40nm SONOS Embedded Select in Trench Memory

Radouane Habhab, Vincenzo Della Marca, Pascal Masson, Nadia Miridi, Clement Pribat, Simon Jeannot, Thibault Kempf, Marc Mantelli, Philippe Lorenzini, Jean-Marc Voisin, et al.

40nm SONOS Embedded Select in Trench Memory

Radouane Habhab[1,3,4], Vincenzo Della Marca[3], Pascal Masson[4], Nadia Miridi[1], Clement Pribat[2], Simon Jeannot[2], Thibault Kempf[1], Marc Mantelli[1], Philippe Lorenzini[4], Jean-Marc Voisin[1], Arnaud Regnier[1], Stephan Niel[2], Francesco La Rosa[1]
STMicroelectronics, [1]Rousset, France, [2]Crolles,France
[3]Aix-Marseille University, CNRS, IM2NP UMR 7334
[4]University of Côte d'Azur, Polytech'Lab UPR UCA 7498

Email: radouane.habhab01@st.com

***Abstract*— In this paper, we discuss a new development of 40nm SONOS eSTM™ (embedded Select in Trench Memory). We present an experimental study based on hot carrier injection mechanism for both programming/erase operations, performed on this new eNVM architecture. The optimization of drain and select gate biases, in order to define the programming and erasing threshold voltages, is also detailed. All the characterizations have been carried out for two different SONOS eSTM™ architectures giving an opportunity to propose different solutions. One of this using a continuous silicon nitride layer for two neighbour cells, taking advantage on the discrete charge trapping nature. As well, we performed endurance tests up to one million cycles for both architectures to evaluate the memory endurance.**

Keywords*— *SiN-based memory, charge trapping mechanisms, SONOS, Hot carrier injection.

I. INTRODUCTION

As the widespread solution for data storage, embedded flash memory using standard floating gate technology deal with integration limitations. Then, the need for cost-effective solutions arises to respond to the growing demand for higher performance data storage across various applications.

Using a charge trapping layer instead of floating gate is a possible solution to address these challenges, the advantages of this solution are described in [1]. In fact, compared with floating gate devices, the fabrication flow complexity of charge trap memory technology is much lower, and the area can be reduced, as well as the production cost. SONOS nonvolatile memory devices are based on silicon nitride (Si_3N_4) storage layer. They have been developed during the last decades [2] and are still under development [3-5].

eSTM™ (embedded Select in Trench Memory) is an innovative floating gate Non-Volatile memory [6-8] specifically developed for various applications such as microcontrollers. The Transmission Electron Microscopy (TEM) picture of this device is reported in Fig. 1. In this paper, we present the very first results of 40nm SONOS eSTM™ using a charge trapping layer and vertical select transistor.

In part II, we present the process and the architecture of this new memory point. Two architectures have been fabricated, one maintaining the standard cell structure with separated memory stacks, and the second using a continuous Si_3N_4 charge trapping layer for two adjacent cells. In part III, the programming/erasing methods are described, while in part IV we propose an optimization of the cell operations. Finally in part V we analyse the programming windows evolution during the endurance experiments.

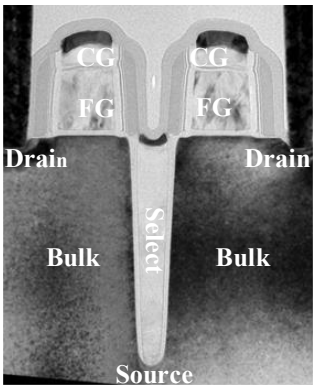

Fig. 1. eSTM™ cell TEM cross section.

II. SONOS eSTM™ ARCHITECTURE

SONOS eSTM is obtained from a p-type silicon substrate where a n-type implant is used to create a buried source area. A trench is etched, and the oxide is grown, before the polysilicon select gate deposition. After that, the trilayer ONO dielectric, where the silicon nitride layer separates two high temperature silicon oxides, is deposed. The memory stack is completed by the polysilicon deposition to make the Control Gate. Another n-implantation is performed for LDD and drain regions. Finally, contact and metal layers had been done as back end of line steps. In Fig.2, we compare the standard eSTM™ process flow versus the SONOS version.

