Revisiting Charge Trapping/Detrapping in Flash Memories From a Discrete and Statistical Standpoint—Part II: On-Field Operation and Distributed-Cycling Effects

Giovanni M. Paolucci, Christian Monzio Compagnoni, *Senior Member, IEEE*, Carmine Miccoli, Alessandro S. Spinelli, *Senior Member, IEEE*, Andrea L. Lacaita, *Fellow, IEEE*, and Angelo Visconti, *Member, IEEE*

I. Introduction

ON-FIELD operation of Flash memory arrays typically involves arbitrary time sequences of program/erase (P/E) cycles and idle periods, with the possibility for temperature to change within a wide range of values. The high electrical stress determined by P/E cycles on the cell tunnel oxide gives rise to charge trapping therein, impacting the array P/E performance [2]–[9] and determining threshold-voltage (V_T) instabilities when cells should instead keep their V_T level, i.e., their datum, in time [7], [10]–[15]. V_T instabilities come from the neutralization of the charge trapped in the cell tunnel oxide, commonly referred to as charge detrapping, representing a sort of damage recovery process taking place when the cell is idle, improving its performance and reliability in the next operations. This process has been shown to be highly thermally activated, following an Arrhenius law with activation energy $E_A \simeq 1.1$ eV [7], [13], [16], [17]. As a consequence, V_T instabilities encountered during a data retention time stretch starting at a certain point of device lifetime depend not only on the number of P/E cycles ($N_{\rm cyc}$) previously performed on the array, but also on how cycles have been distributed along the time axis and on the time-temperature profile of the device. Considering the effect of P/E cycles, idle periods, and temperature on V_T instabilities is, therefore, mandatory for their correct assessment during on-field operation and for the development of accelerated test schemes able to reproduce on-field results in much shorter experimental times [18].

In this paper, starting from the theoretical background on charge detrapping presented in the corresponding Part I [1] and extending the preliminary results presented in [19], we address the impact of idle periods, temperature, and P/E cycles on the spectral distribution of detrapping events, showing how the statistics of V_T displacements (ΔV_T) coming from detrapping can be directly obtained from this distribution. Our results allow, therefore, to deal with whatever on-field usage or testing scheme of the memory array, accounting both for damage recovery through detrapping and for damage creation through P/E cycles. The model is validated against a large number of experimental data, requiring a careful control of the experimental procedures commonly used to assess V_T instabilities and the inclusion of other physical phenomena affecting ΔV_T in the experiments. Results represent a milestone for the modeling and the predictive analysis of V_T instabilities in Flash memories.

II. Spectral Analysis of Detrapping Events

Referring to electron detrapping, two main results achieved in Part I of this paper [1] are that if the number of trapped electrons (N_t) in the tunnel oxide of the array cells is Poisson

Manuscript received March 11, 2014; revised May 6, 2014; accepted May 23, 2014. Date of publication June 19, 2014; date of current version July 21, 2014. The review of this paper was arranged by Editor E. Rosenbaum.

G. M. Paolucci, C. Monzio Compagnoni, and A. S. Spinelli are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan 20133, Italy (e-mail: paolucci@elet.polimi.it; monzio@elet.polimi.it; spinelli@elet.polimi.it).

C. Miccoli is with the Process Research and Development, Micron Technology Inc., Boise, ID 83716 USA (e-mail: cmiccoli@micron.com).

A. L. Lacaita is with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan 20133, Italy, and also with the Consiglio Nazionale delle Ricerche, Istituto di Fotonica e Nanotecnologie, Milan 20133, Italy (e-mail: lacaita@elet.polimi.it).

A. Visconti is with the Process Research and Development, Micron Technology Inc., Agrate Brianza 20864, Italy (e-mail: aviscont@micron.com).

distributed then the number of detrapping events (n_d) in a time t follows a Poisson statistics with average value $\langle n_d \rangle$ and, in turn, the probability density function (pdf—f) of cell ΔV_T is given by

$$f(\Delta V_T) = \mathscr{F}^{-1}\{e^{\langle n_d \rangle [\mathscr{F}\{f(\Delta V_T^1)\}-1]}\} \quad (1)$$

where $\mathscr{F}$ and $\mathscr{F}^{-1}$ represent, respectively, the Fourier and inverse Fourier transform operations and $f(\Delta V_T^1)$ is the pdf of the V_T shift coming from a single detrapping event. Note that (1) does not involve any assumption on the spectral distribution of the single detrapping events, i.e., on the pdf of their time constant τ_d. The impact of this distribution on ΔV_T is, in fact, accounted for via $\langle n_d \rangle$, whose calculation as a function of array on-field use represents, therefore, the critical point for the reliability assessment of V_T instabilities due to charge detrapping.

To calculate $\langle n_d \rangle$, we can start assuming that detrapping events are statistically distributed along the logarithmic time axis according to an arbitrary pdf $f(\log_{10}(\tau_d))$. With N_t that is Poisson distributed among the cells, the number of trapped electrons in an infinitesimal interval $d\log_{10}(\tau_d)$ follows then a Poisson statistics whose average is simply given by $\langle N_t \rangle \cdot f(\log_{10}(\tau_d)) \cdot d\log_{10}(\tau_d)$. The term $\langle N_t^*(\tau_d) \rangle = \langle N_t \rangle \cdot f(\log_{10}(\tau_d))$ represents the average number of trapped electrons per unit log time, i.e., the average spectral density of trapped electrons (units: electrons/decade). This spectral density changes during array operation and the average number of detrapping events in the time stretch between $t = 0$ and t can be straightforwardly calculated from its value at these times ($\langle N_t^*(\tau_d; 0) \rangle$ and $\langle N_t^*(\tau_d; t) \rangle$, respectively) as

$$\langle n_d(t) \rangle = \int_{-\infty}^{+\infty} [\langle N_t^*(\tau_d; 0) \rangle - \langle N_t^*(\tau_d; t) \rangle] \mathrm{d}\log_{10}(\tau_d). \quad (2)$$

In the following, the effect of idle periods, temperature, and P/E cycles on $\langle N_t^*(\tau_d) \rangle$ is addressed.

A. Impact of Idle Periods

Starting from an average spectral density of trapped electrons $\langle N_t^*(\tau_d; 0) \rangle$, the spectral density after an idle time stretch of duration t is given by

$$\langle N_t^*(\tau_d; t) \rangle = \langle N_t^*(\tau_d; 0) \rangle \cdot e^{-t/\tau_d}. \quad (3)$$

Note that the term e^{-t/τ_d} represents the probability that a detrapping event with time constant τ_d does not take place within the time t. Assuming that $\log_{10}(\tau_d)$ at $t = 0$ is uniformly distributed between a minimum ($\tau_d^{\min} = 10^{-5}$ h) and a maximum ($\tau_d^{\max} = 10^6$ h) τ_d value and that $\langle N_t \rangle = 10$, Fig. 1 shows $\langle N_t^*(\tau_d; 0) \rangle$ and $\langle N_t^*(\tau_d; t) \rangle$ for $t = 1$ min and 1 h. Results clearly highlight the existence of a detrapping front located about $\tau_d = t$ and, therefore, moving from left to right in the figure as time elapses, emptying the spectrum from the shortest toward the longest time constants.

