Lateral Profiling of Trapped Charge in SONOS Flash EEPROMs Programmed Using CHE Injection

P. Bharath Kumar, *Student Member, IEEE*, Pradeep R. Nair, *Student Member, IEEE*, Ravinder Sharma, Shiro Kamohara, *Senior Member, IEEE*, and Souvik Mahapatra, *Member, IEEE*

***Abstract*—The lateral profile of trapped charge in a silicon–oxide–nitride–oxide–silicon (SONOS) electrically erasable programmable read-only memory programmed using channel-hot-electron injection is determined using current–voltage (I_D–V_G) measurements along with two-dimensional device simulations and is verified using gate-induced-drain-leakage measurements, charge-pumping (CP) measurements, and Monte Carlo simulations. An iterative procedure is used to match simulated I_D–V_G characteristics with experimental I_D–V_G characteristics at different stages of programming, by sequentially increasing the trapped electron charge in simulations. Fresh cells are found to contain a high laterally nonuniform trapped charge, which (along with large electron injection during the program) make the conventional CP techniques inadequate for extracting the charge profile. This charge results in a nonmonotonous variation of threshold and flat-band voltages along the channel and makes it impossible to simultaneously determine interface and trapped charge profiles using CP alone. The CP technique is modified for application to SONOS cells and is used to verify the charge profile obtained using I_D–V_G and to estimate the interface degradation. This paper enhances the study presented in our earlier work.**

***Index Terms*—Channel-hot-electron (CHE) injection, charge pumping (CP), gate-induced drain leakage (GIDL), nonuniform charge trapping, silicon-oxide-nitride-oxide-silicon (SONOS) electrically erasable programmable read-only memories (EEPROMs), trapped charge profiling.**

I. Introduction

THERE IS A renewed interest in silicon–oxide–nitride–oxide–silicon (SONOS) (including NROM type) electrically erasable programmable read-only memories (EEPROMs) as floating-gate (FG) EEPROMs are approaching their scaling limits [2]–[6]. In addition to their better scalability compared to FG EEPROMs, SONOS EEPROMs are also easier to fabricate making them suitable for embedded applications [4], [5]. They also provide a 2-bit/cell operation, when the charge storage is laterally localized in the nitride by using localized injection mechanisms like channel-hot-electron (CHE) injection and band-to-band-tunneling (BTBT)-induced hot hole injection (HHI) [2], [4]. The knowledge about the location and the spread of this trapped charge is very important to ensure a reliable 2-bit operation. The ability to predict and adjust trapped-electron and trapped-hole profiles is crucial for cell endurance. In addition, the lateral extent of trapped charge also determines the scalability of these SONOS cells.

The charge-pumping (CP) technique is a well-known technique used for profiling trapped charge and interface states in MOSFETs, after a hot-carrier stress [7]–[13]. Recently, a method based on subthreshold I_D–V_G characteristics (I–V) and gate-induced-drain-leakage (GIDL) current along with two-dimensional (2-D) device simulations was proposed for extracting the spatial distribution of CHE injection [14]. However, the lateral spread of trapped charge in a programmed SONOS cell obtained from these methods differs widely [13], [14].

In this paper, we use all the above methods (I–V, GIDL, and CP) to comprehensively establish the lateral profile of trapped electrons in a programmed SONOS cell and remove the ambiguity [1]. We also use Monte Carlo simulations to determine the lateral CHE distribution. We show that the spread of trapped electrons can be estimated by using I–V measurements and 2-D simulations only. Importantly, we demonstrate that the traditional way of profiling trapped charge using CP can give erroneous results in SONOS cells due to large charge trapping involved. The virgin cells also may have a large trapped charge, making it impossible to determine both the trapped charge and interface-trap profiles simultaneously using the traditional CP technique. After making necessary modifications to the extraction technique, we use it to verify that the charge profile obtained using I–V gives back measured CP characteristics. In addition to [1], we show that many distributions of the trapped charge can predict CP characteristics, but all cannot be consistent with measured I–V characteristics. This makes it important to comprehensively determine the lateral charge profile by simultaneously using various available methods. We also show independently from CP results that the lateral spread of trapped electrons cannot be much larger than that obtained from I–V.

Section II presents device and experimental details, along with the simulation method. The first part of Section III describes the charge-profiling method using only I–V measurements and 2-D device simulations. After showing that Monte Carlo simulations also predict similar results, the sensitivity of simulated I–V characteristics to changes in charge spread and magnitude is explored. Finally, GIDL characteristics are used to verify the magnitude of charge trapped in the gate–drain overlap

Manuscript received October 25, 2005. The review of this paper was arranged by Editor R. Shrivastava.

P. B. Kumar and S. Mahapatra are with the Department of Electrical Engineering, Indian Institute of Technology (IIT) Bombay, Mumbai 400076, India (e-mail: bharath@ee.iitb.ac.in).

P. R. Nair was with the Department of Electrical Engineering, IIT Bombay, Mumbai 400076, India. He is now with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285 USA.

R. Sharma was with the Department of Electrical Engineering, IIT Bombay, Mumbai 400076, India. He is now with the STMicroelectronics Pvt. Ltd., Noida 201301, India.

S. Kamohara is with Renesas Technology Corporation, Tokyo 100-6334, Japan.

Digital Object Identifier 10.1109/TED.2006.870533

region. The second part of Section III discusses the application of CP in extracting the lateral profile of trapped charge in SONOS cells. After showing the inadequacy of the traditional method, a modification for SONOS cells is proposed. This is then used to verify that the lateral extent of trapped charge obtained using I–V is consistent with CP measurements.

II. Experimental and Simulation Details

Experiments were performed on isolated SONOS memory cells having top-oxide/nitride/bottom-oxide (ONO) dimensions of 5.8/8/5 nm, width of 2 μm, and ONO stack length of 0.25 μm. Programming was done using CHE injection in small time steps (equi-spaced in logarithmic time scale) with intermediate I–V and GIDL measurements. The bias used was $V_D = 5$ V and $V_S = V_B = 0$ V, with V_G ranging between 7–9 V (fixed for a cell). I–V measurements were done at $V_D = 0.1$ V in the reverse mode [4] by exchanging source and drain terminals (V_D is applied at the terminal acting as the source during program). Reverse-mode I–V shows a higher sensitivity to the trapped electron charge compared to forward read; as a result, less charge injection is needed for similar threshold-voltage shifts [4], [14]. The GIDL current (substrate current) was measured while ramping V_G to negative values with V_D fixed at 2 V. CP measurements were done at the start and at the end of programming using two 400-kHz pulses, one with a fixed base (-4.6 V) and a varying top and the other with a fixed top (4.6 V) and a varying base [7].

