

Compact Modeling of the Program Saturation Regime in Charge Trap Flash Memories

Thomas Hellemans
Imec and KU Leuven
3001 Leuven, Belgium
thomas.hellemans@imec.be

Devin Verreck
Imec and KU Leuven
3001 Leuven, Belgium
devin.verreck@imec.be

Antonio Arreghini
Imec
3001 Leuven, Belgium
antonio.arreghini@kuleuven.be

Geert Van den bosch
Imec
3001 Leuven, Belgium
geert.vandenbosch@imec.be

Maarten Rosmeulen
Imec and KU Leuven
3001 Leuven, Belgium
maarten.rosmeulen@imec.be

Michel Houssa
KU Leuven and Imec
3001 Leuven, Belgium
michel.houssa@kuleuven.be

Jan Van Houdt
Imec and KU Leuven
3001 Leuven, Belgium
jan.vanhoudt@imec.be

Abstract—We propose and assess three distinct methods to include detrapping in compact models for charge trap flash memories. Validation by drift-diffusion TCAD simulations shows that treating detrapped carriers as an additional injection current into the conduction band yields the best agreement. However, the assumption of a uniform trapped-charge distribution in compact models limits their accuracy in the program saturation regime. Calibration to experiments on planar SONOS capacitors demonstrates that including detrapping is essential to reproduce the program saturation behavior across devices with varying layer thicknesses.

Index Terms—Charge Trap Memory, Flash Memory, Compact Modeling, TCAD, Program Saturation, Detrapping, Emission

I. INTRODUCTION

Charge trap flash memories store data by trapping electrons in a nitride or high- κ charge trapping layer (CTL). Accurate simulation of the programming characteristics is essential for technology development, yet full drift-diffusion solutions are computationally expensive. Compact models like Pheido (previously developed by us [1], [2]) and SNU/Integra-3NP [3] address this by deriving a semi-analytical expression for the threshold voltage change rate dV_T/dt during programming. However, the original Pheido formulation considers only trapping and neglects detrapping (emission from traps). As a consequence, it cannot reproduce the program saturation regime, where trapping and detrapping approach a dynamic equilibrium. Therefore, in this work, we extend the Pheido model by incorporating detrapping through three distinct methods. We derive the resulting correction terms analytically and check the validity of the three methods with one-dimensional drift-diffusion TCAD simulations. Finally, we calibrate the extended Pheido model to incremental step pulse programming (ISPP) measurements on planar SONOS capacitors with varying layer thicknesses.

This work was supported by imec's Industrial Affiliation Program on storage memories and Fonds Wetenschappelijk Onderzoek Vlaanderen under grant number 1SS0925N.

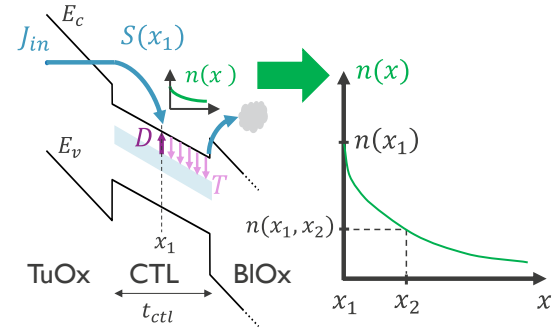


Fig. 1. Schematic band diagram of the device for the derivation of Pheido including detrapping within method 2 or 3, i.e. by including detrapping in the boundary condition for $n(x)$.

II. COMPACT MODEL DERIVATION

We first summarize the original Pheido derivation [1], [2], which starts from the electron continuity and drift-diffusion equations [4]. The device is illustrated in Fig. 1.

$$\frac{dn}{dt} = \frac{1}{q} \nabla \cdot J_n + (G_n - R_n) \quad (1)$$

$$J_n = q\mu_n n \mathcal{E} + qD_n \nabla n \quad (2)$$

where n is the free electron density, t the time, q the elementary charge, J_n the electron current density, μ_n the electron mobility, $\mathcal{E}$ the electric field, D_n the electron diffusion constant, and R_n and G_n the recombination and generation rates, respectively. Throughout this paper, currents and fields are negative when electrons move in the positive x -direction. To obtain a semi-analytical solution, we assume a slowly varying CTL electric field ($\nabla \cdot \mathcal{E} \approx 0$, so $\mathcal{E} \approx \mathcal{E}_{ctl}$), steady state within each pulse ($dn/dt = 0$), a single trap reservoir, i.e. spatially uniform trap and trapped-carrier concentrations (N_T, n_T), and drift-dominated transport ($D_n \nabla n \ll \mu_n n \mathcal{E}$).