Results in Fig. 1 allow to easily explain what happens to the average ΔV_T ($\langle \Delta V_T \rangle$) transient when this is monitored taking a read operation after a delay t_0 since the beginning of the detrapping process as reference, as usually done in experimental tests [14]. The case under study is schematically

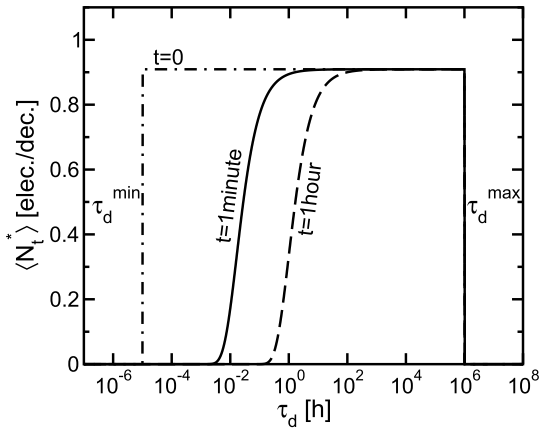

Fig. 1. Calculated spectral density of trapped electrons at $t = 0$, assumed uniformly distributed along the logarithmic τ_d axis between $\tau_d^{\min} = 10^{-5}$ h and $\tau_d^{\max} = 10^6$ h, and after an idle period of duration $t = 1$ min and 1 h. $\langle N_t \rangle = 10$.

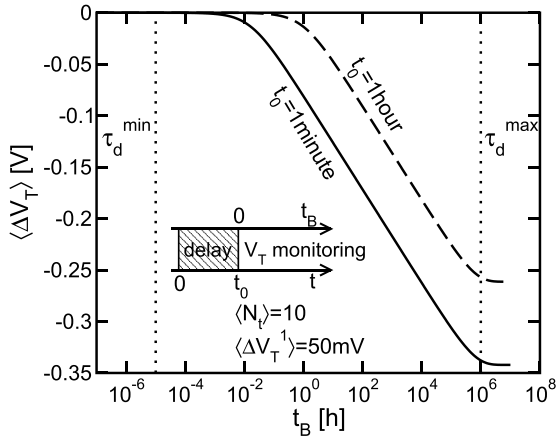

Fig. 2. Calculated $\langle \Delta V_T(t_B) \rangle$ transient for the test schematically reported in the inset: at $t = 0$ detrapping is assumed to begin, but V_T starts to be monitored only after a delay t_0 equal to 1 min or 1 h. ΔV_T is reported as a function of t_B, i.e., the time elapsing since the first read operation on the array. $\Delta V_T^1 = 50$ mV.

shown in the inset of Fig. 2: at $t = 0$, the detrapping process is supposed to begin, with an initial spectral density of trapped electrons $\langle N_t^*(\tau_d; 0) \rangle$, but V_T starts to be monitored only after a delay t_0, referring both ΔV_T and time to the first read operation, i.e., defining ΔV_T as $V_T(t_0 + t_B) - V_T(t_0)$ and considering it as a function of t_B. Fig. 2 shows the $\langle \Delta V_T(t_B) \rangle$ transient in the case of $t_0 = 1$ min and 1 h and assuming for $\langle N_t^*(\tau_d; 0) \rangle$ the uniform distribution at $t = 0$ shown in Fig. 1. Results have been calculated using (2)-(3) and according to

$$\langle \Delta V_T(t_B) \rangle = -\langle \Delta V_T^1 \rangle \cdot (\langle n_d(t_0 + t_B) \rangle - \langle n_d(t_0) \rangle) \quad (4)$$

where $\langle \Delta V_T^1 \rangle$ is the average shift resulting from a single detrapping event. From Fig. 2, the $\langle \Delta V_T(t_B) \rangle$ transient detaches from 0 when t_B reaches t_0, then showing a decreasing rate α up $t_B \simeq \tau_d^{\max}$ and saturating from this time on. This behavior can be explained considering that during the delay period the spectral density of trapped electrons loses almost all of its electrons with $\tau_d < t_0$ (see the detrapping front at 1 min and 1 h in Fig. 1), allowing further detrapping events and,

in turn, the decrease of $\langle \Delta V_T(t_B) \rangle$, only when t_B becomes comparable with or longer than t_0. In addition to that, note that the spectrum of Fig. 1 is not modified for τ_d much longer than 1 min or 1 h, i.e., the values of t_0 considered in Fig. 2, keeping the constant initial density $\langle N_t^*(\tau_d; 0) \rangle$ up to $\tau_d^{\max}$. This explains why both the $\langle \Delta V_T(t_B) \rangle$ transients of Fig. 2 reach the same slope after t_0, corresponding to a decreasing rate α [1]

$$\alpha = \frac{\langle \Delta V_T^1 \rangle \cdot \langle N_t \rangle}{\ln\left(\tau_d^{\max}/\tau_d^{\min}\right)} = \frac{\langle \Delta V_T^1 \rangle \cdot \langle N_t^*(\tau_d > t_0; 0) \rangle}{\ln 10}. \quad (5)$$

This equation highlights a strong connection between the slope of the $\langle \Delta V_T(t_B) \rangle$ transient and the spectral density of trapped electrons in the explored time range, allowing to easily explain the shape of the transient in more complex experimental schemes (see next sections). Finally, note that the saturation of $\langle \Delta V_T(t_B) \rangle$ in Fig. 2 is just the result of the assumption that no detrapping event has a τ_d longer than $\tau_d^{\max}$ in Fig. 1, and the different saturation levels of the curves related to different t_0 come from the higher number of detrapping events taking place during t_0.

From the previous results, the $\langle \Delta V_T(t_B) \rangle$ transients of Fig. 2 can be described for t_B shorter than $\tau_d^{\max}$ by the following relation:

$$\langle \Delta V_T(t_B) \rangle = -\alpha \cdot \ln\left(1 + \frac{t_B}{t_0}\right). \quad (6)$$

This expression has been frequently used to fit experimental results coming from uniform cycling tests [14] and describes very accurately the transition from 0 to the slope α of the $\langle \Delta V_T(t_B) \rangle$ transient when $\langle N_t^*(\tau_d; 0) \rangle$ is constant between $\tau_d^{\min}$ and $\tau_d^{\max}$ (a more rigorous validation of (6) can be derived starting from (10) in Part I of this paper [1]). In this case, the $\langle \Delta V_T(t_B) \rangle$ curves resulting from different delay periods t_0 are just horizontally shifted along the logarithmic t_B axis, as clearly appears from Fig. 2. However, in the more general case where $\langle N_t^*(\tau_d; 0) \rangle$ is not constant, more complex shapes may result for the $\langle \Delta V_T(t_B) \rangle$ transient, as shown in Section III when dealing with distributed-cycling experiments. In these cases, (6) may not be suited to reproduce the $\langle \Delta V_T(t_B) \rangle$ behavior and transients corresponding to different t_0 may not be just horizontally shifted along the logarithmic t_B axis.