The device used for simulations was obtained using Integrated Systems Engineering (ISE) Technology Computer-Aided Design (TCAD) process simulator, DIOS [15]. I–V and GIDL characteristics were simulated using ISE TCAD device simulator DESSIS [15]. Hot electron distributions were simulated in a nonself-consistent full-band Monte Carlo simulator [16] using field information from DESSIS. At the end of a process simulation, the ONO stack was replaced by an oxide of equivalent electrical thickness to facilitate Monte Carlo simulations that could handle only Si and SiO_2. The effect of the charge distributed throughout the ONO stack can be captured using an effective charge at the interface, if second-order effects could be neglected. Here, the charge was placed at the interface of Si and the oxide, using a series of charge packets about 10-nm wide (local grid spacing is about 2 nm). Fig. 1 shows the position of charge packets in a schematic of the cell before replacing the ONO stack with an oxide of equivalent thickness.

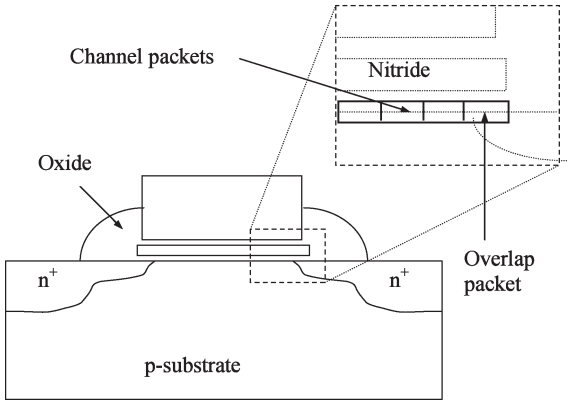


Fig. 1. Schematic of a SONOS cell with the expanded portion showing the scheme used to simulate the effect of trapped charge by placing charge packets at the interface. Channel packets always start at the drain (and source) end and extend into the channel. During simulations, the ONO stack was replaced with an oxide of equivalent electrical thickness.

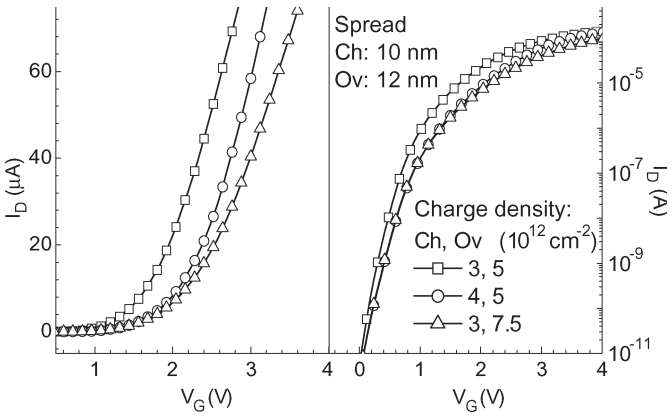


Fig. 2. Simulated I–V with the charge placed in two rectangular packets at the interface (Fig. 1). Increasing the magnitude of a trapped charge (negative) above the channel increases $V_{\rm TH}$ and SS. Increasing the overlap charge magnitude changes the slope in the linear region along with a small increase in $V_{\rm TH}$. A much larger overlap charge is needed compared to the channel charge to get a similar $V_{\rm TH}$ shift. Increasing the overlap charge spread (at a constant charge density) beyond the edge of the ONO stack has a negligible effect on I–V (not shown here).

III. Results and Discussion

A. Trapped Charge Profiling Using I–V

The relation between the lateral extent of charge trapping above the channel region and the subthreshold slope (SS) of I–V has been studied well [14], [17]. This, together with the dependence of GIDL current on the charge trapped above the drain junction (in the gate–drain overlap region, referred to as overlap from now) was used as the basis of trapped-charge-profiling technique in [14]. However, charge trapped in the overlap can also degrade the current in the linear region of I–V, as shown in Fig. 2. Electrons trapped in the overlap region can locally reduce the conductance, which is manifested as a series-resistance effect. We show that this makes it possible to estimate the charge trapped in the overlap using I–V alone (instead of GIDL measurements).

1) Extraction of Program Charge Profile: Charge can also be trapped in the ONO stack of SONOS cells during fabrication [18]. The variation in I–V characteristics of virgin cells measured across a wafer suggests the presence of a spatially nonuniform charge trapping in virgin cells [1]. During simulation of virgin cells, placing a negative charge in the gate stack was found to be essential, as simulated subthreshold characteristics of a virgin cell could not be matched with experimental characteristics by varying simulation parameters alone. For this, charge packets were placed symmetrically on both source and drain ends of the Si-SiO_2 interface (Fig. 1). A match with experimental I–V characteristics was obtained by adjusting the charge in these packets together with mobility parameters. As explained below, being able to consistently match all the

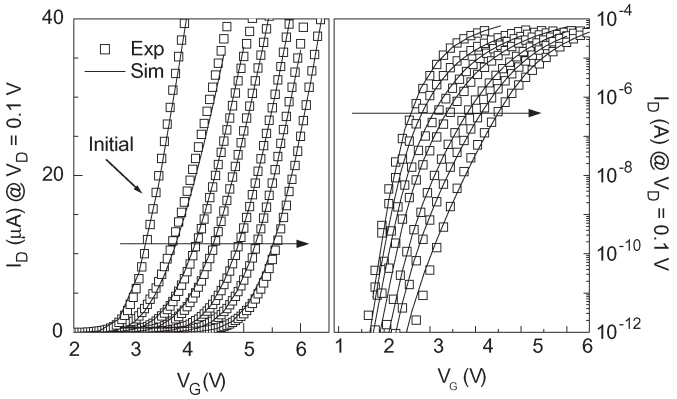


Fig. 3. Measured I–V characteristics during programming ($V_G/V_D/V_B = 8/5/0$ V, total programming time $= 10$ μs) along with the simulated ones. Simulations are done by placing the charge in four rectangular packets (12 nm in the overlap and three 10-nm packets above the channel) along the interface. The initial simulated curve is matched by placing the identical charge on both source and drain sides. Curves during programming are matched by gradually adding a negative charge to the drain side only (iteratively until all curves are matched).

