Current status and future prospects of Non-Volatile Memory modeling

(Invited Paper)

A. Benvenuti*, A. Ghetti*, A. Mauri*, H. Liu†, C. Mouli†

*Micron Technology Inc., Process R&D, Via Olivetti 2, 20864 Agrate Brianza (MB), Italy
†Micron Technology Inc., Process R&D, 8000 S. Federal Way, Boise, ID 83707-006, USA

Abstract

We briefly discuss the evolution of Non-Volatile Memory (NVM) technology in term of macro-trends and their implications for modeling activities in an industrial R&D environment. Some examples of difficult modeling issues for different NVM techologies are mentioned, and finally both present needs and future challanges are critically reviewed.

I. Technology trends and modeling implications

The modeling of NVM cells and arrays has to cope with several trends:

Geometrical scaling This is of course a common trend for all electronic devices, but since several years the pace of Moore's law and the lithography roadmap are driven by NAND Flash, with 15nm products currently available. Memory cells have very high aspect ratios and/or vertical current flow. Any point is within a distance of $\simeq$ 10nm from an edge or an interface, with a huge impact on dopant segregation, impurity cross-contamination, electrostatics, disturbs, and mechanical stress.

Increased statistical variability Smaller dimensions imply for each individual cell a smaller number of dopant atoms / traps / stored electrons in floating gates / polycrystalline grains / magnetic domains, with obvious implications on variability.

Alternative NVM concepts Due to fundamental electrostatic limitations, Floating Gate devices based on the charge control paradigm are becoming almost impossible to further scale down in a planar integration scenario. In parallel to the evolutionary 3D integration approach, alternative NVM concepts based on structural modifications have emerged, including change of crystallographic phase, magnetic domains, or modulation of conduction by migration of structural defects, ions or metallic species.

New materials A growing number of new materials has been introduced along the years, from SiNx to high-k IPD and hybrid metal gates, poly-crystalline Si channel, chalcogenide glasses, solid electrolyte compounds, ferro- and piezo-electric materials, and various magnetic compounds.

Modeling implications are:

3D For preliminary proof-of-concept 1D analysis can be used, but any industrially relevant NVM is fundamentally 3D.

Interface dominance For nano-scale NVM the role of interfaces is huge, not only in process (for dopant dose and loss point-defects injection/sinking), but also during device operation (e.g. grain boundaries barriers to transport, exchange of defects and RedOx processes, thermal boundary resistance effects). Accurate modeling of the role of interfaces (and in some cases also of their evolution) is mandatory.

Statistical simulation methods With monolithic integration capabilities approaching 1Tb, the need to model array distributions and at least intrinsic cell variability down to several σ is obvious. Besides conventional variability sources there is a growing need to model the impact of polycrystalline channel micro-structure and of local compositional fluctuations for multi-specie compound materials.

From process to material modeling Conventional process simulation is becoming less critical in supporting new NVM technologies. 3D NAND is intrinsically no-LDD, and standard etch and oxidation models are unable to effectively link technology parameters with cell morphology. Alternative NVM technologies are based on "exotic" active materials, for which fundamental material modeling based on e.g. DFT is potentially more useful to support both material exploration and device modeling.

From mono- to poly-crystals and amorphous materials In contrast to advanced logic, for NVM we need to model conduction in poly-crystalline or amorphous materials or even in organic molecules. In some cases we also have to model transport through materials while they undergo glass transition, melting and solidification, or a Mott transition.

Multi-physics Traditional TCAD boundaries between process and device simulation, where a structural modification (such as atomic migration) is assumed to occur only during fabrication, fail for alternative NVM technologies. Structural modification during operation is not just a perturbation needed to address long-term reliability, but is a fundamental mechanism used to program the cell. TCAD must be able to fully and self-consistently describe it as a consequence of piezo-magneto-electro-thermo-chemical effects during operation.

Microscopic phenomena and matter granularity In past years modeling successfully dealt with clear manifestations of granularity related to NVM reliability, such as Giant-RTN or 2-TAT limited data retention. In some cases (such as stochastic conductivity fluctuations of aggressively scaled HfOx RRAM current) a statistical and atomistic approach seems to be required to describe the intrinsic variability of each cell. Ultimately a modeling framework allowing to seamlessly bridge a continuum with an atomistic description of nano-sized NVM will be required.

