TCAD Modeling and Optimization of 28nm HKMG ESF3 Flash Memory

Alban Zaka
GLOBALFOUNDRIES Dresden
Dresden, Germany
alban.zaka@globalfoundries.com

Tom Herrmann
GLOBALFOUNDRIES Dresden
Dresden, Germany
tom.herrmann@globalfoundries.com

Ralf Richter
GLOBALFOUNDRIES Dresden
Dresden, Germany
ralf.richter@globalfoundries.com

Stefan Duenkel
GLOBALFOUNDRIES Dresden
Dresden, Germany
stefan.duenkel@globalfoundries.com

Ruchil Jain
GLOBALFOUNDRIES Dresden
Dresden, Germany
ruchil.jain@globalfoundries.com

***Abstract*—The paper presents a TCAD modeling approach of the 28nm HKMG ESF3 Flash Cell. The methodology encompasses both DC and transient simulations with focus on hot carrier injection modeling. The ensuing Floating Gate Spacer optimization presents the trade-off between the various figures of merit and highlights the need for a comprehensive DC/transient simulation approach during Flash cell optimization.**

Keywords—ESF3, TCAD, Hot Carrier Modeling

I. Introduction

Split-gate cells have been object of numerous studies in the last 30 years [1-10]. In the recent years, the third generation of SuperFlash® Cell has been successfully demonstrated on planar 28nm Gate-First HKMG [1, 10]. State of the art TCAD has been employed to model and optimize the 28nm HKMG ESF3 device. Section II introduces the flash cell under study, whereas Sections III, IV present the DC and hot-carrier modeling of the cell, respectively. Finally Section V discusses the optimization metrics of the Floating Gate Spacer (FGSpc) within the Cell.

II. Cell Description

The split-gate ESF3 cell is integrated together with HV-MOS devices on the 28nm HKMG gate-first Super Low Power platform [1]. The ESF3 fabrication process includes the formation of the Floating Gate (FG), Control Gate (CG) and of the Erase Gate (EG) as highly-doped Polysilicon regions, followed by the Wordline (WL) formation which is composed of the medium-thick oxide and the HKMG module of the baseline technology (Fig.1). The Silicon doping below the FG, WL as well as around BL and SL can be individually tuned.

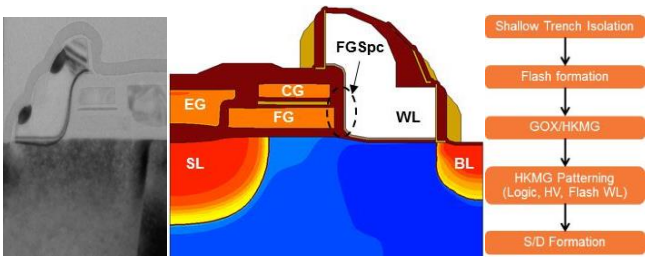

Fig.1: TEM and TCAD structure of the investigated 28nm HKMG ESF3 cell as well as the major fabrication steps, whereby the Wordline (WL) contains the HKMG stack, while Floating Gate (FG), Control Gate (CG) and Erase Gate (EG) are built with Polysilicon. The encircled are "FGSpc" is the object of the optimization section.

Table 1 summarizes the biasing schemes employed to program and erase the cell as well as to retrieve the threshold voltage of the cell (VtCG) and the BL current after program and erase operations (IR0 and IR1, respectively). While the erase operation is included in the present modeling, it is not object of a detailed analysis.

	Program	Erase	V_TCG (P/E)	V_TWL	IR (0:P, 1:E)
time	10us	10ms	-	after erase	-
SL	4.5V	0	0	0	0
BL	1uA	0	0.1V	0.1V	0.8V
EG	4.5V	11V	0	0	0
CG	10.5	0	ramp	1.8	1.8V
WL	0.8	0	1.8	ramp	1.8V

Table1: Nominal bias conditions used in this study. During programming, BL voltage is set such as to have 1uA flowing in the cell at the program start and then kept constant. The Programmed and Erased states are characterized by their threshold voltage (VtCG-Prog, VtCG-Erase) and the current which can flow in the channel after the operation (IR0, IR1).

