Received 1 July 2024; accepted 16 August 2024. Date of publication 21 August 2024; date of current version 30 August 2024.
The review of this article was arranged by Editor L. Wei.

Digital Object Identifier 10.1109/JEDS.2024.3447150

Discrete-Trap Effects on 3-D NAND Variability – Part II: Random Telegraph Noise

GERARDO MALAVENA[1] (Member, IEEE), SALVATORE M. AMOROSO[2], ANDREW R. BROWN[2], PLAMEN ASENOV[2], XI-WEI LIN[2] (Member, IEEE), VICTOR MOROZ[2], MATTIA GIULIANINI[1], DAVID REFALDI[1], CHRISTIAN MONZIO COMPAGNONI[1] (Senior Member, IEEE), AND ALESSANDRO S. SPINELLI[1] (Senior Member, IEEE)

1 Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milan, Italy
2 Synopsys Inc., Mountain View, CA 94043, USA

CORRESPONDING AUTHOR: G. MALAVENA (e-mail: gerardo.malavena@polimi.it)

ABSTRACT In Part II of this article we discuss the impact of a discrete treatment of traps on 3-D NAND Flash random telegraph noise (RTN). A higher RTN results when discrete traps are taken into account, that can only be explained by a stronger influence of the discrete charged traps on the current conduction, leading to more percolation. The effects are then investigated as a function of the cell parameters, showing that a continuous model for traps cannot reproduce the correct dependence.

INDEX TERMS 3-D NAND Flash memories, variability, random telegraph noise, discrete traps.

I. INTRODUCTION

In Part I of this work we discussed the improvement in device modeling that follows the adoption of a discrete treatment of the traps in the polysilicon channel of 3-D NAND cells, showing that models relying on a continuous trap (CT) approach overestimate V_T, as a discrete-trap (DT) treatment allows for additional current percolation and fewer occupied traps. Here, we extend such analysis to random telegraph noise (RTN), *i.e.,* V_T fluctuations following the capture of an electron by a trap. RTN is an important reliability issue in 3-D NAND cells [1], [2], [3], affecting the V_T distribution width [4], [5] while acting as a process monitor [6], [7]. Indeed, RTN has been already investigated by several authors in a CT [8], [9], [10], [11], [12], [13], [14], [15], [16] and DT [17], [18] approach, but a thorough comparison between CT and DT results is still missing. As with Part I, preliminary results have been presented in [19]; here we expand the physical analysis and discuss the dependence on device parameters and temperature. Results of this work make clear that a CT model cannot reproduce the correct dependence on device parameters provided by a full DT approach.

II. SIMULATION RESULTS

We refer the reader to Part I [20] for details and numerical values of the device parameters and just recall here the main features of our simulations: The structure is a 3-D NAND string consisting of a central cell plus two half-cells per side. Memory cells have a Macaroni structure [21], made of an inner filler oxide, a thin layer of polysilicon channel with 10 nm thickness, an oxide/nitride/oxide stack, and a metal wordline with length $L = 35$ nm (see Fig. 1 in Part I). Polysilicon is described by random grains with average size $D_g = 30$ nm and traps located at their grain boundaries (GBs), having an energy distribution derived from [22]. A polysilicon/oxide interface state density has been also considered.

3-D Monte Carlo simulations were performed using Sentaurus Process [23] and Sentaurus Device [24] for processing the typical cell structure and Garand VE [25] for studying the statistical behavior of the cells. For the simulation of RTN we follow a standard procedure [26] and begin with performing a drain current vs. wordline voltage $I-V$ simulation up to the threshold level of 10 nA. Then, one additional trap is filled with an electron and a second $I-V$ curve is repeated, leading to a higher V_T. The RTN fluctuation ΔV_T is the difference value. This procedure is repeated with random grain number and positions, randomly locating the RTN traps either at the polysilicon/oxide interfaces or at GBs, depending on the relative densities. It is important to note that the RTN