We first consider the case without detrapping, in which the net recombination-generation term reduces to

$$G_n - R_n = -R_c n \quad (3)$$

where R_C is the electron trapping (capture) rate. Substituting (2) and (3) into (1) under the above assumptions yields, in one dimension

$$\mu\mathcal{E}_{ctl} \frac{\partial n}{\partial x} = R_C n \quad (4)$$

With injection current J'_{in} at position x_1 , the boundary condition for this differential equation follows from (2):

$$n(x_1) = \frac{J'_{in}}{q\mu\mathcal{E}_{ctl}} \quad (5)$$

For spatially uniform R_C the solution of differential equation(4) in combination with boundary condition (5) is

$$n(x) = \frac{J'_{in}}{q\mu\mathcal{E}_{ctl}} \exp\left(\frac{R_C}{\mu\mathcal{E}_{ctl}}(x - x_1)\right) \quad (6)$$

To account for energy relaxation, the convolution injection model distributes the injection current J_{in} from the channel through the tunnel oxide (TuOx) into the CTL over the whole CTL thickness according to a normalized shape function $S(x)$ [2], [5] (Fig. 1). Hence, the injection current per unit length at position x_1 can be written as $J_{in}S(x_1)$. The carrier density per unit length injected at x_1 that reaches position x_2 without being trapped is therefore

$$n(x_1, x_2) = \frac{J_{in}S(x_1)}{q\mu\mathcal{E}_{ctl}} \exp\left(\frac{R_C}{\mu\mathcal{E}_{ctl}}(x_2 - x_1)\right) \quad (7)$$

Integrating over the CTL thickness, i.e. all injection points, with a uniform shape function $S(x) = 1/t_{ctl}$ gives the total carrier density reaching the CTL - blocking oxide (BIOx) interface:

$$n_{blox} = \int_{x_{ctl}}^{x_{blox}} n(x, x_{blox}) dx \quad (8)$$

The outgoing current density at the BIOx interface is

$$J_{out} = -qv_{blox}n_{blox} = q\mu\mathcal{E}_{ctl}n_{blox} \quad (9)$$

Now, the threshold voltage change rate is obtained via charge conservation, i.e. by relating the trapped charge increase to the injection and output currents. This yields the original Pheido expression [1], [2]:

$$\begin{aligned} \frac{dV_T(t)}{dt} \Big|_{R_E=0} &= -\frac{J_{in} - J_{out}}{C_{top}} = -\frac{J_{in}}{C_{eq}}(1 - \alpha) \\ \alpha &= -\frac{\mu\mathcal{E}_{ctl}}{R_C t_{ctl}} \left(1 - \exp\left(\frac{R_C t_{ctl}}{\mu\mathcal{E}_{ctl}}\right)\right) \end{aligned} \quad (10)$$

where the subscript $R_E = 0$ indicates that detrapping is absent, C_{eq} is the equivalent capacitance per unit area between the trapped charge centroid and the gate [2], [6], and α is the ratio of carriers reaching the blocking oxide (BIOx) interface without being trapped. The electric fields in the CTL and BIOx are evaluated from a simple capacitor model [6].

The threshold voltage shift after i ISPP pulses is obtained from (10) as

$$\Delta V_{T,i} = \sum_{p=1}^i \Delta V_{T,p} = \sum_{p=1}^i \int_0^{t_p} \frac{dV_T(t)}{dt} dt \quad (11)$$

with t_p the pulse duration of pulse p [2]. The integral can be solved as an initial value problem.