III. Cell DC Modeling

Process and device simulations have been put in place using Sentaurus TCAD tools. The first step of the modeling concerns the correct description of the capacitances and doping in the cell. To achieve this, 2D Process simulations including the relevant fabrication steps and many TEMs taken after different process steps allow achieving the structural matching of the cell. The calibration of the dopant profiles in Silicon is facilitated by having several implant splits affecting the different Silicon areas of the device, i.e. below the FG, below the WL,etc.

Additionally, there are two aspects to consider. Firstly and more importantly, since the cell is a 3D device, an additional capacitance between FG and CG has to be added in the electrical simulations to match the CG-FG coupling ratio (separately measured, defined as $C_{FG\text{-}CG}/C_{tot}$), which is a key parameter for the cell operation. Secondly, one needs to adjust the charge in the FG electrode such as to have the V_TCG-Erase match the HW readout for a given initial structure. This value corresponds to a deep erase state (V_TCG-Erase ~ -8V) of the FG and it is important to note that this is a unique value, after fixing the additional FG-CG capacitance.

TCAD DC device simulations are performed using a drift-diffusion transport approach with doping, field and surface mobility degradation components. The FG terminal remains

floating throughout the modeling, i.e. its potential is defined from the capacitive coupling to the other cell terminals and from the amount of charges flowing in/out of the FG.

The overall goodness of the electrostatic modeling of the cell has been established by comparing the simulation results to Silicon in terms of VtWL and IR1 (Fig.2) as well as to cell IV curves (Fig.3) for a range of doping split experiments.

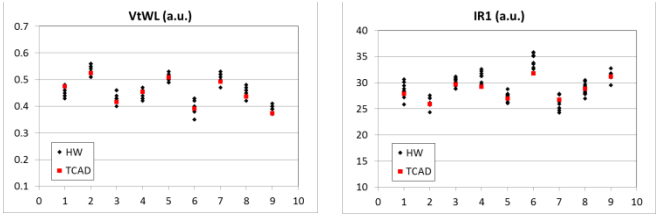

Fig.2: HW vs. TCAD comparison of VtWL and IR1 (c.f. definition in Table.1) for different Silicon doping splits affecting the cell electrostatics. Good agreement has been achieved around typical cell operation range.

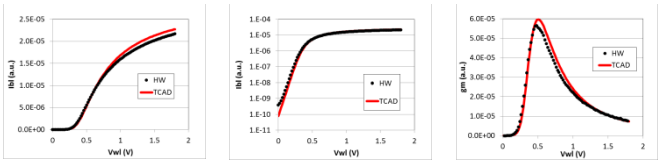

Fig.3: Typical example of HW vs TCAD comparison in terms of I_{BL} vs. V_{WL} and g_m vs. V_{WL} of the cell at V_{BL}=0.1V and V_{CG}=1.8V at erased state. Reasonable agreement is achieved below and above the threshold voltage.

The agreement achieved in these comparisons illustrate that the 2D TCAD methodology is able to capture reasonably well the 3D cell behavior and is ready to be used in transient simulations.

IV. Hot Carrier Injection Modeling in 28nm HKMG ESF3 cells

The analysis of the carrier transport and in particular of the hot carrier population and the ensuing injection current into the FG is performed using the Spherical Harmonics Expansion (SHE) method proposed in [11]. The accuracy of this method compared to Monte Carlo reference has been discussed in [12] and in the references therein. It is worth noting that the SHE method used in this study does not include Carrier-Carrier Scattering (CCS) phenomena.

At program start, the cell is in the erased state with the considerations of Section III. The initial programming current of I_{BL}=1uA is set by adjusting the V_{BL} (typically in a range of 0.1 – 0.5V), whereas the other terminals follow Table 1. Under these bias conditions, the portion of the channel under the WL is in saturation since V_{WL}~ $V_{BL}+V_{tWL}$. Since FG is in deep erase state and the adjacent terminals are high, the VFG is also pulled high. Hence, the channel under the FG is in strong inversion since it satisfies V_{FG} >> V_{SL} + V_{tFG}, where Vt_{FG} is the threshold voltage of the channel area under the FG.