Finally, note that the existence of delay periods between the beginning of the detrapping process and the first read operation on the array makes the $\Delta V_T(t_B)$ results independent of $\tau_d^{\min}$. $\Delta V_T(t_B)$ depends, in fact, on the spectral density of trapped electrons at times $t \geq t_0$, i.e., when the detrapping front has already removed almost all of the electrons with $\tau_d < t_0$. Therefore, $\tau_d^{\min}$ can be arbitrarily chosen provided that the spectral density of trapped electrons is not modified for τ_d close or longer than t_0. Moreover, note also that even $\tau_d^{\max}$ can be arbitrarily chosen in the absence of a clear saturation of the $\langle \Delta V_T(t_B) \rangle$ transient measured in an experimental test, with the only constraint that this time constant is longer than the maximum experimental time and that, therefore, $\langle N_t^*(\tau_d; 0) \rangle$ is not modified up to that time. This explains why we used a very high $\tau_d^{\max} = 10^6$ h in this and in the next section, leading to the saturation of the $\langle \Delta V_T(t_B) \rangle$ transient only at extremely long, and experimentally unapproachable, times. Saturation in Figs. 2 and 4 should, therefore, be considered as proof that our model would be capable of reproducing, if experimental evidence were provided, this feature of the detrapping process also, but not as a theoretical prediction of any upper boundary to τ_d. From these considerations, $\tau_d^{\min}$ and $\tau_d^{\max}$ appear just as arbitrary parameters defined for mathematical convenience but without any meaningful physical role.

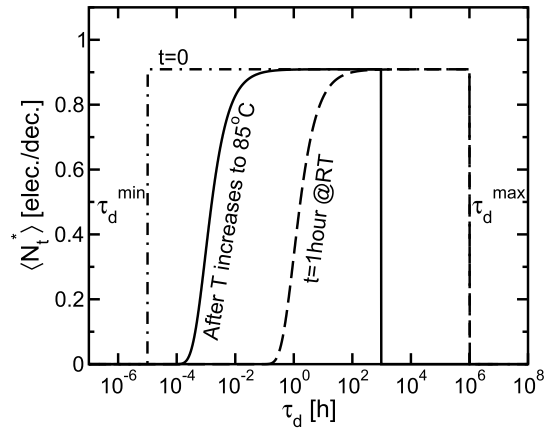

Fig. 3. Calculated spectral density of trapped electrons at $t = 0$ (same assumptions of Fig. 1), after a delay period of duration $t_0 = 1$ h at room temperature (RT) and after an increase of temperature to $T_B = 85$ °C.

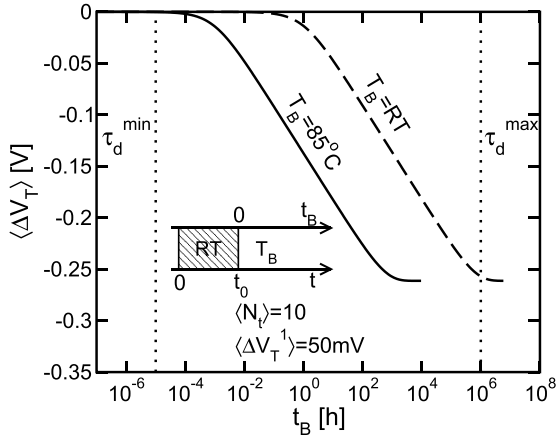

Fig. 4. Same as in Fig. 2, but with an increase of temperature to $T_B = 85$ °C after a delay of duration $t_0 = 1$ h at RT. The inset shows schematics for the test.

B. Impact of Temperature

A general agreement exists in that thermal activation of detrapping in mainstream Flash technologies can be described by an Arrhenius law with activation energy $E_A \simeq 1.1$ eV [7], [13], [16]–[18]. To consider this effect, we assumed that changing temperature from a value T_1 to a value T_2 modifies the time constant τ_d of detrapping events according to

$$\tau_d(T_2) = \tau_d(T_1) \cdot e^{-E_A(1/kT_1 - 1/kT_2)} \quad (7)$$

then resulting in a horizontal shift of $\langle N_t^*(\tau_d; t) \rangle$ along the logarithmic τ_d axis.

The effect of temperature on the spectral density of trapped electrons is shown in Fig. 3 assuming that $\langle N_t^*(\tau_d; 0) \rangle$ is

uniform between $\tau_d^{\min} = 10^{-5}$ h and $\tau_d^{\max} = 10^6$ h and that after an idle period of duration $t_0 = 1$ h at RT temperature is increased to $T_B = 85$ °C (see the inset of Fig. 4 for the considered test scheme). Due to the temperature increase and the thermal activation of detrapping events, the spectral density at the end of the idle period (dashed line) is leftward shifted of an amount given by the exponential term in (7) (solid line). As a consequence, the resulting $\langle \Delta V_T(t_B) \rangle$ transient during the data retention phase at $T_B = 85$ °C in Fig. 4 is just a leftward-shifted replica of the transient that would be obtained at $T_B = $ RT (note that this would be true even if $\langle N_t^*(\tau_d; 0) \rangle$ were not uniform), in agreement with experimental observations [14]. From these results, the duration t_B and the temperature T_B of idle periods can be arbitrarily modified provided that they follow the Arrhenius law with $E_A \simeq 1.1$ eV, giving the possibility to reproduce in reasonable experimental time stretches the same detrapping dynamics taking place over a much longer timescale by means of a higher temperature [7], [13], [18].