Our study investigates two distinct devices. In Fig. 3a the TEM picture of dual gate SONOS eSTM™ is presented. This cell maintains the key features of the standard floating gate device using two separate memory stacks for adjacent devices. In contrast, the second studied architecture (Fig. 3b) called overlap cell [9] is developed by skipping the etching above the select transistor. We can note that the selection is performed by drain activation for both cells. Hence, the ONO layer overlaps the trench, enabling the fabrication process cost reduction. This is possible thanks to the Si_3N_4 properties that confines the charge trapping around the hot carrier injection region between the select transistor and the memory called sense transistor. The primary interest of this cell architecture is the enhanced scalability and select-sense alignment issues, which are essential for advancing future technological embedded applications. Another main difference is the vacancy of the floating implant between the sense and select transistor for the overlap device, that represents a key point for the programming efficiency [10]. As reported in [11], it is known that the efficiency of charge trapping is directly correlated with the portion of silicon in silicon nitride. In fact, silicon rich SiN induces a large amount of silicon dangling bonds (SiDBs) which can capture/emit electrons or holes presents at the transistor silicon interface. These charge

polarizations correspond respectively to the programmed or erased memory state.

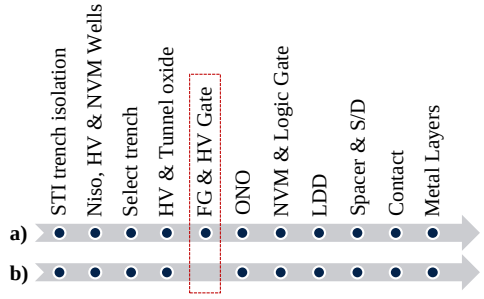

Fig. 2. Process flow for eSTM™ (a) and SONOS eSTM™ (b).

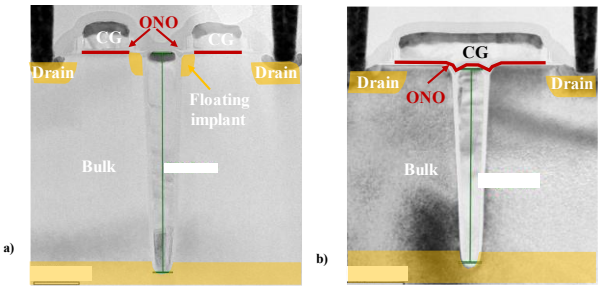

Fig. 3. SONOS eSTM™ TEM cross sections: (a) dual gate and (b) Overlap cells.

Also, X-ray Photoelectron Spectroscopy (XPS) measurements have been performed to obtain the profiles in atomic concentration of nitride and silicon elements presented in Fig. 4. As reported in Table I, we show that our layer has a N/Si ratio equal to 1.27, very close to 1.33 reported in [12] for a stoichiometric Si_3N_4.

III. Cell Electrical Activation

Several physical mechanisms can be used to program and erase SONOS memories. In this study, we used the source side injection (SSI) as programming mechanism [13] resulting in an electrons injection within the silicon nitride layer. Select and sense transistors are both biased in strong inversion regime and with a sufficient drain potential. The hot electrons are generated in the select transistor channel pinch-off region and accelerated toward the floating implant. The vertical electric field generated by the positive control gate bias, allows to inject a part of those electrons in the trapping layer. The large number of hot electrons allows a short programming time. The erasing is performed by hot holes injection (HHI) [14-17].

Even if the sense transistor is pushed into the accumulation regime, the drain current enables the impact ionization generation creating hot electron-hole pairs. In this case, the hot holes are attracted toward the interface of the sense transistor thanks to the negative control gate potential. A part of these holes are injected in the nitride layer and captured by silicon dangling bonds to erase the memory cell.

TABLE I. NITRIDE/SILICON RATIO VALUES FOR SI_3N_4:

[N]/[S] concentration ratio	XPS	[12]
	1.27	1.33

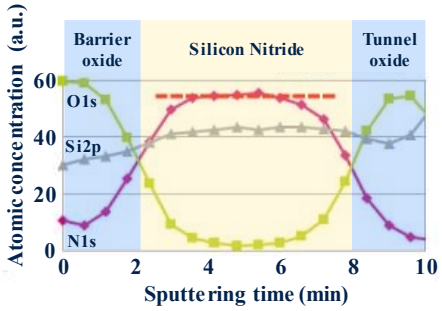

Fig. 4. Oxygen, silicon and nitride XPS profiles in the ONO stack.

IV. Operations Optimization

A. Programming

The efficiency of the programming state strongly depends on the potentials applied to the drain and select gate We need to determine the best trade-off to program our cell by adjusting these parameters (V_D and V_{Sel}). For a programming time fixed at 10 µs and the control gate voltage at V_{CG} = 10 V.