We now extend the Pheido model to include detrapping with emission rate R_E from traps to the CTL conduction band, assuming R_E is spatially uniform. We consider three different methods, each adding a correction $\delta_k < 0$ to the threshold voltage change rate

$$\frac{dV_T}{dt} \Big|_{R_E \neq 0} = \frac{dV_T}{dt} \Big|_{R_E=0} + \delta_k \quad (12)$$

The most straightforward approach replaces (3) with $G_n - R_n = R_E n_T - R_C n$. This corresponds to the implementation of detrapping in numerical drift-diffusion simulators. Re-deriving (3) to (12) results in:

$$\delta_1 = \frac{q\mu\mathcal{E}_{ctl}}{C_{eq}} \frac{R_E}{R_C} n_T (1 - \alpha) \quad (13)$$

with $\mu\mathcal{E}_{ctl}$ the drift velocity of free carriers and $R_E n_T / R_C$ the equilibrium free carrier density sustained by detrapping. This equilibrium density can also be obtained as the steady-state solution of the coupled continuity equations for free and trapped carriers (neglecting currents):

$$\begin{cases} \partial n / \partial t = R_E n_T - R_C n \\ \partial n_T / \partial t = -R_E n_T + R_C n \end{cases} \Rightarrow n = \frac{R_E}{R_C} n_T \text{ for } 0 \ll t \quad (14)$$

Counterintuitively, (13) includes $1 - \alpha$, i.e. the probability for carriers injected into the CTL conduction band to become trapped.

Methods 2 and 3 leave the continuity equation (3) unchanged and instead model detrapping as an additional free-carrier source in the boundary condition for $n(x)$ (5). Assuming local equilibrium between free and trapped carriers, the boundary condition becomes $n(x_1) = J_{in}/(q\mu\mathcal{E}_{ctl}) + R_E n_T / R_C$. The first term is the injection driven free-carrier density, while the second term is the equilibrium free carrier density sustained by detrapping (14). This adapted boundary condition yields

$$\delta_2 = \frac{q\mu\mathcal{E}_{ctl}}{C_{eq}} \frac{R_E}{R_C} n_T \alpha \quad (15)$$

which is very similar to (13), but includes α , the probability for free carriers to reach the BIOx interface, instead of $1 - \alpha$.

Alternatively, the detrapped carriers can be treated as an additional injection current $J_{detrap} = qR_E t_{ctl} n_T$ into the CTL conduction band, modifying the boundary condition (5) to $n(x_1) = (J_{in} + qR_E t_{ctl} n_T)/(q\mu\mathcal{E}_{ctl})$, which yields

$$\delta_3 = -\frac{qR_E t_{ctl}}{C_{eq}} n_T \alpha \quad (16)$$

Here, the drift velocity term $\mu\mathcal{E}_{ctl}$ is replaced by an effective detrapping velocity $R_E t_{ctl}$, which is multiplied by the trapped charge density n_T and α to obtain the detrapping current density reaching the BIOx interface. Note that (16) is similar to the detrapping implementation in [3], but includes an additional factor α . This can be understood as accounting for

the possibility that detrapped carriers are recaptured before reaching the CTL-BIOx interface. This is particularly relevant for devices with thick CTLs.

III. COMPACT MODEL VALIDATION

Now we proceed by validating the three detrapping methods with one-dimensional drift-diffusion TCAD simulations. For illustration we use the following models for carrier injection, capture and emission for both the compact model and the TCAD simulations. Note however that any carrier injection, capture or emission model can be used in this compact modeling framework, as long as it depends solely on constants, the TuOx and BIOx electric field and averaged quantities within the CTL, e.g. $\mathcal{E}_{ctl}$ or n_T . First, the injection current density J_{in} is described by a Fowler-Nordheim based tunneling model

$$J_{in} = qv_{th}n_{ch} \exp\left(-\frac{4\sqrt{2m_{tuox}^*}E_B^{3/2}}{3\hbar q\mathcal{E}_{tuox}}\right) \quad (17)$$

with v_{th} the thermal velocity, n_{ch} the carrier density in the inversion layer at the channel-TuOx interface, m_{tuox}^* the effective mass in the TuOx, $\hbar$ the reduced Planck constant, $\mathcal{E}_{tuox}$ the electric field in the tunnel oxide and E_B the conduction band offset between the channel and TuOx. Second, the capture rate is given by