C. Impact of P/E Cycles

Charge trapping in mainstream Flash technologies has been shown to follow approximately a square root dependence on the number of P/E cycles [7], [13], leading to $\langle \Delta V_T(t_B) \rangle$ transients of higher slope α for increasing $N_{\rm cyc}$ [14], in agreement with the link between α and $\langle N_t^*(\tau_d; 0) \rangle$ given by (5). To reproduce the storage of new electrons in the cell tunnel oxide resulting from each P/E cycle, we assumed that $\langle N_t^*(\tau_d) \rangle$ grows with $N_{\rm cyc}$ according to

$$\langle N_t^*(\tau_d; N_{\rm cyc}) \rangle = \langle N_t^*(\tau_d; N_{\rm cyc} - 1) \rangle + \langle \Delta N_t^*(\tau_d) \rangle \quad (8)$$

with

$$\langle \Delta N_t^*(\tau_d) \rangle = \frac{\eta^2}{2 \cdot \langle N_t^*(\tau_d; N_{\rm cyc} - 1) \rangle}, \quad \langle N_t^*(\tau_d; N_{\rm cyc} - 1) \rangle > \eta \quad (9)$$

$$\langle \Delta N_t^*(\tau_d) \rangle = \eta - \frac{1}{2} \cdot \langle N_t^*(\tau_d; N_{\rm cyc} - 1) \rangle, \quad \langle N_t^*(\tau_d; N_{\rm cyc} - 1) \rangle < \eta \quad (10)$$

where η is a parameter (similar expressions could be obtained if charge trapping grew according to slightly different power laws of $N_{\rm cyc}$). Note that (10) represents just an approximation of (9) used to deal with cases where the density of trapped electrons is very low and that the combination of these equations leads to $\langle N_t^*(\tau_d, N_{\rm cyc}) \rangle \simeq \eta \cdot (N_{\rm cyc})^{1/2}$ as $N_{\rm cyc}$ grows, reproducing the experimental observations [7], [13].

An important remark is that, using (8)–(10), we completely avoided any microscopic investigation of the charge trapping process and we did not aim at looking for the physical origin of the square root dependence of the amount of trapped charge in the cell tunnel oxide on $N_{\rm cyc}$. Though these points can be considered to be of extreme importance for a comprehensive understanding of Flash memory reliability, their analysis is out of the scope of this paper, deserving dedicated microscopic studies on defects generation and carrier trapping in the oxide. We assumed, therefore, (8)–(10) as phenomenological and justified their validity in that they can well reproduce data obtained from a large variety of experimental tests (Section III).

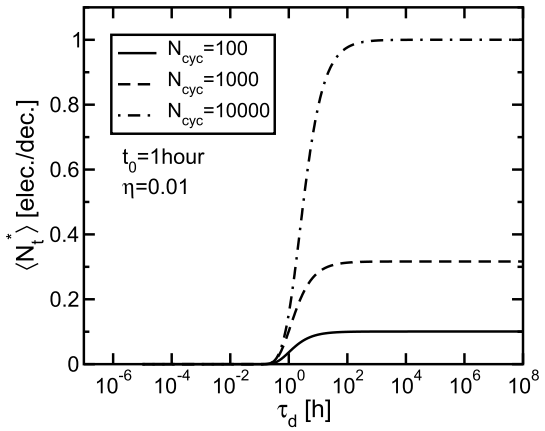

Fig. 5. Calculated spectral density of trapped electrons after an RT delay period of duration $t_0 = 1$ h since the end of cycling, as resulting from (8)–(10) assuming zero initial trapped electrons, $\eta = 0.01$ and different $N_{\rm cyc}$.

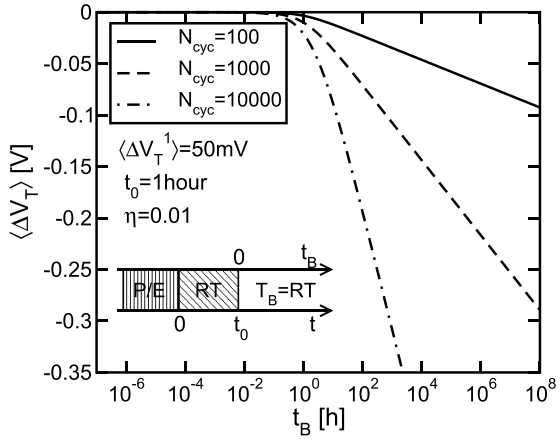

Fig. 6. Calculated $\langle \Delta V_T(t_B) \rangle$ transient for the test schematically reported in the inset: $N_{\rm cyc}$ P/E cycles are first performed on the cells, and after an RT delay period of duration $t_0 = 1$ h, V_T starts being monitored, keeping $T_B = RT$. $\eta = 0.01$ and $\Delta V_T^1 = 50$ mV.

The effect of cycling on the spectral density of trapped electrons is investigated in Fig. 5 considering the simple case of a uniform cycling scheme with constant pace of 1 P/E cycle per minute, in the case of $N_{\rm cyc} = 100$, 1000, and 10 000, and $\eta = 0.01$. The spectral density is shown after a delay period of duration $t_0 = 1$ h since the end of cycling and assuming zero trapped electrons at the first P/E cycle. Note that this latter assumption results in a uniform electron trapping along the τ_d axis as cycling proceeds, with the need neither of a $\tau_d^{\min}$ nor of a $\tau_d^{\max}$ but for numerical implementations. $\langle N_t^*(\tau_d) \rangle$ in Fig. 5 reaches, therefore, a constant value for $\tau_d \gg t_0$, i.e., for τ_d longer than the position of the detrapping front, increasing with $(N_{\rm cyc})^{1/2}$. Fig. 6 shows the $\langle \Delta V_T(t_B) \rangle$ transients resulting from detrapping taking place from t_0 on (see the inset for the considered test scheme), displaying the increase of α with $(N_{\rm cyc})^{1/2}$ expected from (5) (note that all of the curves detach from the time axis, instead, at the same time $t_B = t_0$).

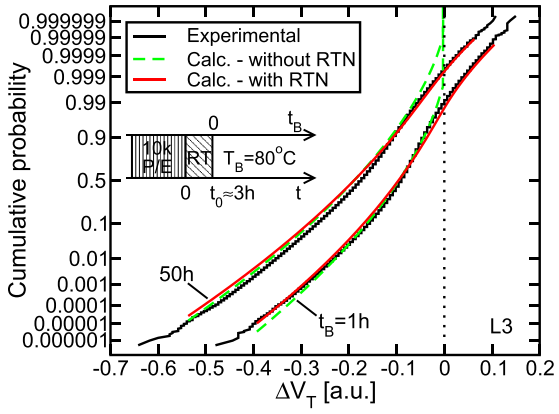

Fig. 7. Measured and calculated ΔV_T distribution for the test schematically reported in the inset: $N_{\rm cyc} = 10$ k P/E cycles are performed at RT on the array and after an RT delay of duration $t_0 \simeq 3$ h V_T starts being monitored, accelerating the detrapping dynamics by increasing temperature to $T_B = 80$ °C. Results are shown for $t_B = 1$ and 50 h. Calculations are performed either considering or neglecting random telegraph noise (RTN). Data retention is on level $L3$.