In Fig. 5a and 5b, we present the programming threshold voltage (V_T) dependence with the drain and select potentials for the dual gate and overlap cells. Before each programming operation, we erase the memory to reach an initial threshold voltage V_T for both architectures. We can note that the voltage applied to the drain shows a strong influence on V_T than the voltage on the select. This latter is used to control the current consumption during programming. We can also observe a slight difference of the overlap cell threshold voltages compared to the dual gate. This is mainly due to the vacancy of the floating implant, that induces a charge trapping closer to the select gate as well as an increasing of the effective cell channel length. Moreover, to optimize the programming operation the select gate bias can be decreased maintaining a high programming window (V_D = 4.5 V), while the current consumption is reduced as detailed in section IV.C.

B. Erasing

Using the programming bias setup found previously to fix the higher programming threshold voltage the dual gate and overlap cells, we present an optimization scheme to HHI erase. The erase time is fixed at 10 ms, and the V_{CG} = -10 V.

Fig. 6a and 6b present the dependence of the erase V_T as a function of the drain and the select gate voltages for the two studied architectures. We can notice that the erase efficiency is improved with higher drain voltage (lower V_T).

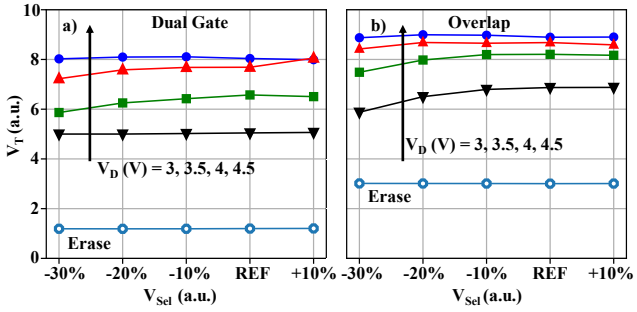

Fig. 5. SONOS eSTM™ program threshold voltage (V_T) evolution versus V_{Sel} bias at various V_D for dual gate (a) and overlap (b)

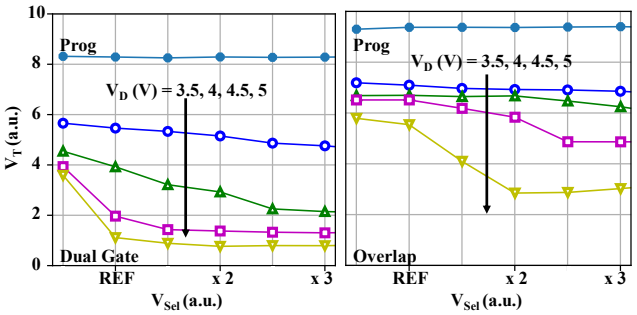

Fig. 6. SONOS eSTM™ erase threshold voltage (V_T) evolution versus V_{Sel} bias at various V_D for dual gate (a) and overlap (b)

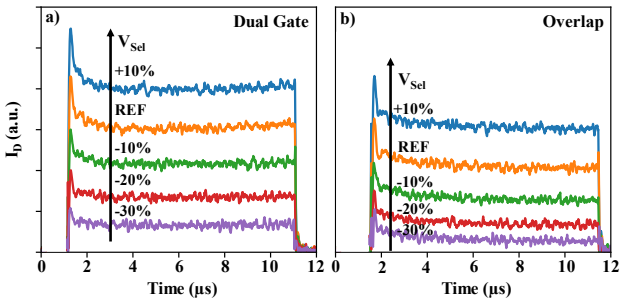

Fig. 7. Drain current evolution during programming for dual gate (a) and overlap (b).

Still, the overlap cell shows higher threshold voltage amounts. In addition, the erase threshold voltage reaches a plateau for a value of the select voltage which is all the weaker as the drain voltage is large. This point is of great importance to avoid an over-erased memory cell. Thanks to these results, we can thus determine the optimal bias to erase our cells. As during the programming phase, the final threshold voltage also depends on the memory cell architecture. In order to reach the larger programming window, we consider V_D = 4.5 V and V_D = 5 V for dual gate and overlap cell, respectively. We point the fact that for the overlap cell the select gate bias has to be increased by two times with respect to the REF condition, to complete the erase operation.

C. Programming current consumption and V_T dispersion

In this part, we propose to study the consumption performances on single cells. In order to measure this key performance parameter, it is necessary to measure the dynamic drain current consumption (I_D) at various bias conditions. The results are reported in Fig. 7.

As expected, the drain current consumption is influenced by V_{Sel} values. During the programming state, there is an initial peak followed by a constant drain current, which is a defining feature of charge trapping memory behaviour [18]. To achieve the best programming performance and maximize the programming window for our memory cells, we analyse the threshold voltages dispersion during the programming operation as reported in Fig.8. Programming with various V_D, we can note firstly that the overlap architecture is more dispersed than dual gate. This can be explained by the vacancy of floating implant that ensures the current continuity between the channel of select and sense transistors. Thus, the potential needs to be transferred from the drain contact to the channels intersection.