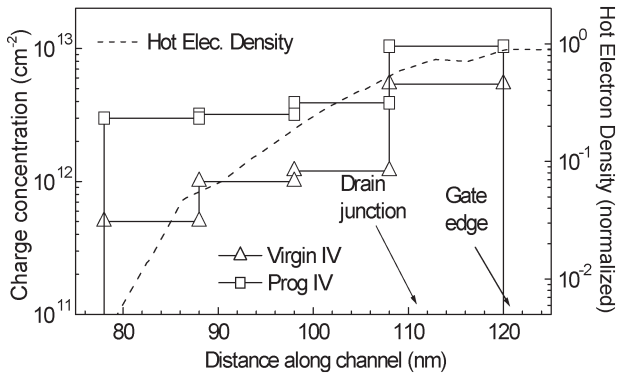


Fig. 4. Charge (negative) profile on the drain side in virgin and programmed cells obtained from I–V simulations (Fig. 3). An additional uniform sheet of charge present throughout the interface in both virgin and programmed devices is not shown for clarity. The CHE profile obtained from the Monte Carlo simulation of the virgin device at programming bias is also shown.

subsequent programming I–Vs makes this profile more likely, as any arbitrary profile would not achieve this. The charge in a virgin cell was found to be highest at the gate edges, gradually decreasing towards the center of channel. This charge can have major implications in applying the CP technique for charge profiling, as shown in Section III-B.

I–V characteristics measured at various intermediate stages while programming the SONOS cell are shown in Fig. 3. Distribution of electrons trapped in the ONO stack of a programmed cell was found using an iterative procedure [14] by matching simulated and experimental I–Vs while satisfying the following conditions: 1) I–V characteristics at intermediate stages of programming are simulated sequentially and 2) only a negative charge is added (in the drain half of the channel) to proceed from one programmed state to the next. The latter condition means that any lateral redistribution of trapped electrons during programming or intermediate I–V measurements is neglected here. The process of sequentially matching all intermediate I–Vs makes the profile more likely. Although more than one charge profile can simulate the final I–V, all I–Vs starting from the virgin state cannot be obtained consistently unless that profile is close to reality. As a further check, the sensitivity of simulated I–V values to changes in the obtained charge profile is discussed in next part of this section.

Programming I–V characteristics simulated using the above method are shown in Fig. 3, for a cell programmed at $V_G/V_D/V_B = 8/5/0$ V. During the initial stages of programming, most of the charge (electron) trapping was found to occur in the overlap region. This can also be observed from the sudden decrease of the slope in the linear region of I–V after the very first program pulse. Fig. 4 plots charge distributions corresponding to the virgin and final programmed states. The figure shows a charge spread of about 40 nm (in agreement with [14]). For the same $V_{\rm TH}$ shift, a similar charge spread (40–50 nm) was obtained for cells programmed (separately) at biases of 7/5/0 and 9/5/0 V. (A uniform sheet of charge extending throughout the interface was present in both virgin and programmed states and is not shown, as it only shifts the I–V curves in parallel. The magnitude of this charge is less than the injected charge.)

Monte Carlo simulations are used to separately get an estimate of the hot electron distribution in the channel. For this purpose, a virgin device (along with its charge) was first simulated at a program bias in DESSIS. The Monte Carlo simulator was subsequently used to get the hot electron distribution along the channel. Fig. 4 plots this hot electron distribution in a virgin cell under the same program bias and it predicts a similar spread of trapped charge as the I–V method. However, it should be noted that injected charge depends on the fraction of these electrons directed towards interface and the presence of a favorable oxide field. This field dynamically changes with charge trapping during programming and affects subsequent charge injection. The simulation of a programmed cell (with the corresponding charge) at the same bias shows an increase in the region of opposing oxide field and a slight decrease in the CHE density. While the exact dynamics of injection of a charge into an ONO stack and its subsequent transport are beyond the scope of this paper, it is clear that for reasonable shifts in $V_{\rm TH}$ the lateral spread of a trapped charge cannot be significantly larger than that obtained above.

2) Correctness of the Charge Profile: To verify the charge profile obtained above, the sensitivity of simulated I–V characteristics to changes in this profile is studied and an independent estimation of the overlap charge is made using GIDL current. Increasing the charge spread inside channel by adding a 10-nm charge packet (having the same charge as the neighboring packet) decreased the SS as shown in Fig. 5. To get back the actual SS, charge in the new packet had to be decreased to a very small value, which has a negligible effect on I–V. Decreasing the original spread by removing the packet towards the center of the channel increased the SS as shown in Fig. 5. Modifying the charge in other packets could not revert back the SS to experimental value. Furthermore, using a profile with a much different channel spread (with magnitude changed accordingly) cannot consistently simulate all the intermediate I–V characteristics. Varying the magnitude of charge trapped above the channel mainly shifts I–V characteristics along the

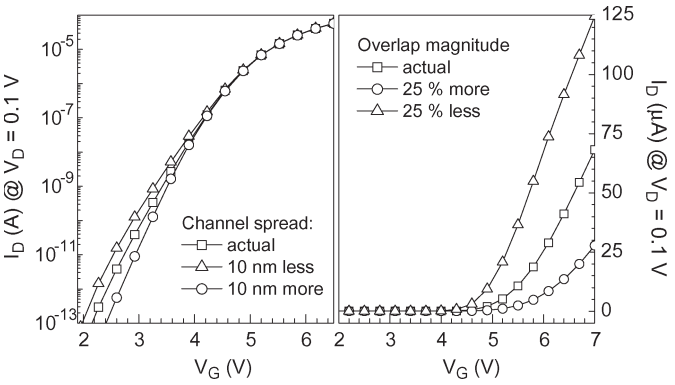


Fig. 5. Changing the spread of charge in the channel of programmed cell mainly affects SS of simulated I–V. Without the last packet (78–88 nm), the slope could not be recovered by changing the charge in existing packets. Matching the slope using an additional packet (68–78 nm) required a negligibly small charge in that packet. Changing the magnitude of the overlap charge (negative) changes the slope in linear region along with a slight shift in $V_{\rm TH}$.

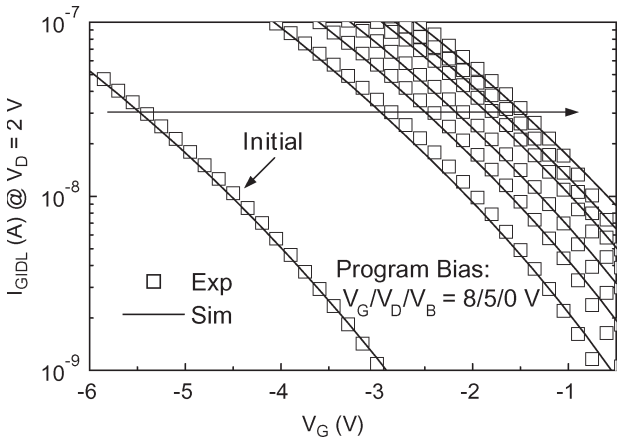


Fig. 6. Measured and simulated GIDL currents at various stages of programming. Charge packets are used during simulations as in Fig. 3 (similar to [14]). To match the initial curve, simulation parameters are adjusted using the charge distribution from I–V. Programmed curves are matched by adding the charge on the drain side (independent of I–V). A large change in $I_{\rm GIDL}$ at the start of a program indicates a large electron injection into the overlap region and supports a similar observation from I–V characteristics.