As a consequence, at the conjunction of these two regions both a high lateral and vertical field are present (Fig.2).

Fig.5 show the simulated Electron Distribution Functions (EDF) obtained with SHE along the channel of the cell at the beginning (t=0) and at the end of the programming (t=10us). The former are the object of the following discussion.

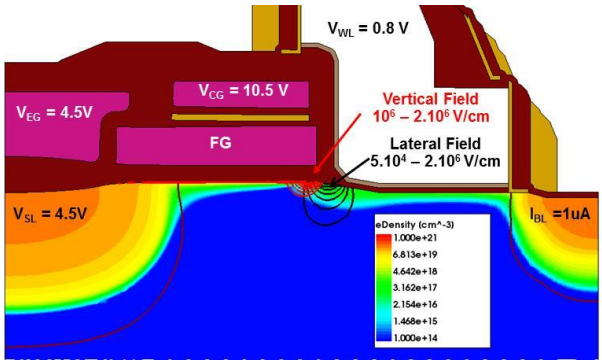

Fig.4: Electron density in the channel of the cell and Electric Fields at the Program start condition. The peak lateral field (along carrier transport) is located below the FG Spacer; the peak vertical field is located at the edge of the FG.

Analyzing the electron flow from BL to SL one can distinguish several regions of interest. As the "cold" electrons (Pos1) of the BL enter into and travel in the channel, they acquire the kinetic energy under the work of the Lateral Field. Note that nothing really happens till ~2/3 of the WL (Pos2), but then a rapid shift of the electron population towards the higher kinetic energies is observed (Pos3, 4, 5). This closely follows the pattern of the Lateral Field (Fig.4), albeit with ~10-20nm spatial delay, whereby the most heated electron population is located a few nm inside the FG (Pos5).

The "heated" distributions are described by a "plateau" and a Maxwellian slope (~1/kT), whereby i) the plateau is a result of the rearrangement in energy of the electrons after being subject to isotropic elastic and inelastic phenomena (such as acoustic and optical phonon scattering and impact ionization), ii) the steep slope is a reflection of the scattering phenomena which involve energy absorption (such as optical phonons – included, or CCS – not included here). The virtual intersection of these two parts is also called the knee point. It is interesting to notice that the knee point roughly corresponds to the amount of the available kinetic energy in the system under consideration ~q.(V_{SL}-V_{BL}).

This means that there is a non-negligible amount of electrons which were subject to only a few inelastic processes up to Pos5. As the electrons move further towards the SL (Pos6 and 7), the "hot" plateau starts cooling off, even though hot electrons are still present in the SL. At the same time the cold electrons population of the SL is clearly visible at the lower energies.

Fig.6 shows the distributions at Pos5 for different V_{SL}/V_{CG} biases at program start. The overall shape remains the same, but the knee point is shifted according to the available energy with a clear dependence to V_{SL}. It is interesting to notice that the V_{CG} also affects the maximum available carrier energy, by modulating the strong inversion under the FG, which modulates the distance needed to drop the lateral potential. However the vertical field shows a second-order effect compared to the lateral field in the hot carrier generation.