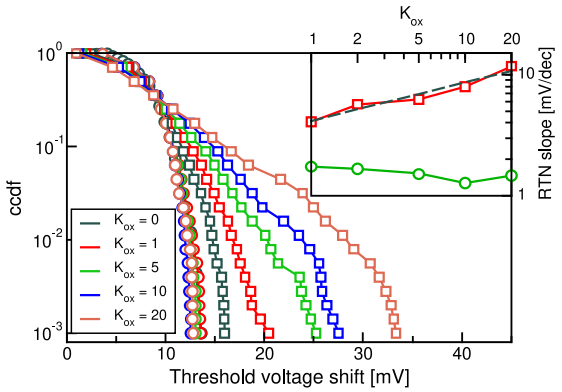

FIGURE 1. ΔV_T ccdf for different values of the polysilicon/oxide interface trap density. Inset shows the RTN slope as a function of K_{ox}, with the dashed line highlighting a power law with exponent 0.32. In both graphs: Circles = CTs, squares = DTs.

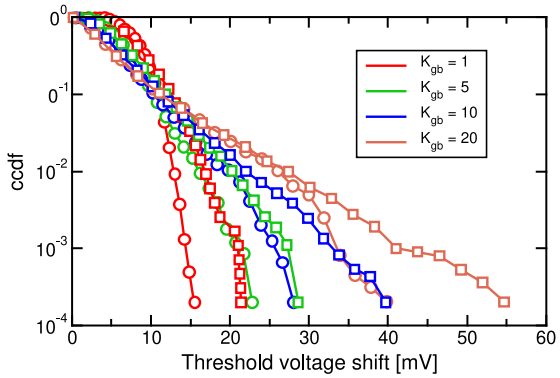

FIGURE 2. ΔV_T ccdf for different values of the GB trap density with $K_{ox} = 0$. Circles = CTs, squares = DTs.

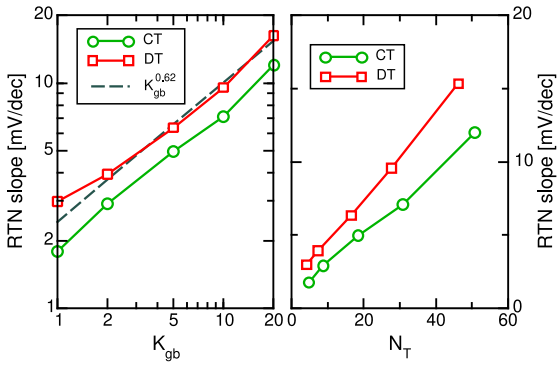

FIGURE 3. Left: λ as a function of the GB trap density. The dashed line shows a power law with exponent 0.62. Right: λ as a function of the number of filled traps in the cell channel region. In both graphs: Circles = CTs, squares = DTs.

trap is always treated as discrete, irrespective of the model adopted for all the other ones. Note also that this procedure amounts to computing the single-trap RTN distribution, that can however be easily linked to the real multi-trap one [27].

A. POLYSILICON/OXIDE INTERFACE TRAPS

Fig. 1 shows the ΔV_T complementary cumulative distribution function (ccdf) when the polysilicon/oxide interface state density is multiplied by a factor K_{ox}. Note that such increase has no effect on the CT ΔV_T distributions (circles), that present vanishingly small exponential tails with inverse slope λ of less than 2 mV/decade, but results in clear exponential distributions with λ increasing with K_{ox} whenever DTs are accounted for. In fact, in the CT case the trap distribution is uniform along the device, and the only percolation source is the GB traps, resulting in λ being independent of K_{ox}. In the DT case, instead, the Coulomb peaks at the filled trap sites generate a percolative current flow, by analogy with the influence of DTs on V_T already discussed in Part I. The magnitude of such effect is displayed in the inset of Fig. 1: λ in the DT case (λ_{DT}) changes from about 4 to 12 mV/decade in the investigated range, and follows a power-law with exponent 0.32. For $K_{ox} = 20$, λ_{DT} is about a factor of 8 larger than its CT counterpart (λ_{CT}).