V_G axis. However, it should also be noted that the actual trapped charge magnitude depends on the vertical distribution of the charge in the ONO stack, which cannot be determined by this method.

The magnitude of the overlap charge largely affects the slope of I–V in the linear region (Fig. 5). On the other hand, the spread of the overlap charge beyond the ONO stack has no effect on I–V, as mentioned earlier. The GIDL current is highly sensitive to any charge trapped in the overlap region [14], as it originates from BTBT in the drain region under the gate–drain interface [19], [20]. Here, a separate estimate of the overlap charge was obtained using GIDL measurements and simulations. The GIDL characteristics of a virgin cell were simulated using the virgin trapped charge profile obtained from I–V matching and by adjusting BTBT model parameters. Once a good match was obtained, the parameter values were fixed and program characteristics were simulated by adding a negative charge at the drain end as in the I–V method discussed earlier. Fig. 6 shows experimental and simulated GIDL currents for the programming shown in Fig. 3. Even GIDL curves suggest a large injection of negative charge into the overlap at the start of programming. In this case, the overlap charge obtained for the final programmed state closely matches the charge obtained using I–V (within 5%), indicating that the I–V method alone is adequate to predict this. Even though I–V is comparatively less sensitive to overlap charge in these cells, the large magnitude of the overlap charge helps in a reasonably accurate prediction.

B. Profiling Using CP

1) Traditional CP Method and its Limitations: CP method is a well-known technique used to extract the spatial distribution of oxide trapped charge and interface degradation, due to hot-carrier stress in MOSFETs [7]–[13]. Ideally, CP characteristics measured before and after the stress can be used to directly determine the lateral distribution of trapped charge $(N_{\rm OT})$ along with generated interface traps $(\Delta N_{\rm IT})$. In brief, CP measurement involves using a gate pulse to repeatedly drive parts of the channel (CP region) between inversion and accumulation. The two levels of gate pulse ($V_{\rm BASE}$ and $V_{\rm TOP}$) along with local threshold $(V_{\rm TH})$ and flat-band $(V_{\rm FB})$ voltages of a MOSFET determine the CP region (region with $V_{\rm BASE} < V_{\rm FB}$ and $V_{\rm TOP} > V_{\rm TH}$). Different modes of CP are possible based on how $V_{\rm BASE}$ and $V_{\rm TOP}$ are varied [7]. The CP current $(I_{\rm CP})$ is directly proportional to the number of interface traps $(N_{\rm IT})$ in the CP region. A spatial distribution of $N_{\rm OT}$ and $\Delta N_{\rm IT}$ can be extracted from CP characteristics as $N_{\rm OT}$ changes local $V_{\rm TH}$ and $V_{\rm FB}$ profiles and $\Delta N_{\rm IT}$ changes the current contribution from a given region. This is the basic principle behind various CP techniques used for profiling. However, all the techniques proposed so far explicitly or implicitly assume that $V_{\rm TH}$ and $V_{\rm FB}$ profiles vary monotonously in either half of the channel, having a one-to-one correspondence with the position in channel [7]–[13].

This assumption becomes questionable when the large charge trapping in SONOS cells (also referred to as $N_{\rm OT}$) is considered. It can become invalid even in fresh cells, which may have very high charge trapping near the gate edges (Fig. 4). This highly nonuniform trapped charge increases further with programming. As a check, $V_{\rm TH}$ and $V_{\rm FB}$ profiles in a virgin device were simulated using 2-D device simulations. Initially, profiles in a cell with no charge were simulated (using channel electron and hole concentrations of 10^{16} cm^{-3} to define $V_{\rm TH}$ and $V_{\rm FB}$, respectively), and the contribution of the trapped charge was added to these ([1, Fig. 8]). The cell without charge has peak $V_{\rm FB}$ and $V_{\rm TH}$ at the center of the channel, as expected. However, in the presence of charge $V_{\rm FB}$ shows two peaks in either half of the channel and $V_{\rm TH}$ has peaks near either end of the channel. Thus, it may not be always possible to unambiguously determine $V_{\rm FB}$ profile from experiments, unless the charge distribution is known. Here, $V_{\rm TH}$ profile can still be determined under some approximations, as it largely remains monotonous except towards the end of the channel. However, this may not be true in general, and the initial charge is also not always available to perform a check.

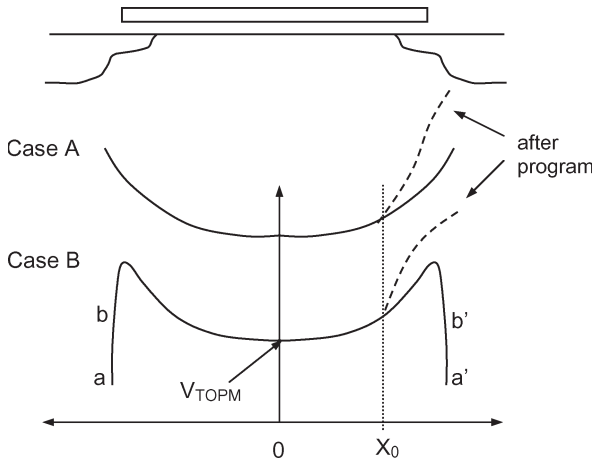


Fig. 7. Schematics of two possible $V_{\rm TH}$ profiles along the channel in virgin and programmed cells. In the virgin cell, with an increasing $V_{\rm TOP}$ (when $V_{\rm BASE}$ is below $V_{\rm FB}$), the initial contribution to $I_{\rm CP}$ comes from the center of the channel in case A and from the ends of channel (a–b and a′-b′) in case B. After programming, in case A, $I_{\rm CP}$ remains the same until X_0 is reached; whereas in case B, $I_{\rm CP}$ becomes half (a′–b′ does not contribute) until the center of the channel starts contributing.