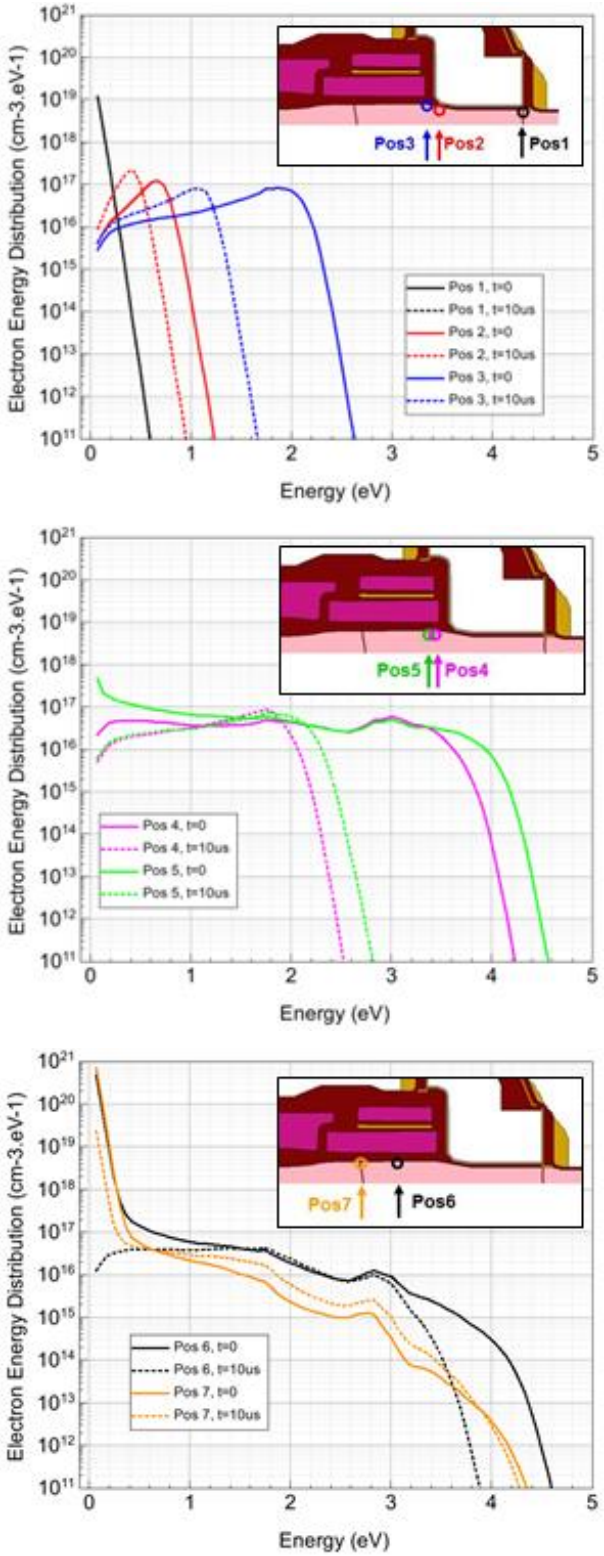

Fig.5: Electron Distribution Function (EDF) at the Silicon / Oxide interface plotted at the beginning (plain) and at the end (dashed) of the programming sequence (Table 1), respectively, ***a)*** in the channel region from BL to FGSpc, ***b)*** below the FG corner, ***c)*** below the FG and at the SL junction.

Assuming an isotropic carrier energy distribution, the current density tunneling into the FG is then calculated following [11]. Electrons having a kinetic energy higher than ~3eV will have a higher probability of being injected into the FG due to thermionic emission process. This illustrates the importance of a correct description of the energy distributions.

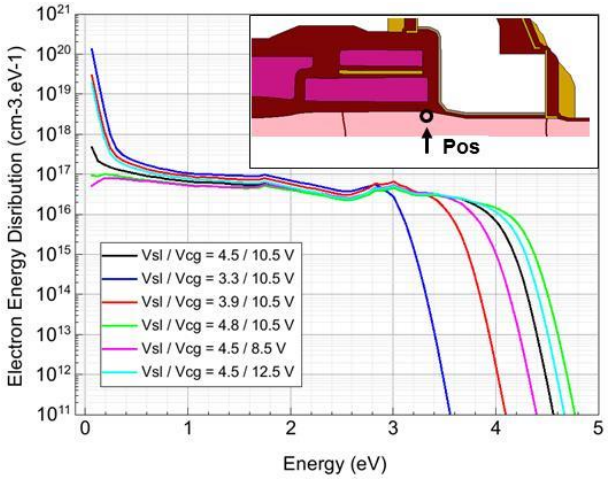

Fig.6: EDF plotted below the FG corner for varying Vsl/Vcg bias conditions at program start.

Fig.7 shows the calculated injected current density together with the electric fields along the channel position beginning (t=0) and at the end of the programming (t=10us). Initially the hot carrier injection into the FG occurs at the FG edge with a high injection length of ~30nm, and then falls exponentially towards the SL. As more electrons are injected into the FG, its potential drops which continually brings the Silicon channel region below the FG in saturation. This simultaneously causes a decrease i) of the Iprog (current flowing in the channel), ii) of the lateral field, iii) of the vertical field. All these effects limit the hot carrier injection. As the fields are modified, the EDF in Fig.5 (dashed) are also modified and at the end of the programming at 10us, the hottest distribution is found at around Pos.6. Eventually, the injection location is also shifted towards the SL, while still conserving the overall shape.