Moreover, we can note that for highest value, V_D=4.5V, we can reduce the bias applied on the select gate allowing a reduction of current consumption and reaching the best programming window for both architectures. Based on our results, we can determine the optimal programming conditions for the two structures. In summary, the optimal programming condition is achieved when V_{CG} is set to 10V, V_D is set to 4.5V, while V_{Sel} can be reduced by 30% compared to the reference.

V. ENDURANCE OF SONOS eSTM ™

Endurance is one of the key points to validate the reliability of silicon nitride memory cells. Fig. 9 shows the V_T window evolution during cycling at optimal biasing conditions found in part IV and summarized in Table II

It is demonstrated that the endurance, at room temperature, can reach 10^6 cycles with a programming window maintained at more than 70 % for both architectures. As represented in Fig.9a, it is interesting to note that the degradation of the erase V_T is present only for the overlap cell due to specific architecture such as the non-planar topology of the Si_3N_4 layer and the vacancy of the floating implant. This makes it more difficult to erase memory during the endurance test. Also, we can note the shift of both programming threshold voltages, indicating a weak parasitic electron trapping in the tunnel oxide. On the dual gate cell, we can observe a programming threshold voltage decrease during cycling, revealing a small loss in terms of programming efficiency due to the degradation of the tunnel oxide. As we can see in Fig.9b, endurance can be achieved at a lower erase time with smaller programming window. Finally, this figure shows the same behavior for the programmed state of the two architectures and also shows a lower variation of erased threshold voltage for the overlap architecture. This result is interesting to reduce erase time and current consumption.

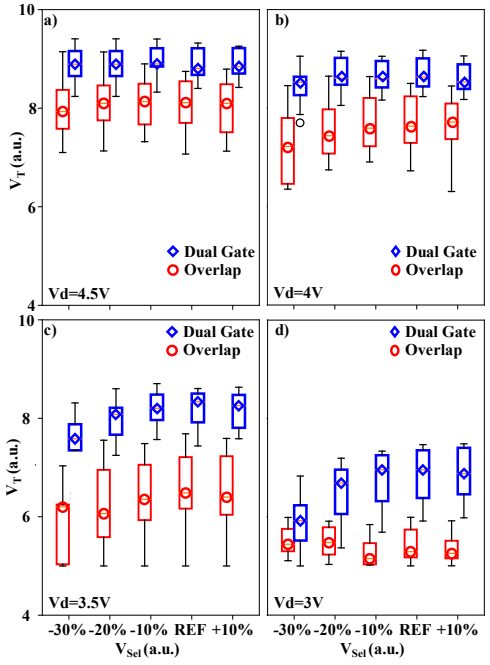

Fig. 8. Program V_T dispersion (10 samples) at various V_{Sel} and for V_D =4.5 V (a), 4 V (b), 3.5 V (c), 3 V (d)

TABLE II. OPTIMAL PROGRAMMING AND ERASING CONTIONS FOR BOTH ARCHITECTURES.

Devices	States	V_{CG}	V_D	V_{Sel}
Dual Gate	Prog	10 V	4.5 V	REF - 30%
	Erase	- 10 V		REF x 2
Overlap	Prog	10 V	4.5 V	REF - 30%
	Erase	- 10 V	5 V	REF x 2

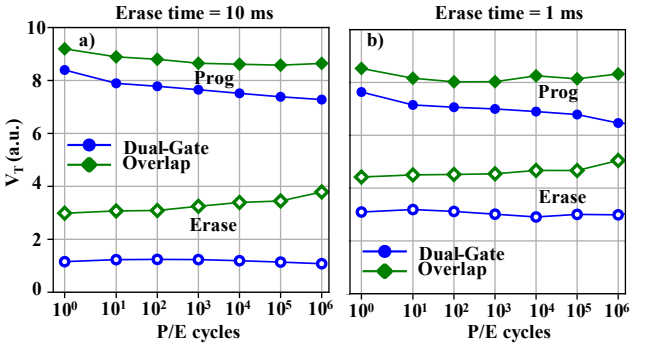

Fig. 9. SONOS eSTM™ program/erase threshold voltage

CONCLUSION

Results presented in this paper demonstrate the feasibility of a silicon nitride-based memory cells in eSTM™ technology Two architectures have been studied, the dual gate and the overlap, this latter is characterized by the continuity of the ONO layer between the 2 senses and the absence of the floating implant between the sense and select transistors. The measurements realized for programming optimization allow to maintain a high programming window and a decrease of the select gate bias. This condition enables to reach a low dispersion with an energy consumption reduction.