B. GRAIN BOUNDARY TRAPS

To study the GB trap density dependence, we first removed the oxide traps from the following simulations, in order not to overshadow the low-density data, where they become dominant. ΔV_T ccdfs are shown in Fig. 2 when the GB trap density is multiplied by a factor K_{gb}: both distributions are now affected by K_{gb}, although larger ΔV_T values are reached in the DT case. The λ dependence on K_{gb} is plotted in Fig. 3 (left): both models return a power-law with exponent of about 0.62 (dashed line), with λ_{DT} remaining higher by about 35% with respect to λ_{CT}. In terms of trap density, this means that an increase by a factor of 58% is required to a CT approach in order to obtain the correct slope. For $K_{gb} = 1$, our data show a deviation from the power law. However, the exponential tail of the ΔV_T distribution cannot be neatly distinguished here, and a larger number of simulations would be required for a reliable extraction of λ.

Fig. 3 (right) shows λ versus the average number of occupied traps N_T. As a difference from V_T in Part I, no unique behavior is observed, and λ_{DT} is higher by about 50% with respect to λ_{CT}, confirming the stronger percolation through discrete charged traps. In the CT case, instead, the RTN is due to the randomness in the GB configuration [11].

C. GRAIN SIZE

Fig. 4 (left) shows the dependence of λ on D_g for different values of K_{gb}. The increase in λ with smaller grains has been reported in [15] and can be understood recalling (see Part I) that the GB surface trap density is independent of D_g, meaning that decreasing D_g we obtain more grains and GBs, hence more traps. More important, results show that the difference between λ_{CT} and λ_{DT} becomes less relevant for smaller D_g. To clarify this point, we can look at data as a function of N_T (right), showing that the impact of traps on λ decreases for smaller grains, where we have more GBs, hence a more even distribution of traps and weaker percolation, *i.e.,* smaller λ and smaller differences between

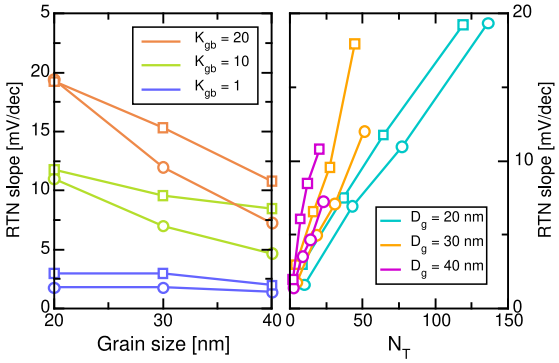

FIGURE 4. RTN slope as a function of the grain size (left) and N_T (right). Circles = CTs, squares = DTs.

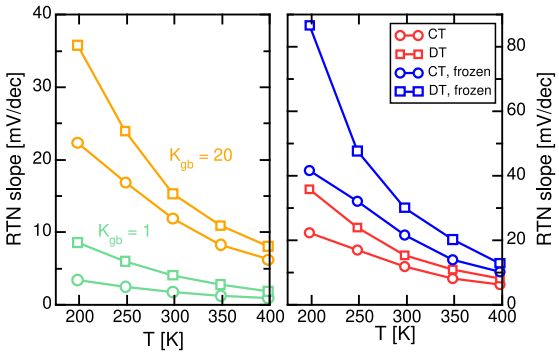

FIGURE 5. Left: λ as a function of temperature for K_{gb} = 1 and 20. Right: Impact of traps equilibrium condition on the temperature dependence of λ for K_{gb} = 20. In both graphs: Circles = CTs, Squares = DTs.

λ_{CT} and λ_{DT}. Finally, when data are plotted as a function of K_{gb} (not shown), a behavior similar to Fig. 3 arises, with a power-law exponent changing from 0.56 ($D_g = 40$ nm) to 0.71 ($D_g = 20$ nm).

D. TEMPERATURE

Fig. 5 (left) shows λ as a function of temperature T for the limiting cases of $K_{gb} = 1$ and 20. The increase of λ at low T has been already discussed in [11], but the main result is the magnitude of the difference between λ_{CT} and λ_{DT}, that grows up to about 100% for $K_{gb} = 1$ and 60% for $K_{gb} = 20$, at the lowest considered T. This result makes the adoption of a DT approach even more compelling in low-temperature applications.