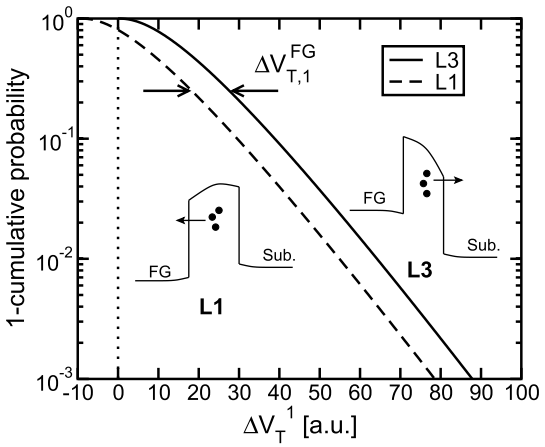

Fig. 8. ΔV_T^1 distribution used for calculations in Section III, referring to cells on level $L3$ and $L1$ during data retention. The two distributions are just horizontally shifted of an amount equal to $\Delta V_{T,1}^{FG}$, i.e., the V_T shift corresponding to one electron in the floating gate.

In the next section, more complex cycling schemes than that considered in Figs. 5 and 6 will be addressed, leading to more complex behaviors of $\langle N_t^*(\tau_d)\rangle$ and $\langle \Delta V_T(t_B)\rangle$.

III. Model Validation

We validated our statistical model for charge trapping/detrapping in nanoscale Flash memories on our 20-nm multilevel NAND technology [20]. We started considering the simple case of a uniform cycling experiment ($N_{\rm cyc} = 10$ k, cycling duration $t_{\rm cyc} \simeq 24$ h, cycling temperature $T_{\rm cyc} =$ RT) followed by a data retention phase at constant temperature $T_B = 80$ °C, with a $t_0 = 3$ h delay period at RT in between required by the experimental procedure (see the inset of Fig. 7). Cycling was performed with a random programming pattern, i.e., randomly moving the erased cells to one of the four possible V_T levels of the multilevel device, namely, from lowest to highest, E, $L1$, $L2$, and $L3$. At t_0, the reference read operation on the array was performed, gathering the V_T of all the cells in a block, and the sample was then warmed to T_B to accelerate charge detrapping. ΔV_T was, finally, evaluated for each cell by periodically cooling the sample to RT and gathering again the block V_T map. The resulting experimental distribution of ΔV_T is shown in Fig. 7, referring to cells being on level $L3$ during data retention and in the case of $t_B = 1$ and 50 h.

To calculate the statistical distribution of ΔV_T resulting from the experimental test considered in Fig. 7, we carefully reproduced the test sequence in our model, starting from zero trapped electrons and using (8)–(10) to increase $\langle N_t^*(\tau_d)\rangle$ after each P/E cycle, (3) to account for electron detrapping during the idle time stretches in between the cycles and during t_0 and (7) to manage the temperature increase during data retention. From the resulting spectral density of trapped electrons at t_0 and $t_0 + t_B$, we calculated the average number of detrapping events during t_B by means of (2) as $\langle n_d\rangle = \langle n_d\,(t_0 + t_B)\rangle - \langle n_d\,(t_0)\rangle$ (note that $t = 0$ is now the time corresponding to the end of the cycling phase) and used this number in (1) to find the statistical distribution of ΔV_T. Fig. 7 shows that modeling results (dashed green curves) can nicely reproduce the experimental data for $\Delta V_T < 0$ after tailoring η and using the ΔV_T^1 distribution reported in Fig. 8 (continuous line). The functional form of this distribution was set equal to that of a Gamma function, introducing, therefore, two free parameters in the model besides η. The good agreement between modeling and experimental data in Fig. 7 with such a low number of fitting parameters, namely, three, represents a first validation of our model for charge detrapping, which will be further challenged with other test schemes in the following. However, before considering different experimental schemes, in the next section, we address the possibility to extend our model to account also for positive ΔV_T during data retention.

A. Additional Contributions to ΔV_T

Despite our model proved itself able to reproduce the negative part of the ΔV_T statistics in Fig. 7, there is no possibility for it to explain positive ΔV_T including only electron detrapping (see the discussion at the end of Part I of this paper [1]). As a consequence, additional physical effects must be considered in the model, introducing the possibility that a cell increases its V_T during data retention. Hole detrapping and RTN can be considered two of the most relevant of these effects and can be straightforwardly introduced in our model under the assumption that all these contributions to ΔV_T are independent. Under this assumption, in fact, the pdf of the total ΔV_T ($f_{\rm tot}(\Delta V_T)$) can be obtained from the convolution of the single contributions as

$$f_{\rm tot}(\Delta V_T) = f_{\rm ED}(\Delta V_T) \otimes f_{\rm HD}(\Delta V_T) \otimes f_{\rm RTN}(\Delta V_T) \otimes f_{GN}(\Delta V_T) \qquad (11)$$

where $f_{\rm ED}(\Delta V_T)$ and $f_{\rm HD}(\Delta V_T)$ are the pdf of ΔV_T coming from electron and hole detrapping, respectively, and can be calculated from (1) and the spectral analysis methodology given in Section II applied both to electrons and to holes; $f_{\rm RTN}(\Delta V_T)$ is, instead, the pdf of the RTN contribution to

ΔV_T and $f_{GN}(\Delta V_T)$ is that of a Gaussian noise (standard deviation σ_{GN}, zero average value) keeping into account the experimental setup noise.

From [21], a good functional shape for $f_{RTN}(\Delta V_T)$ is the following:

$$f_{RTN}(\Delta V_T) = c \cdot \delta(\Delta V_T) + \frac{1-c}{2\sigma_{RTN}} e^{-|\Delta V_T|/\sigma_{RTN}} \quad (12)$$

where c represents the probability that RTN does not change cell V_T between the two read operations used to evaluate ΔV_T, i.e., that its ΔV_T due to RTN equals 0. The second term on the right-hand side of (12) accounts, in turn, for the possibility that RTN modifies cell V_T between the two read operations. This term introduces two exponential tails departing toward positive and negative ΔV_T, with spread given by σ_{RTN}, following the typical trend of the RTN amplitude statistics [22]–[26]. Note that c, σ_{RTN}, and σ_{GN} were extracted from the comparison of the ΔV_T distribution coming from the convolution of $f_{RTN}(\Delta V_T)$ and $f_{GN}(\Delta V_T)$ with the ΔV_T distribution experimentally obtained with two close read operations on the array at very long t_B, preventing any possible contribution of charge detrapping on the results. These parameters, therefore, are not free when comparing our detrapping model with experimental data.

Despite some experimental evidence of hole detrapping has been reported in [15], including only the RTN contribution to ΔV_T when modeling the experimental test of Fig. 7 turned out to be enough to achieve a good agreement between calculated (solid red lines) and experimental results, both for negative and for positive ΔV_T values. In particular, RTN appears to introduce the high tail of the ΔV_T distribution, while negligibly affecting its low tail. A higher impact of RTN on the low tail of the ΔV_T distribution may be expected at very low t_B or for low N_{cyc}, i.e., for small contributions of detrapping to V_T instabilities.