$$R_C = \sigma v_{th}(N_T - n_T) \quad (18)$$

where σ is the capture cross section. Third, the detrapping rate is modeled using a Poole-Frenkel emission model:

$$R_E = \nu \exp\left(\frac{\beta\sqrt{\mathcal{E}_{ctl}} - E_t}{kT}\right) \quad \text{with} \quad \beta = \sqrt{\frac{q^3}{\pi\epsilon_0\epsilon_\infty}} \quad (19)$$

with ν the escape frequency, E_T the trap energy level with respect to the conduction band, T the temperature and ϵ_∞ the high-frequency dielectric constant [7]. We draw three main conclusions:

First, among the three detrapping extensions, method 3 (16) provides the closest agreement with one-dimensional drift-diffusion TCAD simulations with a single trap reservoir (Figs. 2, 3). This single trap reservoir setup in the numerical drift-diffusion simulations is chosen to match Pheido's assumption of a spatially uniform trapped-charge distribution (single trap reservoir). For method 1, the threshold voltage correction is underestimated due to double counting errors in the trapping and detrapping contributions when integrating over the CTL to calculate the carrier density reaching the CTL-BIOx interface (8). For method 2, the threshold voltage shift is overestimated because its derivation assumes local equilibrium between free and trapped carriers. Such an equilibrium is not attained within a single programming pulse of 100 μs . By contrast, method 3 models detrapping as an additional injection current into the CTL and reproduces the drift-diffusion results within the single-reservoir framework, confirming that δ_3 is the optimal detrapping correction for compact models of this type.

Second, we conclude that the assumption of a single trap reservoir significantly limits Pheido's accuracy, with or without detrapping, in the program saturation regime (Figs.

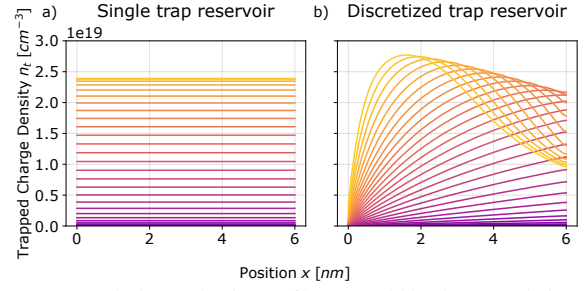


Fig. 2. Trapped charge density profiles n_T within the CTL during ISPP (Fig. 3 and 4). (a) The single trap reservoir approach, which assumes a spatially uniform n_T . (b) The spatially discretized trap reservoir approach, which allows for a non-uniform n_T within the CTL.

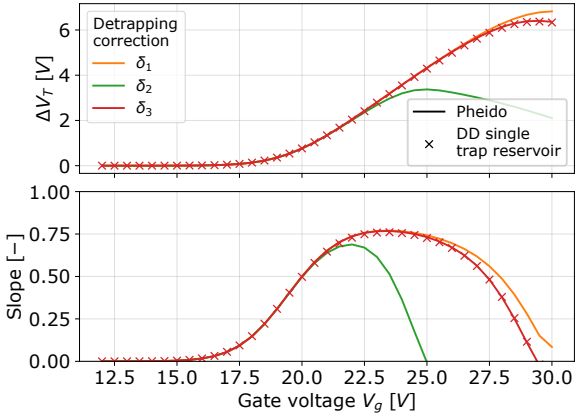


Fig. 3. Simulated ISPP curves for a planar SONOS capacitor ($t_{tuox} = 4 \text{ nm}$, $t_{ctl} = 6 \text{ nm}$, $t_{blox} = 8 \text{ nm}$), comparing different Pheido corrections for detrapping δ_i to TCAD drift-diffusion (DD) simulations with a single trap reservoir. Correction δ_3 (16) provides the best agreement with the drift-diffusion results.

2, 4). This becomes apparent from the difference in ISPP curves obtained with a single trap reservoir versus a spatially discretized trap reservoir in drift-diffusion simulations. Therefore, compact models that rely on a single trap reservoir are inherently unable to capture the full program saturation behavior.