As T is lowered, however, the role played by the time constants of capture/emission to/from traps should be considered. In fact, as an RTN trap is filled by an electron, the occupation probability of nearby traps lowers slightly, as the conduction band is locally raised. Traps will then reach the new equilibrium condition within a few capture/emission time constants. However, at low T such time constants increase [12], and it becomes harder for traps to reach the new equilibrium condition. A full treatment of this phenomenon would require a time-dependent Monte Carlo, but the magnitude of the effect can be grasped by comparing

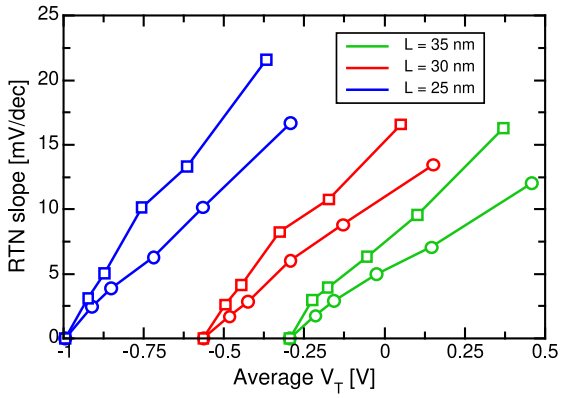

FIGURE 6. RTN slope as a function of the average V_T for different values of L. Circles = CTs, squares = DTs.

two approaches: one in which all trap occupations reach the equilibrium condition after filling of the RTN trap (our default case), and one where trap occupations are "frozen" to the pre-capture condition: As T is reduced, we could envisage a transition from the former to the latter regime. Data for $K_{gb} = 1$ (not shown) feature no difference in the CT case and an increase by 17% in λ_{DT} at 198 K; data for $K_{gb} = 20$ are reported in Fig. 5 (right), showing a very significant difference, with slopes increasing by about 60% at high T and by more than a factor of 2 in λ_{DT} at 198 K (where the increase in λ_{CT} is about 86%).

III. DISCUSSION

It remains now to be seen whether DT effects can be reproduced by a CT model via a change in the trap density or not. In such case, a CT approach could still be employed, although with the wrong value for the trap density; otherwise, a DT approach would become necessary to reproduce the correct dependences. The latter case is obviously the one for the interface trap density, where the RTN slopes are very different (see inset of Fig. 1). For the K_{gb} dependence, we reported λ as a function of the average V_T (taken from Part I) in Fig. 6, for all values of K_{gb}, extending the study to different wordline lengths L. So, each curve represents the possible values of V_T and λ that can be achieved with a given model when the trap density is changed. As CT and DT curves have different slopes, a DT approach is needed to capture the correct dependences.

As for the dependence on L, results for crystalline planar [28] and nanowire devices [26] already showed that λ increases in shorter channels, here by a factor of about $35-40\%$ (similar for λ_{CT} and λ_{DT}) as L is shrunk from 35 to 25 nm. Such trends are the results of different phenomena, including an increased impact of charged traps in shorter channels and a reduction in their number. Moreover, short-channel effects also play a contribution, making it harder for the gate to control the channel conduction and offset the impact of the RTN trap.

Finally, Fig. 7 shows results similar to Fig. 6, but for different values of D_g (see Fig. 4). In this case as well,

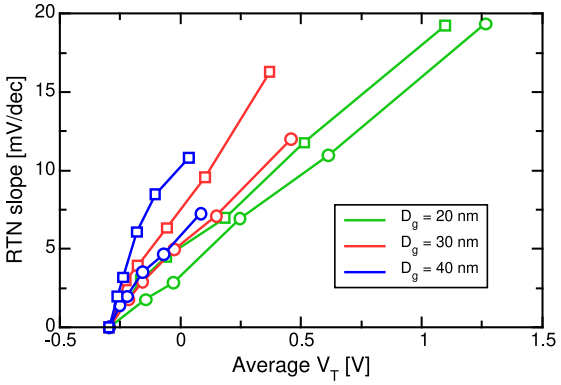

FIGURE 7. RTN slope as a function of the average V_T for different values of D_g. Circles = CTs, squares = DTs.

no overlap can be seen between the CT and DT curves, strengthening the need for a treatment of the discrete nature of traps.