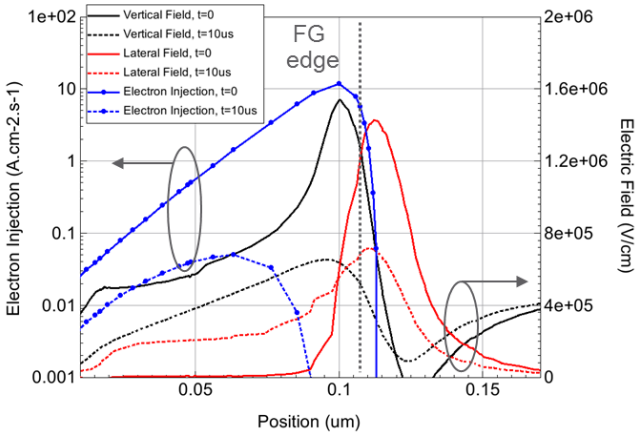

Fig.7: Lateral and vertical electric fields and electron injection into the FG along the Silicon / Oxide interface.

Finally Fig.8 shows the HW vs. TCAD comparison of the V_{tCG} after 10us of programming for a wide range of SL and CG biases. In particular, the SL bias sensitivity is well captured. This validates the injection and the overall cell simulation methodology.

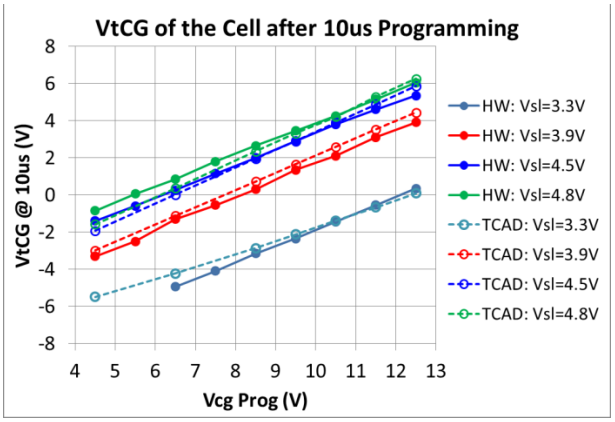

Fig.8: Vt_{CG} as a function of the V_{CGprog} at different V_{SLprog} for HW (solid) and TCAD (dashed).

V. Cell Optimization

The impact of the FGSpc thickness on the Cell performance is analyzed hereafter. BV aspect, as discussed in [9] is outside of our scope. Fig.9 shows that from a purely injection perspective, the injection efficiency (I_{FG}/I_{BL}) is rather constant or slightly reducing when increasing the FG Spacer thickness, at a given V_{FG}. This is in agreement with the MC findings of [7].

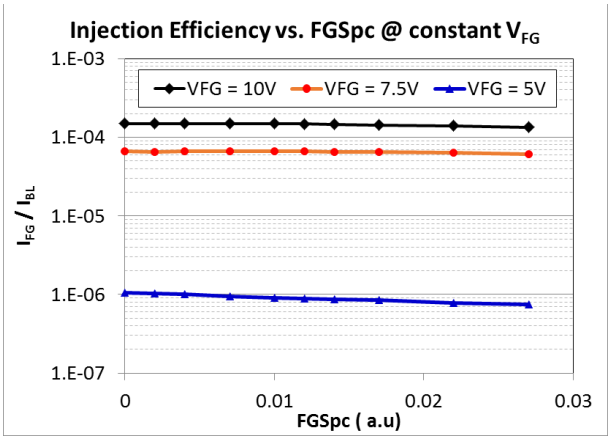

Fig.9: Injection efficiency for various FGSpc Thk. and FG biases.