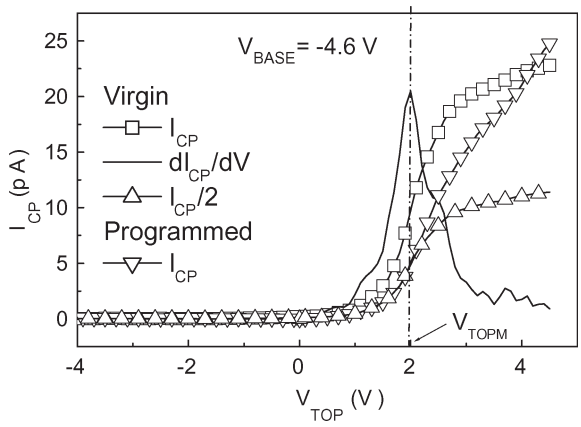


Fig. 8. CP current measured as a function of $V_{\rm TOP}$ before and after programming. $I_{\rm CP}$ after the program is half of the virgin $I_{\rm CP}$ until the center of the channel starts participating in the CP (peak of $dI_{\rm CP}/dV$ occurs at this $V_{\rm TOP}$) indicating that $V_{\rm TH}$ profile is closer to case B than case A (Fig. 7).

Both $V_{\rm TH}$ and $V_{\rm FB}$ profiles are needed to determine $N_{\rm OT}$ and $\Delta N_{\rm IT}$ profiles simultaneously. This is not possible here as only the virgin $V_{\rm TH}$ profile (and not the $V_{\rm FB}$ profile) can be obtained from measurements. Some techniques using a single profile require an additional step for neutralizing the trapped charge [12], so that only $\Delta N_{\rm IT}$ changes between measurements. However, in this case, there is no known way to selectively neutralize $N_{\rm OT}$. $N_{\rm OT}$ can be also be extracted when $\Delta N_{\rm IT}$ is negligible [13]; even this was not true here. Therefore, we use CP only to verify the $N_{\rm OT}$ profile obtained from I–V. Even for this, the CP technique required a modification that is shown in the next part of this section. After modification, it was used to verify if the experimental CP characteristics are consistent with the $N_{\rm OT}$ profile obtained from the I–V method. This technique can also be used to determine the $\Delta N_{\rm IT}$ profile when $N_{\rm OT}$ is known.

2) CP Technique for a Programmed SONOS Cell: $N_{\rm OT}$ in the virgin SONOS cells can result in two different $V_{\rm TH}$ profiles shown in Fig. 7, with the peak either at the edge of the ONO stack or slightly inside. In either case, programming shifts this peak towards the edge of the ONO stack due to large charge injection in the overlap region. Comparing experimental $I_{\rm CP}$ before $(I_{\rm CPI})$ and after $(I_{\rm CPP})$ programming distinguishes the two possible virgin-cell profiles. Let the charge injection take place starting from X_0, as shown in Fig. 7, and let the corresponding $V_{\rm TH}$ be V_{T0}. When virgin $V_{\rm TH}$ peaks at the gate edge (case A), the CP region starts from the center of the channel as $V_{\rm TOP}$ starts increasing. Thus, $I_{\rm CPP}$ remains the same as $I_{\rm CPI}$ for $V_{\rm TOP} < V_{T0}$. However, when $V_{\rm TH}$ peaks inside the channel (case B), the CP region initially starts from edges (regions a–b and a′–b′ in Fig. 7) before the center of the channel participates. After programming, as the CP region starts only from source side (a–b), $I_{\rm CPP}$ remains half of $I_{\rm CPI}$ until the center of channel takes part in the CP. As the $V_{\rm TH}$ profile is flat near the center of the channel, $dI_{\rm CP}/dV_{\rm TOP}$ peaks at this point and the corresponding $V_{\rm TOP}$ value gives the local $V_{\rm TH}$ $(V_{\rm TOPM})$. Fig. 8 shows experimental CP characteristics before and after program, which support case B for the cells used here.

The $V_{\rm TH}$ profile inside the channel of a virgin cell $(V_{\rm TI})$ can be found if following assumptions are made. 1) Before programming, $N_{\rm IT}$ is uniform along the channel ($N_{\rm IT}$ (x) is constant). 2) Eliminating the contribution to $I_{\rm CP}$ from regions with $V_{\rm TOP} < V_{\rm TOPM}$ gives the contribution from the region between the two peaks of $V_{\rm TI}$ [or $I_{\rm CPI,EFF}(V_{\rm TOP}) = I_{\rm CPI}(V_{\rm TOP}) - I_{\rm CPI}(V_{\rm TOPM})$]. 3) The peak of $V_{\rm TI}$ occurs very close to the junction so that the extent of the region between peaks of $V_{\rm TI}$ is the same as $L_{\rm EFF}$. 4) CP covers the full channel ($V_{\rm TOP}$ is increased up to the peak $V_{\rm TI}$, after which $I_{\rm CP}$ stops increasing).

The following expression gives $V_{\rm TI}$ as a function of x ($V_{\rm TI} = V_{\rm TOP}$ when $I_{\rm CP}$ varies with $V_{\rm TOP}$):

$$I_{\rm CPI,EFF}(V_{\rm TOP}) = 2qWfN_{\rm IT}x \tag{1}$$

where, f is the frequency of gate pulse. Here, the center of channel is located at $x = 0$, and $V_{\rm TI}$ is symmetric on either side. Using this estimated $V_{\rm TI}(x)$ and $N_{\rm OT}(x)$ obtained from I–V, the $V_{\rm TH}$ profile after programming $[V_{\rm TP}(x)]$ can be found. Here, a rectangular box profile of $N_{\rm OT}$ is converted to an approximate piecewise-linear function of x. In addition, any contribution to $V_{\rm TP}$ from $\Delta N_{\rm IT}$ is neglected.

$I_{\rm CPP}$ is related to $I_{\rm CPI}$, $V_{\rm TP}$, and $\Delta N_{\rm IT}$ and can be calculated using the following relations.

For $V_{\rm TOP} < V_{\rm TOPM}$

$$I_{\rm CPP}(V_{\rm TOP}) = \frac{I_{\rm CPI}(V_{\rm TOP})}{2}. \tag{2}$$

For $V_{\rm TOPM} < V_{\rm TOP} < V_{T0}$ (or equivalently, for $x < X_0$)

$$I_{\rm CPP}(V_{\rm TOP}) = I_{\rm CPI}(V_{\rm TOP}) - \frac{I_{\rm CPI}(V_{\rm TOPM})}{2} \tag{3a}$$

$$V_{\rm TOP} = V_{\rm TP}(x) = V_{\rm TI}(x). \tag{3b}$$

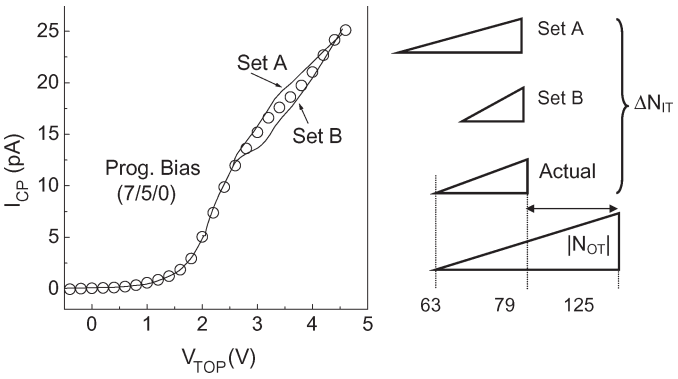


Fig. 9. Best fit simulated $I_{\rm CP}$ when the $\Delta N_{\rm IT}$ spread (inside the channel) is 5 nm more (A) or less (B) than the $N_{\rm OT}$ spread. $\Delta N_{\rm IT}$ is shown only in the region accessible to the CP after programming (up to 79 nm). $I_{\rm CP}$ is always underestimated or overestimated compared to the actual value. ($N_{\rm OT}$ is schematically shown as a triangle but is actually a piecewise linear profile).