All the above trends coexist and interact, making modeling of present and future NVM cells extremely challenging.

In the next section we will just touch upon a few examples of difficult modeling issues for different technologies.

978-1-4799-5288-5/14/$31.00 © 2014 IEEE

II. **Modeling status and open issues**

A. Process simulation

Historically, oxidation modeling focused on the description of the kinetics, adding Massoud corrections to describe the initial regime, and conceptually reverting 2D and 3D structures oxidation to a 1D-like description. Non-physical models for radical-based oxidation and the native oxide are in use, and concurrent oxidation of silicon and silicon nitride is not properly described. Especially when strongly non-planar geometries (such as in 3D NAND, but also simply with a recessed STI filling) are subject to a thin oxidation, resulting morphologies are far from realistic and artifacts appear near triple points. A physically-based, truly 3D diffusion/reaction stress-dependent oxidation model is expected to relieve these issues. The focus should not be on 1D kinetics (which can be easily experimentally found and empirically adjusted), but on a correct morphology evolution. Predictive evolution of effective doping, grain structure and trap density in polysilicon is insufficient or completely missing, while it would be very important in particular for 3D NAND and vertically integrated NVM. Accurate prediction of dopant dose loss due to outgassing and segregation in narrow regions is very critical; for instance the effective FG and CG doping (hence the Program/Erase efficiency) of wrap-type sub-20nm 2D NAND strongly depends on on the IPD and stack sealing process flow.

B. NOR Flash

Despite several alternative models have been proposed, an accurate modeling of NOR Flash cell programming by hot electrons is best achieved by means of Full-Band MonteCarlo simulations [1]. The Spherical Harmonics Expansion (SHE) method is an appealing alternative in order to reduce CPU time and to allow full transient simulations. We benchmarked an higher-order SHE against our MC code reference and experimental data on several structures, including a NOR Flash cell [2]. In general SHE does a good job in reproducing the normalized electron distribution function down to $\simeq$ 1e-6, but does not exhibit the bump in the high energy tail on the source side due to CHISEL (Fig. 1). While the normalized body current is reasonably well reproduced also by SHE, the normalized gate current dependence on gate voltage is too high (Fig. 1), and Ig/Id is strongly underestimated at low Vgs (possibly again due to the missing CHISEL contribution). Unfortunately a coupled solution of SHE for both electrons and holes, which we expect could solve or alleviate such discrepancy, is not currently available.

C. NAND Flash

One of the most relevant goals for NAND Flash modeling is to accurately predict Program/Erase (P/E) performance. In particular in conventional Poly FG cells programmed Vt saturation primarily occurs when (due to the increased IPD field) the outgoing electron tunneling flow from FG to CG balances the incoming electron tunneling from the channel to the FG, and vice-versa for Incremental Step Pulse Erase

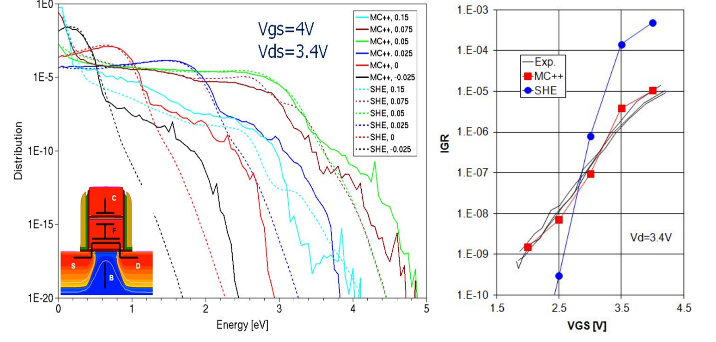

Fig. 1. (Left) Comparison of electron distribution functions obtained with SHE and with a full-band MC code [1] on a NOR cell [2]. Coordinates are referred to the metallurgical drain junction position. (Right) Comparison of normalized gate currents ($IGR = Ig/Id$) obtained with SHE and with MC.