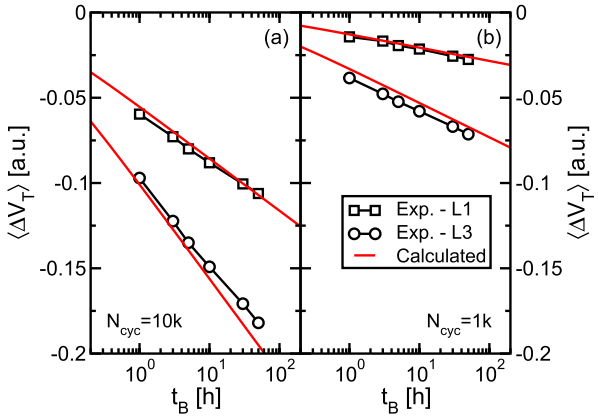

Fig. 9. (a) $\langle \Delta V_T(t_B) \rangle$ transient for the test considered in Fig. 7, in the case of cells on level $L3$ and $L1$ during data retention. (b) Same results but for $N_{cyc} = 1$k.

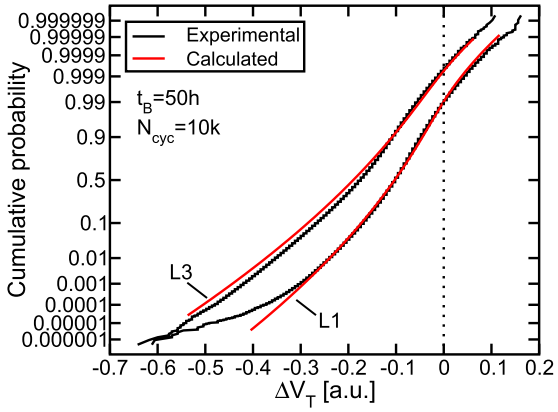

Fig. 10. ΔV_T cumulative distribution for the test considered in Fig. 7, in the case of cells on level $L3$ and $L1$ during data retention; $t_B = 50$ h and $N_{cyc} = 10$ k.

B. Further Validations

To further validate our model for charge detrapping, Fig. 9(a) shows the experimental and calculated $\langle \Delta V_T(t_B) \rangle$ transient for the uniform cycling test considered in Fig. 7. Our model appears to correctly reproduce the experimental trend both for cells on level $L3$ and $L1$ during data retention. To catch the level dependence, we assumed that the detrapping dynamics is negligibly affected by the electric field in the cell tunnel oxide [7], [13], meaning that $\langle N_t^*(\tau_d) \rangle$ and $\langle n_d \rangle$ do not change with the programmed level set at the beginning of the idle periods. We assumed, instead, that charge is exchanged either with the substrate or with the floating gate depending on the cell V_T level, due to a different field direction in the tunnel oxide. This is schematically shown in the inset of Fig. 8 referring to electron detrapping from level $L3$ and $L1$: in the former case, we assumed that electron detrapping involves an interaction with cell channel, while with the floating gate in the latter. In so doing, the key difference between V_T instabilities on $L3$ and $L1$ is not in the detrapping dynamics but in the impact that each detrapping event has on cell V_T, i.e., on ΔV_T^1. As a result of this physical picture, the ΔV_T^1 distribution for cells on $L1$ was assumed equal to that on $L3$ but for a leftward horizontal shift corresponding to the change of V_T following the storage of a single electron in the floating gate ($\Delta V_{T,1}^{FG}$), as shown in Fig. 8. This latter parameter was extracted from measurements of the electron injection spread during incremental step pulse programming [27]–[29] and, therefore, does not represent an additional fitting parameter in our model. Fig. 10 shows that, following this approach, not only the average value, but also the entire ΔV_T distribution coming from the detrapping experiment on level $L1$ can be nicely reproduced keeping all the parameters but ΔV_T^1 identical.

Figs. 11 and 9(b) prove that our model correctly catches the cycling dependence of V_T instabilities coming from detrapping, agreeing with data obtained from a uniform cycling experiment similar to that shown in the inset of Fig. 7 but with $N_{cyc} = 1$k. Both the $\langle \Delta V_T(t_B) \rangle$ transient and the whole ΔV_T distribution are well reproduced by our model, confirming that charge detrapping is the dominant source of V_T instabilities for the vast majority of cells in the investigated test conditions [30]. Note, however, that the mismatch between data and calculations appearing at very low probabilities for some of the cases dealt with in Figs. 11 and 10 may be

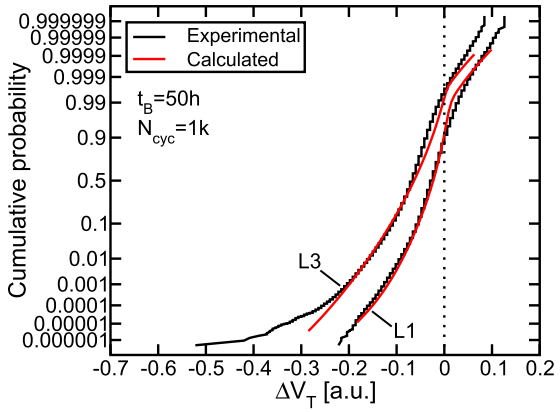

Fig. 11. Same as in Fig. 10, but for $N_{cyc} = 1$ k.

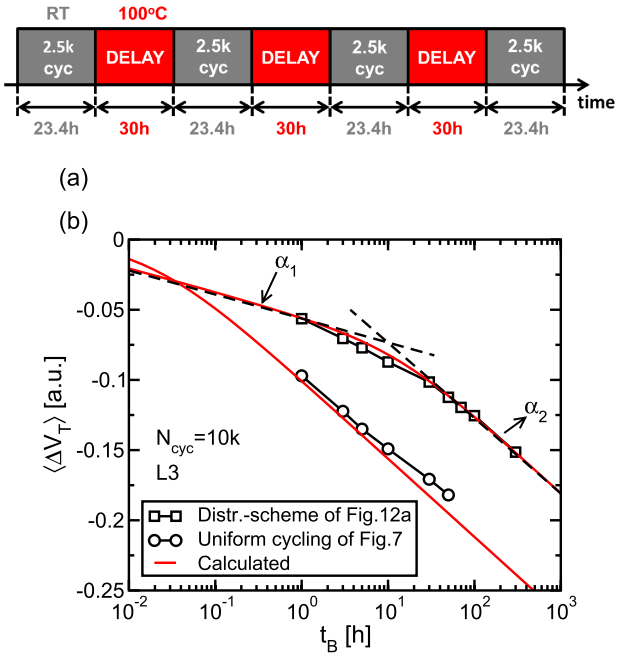

Fig. 12. (a) Schematics for the distributed-cycling test considered in Section III-C: four groups of 2.5 k P/E cycles are performed at RT in $\simeq$ 23.4 h, with three delay periods $\simeq$ 30-h long at 100 °C in between. At the end of this cycling phase, V_T starts being monitored after an RT idle period of duration $t_0 \simeq 20$ h, then accelerating charge detrapping by increasing the temperature to $T_B = 125$ °C. (b) Experimental and calculated $\langle \Delta V_T(t_B) \rangle$ transient resulting from the test in (a). Results from the test in Fig. 7 are also shown for comparison.

considered the evidence of additional physical mechanisms coming into play, which are currently under investigation.