However, applied to our investigated cell and bias conditions (Fig.1, Table1), a thicker FGSpacer results in an increased CG-FG coupling ratio (Fig.10), since the capacitance to the WL is reduced. This in turn caused the FG potential to be higher during the programming, which in turn increases the injection efficiency (Fig.9). Eventually a higher number of electrons are injected into the FG, thus resulting in a higher Vt_{CG}. On the other hand, a thicker FGSpc degrades the read current of the cell in the ereased state (Fig.11), due to an increased resistance between channel areas under WL and FG. Inversely, having a too thin FGSpc, despite reliability concerns, also increases the SL-BL leakage after programming. This leads to a trade-off value for the FGSpc.

The discussed application highlights the need to consider comprehensive simulation methodologies which capture the interactions between FEoL processing, DC device operation and transient hot carrier injection modeling.

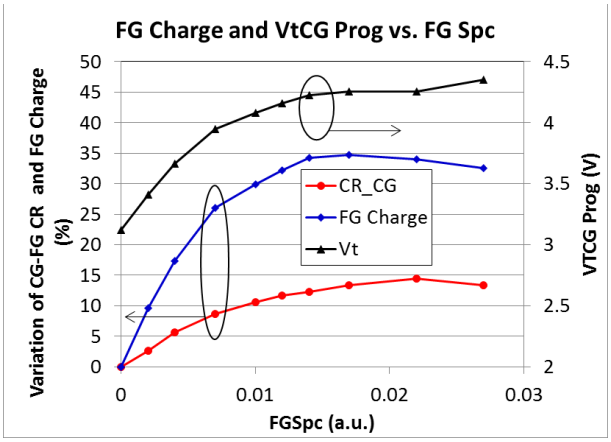

Fig.10: Variation of FG-CG coupling ratio,FG charge (left) as well Vt_{CG} (right) vs. FGSpc Thk.

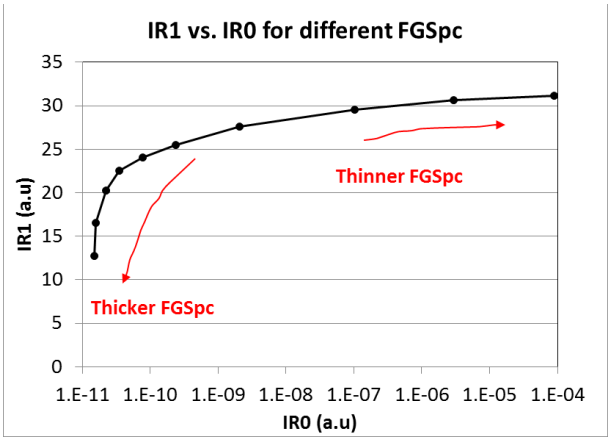

Fig.11: Cell read current after Erase (IR1) and Program (IR0) for different FGSpc thicknesses.

VI. Conclusion

28nm HKMG ESF3 Flash Cell operation has been modeled with TCAD in DC and transient operation with focus on hot carrier injection mechanism. The developed methodology has been employed to illustrate on a specific case the various metrics which are interwoven and which should be considered during cell architecture optimization.

Acknowledgment

The authors would like to thank SST team for providing the presented measurements as well as for the very insightful discussions.

References

[1] R. Richter et. al., IEDM, 2018, pp. 18.5.1-18.5.4.
[2] Y. K. Lee et al., VLSI, 2017, pp. T202-T203.
[3] L.Q. Luo et. al., IMW, 2017, pp. 1-4.
[4] N. Do et al., IMW, 2018, pp. 1-3.
[5] Y. Tkachev, et al., IMW, pp. 1-4, 2017.
[6] W. Stefanutti, et al., IEEE TED, vol. 53, no. 1, pp. 89-96, Jan. 2006.
[7] P. Palestri, et al., IEEE TED, vol. 53, no. 3, pp. 488-493, March 2006.
[8] J. Van Houdt et al., IEEE TED, vol. 39, no. 5, pp. 1150-1156, May 1992.
[9] M. Slotboom et al., ESSDERC, 2003, pp. 159-162.
[10] S.Jourba et al., IMW, 2020.
[11] S. Jin, et al., SISPAD, 2009, pp. 1-4.
[12] A.Zaka et al., 2014, in Grasser T. (ed.), ISBN 978-3-319-08994-2, pg.151-196