Third, we find that despite these limitations, Pheido with the detrapping correction can still yield good calibration results on experiments (Fig. 5). The fitted parameters (Table I) only approximate the physical values, while more accurate values require spatially discretized drift-diffusion simulations. The experiments are performed on planar capacitors with varying TuOx, CTL and BIOx thicknesses. In the simulations, all layer thicknesses were scaled by the ratio of the measured capacitance equivalent thickness (CET) and targeted effective oxide thickness (EOT). This ratio ranged between 0.89 and 0.92. For the calibration of a single device, without detrapping, the ratio N_T/μ determines the programming onset ((10) and (18)), while the absolute value of N_T can be varied independently to fit the program saturation regime [2]. However, this method fails when applied to multiple devices with different CTL thicknesses simultaneously (Fig. 5, dashed line). For

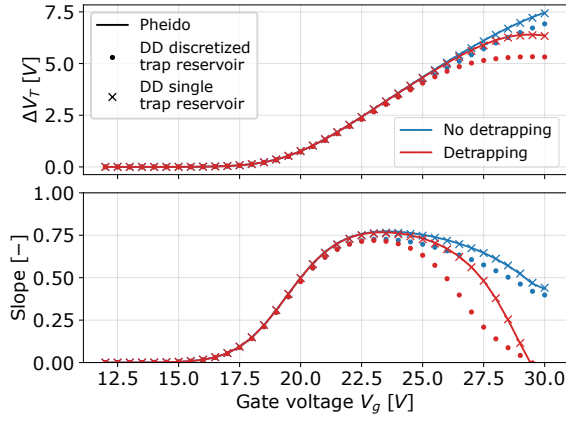


Fig. 4. Simulated ISPP curves for a planar SONOS capacitor ($t_{tuo} = 4 \text{ nm}$, $t_{ctl} = 6 \text{ nm}$, $t_{blo} = 8 \text{ nm}$), comparing Pheido and TCAD drift-diffusion (DD) simulations with and without detrapping. In the drift-diffusion results, both a single trap reservoir (uniform trap filling in the CTL, cf. Fig. 2a) and a spatially discretized trap reservoir (cf. Fig. 2b) are considered. This illustrates that the single trap reservoir assumption significantly limits the accuracy of compact models in the program saturation regime.

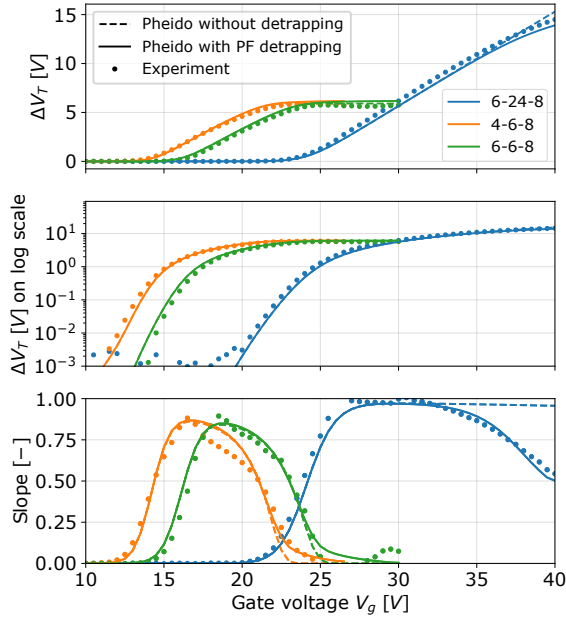


Fig. 5. Experimental and simulated ISPP curves for planar SONOS capacitors with varying layer thicknesses $t_{tuo} - t_{ctl} - t_{blo}$ [nm]. Including detrapping is necessary to obtain a good calibration of the program saturation regime across all devices.

thick CTL, the N_T calibrated to the thin CTL devices is too high, and detrapping, rather than trap filling, governs program saturation. Including detrapping for all devices with a recalibrated parameter set (different N_T and μ , same N_T/μ) reproduces the full ISPP behavior across all devices (Fig. 5, full line). The updated compact model captures both the programming onset (log-scale ISPP) and the saturation regime (slope degradation in the ISPP).