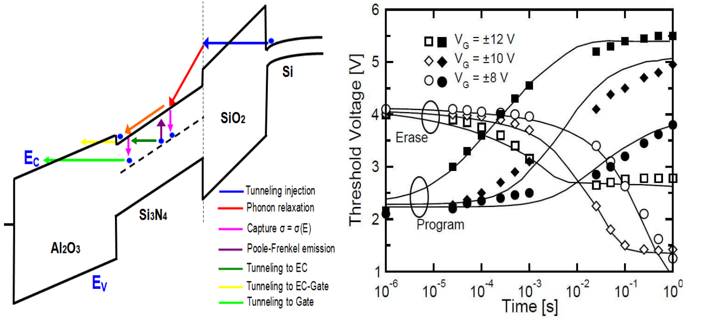

Fig. 2. (Left) Schematic of physical phenomena involved in SONOS/TANOS programming. (Right) Comparison between experimental SONOS capacitors program/erase data (symbols) and the corresponding transient simulation results (lines) using the model of [3].

(ISPE). Despite such physical process being conceptually simple, standard TCAD tools use simplistic assumptions in the calculation of the 3D tunneling paths, and do not accurately describe the huge role of fringing fields on scaled cells.

Furthermore in SONOS or TANOS cells, or more generally in Charge Trap architectures, the FG is replaced by a suitable dielectric trapping material, such as SiNx. In such a case the physical picture becomes more complex, due to the need to describe electron/hole capture/emission by traps due to several mechanisms, and transport both by drift/diffusion and tunneling [3] (Fig. 2). By implementing a 3D WKB calculation along the electric field streamtraces of incoming and outgoing tunneling contributions, coupled with the energy resolved trapping/detrapping module of Fig. 2, we succeeded to accurately model the ISPP/ISPE slopes and saturation for a variety of FG and Charge Trap architectures, both for planar capacitors (Fig. 2) and for 2D and vertically-integrated cells. Also the parasitic charge trapping taking place e.g. in a Band-engineered IPD AlOx layer or in the SiNx IPD layer of conventional NOR and NAND can be successfully described with the above approach; similarly for charge sharing along the continuous trapping layer of Mirror-bit NOR cells or in 3D NAND [4], [5].

With respect to planar integration, 3D NAND introduces

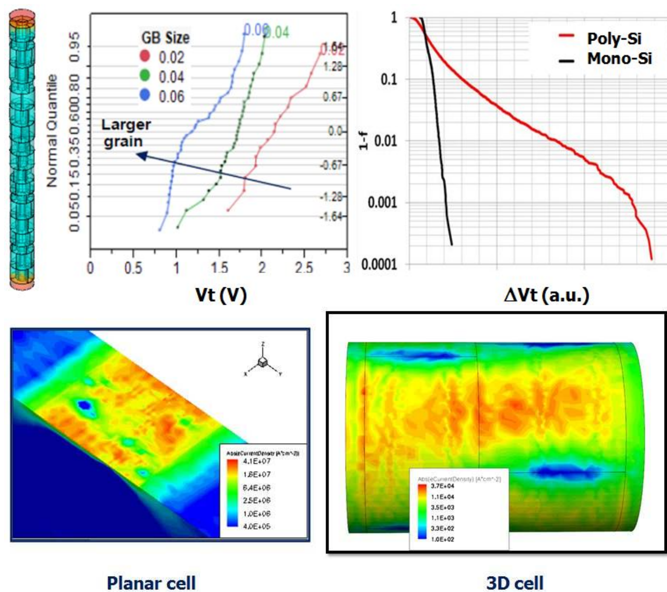

Fig. 3. (Top) Statistical distributions of string read current (Left) and Vt distribution (Right) for a poly-silicon channel with three different grain sizes. (Bottom) Percolative current distribution in 2D NAND mono-silicon channel (due to RDF) vs. 3D NAND poly-silicon channel (due to grain structure).