Finally, we show that it is possible to achieve an endurance of 1 Million cycles maintaining a sufficient programming window for both architectures while erase time reduction.

We can conclude that if the overlap architecture is better for high density solution, some details can be improved like the ONO layer planarity above the select transistor. On the other hand, the dual gate architecture is more adapted for low-cost embedded applications operating with low bias voltages and dispersion.

ACKNOWLEDGMENT

The authors gratefully acknowledge all manufacturing teams of STMicroelectronics Crolles for supporting this program.

REFERENCES

[1] K. Ramkumar et al., "A scalable, low voltage, low cost SONOS memory technology for embedded NVM applications," 2013 5th IEEE International Memory Workshop, 2013.

[2] P. Chen, “Threshold-alterable Si-gate MOS devices," IEEE Transactions on Electron Devices, vol. 24, no. 5, pp. 584-586, 1977.

[3] K. Ramkumar et al., "Reliability Aspects of SONOS Based Analog Memory for Neuromorphic Computing," 2020 IEEE International Reliability Physics Symposium (IRPS), 2020, pp. 1-5

[4] Jeong, J.-K.; Sung, J.-Y.; Ko, W.-S.; Nam, K.-R.; Lee, H.-D.; Lee, G.-W. Physical and Electrical Analysis of Poly-Si Channel Effect on SONOS Flash Memory. Micromachines 2021.

[5] V. Agrawal et al., "Subthreshold operation of SONOS analog memory to enable accurate low-power neural network inference," 2022 International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2022, pp. 21.7.1-21.7.4,.

[6] F. La Rosa, S. Niel, and A. Regnier. “Read performance of a non-volatilememory device, in particular a non- volatile memory device with buriedselect transistor”. US Patent No 9825186. (2017,Nov)

[7] F. L. Rosa et al., "40nm embedded Select in Trench Memory (eSTM) Technology Overview," 2019 IEEE 11th International Memory Workshop (IMW), 2019, pp. 1-4.

[8] S. Niel et al., "Embedded Select in Trench Memory (eSTM), best in class 40nm floating gate-based cell: a process integration challenge," 2018 IEEE International Electron Devices Meeting (IEDM), 2018, pp. 7.4.1-7.4.4

[9] F.La Rosa et al ., “Compact non-volatile memory device of the type with charge trapping in a dielectric interface”, US Patent No 10438960 (2019,Octobre)

[10] F.La Rosa, S.Niel, A.Regnier, J.Delalleau, “Hot-carrier injection programmable memory and method of programming such a memory.US Patent No 9224482 (2015,Decembre)

[11] B. Eitan, P. Pavan, I. Bloom, E. Aloni, A. Frommer V.A. Gritsenkoet al., “Excess silicon at the silicon nitride/thermal oxide interface in oxide–nitride–oxide structures”, J. Appl. Phys.,1999.

[12] J. Robertson, W.L. Warren, J. Kanicki, “Nature of the Si and N dangling bonds in silicon nitride”, Journal of Non-Crystalline Solids, Volume 187,1995, Pages 297-300

[13] F. Melul et al., "Hot Electron Source Side Injection Comprehension in 40nm eSTM™," 2021 IEEE International Memory Workshop (IMW), Dresden, Germany, 2021, pp. 1-4

[14] K. Yoshikawa et al., «Lucky-hole injection induced by band-to-band tunneling leakage in stacked gate transistors,» International Technical Digest on Electron Devices, pp. 577 - 580, 1990.

[15] N. Akil et al., «New Punch-through Assisted Hot Holes Programming Mechanism for Reliable SONOS FLASH Memories with Thick Tunnel Oxide,» IEEE Non-Volatile Semiconductor Memory Workshop, pp. 92 - 93, 2007

[16] D. Finzi, "NROM: A novel localized trapping, 2-bit nonvolatile memory cell," in IEEE Electron Device Letters, vol. 21, no. 11, pp. 543-545, Nov. 2000

[17] M. van Duuren et al., «New writing mechanism for reliable SONOS embedded memories with thick tunnel oxide,» IEEE International Conference on Integrated Circuit Design and Technology and Tutorial, pp. 181 - 184, 2008

[18] V. Della Marca et al., «Impact of endurance degradation on the programming efficiency and the energy consumption of NOR flash memories,» Microelectronics Reliability, pp. 2262 - 2265, 2014