C. Distributed-Cycling Schemes

To challenge our detrapping model with more complex experimental schemes than the uniform cycling test investigated in the previous section, we considered the distributed-cycling experiment shown in Fig. 12(a). A total number $N_{cyc} = 10$ k P/E cycles were performed on the array at RT as in the test of Fig. 10, but in this case, cycles were not uniformly distributed in time, but gathered in four groups of 2.5 k with three idle periods 30-h long in between at 100 °C.

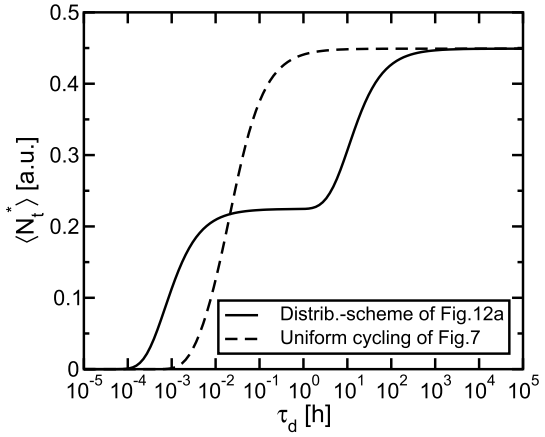

Fig. 13. Calculated spectral density of trapped electrons at the first read operation for the test in Figs. 12(a) (shown at $T_B = 125$ °C) and 7 (shown at $T_B = 80$ °C).

At the end of cycling, the first read operation on the array was performed after an RT delay with duration $t_0 \simeq 20$ h, then starting a bake at $T_B = 125$ °C during which V_T was periodically monitored. Fig. 12(b) shows that, after carefully reproducing the sequence of this new test in our model, this can correctly reproduce the experimental results with the same set of parameters extracted for the test of Fig. 10. Considering that the cycling scheme of Fig. 12(a) involves changes of temperature and idle periods during cycling, this represents a further strong proof of the validity of our detrapping model, which represents a powerful tool for the reliability analysis of nanoscale Flash memories. Note that the model allows, moreover, to easily explain the shape of the $\langle \Delta V_T(t_B) \rangle$ transients of Fig. 12(b), which do not appear to follow a simple logarithmic behavior like that predicted by (6). In particular, the slope of the transient seems to increase for t_B longer than 10 h, reaching a value that matches the one obtained from the uniform cycling test of Fig. 10. This can be explained considering the spectral distribution of trapped electrons at the first read operation on the array, shown in Fig. 13 at $T_B = 80$ °C for the test of Fig. 10 and at $T_B = 125$ °C for the test of Fig. 12(a). In the case of the uniform cycling experiment, the spectrum (dashed line) grows from zero to the value corresponding to $N_{cyc} = 10$ k with a detrapping front about $\simeq 10^{-2}$ h, corresponding to the t_0 value of this experiment converted to $T_B = 80$ °C. This leads to the purely logarithmic behavior of $\langle \Delta V_T(t_B) \rangle$ appearing in Figs. 9–12, which can be well described by (6). In the case of the distributed-cycling scheme of Fig. 12(a), instead, the spectral distribution of trapped electrons (solid line) is more complex, displaying a two-step behavior. To understand this result, note that the three high-temperature delay periods in the distributed-cycling scheme of Fig. 12(a) allow the loss of electrons trapped in the previous groups of P/E cycles up to a maximum τ_d of $\simeq$10 h, corresponding to their duration (90 h) converted to the bake temperature $T_B = 125$ °C. Electrons trapped during the last group of 2.5 k P/E cycles in Fig. 12(a) may be detrapped, instead, only by the last idle period of duration t_0 in between the cycling and the retention phase, i.e., for τ_d shorter than $\simeq 10^{-3}$ h (corresponding to t_0 converted

to $T_B = 125$ °C). As a result, $\langle N_t^*(\tau_d)\rangle$ grows from zero to the value corresponding to 2.5 k P/E cycles at $\simeq 10^{-3}$ h and then reaches the value corresponding to 10 k P/E cycles only above $\simeq 10$ h, i.e., for τ_d so long to require much more time than the idle periods involved in Fig. 12(a) to achieve a detrapping event. From this spectral distribution of trapped electrons, $\langle \Delta V_T(t_B)\rangle$ in Fig. 12(b) is expected to start decreasing with a first slope α_1 at about $\simeq 10^{-3}$ h and to increase its slope to α_2 above $\simeq 10$ h, with α_1 and α_2 corresponding to 2.5 k and 10 k P/E cycles, respectively. This reveals that the effect of distributed cycling on the $\langle \Delta V_T(t_B)\rangle$ transient may be more complex than just a horizontal shift of the curve along the logarithmic t_B axis, as resulting, instead, when increasing $T_{\rm cyc}$ but keeping P/E cycles uniformly distributed in time [14], [17].

IV. Conclusion

This paper addressed the effect of idle periods, temperature, and P/E cycles on the spectral distribution of detrapping events and, in turn, on the consequent data retention V_T instabilities in nanoscale Flash memories. The resulting model represents a powerful tool for the investigation and predictive analysis of the impact of charge trapping/detrapping on the reliability of Flash technologies, being able to deal with whatever on-field usage or testing scheme of the memory array.

References

[1] G. M. Paolucci, C. Monzio Compagnoni, C. Miccoli, A. S. Spinelli, A. L. Lacaita, and A. Visconti, "Revisiting charge trapping/detrapping in flash memories from a discrete and statistical standpoint—Part I: V_T instabilities," *IEEE Trans. Electron Devices*, vol. 61, no. 8, pp. 2802–2810, Aug. 2014.

[2] S. Yamada, Y. Hiura, T. Yamane, K. Amemiya, Y. Ohshima, and K. Yoshikawa, "Degradation mechanism of flash EEPROM programming after program/erase cycles," in *IEDM Tech. Dig.*, Dec. 1993, pp. 23–26.

[3] Y.-B. Park and D. K. Schroder, "Degradation of thin tunnel gate oxide under constant Fowler–Nordheim current stress for a flash EEPROM," *IEEE Trans. Electron Devices*, vol. 45, no. 6, pp. 1361–1368, Jun. 1998.

[4] J.-D. Lee, J.-H. Choi, D. Park, and K. Kim, "Degradation of tunnel oxide by FN current stress and its effects on data retention characteristics of 90 nm NAND flash memory cells," in *Proc. 41st Annu. IEEE IRPS*, Mar./Apr. 2003, pp. 497–501.