TABLE I
SIMULATION PARAMETERS: CALIBRATED TO EXPERIMENTS IN FIG. 5

Symbol	Meaning	Value (without with detrapping)
ϵ_{ox}	Permittivity oxide	$3.9 \epsilon_0$
ϵ_{ctl}	Permittivity CTL	$7.4 \epsilon_0$
t_p	Pulse duration	$100 \mu s$
T	Temperature	300 K
n_{ch}	Channel electron density	$1.0e21 \text{ cm}^{-3}$
v_{th}	Thermal velocity	$1.0e7 \text{ cm/s}$
m_{TuOx}^*	Effective mass TuOx	$0.5 m_e$
E_B	Channel-CTL band offset	3.25 eV
σ	Capture cross-section	$1.0e-16 \text{ cm}^2$
N_T	Trap density	$2.8e19 \mid 3.0e19 \text{ cm}^{-3}$
μ	Mobility	$3.49e-3 \mid 3.74e-3 \frac{\text{cm}^2}{\text{Vs}}$
E_t	Trap energy	$- \mid 1.6 \text{ eV}$
ν	Attempt-to-escape frequency	$- \mid 2.5e13 \text{ Hz}$
ϵ_∞	High-frequency permittivity	$- \mid 5.0 \epsilon_0$

IV. CONCLUSION

In summary, we derived three analytical detrapping corrections for compact models based on averaged quantities in the CTL and found that modeling carrier emission as an additional injection current into the CTL conduction band best matches 1D drift-diffusion TCAD. The remaining saturation-regime error stems from the single trap reservoir assumption, so δ_3 is the best possible correction within the limitations of these frameworks. It extends Pheido's applicability to the program saturation regime for devices with varying layer thicknesses. Further accuracy on the ISPP and calibrated parameters requires a spatially resolved trapped charge distribution, which can be achieved by numerical drift-diffusion simulations.

REFERENCES

- [1] D. Verreck, A. Arreghini, F. Schanovsky, G. Rzepa, Z. Stanojević, F. Mitterbauer, C. Kernstock, O. Baumgartner, M. Karner, G. Van den Bosch, and M. Rosmeulen, "Understanding the ISPP Slope in Charge Trap Flash Memory and its Impact on 3-D NAND Scaling," in *2021 IEEE International Electron Devices Meeting (IEDM)*, Dec. 2021, pp. 1-4.
- [2] D. Verreck, F. Schanovsky, A. Arreghini, G. Van den Bosch, Z. Stanojević, M. Karner, and M. Rosmeulen, "Modeling the Operation of Charge Trap Flash Memory—Part II: Understanding the ISPP Curve With a Semianalytical Model," *IEEE Trans. Electron Devices*, vol. 71, no. 1, pp. 554-559, Jan. 2024.
- [3] J. Yoo, H. Choi, and H. Shin, "A semi-analytical physics-based transient model for program (PGM) operation of charge-trap-based 3-D NAND flash memories," *J Comput Electron*, vol. 24, no. 5, p. 148, Jul. 2025.
- [4] S. M. Sze and M. K. Lee, *Semiconductor Devices, Physics and Technology*, 3rd ed. Hoboken, N.J.: Wiley, 2012.
- [5] F. Schanovsky, D. Verreck, Z. Stanojević, S. Schallert, A. Arreghini, G. van den Bosch, M. Rosmeulen, and M. Karner, "Modeling the Operation of Charge Trap Flash Memory—Part I: The Importance of Carrier Energy Relaxation," *IEEE Trans. Electron Devices*, vol. 71, no. 1, pp. 547-553, Jan. 2024.
- [6] S. M. Amoroso, C. M. Compagnoni, A. Mauri, A. Maconi, A. S. Spinelli, and A. L. Lacaita, "Semi-Analytical Model for the Transient Operation of Gate-All-Around Charge-Trap Memories," *IEEE Trans. Electron Devices*, vol. 58, no. 9, p. 3116, 2011.
- [7] J. Frenkel, "On Pre-Breakdown Phenomena in Insulators and Electronic Semi-Conductors," *Phys. Rev.*, vol. 54, no. 8, pp. 647-648, Oct. 1938.