IV. CONCLUSION

We have studied the impact of a DT approach on RTN in 3-D NAND with polysilicon channel. DT simulations lead to larger RTN and stronger dependence on the trap density, which is explained in terms of a stronger percolation through discretely charged traps. The dependence on the cell parameters shows that a model based on CTs cannot reproduce the correct dependences predicted by a DT-based one.

REFERENCES

[1] A. S. Spinelli, C. Monzio Compagnoni, and A. L. Lacaita, "Reliability of NAND Flash memories: Planar cells and emerging issues in 3D devices," *Computers*, vol. 6, p. 16, 2017, doi: 10.3390/computers6020016.
[2] A. Goda, "3D NAND technology achievements and future scaling perspectives," *IEEE Trans. Electron Devices*, vol. 67, no. 4, pp. 1373–1381, Apr. 2020, doi: 10.1007/TED.2020.2968079.
[3] A. S. Spinelli, G. Malavena, A. L. Lacaita, and C. Monzio Compagnoni, "Random telegraph noise in 3D NAND Flash memories," *Micromachines*, vol. 12, no. 6, p. 703, 2021, doi: 10.3390/mi12060703.
[4] C. Monzio Compagnoni, M. Ghidotti, A. L. Lacaita, A. S. Spinelli, and A. Visconti, "Random telegraph noise effect on the programmed threshold-voltage distribution of Flash memories," *IEEE Electron Device Lett.*, vol. 30, pp. 984–986, Sep. 2009, doi: 10.1109/LED.2009.2026658.
[5] C. C. Lu et al., "Analysis and realization of TLC or even QLC operation with a high performance multi-times verify scheme in 3D NAND Flash memory," in *IEDM Tech. Dig.*, 2018, pp. 31–34, doi: 10.1109/IEDM.2018.8614548.
[6] M. Andrade et al., "RTN assessment of traps in polysilicon cylindrical vertical FETs for NVM application," *Microelectron. Eng.*, vol. 109, pp. 105–108, 2013, doi: 10.1016/j.mee.2013.03.019.
[7] H.-J. Kang et val., "Effect of traps on transient bit-line current behavior in word-line stacked NAND Flash memory with poly-Si body," in *Proc. Symp. VLSI Tech. Dig.*, 2014, pp. 24–25, doi: 10.1109/VLSIT.2014.6894348.
[8] M.-K. Jeong et al., "Analysis of random telegraph noise and low frequency noise properties in 3-D stacked NAND Flash memory with tube-type poly-Si channel structure," in *Proc. Symp. VLSIT Tech. Dig.*, 2012, pp. 55–56, doi: 10.1109/VLSIT.2012.6242458.
[9] P.-Y. Wang and B.-Y. Tsui, "A novel approach using discrete grain-boundary traps to study the variability of 3-D vertical-gate NAND Flash memory cells," *IEEE Trans. Electron Devices*, vol. 62, no. 8, pp. 2488–2493, Aug. 2015, doi: 10.1109/TED.2015.2438001.
[10] A. Subirats et al., "Experimental and theoretical verification of channel conductivity degradation due to grain boundaries and defects in 3D NAND," in *IEDM Tech. Dig.*, 2017, pp. 517–520, doi: 10.1109/IEDM.2017.8268433.
[11] G. Nicosia et al., "Characterization and modeling of temperature effects in 3-D NAND Flash arrays—Part II: Random telegraph noise," *IEEE Trans. Electron Devices*, vol. 65, no. 8, pp. 3207–3213, Aug. 2018, doi: 10.1109/TED.2018.2839904.
[12] G. Nicosia, A. Goda, A. S. Spinelli, and C. Monzio Compagnoni, "Investigation of the temperature dependence of random telegraph noise fluctuations in nanoscale polysilicon-channel 3-D Flash cells," *Solid-State Electron.*, vol. 151, pp. 18–22, 2019, doi: 10.1016/j.sse.2018.10.010.