For $V_{\rm TOP} > V_{T0}$ (for $x > X_0$)

$$I_{\rm CPP}(V_{\rm TOP}) = \frac{I_{\rm CPI}(V_{\rm TOP})}{2} + \frac{I_{\rm CPI}(V_{\rm TI})}{2} - \frac{I_{\rm CPI}(V_{\rm TOPM})}{2} + \Delta N_{\rm IT}\ \text{contribution (from } X_0 \text{ to } x) \quad (4a)$$

$$V_{\rm TOP} = V_{\rm TP}\,(x) = V_{\rm TI}\,(x) + \frac{q\,N_{\rm OT}\,(x)}{\rm C_{OX}}. \quad (4b)$$

In (4a), $I_{\rm CPI}\,(V_{\rm TOP})/2$ is the contribution from the source side of channel while the rest indicates contribution from the drain side.

$\Delta N_{\rm IT}$ is usually assumed to have the same shape as $N_{\rm OT}$ [8], [12]. In this paper, to calculate $I_{\rm CPP}$, $\Delta N_{\rm IT}$ is approximated to have a triangular shape and to exist only in the region of charge injection. The best fit $I_{\rm CPP}$ is obtained by iteratively adjusting the magnitude of $\Delta N_{\rm IT}$ and its spread towards the gate edge. The latter is related to the extent of CP region towards the gate edge. An accurate $I_{\rm CPP}$ could be calculated this way, and Fig. 9 shows the schematic of $\Delta N_{\rm IT}$ for a device programmed at 7/5/0 V. It was also seen that $I_{\rm CPP}$ could not be obtained when the $\Delta N_{\rm IT}$ profile starts at a different point in the channel compared to $N_{\rm OT}$. The best fit $I_{\rm CPP}$ for two cases, with the starting point of $\Delta N_{\rm IT}$ shifted by 5 nm, is also shown in Fig. 9. $I_{\rm CPP}$ is always underestimated or overestimated at some parts of the curve.

3) Accuracy of the CP Method: CP technique alone may not be sufficient for profiling charge in SONOS cells and different combinations of $N_{\rm OT}$ and $\Delta N_{\rm IT}$ can give well-matched CP characteristics as shown below. But all such profiles are not realistic and do not give the expected I–V curves. Moreover, CP can probe only to a limited extent into the charge-trapping region (up to the point where $V_{\rm TP} < V_{\rm TOP}$), depending on maximum $V_{\rm TOP}$ used. Therefore, using trapped charge obtained from I–V method to match CP characteristics cannot guarantee accuracy of this charge in regions beyond the reach of CP ($V_{\rm TP} > V_{\rm TOP}$). Nevertheless, CP can be used to predict the charge spread towards the center of channel as this region participates in CP even after programming.

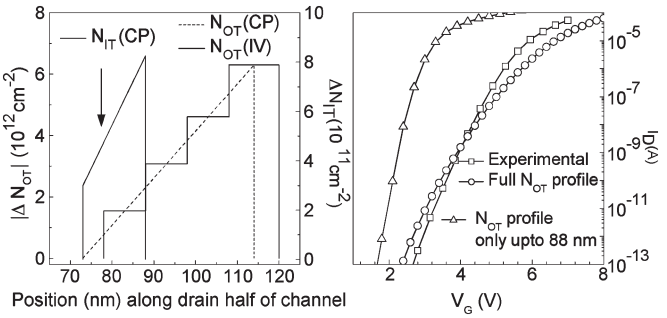


Fig. 10. A triangular (instead of piecewise linear) $N_{\rm OT}$ profile gives well-matched CP characteristics (not shown) with the $\Delta N_{\rm IT}$ profile shown. Here, CP probes only up to 88 nm into channel, and the $N_{\rm OT}$ beyond that point has no effect on $I_{\rm CP}$. The I–V simulated using this profile (after converting into a box profile as shown) did not match with the experimental I–V. The $\Delta N_{\rm IT}$ profile is also unphysical, with its magnitude abruptly increasing at a point in channel.

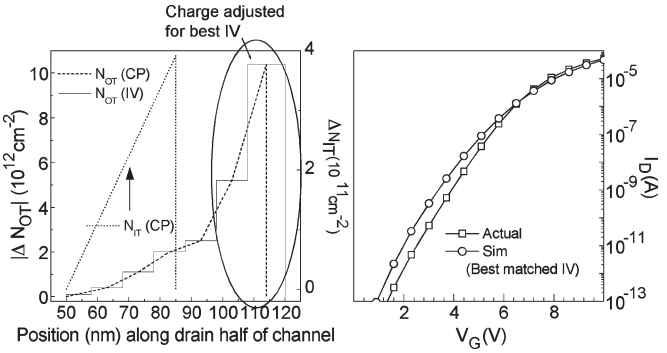


Fig. 11. $N_{\rm OT}$ profile with a large lateral spread gives well-matched CP characteristics with the triangular $\Delta N_{\rm IT}$ profile shown ($\Delta N_{\rm IT}$ is assumed to start at the same point in the channel as $N_{\rm OT}$). A much-less charge was required in packets between 50–70 nm, showing that the charge spread inside the channel cannot be much beyond 70 nm. The best matched I–V obtained after adjusting the charge beyond the CP region is also shown.

First, a triangular $N_{\rm OT}$ profile having same peak value and lateral extent of charge profile obtained from I–V is assumed. It needed a trapezoidal $\Delta N_{\rm IT}$ for matching $I_{\rm CPP}$, as shown in Fig. 10(a). However, such a profile with $\Delta N_{\rm IT}$ abruptly increasing at a point is unrealistic. The extent of $N_{\rm OT}$ beyond the CP region ($>$ 88 nm, here) has no effect on $I_{\rm CP}$ but can significantly affect the I–V characteristics. The wide variation in I–V curves for two different spreads of $N_{\rm OT}$ is shown in Fig. 10(b).

