ph.D. in the Department of Electrical and Computer Engineering, Vanderbilt University, Nashville, TN, working on light emission from silicon-based devices.

From 1999 to 2000, he was a Postdoctoral Associate with University of Twente, Enschede, The Netherlands, working on nanoscale silicon-based devices for applications in micrototal analysis systems. In 2000, he was with Philips Research Laboratories, Eindhoven, The Netherlands, and in 2001, he moved to Philips Research Leuven (now NXP Semiconductors), Leuven, Belgium, where he is currently working on modeling and characterization of floating gate and charge trapping memories.



Rob van Schaijk was born in Schaijk, The Netherlands, in 1971. He received the M.S. degree in physics from the Technical University of Eindhoven, Eindhoven, The Netherlands, in 1995, working on optical and transport properties of 2-D hole-gases in GaAs, and the Ph.D. degree from the University of Amsterdam, Amsterdam, The Netherlands, in 1999. His Ph.D. work was on magneto transport in low-dimensional semiconductor systems.

In 1999, he was with Philips Research, Eindhoven, and in 2000, he moved to Leuven, Belgium, to join Philips Research Leuven (now NXP semiconductors), which is in partnership with the Interuniversity Microelectronics Center. His research activities are in the field of process integration of embedded nonvolatile memories. In particular, the work is focused on integration of floating gate Flash and SONOS memory cells.