some peculiar effects. 2D cylindrical coordinates or full 3D simulation must be used to capture the increased electric field on the tunnel and blocking dielectrics due to the gate-all-around architecture, resulting in an higher effective Gate Coupling Ratio (GCR) for a smaller radius. Due to the lack of a contact to the string body, accumulation of charges during transients can lead to floating body effects akin to SOI CMOS. During read, undesired capacitive boosting on the Vpass trailing edge causes a large lateral field and spurious hot electron programming on adjacent cells [6]. Because of the 3D array architecture with multiple strings connected to a single physical block, program disturb can be more severe than in 2D NAND, and additional disturb modes appear. Optimizing the WL bias and the Vt of the dummy cells the electric field between the selector and the first non-dummy cell can be minimized, reducing the disturb [7]. Another peculiar feature due to the no-LDD cell architecture is a reduced electrostatic control of the gate on the portion of the channel between two cells, which can become the bottleneck for the string current in reading an erased cell. In such a way a premature saturation of the erase can appear [8]. The impact of the polycrystalline structure on string current has been extensively characterized and described either with statistical analytical models [9], based on simple resistive networks, or numerically, based on random grain distributions. In Fig. 3 an example of the grain size impact on the Vt distribution and subthreshold behaviour is shown; while the SS slope degrades with smaller grains, the tighter Vt distribution is obtained with a medium grain size. Furthermore we used the Impedence Field Method [10] to evaluate the expected impact of trapping on the grain boundaries on Random Telegraph Noise (RTN): in Fig. 3 we show the impact of the grain boundary trap concentration and the potential impact of an unoptimized process with a large trap concentration compared to mono-crystalline silicon. The root cause of Giant RTN in planar vs. 3D PolySi channel cells is qualitatively highlighted. In both cases the conduction is non-uniform: in the planar case due to Random Dopant Flutuations (RDF), while in the 3D Poly case due to the barrier associated to trapped electrons on grain boundaries. Giant RTN occurs when a trap is located on a critical position able to switch-off the main percolation path for a dopant/grain configuration with a highly non-uniform conduction.

D. PCM

PCM memories [11] feature several unique phenomena, including a thermally-activated exponential off-state I-V regime, a field-activated electronic switching leading to an NDR region, super-linear on-state I-V above hold, and of course the phase transitions that are the foundation of their memory effect, with crystallization due to stochastic nucleation and growth processes, melting above about 620 C, and solidification into a glassy amorphous phase upon fast quenching. Furthermore the amorphous phase exhibits a slow evolution of its electronic properties (drift), and the cell features a bias polarity dependence due to thermoelectric effects. Driven by the challenge to describe such a complex scenario, PCM memories represent a clear example of the need to describe the coupling between multiple physical phenomena in a TCAD environment. We developed along the years [12]–[15] a comprehensive modeling framework (Fig. 4) able to address most of the above phenomena in a consistent way, allowing us to improve device understanding and support product optimization. Besides being able to describe the cell I-V behavior in both states and its programming characteristic (most notably the so called “R-I” characteristic) , the model describe both the reset state data retention under isothermal conditions and the thermal disturb due to the programming (reset) operation on neighbor cells [15] (Fig. 4, Top right). Furthermore we recently addressed another phenomenon which has a fundamental impact on PCM performance and reliability, namely the atomic migration of the various species during program. The model is based on a diffusion-convection PDE for each of the three species in the ternary compound, including, beside concentration, electrostatic potential and temperature gradient driving forces, also a phase segregation contribution at the moving solid-molten phase interface. The accumulation of antimony in the active region near the heater interface (Fig. 4, Bottom right) is explained by a “snowplough” effect due to preferential segregation on the liquid phase [14].

E. RRAM

RRAM is a broad class of NVM using a plethora of material systems, with different underlying physical mechanisms and operation mode (e.g. unipolar vs. bipolar switching); the common characteristic is a resistance modulation not due to a phase change, but related to the formation of a conductive filament/bridge assisted by defects and/or metal ions migration.