[13] N. Choi, H.-J. Kang, and J.-H. Lee, "Analysis of random telegraph noise characteristics with two different cell states in 3-D NAND Flash memory," *J. Semicond. Technol. Sci.*, vol. 19, pp. 153–157, Apr. 2019, doi: 10.5573/JSTS.2019.19.2.153.
[14] A. S. Spinelli, C. Monzio Compagnoni, and A. L. Lacaita, "Random telegraph noise in Flash memories," in *Noise in Nanoscale Semiconductor Devices*, T. Grasser, Ed., Cham, Switzerland: Springer, 2020, ch. 6, pp. 201–227.
[15] B. Song, H. Liu, L. Jin, X. Fu, F. liu, and Z. Huo, "Impact of stacking layers on RTN in 3D charge trapping NAND Flash memory," *Microelectron. Rel.*, vol. 127, Dec. 2021, Art. no. 114415, doi: 10.1016/j.microrel.2021.114415.
[16] X. Jia et al., "Investigation of random telegraph noise under different programmed cell V_t levels in charge trap based 3D NAND Flash," *IEEE Electron Dev. Lett.*, vol. 43, pp. 878–881, Jun. 2022, doi: 10.1109/LED.2022.3171176.
[17] R. Degraeve et al., "Statistical poly-Si grain boundary model with discrete charging defects and its 2D and 3D implementation for vertical 3D NAND channels," in *IEDM Tech. Dig.*, 2015, pp. 121–124, doi: 10.1109/IEDM.2015.7409636.
[18] D. Verreck et al., "Quantitative 3-D model to explain large single trap charge variability in vertical NAND memory," in *IEDM Tech. Dig.*, 2019, pp. 755–758, doi: 10.1109/IEDM19573.2019.8993552.
[19] S. M. Amoroso et al., "Understanding the impact of polysilicon percolative conduction on 3D NAND variability," in *Proc. SISPAD*, 2023, pp. 5.2.1–5.2.4, doi: 10.23919/SISPAD57422.2023.10319604.
[20] G. Malavena et al., "Discrete-trap effects on 3-D NAND variability—Part I: Threshold voltage," *IEEE J. Electron Dev. Soc.*, vol. 12, pp. 651–657, 2024.
[21] Y. Fukuzumi et al., "Optimal integration and characteristics of vertical array devices for ultra-high density, bit-cost scalable Flash memory," in *IEDM Tech. Dig.*, 2007, pp. 449–452, doi: 10.1109/IEDM.2007.4418970.
[22] M. D. Jacunski, M. S. Shur, and M. Hack, "Threshold voltage, field effect mobility, and gate-to-channel capacitance in polysilicon TFTs," *IEEE Trans. Electron Devices*, vol. 43, pp. 1433–1440, Sep. 1996, doi: 10.1109/16.535329.
[23] *Sentaurus Process User'S Manual V. V-2023.12*, Synopsys Inc., Mountain View, CA, USA, 2023.
[24] *Sentaurus Device user's Manual V. V-2023.12*, Synopsys Inc., Mountain View, CA, USA, 2023.
[25] *Garand VE User'S Manual Supplement for 3D-NAND V. V-2023.12*, Synopsys Inc., Mountain View, CA, USA, 2023.
[26] A. S. Spinelli, C. Monzio Compagnoni, and A. L. Lacaita, "Variability effects in nanowire and macaroni MOSFETs—Part II: Random telegraph noise," *IEEE Trans. Electron Devices*, vol. 67, no. 4, pp. 1492–1497, 2020, doi: 10.1109/TED/2020.2976630.
[27] 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, pp. 388–395, Jan. 2008, doi: 10.1109/TED.2007.910605.
[28] 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, pp. 1746–1752, Aug. 2009, doi: 10.1109/TED.2009.2024031.

Open Access funding provided by 'Politecnico di Milano' within the CRUI CARE Agreement