[5] J.-D. Lee, J.-H. Choi, D. Park, and K. Kim, "Data retention characteristics of sub-100 nm NAND flash memory cells," *IEEE Electron Device Lett.*, vol. 24, no. 12, pp. 748–750, Dec. 2003.

[6] J.-D. Lee, J.-H. Choi, D. Park, and K. Kim, "Effects of interface trap generation and annihilation on the data retention characteristics of flash memory cells," *IEEE Trans. Device Mater. Rel.*, vol. 4, no. 1, pp. 110–117, Mar. 2004.

[7] N. Mielke *et al.*, "Flash EEPROM threshold instabilities due to charge trapping during program/erase cycling ," *IEEE Trans. Device Mater. Rel.*, vol. 4, no. 3, pp. 335–344, Sep. 2004.

[8] M. Park, K. Suh, K. Kim, S.-H. Hur, K. Kim, and W.-S. Lee, "The effect of trapped charge distributions on data retention characteristics of NAND flash memory cells," *IEEE Electron Device Lett.*, vol. 28, no. 8, pp. 750–752, Aug. 2007.

[9] A. Fayrushin, K. Seol, J. Na, S. Hur, J. Choi, and K. Kim, "The new program/erase cycling degradation mechanism of NAND flash memory devices," in *IEDM Tech. Dig.*, 2009, pp. 823–826.

[10] M. Kato *et al.*, "Read-disturb degradation mechanism due to electron trapping in the tunnel oxide for low-voltage flash memories," in *IEDM Tech. Dig.*, Dec. 1994, pp. 45–48.

[11] R.-I. Yamada, Y. Mori, Y. Okuyama, J. Yugami, T. Nishimoto, and H. Kume, "Analysis of detrap current due to oxide traps to improve flash memory retention," in *Proc. 38th Annu. IEEE IRPS*, 2000, pp. 200–204.

[12] R. Yamada, T. Sekiguchi, Y. Okuyama, J. Yugami, and H. Kume, "A novel analysis method of threshold voltage shift due to detrap in a multi-level flash memory," in *Symp. VLSI Technol. Dig. Tech. Papers*, Jun. 2001, pp. 115–116.

[13] N. Mielke, H. P. Belgal, A. Fazio, Q. Meng, and N. Righos, "Recovery effects in the distributed cycling of flash memories," in *Proc. 44th IEEE IRPS*, Mar. 2006, pp. 29–35.

[14] C. Miccoli, C. Monzio Compagnoni, S. Beltrami, A. S. Spinelli, and A. Visconti, "Threshold-voltage instability due to damage recovery in nanoscale NAND flash memories," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2406–2414, Aug. 2011.

[15] C. Miccoli *et al.*, "Resolving discrete emission events: A new perspective for detrapping investigation in NAND flash memories," in *Proc. IEEE IRPS*, Apr. 2013, pp. 3B.1.1–3B.1.6.

[16] G. Verma and N. Mielke, "Reliability performance of ETOX based flash memories," in *Proc. 26th Annu. IRPS*, Apr. 1988, pp. 158–166.

[17] C. Monzio Compagnoni *et al.*, "Investigation of the threshold voltage instability after distributed cycling in nanoscale NAND flash memory arrays," in *Proc. IEEE IRPS*, May 2010, pp. 604–610.

[18] *Electrically Erasable Programmable ROM (EEPROM) Program/Erase Endurance and Data Retention Stress Test*, JEDEC Solid State Technology Association, Arlington, VA, USA, 2011.

[19] G. M. Paolucci *et al.*, "A new spectral approach to modeling charge trapping/detrapping in NAND flash memories," in *Proc. IRPS*, 2014.

[20] A. Goda and K. Parat, "Scaling directions for 2D and 3D NAND cells," in *Proc. IEEE IEDM*, Dec. 2012, pp. 2.1.1–2.1.4.

[21] C. Monzio Compagnoni, R. Gusmeroli, A. S. Spinelli, A. L. Lacaita, M. Bonanomi, and A. Visconti, "Statistical model for random telegraph noise in flash memories," *IEEE Trans. Electron Devices*, vol. 55, no. 1, pp. 388–395, Jan. 2008.

[22] A. Ghetti, M. Bonanomi, C. Monzio Compagnoni, A. S. Spinelli, A. L. Lacaita, and A. Visconti, "Physical modeling of single-trap RTS statistical distribution in flash memories," in *Proc. IEEE IRPS*, Apr./May 2008, pp. 610–615.

[23] A. Ghetti *et al.*, "Scaling trends for random telegraph noise in deca-nanometer flash memories," in *IEDM Tech. Dig.*, pp. 835–838, Dec. 2008.

[24] A. Ghetti, C. Monzio Compagnoni, A. S. Spinelli, and A. Visconti, "Comprehensive analysis of random telegraph noise instability and its scaling in deca–nanometer flash memories," *IEEE Trans. Electron Devices*, vol. 56, no. 8, pp. 1746–1752, Aug. 2009.

[25] S. M. Amoroso *et al.*, "Investigation of the RTN distribution of nanoscale MOS devices from subthreshold to on-state," *IEEE Electron Device Lett.*, vol. 34, no. 5, pp. 683–685, May 2013.

[26] K. Fukuda, Y. Shimizu, K. Amemiya, M. Kamoshida, and C. Hu, "Random telegraph noise in flash memories—Model and technology scaling," in *Proc. IEEE IEDM*, Dec. 2007, pp. 169–172.

[27] C. Monzio Compagnoni, A. S. Spinelli, R. Gusmeroli, S. Beltrami, A. Ghetti, and A. Visconti, "Ultimate accuracy for the NAND flash program algorithm due to the electron injection statistics," *IEEE Trans. Electron Devices*, vol. 55, no. 10, pp. 2695–2702, Oct. 2008.

[28] C. Monzio Compagnoni, R. Gusmeroli, A. S. Spinelli, and A. Visconti, "Analytical model for the electron-injection statistics during programming of nanoscale NAND flash memories," *IEEE Trans. Electron Devices*, vol. 55, no. 11, pp. 3192–3199, Nov. 2008.

[29] C. Monzio Compagnoni, C. Miccoli, A. L. Lacaita, A. Marmiroli, A. S. Spinelli, and A. Visconti, "Impact of control-gate and floating-gate design on the electron-injection spread of decananometer NAND flash memories," *IEEE Electron Device Lett.*, vol. 31, no. 11, pp. 1196–1198, Nov. 2010.

[30] G. M. Paolucci, C. Miccoli, C. Monzio Compagnoni, A. S. Spinelli, and A. L. Lacaita, "String current in decananometer NAND flash arrays: A compact-modeling investigation," *IEEE Trans. Electron Devices*, vol. 59, no. 9, pp. 2331–2337, Sep. 2012.