The possibility of $N_{\rm OT}$ having a large lateral extent is explored next. A piecewise linear $N_{\rm OT}$ profile (which can be easily converted into a box profile for simulating I–V values) starting deep inside the channel and a triangular $\Delta N_{\rm IT}$ profile starting at the same point are assumed. The $N_{\rm OT}$ profile was adjusted along with the magnitude of $\Delta N_{\rm IT}$ to get a well-matched $I_{\rm CPP}$. After converting to a box profile, $N_{\rm OT}$ in the region beyond the reach of CP was adjusted to get a best fit with experimental I–V. The resultant $N_{\rm OT}$ and $\Delta N_{\rm IT}$ profiles along with the I–V characteristics are shown in Fig. 11. While a well-matched $I_{\rm CPP}$ could be obtained (not shown), I–V characteristics could not be matched well. But importantly, to match $I_{\rm CPP}$, the magnitude of $N_{\rm OT}$ deep inside the channel (between 50–70 nm in Fig. 11) had to be made very small (these

magnitudes have negligible effects on I–V characteristics). This independently shows that the lateral extent of the trapped charge cannot be deeper inside the channel, than predicted by I–V.

In CP measurements, the extent of the channel probed in both virgin and programmed cells is restricted by the maximum pulse voltage used. Due to large charge trapping in programmed cells, scanning the full channel may require high voltages, which can disturb the existing charge distribution. Therefore, CP measurements have limited utility in extracting a full charge profile. They can be used to predict the lateral extent of a trapped charge inside a channel, in cases where charge injection is minimum towards center of the channel and gradually increases toward the gate edges. They can also be used in cases where the magnitude of a trapped charge is low or can be selectively decreased by neutralizing. When applicable, CP is a better choice compared to charge profiling using I–V and 2-D device simulations. The complex nature of the I–V method also makes it difficult to guarantee the uniqueness of the profile. CP is also more sensitive to localized hole injection compared to I–V. At the same time, neither of these methods gives any information of charge distribution in the vertical direction (in a gate stack) but can only give an estimate of the lateral spread.

IV. Conclusion

The lateral profile of trapped electrons in a SONOS cell (programmed by CHE injection) is determined using I–V measurements with 2-D device simulations, and verified using GIDL and CP characteristics, and Monte Carlo simulations. An iterative procedure is used to simulate I–V characteristics while varying the charge placed at the Si-SiO_2 (bottom oxide) interface to obtain a match with experimental I–V characteristics. Virgin cells used in this study are shown to have a large, non-uniform trapped charge (negative and symmetric on either side of the channel) with magnitude gradually decreasing from gate edge to the center of the channel. These cells also show a large injection of electrons into gate–drain overlap during the initial stages of programming, which is separately confirmed from GIDL characteristics. The lateral extent of trapped electrons is about 40–50 nm for a $V_{\rm TH}$ shift of 2.5 V, and the cell I–V characteristics are found to be sensitive to changes in this charge spread. The CHE distribution simulated using the Monte Carlo method also predicts a similar spread.

Conventional CP-based techniques for profiling $N_{\rm OT}$ and $\Delta N_{\rm IT}$, which require monotonous $V_{\rm TH}$ and $V_{\rm FB}$ profiles along the interface in either half of the channel, are shown to fail in SONOS cells due to the large amount of trapped charge in both virgin and programmed cells. The nonmonotonous $V_{\rm TH}$ and $V_{\rm FB}$ profiles in SONOS cells make it impossible to simultaneously determine $N_{\rm OT}$ and $\Delta N_{\rm IT}$ using normal CP measurements alone. The technique is modified for the case where $V_{\rm TH}$ of a virgin SONOS cell largely remains monotonous, except near the gate edges. This modified technique is used to verify the $N_{\rm OT}$ profile (in the channel region) obtained from I–V. Results show that while the obtained $N_{\rm OT}$ profile may not be unique, its spread inside the channel cannot be much wider than that predicted by I–V. The modified technique can also be used to find either $N_{\rm OT}$ or $\Delta N_{\rm IT}$ in SONOS cells, when one of them is known.

Acknowledgment

The authors would like to thank J. D. Bude who was with Agere Systems, Allentown, PA, for providing the Monte Carlo simulator.

References

[1] P. R. Nair, P. B. Kumar, R. Sharma, S. Kamohara, and S. Mahapatra, "A comprehensive trapped charge profiling technique for SONOS flash EEPROMs," in *IEDM Tech. Dig.*, 2004, pp. 403–406.

[2] T. Y. Chan, K. K. Young, and C. Hu, "A true single-transistor oxide-nitride-oxide EEPROM device," *IEEE Electron Device Lett.*, vol. EDL-8, no. 3, pp. 93–95, Mar. 1987.

[3] M. H. White, Y. L. Yang, A. Purwar, and M. L. French, "A low voltage SONOS nonvolatile semiconductor memory technology," *IEEE Trans. Compon., Packag., Manuf. Technol. A*, vol. 20, no. 2, pp. 190–195, Jun. 1997.

[4] B. Eitan, P. Pavan, I. Bloom, E. Aloni, A. Frommer, and D. Finzi, "Can NROM, a 2-bit, trapping storage NVM cell, give a real challenge to floating gate cells?," in *Proc. SSDM*, 1999, pp. 522–524.

[5] E. J. Prinz, G. L. Chindalore, K. Harber, C. M. Hong, C. B. Li, and C. T. Swift, "An embedded 90 nm SONOS flash EEPROM utilizing hot electron injection programming and 2-sided hot hole injection erase," in *Proc. IEEE NVSMW*, 2003, pp. 56–57.

[6] G. Xue, J. Van Houdt, D. Wellekens, L. Haspeslagh, P. Hedrickx, J. De Vos, and H. E. Maes, "Low voltage low cost nitride embedded flash memory cell," in *Proc. IEEE NVSMW*, 2003, pp. 62–64.

[7] G. Groeseneken, H. E. Maes, N. Beltran, and R. F. De Keersmaecker, "A reliable approach to charge-pumping measurements in MOS transistors," *IEEE Trans. Electron Devices*, vol. ED-31, no. 1, pp. 42–53, Jan. 1984.

[8] M. Tsuchiaki, H. Hara, T. Morimoto, and H. Iwai, "A new charge pumping method for determining the spatial distribution of hot-carrier-induced fixed charge in p-MOSFETs," *IEEE Trans. Electron Devices*, vol. 40, no. 10, pp. 1768–1779, Oct. 1993.

[9] W. K. Chim, S. E. Leang, and D. S. H. Chan, "Extraction of metal-oxide-semiconductor field-effect-transistor interface state and trapped charge spatial distributions using a physics-based algorithm," *J. Appl. Phys.*, vol. 81, no. 4, pp. 1992–2001, Feb. 1997.

[10] S. Mahapatra, C. D. Parikh, J. Vasi, V. R. Rao, and C. R. Viswanathan, "Direct charge pumping technique for spatial profiling of hot-carrier-induced interface and oxide traps MOSFETs," *Solid State Electron.*, vol. 43, no. 5, pp. 915–922, May 1999.

[11] Y. L. Chu, D. W. Lin, and C. Y. Wu, "A new charge-pumping technique for profiling the interface-states and oxide-trapped charges in MOSFETs," *IEEE Trans. Electron Devices*, vol. 47, no. 2, pp. 348–353, Feb. 2000.

[12] A. M. Martirosian and T. P. Ma, "Improved charge-pumping method for lateral profiling of interface traps and oxide charge in MOSFET devices," *IEEE Trans. Electron Devices*, vol. 48, no. 10, pp. 2303–2309, Oct. 2001.

[13] M. Rosmeulen, L. Breuil, M. Lorenzini, L. Haspeslagh, J. Van Houdt, and K. De Meyer, "Characterization of the spatial charge distribution in local charge-trapping memory devices using the charge-pumping technique," *Solid State Electron.*, vol. 48, no. 9, pp. 1525–1530, 2004.

[14] E. Lusky, Y. Shacham-Diamand, G. Mitenberg, A. Shappir, I. Bloom, and B. Eitan, "Investigation of channel hot electron injection by localized charge-trapping nonvolatile memory devices," *IEEE Trans. Electron Devices*, vol. 51, no. 3, pp. 444–451, Mar. 2004.

[15] *ISE TCAD Manuals, Release 8.0*, Zurich, Switzerland: Synopsys, Inc. [Online]. Available: www.ise.ch

[16] J. D. Bude, M. R. Pinto, and R. K. Smith, "Monte Carlo simulation of CHISEL flash memory cell," *IEEE Trans. Electron Devices*, vol. 47, no. 10, pp. 1873–1881, Oct. 2000.

[17] L. Larcher, G. Verzellesi, P. Pavan, E. Lusky, I. Bloom, and B. Eitan, "Impact of programming charge distribution on threshold voltage and subthreshold slope of NROM memory cells," *IEEE Trans. Electron Devices*, vol. 49, no. 11, pp. 1939–1946, Nov. 2002.

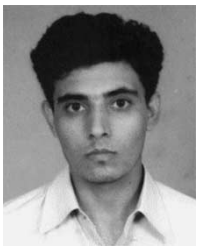
[18] Y. Roizin, M. Gutman, S. Alfassi, and R. Yosefi, "In-process charging in MicroFLASH memory cells," in *Proc. IEEE NVSMW*, 2003, pp. 83–84.
[19] C. Chang and J. Lien, "Corner-field induced drain leakage in thin oxide MOSFETs," in *IEDM Tech. Dig.*, 1987, pp. 714–717.
[20] T. Y. Chan, J. Chen, P. K. Ko, and C. Hu, "The impact of gate-induced drain leakage current on MOSFET scaling," in *IEDM Tech. Dig.*, 1987, pp. 718–721.

P. Bharath Kumar (S'04) received the M.Sc. degree in physics from Sri Sathya Sai Institute of Higher Learning, Prasanthinilayam, India, and the M.Tech. degree in electrical engineering from the Indian Institute of Technology (IIT) Bombay, Mumbai, India, in 2001 and 2003, respectively. He is currently working toward the Ph.D. degree at IIT.

His main research interests are in the areas of physics, simulation and characterization of semiconductor devices. His doctoral work involves studying the reliability of flash memory cells. He has worked on bias temperature instability in pMOS devices.

Pradeep R. Nair (S'04) received the B.Tech. degree in electronics and communication engineering from Calicut Regional Engineering College, Calicut, India, and the M.Tech. degree in electrical engineering from the Indian Institute of Technology (ITT), Bombay, Mumbai, India, in 2002 and 2004, respectively. He is currently working toward the Ph.D. degree in electrical engineering at Purdue University, West Lafayette, IN.

His current interests include characterization and numerical simulation of semiconductor devices and computational biophysics.



Ravinder Sharma received the M.Sc. degree in physics from Panjab University, Chandigarh, India, and the M.Tech. degree in electrical engineering from the Indian Institute of Technology Bombay, Mumbai, India, in 2002 and 2004, respectively.

He is currently with STMicroelectronics Pvt. Ltd., Noida, India. His area of interest includes MOS transistors and flash memories in devices and system performance in design.

Shiro Kamohara (M'98–SM'01) received the B.S. degree in physics from Keio University, Kanagawa, Japan, and the M.E. degree from the Tokyo Institute of Technology, Tokyo, Japan, in nuclear engineering, in 1986 and 1988, respectively.

He joined the Central Research Laboratory of Hitachi Ltd. and was involved in the development of in-house technology computer-aided design (TCAD) system. Since 1995, he has been a member of the Semiconductor and Integrated Circuit Division of Hitachi Ltd., where he has been working on the modeling of process, device, reliability, yield, and so on. He has been also involved in the device design for the MOSFET for dynamic RAM (DRAM), static RAM (SRAM), and logic and RF application. He has developed a variety of the physical models, including the oxide soft breakdown, the abnormal junction leakage of DRAM, and so on. He is currently the leader of a group responsible for the yield analysis, the device design using the modeling, and the simulation approach. His current activities include data mining for yield improvement and the robust design of an MOSFET. He was a Visiting Industrial Fellow of the University of California at Berkeley in 1996. He is the author or coauthor of more than 20 technical papers.

Mr. Kamohara is a member of the Applied Physics Society of Japan. He received the Best Paper Award from Asia and South Pacific Design Automation Conference in 2000.

Souvik Mahapatra (S'99–M'99) received the Ph.D. degree in electrical engineering from the Indian Institute of Technology (IIT) Bombay, Mumbai, India, in 1999.

From 2000 to 2001, he was with Bell Laboratories, Lucent Technologies, Murray Hill, NJ. Since 2002, he has been with the Department of Electrical Engineering, IIT, where he is currently an Associate Professor. His research interests include electrical characterization of defects in dielectric-semiconductor interfaces, hot-carrier and bias temperature instability in CMOS devices, high-k and novel dielectrics for CMOS, and Flash electrically erasable programmable read-only memories. He is the author or coauthor of more than 50 papers in refereed international journals and conferences. He was a Reviewer for many international journals and conferences.

Dr. Mahapatra was an Invited Speaker at several major international conferences including the International Electron Devices Meeting.