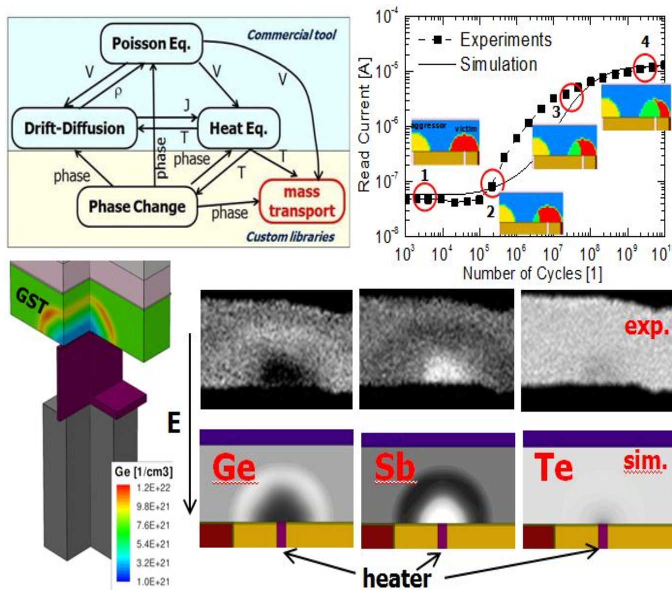

Fig. 4. (Top left) Main blocks of our PCM simulation environment. The last enhancement is the mass transport module [14], added to the previously calibrated electro-thermal-phase change model of [12], [13] and solved as a post-processing; all other equations are solved self-consistently. (Top right) Simulated accelerated thermal disturb stress vs. experiment [15]. (Bottom left) Simulated final Ge concentration in a 45nm PCM "Wall Heater" cell after a slow quench direct polarity SET pulse starting from above melt (Dielectrics and 1/4 of the simulation domain are not shown). (Bottom right): experimental (top row) and simulated (bottom row) concentration profiles extracted on the central 2D plane of the simulation. A darker shade corresponds to a lower concentration.

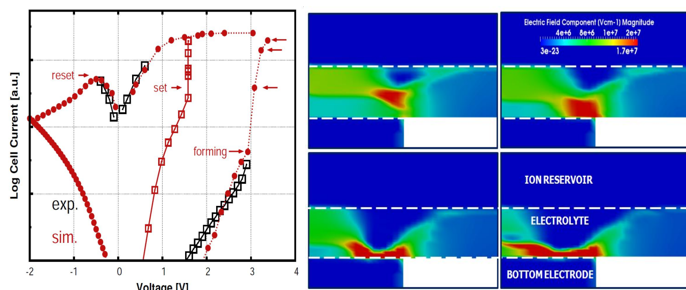

Fig. 5. (Top) Simulation of forming and set/reset "butterfly" IV characteristics for a copper RRAM cell [21]. Red symbols = TCAD results, black symbols = measurements. (Bottom) Snapshots of the electric field (taken at the currents indicated by the arrows on the simulated forming IV), showing a redistribution of the electric field inside the electrolyte. Only half cell simulated (left boundary ≡ symmetry plane).

RRAM behavior can be described as a charged multi-species system moving under a combined effect of electrical, chemical and thermal force. Although several groups proposed various models [16]–[20], no off-the-shelf implementation is available; therefore we developed, in-house, a 3D TCAD modeling environment allowing to treat "multi-physics" interactions. In Fig. 5 the simulated forming, reset and set IV is shown for a Copper RRAM cell [21], compared with experimental results. Various snapshots of the internal electric field on a 2D vertical cutplane during the forming transient are shown in Fig. 5; as Cu ions neutralize, a broad filament grows from the ion reservoir towards the bottom electrode, causing a progressive shrink of the high field region.

F. Magnetic NVM

The theoretical framework for magnetic devices is well established; from the computational modeling point of view a suitable environment for MTJ and STT-RAM physically-based TCAD modeling has been recently demonstrated [22], incorporating spin-selective tunneling and the Landau-Lifshitz-Gilbert equation to describe the interaction between the magnetic field, the magnetization vector and the spin torque. By incorporating an exchange term in the effective magnetic field some micro-magnetic effects can also be described.

III. Future needs

While long term requirements are still mainly at the conceptual level, and their priorities will depend on the evolution of the market, some present and near term needs are very clear and urgent, and are shortly summarized here. Relatively simple enhacments include a truly-3D diffusion/reaction oxidation model, coupled SHE for electrons and holes, and an accurate calculation of tunneling along 3D streamtraces, to be coupled with energy-resolved charge trapping models in dielectrics.

Present commercial tools are not suitable to describe complex dynamic coupling phenomena; urgent improvements are needed to describe statistical crystallization (nucleation+growth) and domain switching dynamics, coupling with ions/neutrals/defect migration, chemical reactions and clustering, and mutual interactions with mechanical stress (including phase change induced density variations).

Eventually a seamless bridge between continuum and atomistic description will be required.

Acknowledgments

The authors acknowledge the Micron TCAD team for technical contributions, C.Lombardi, P. Cappelletti and R.Bez for guidance and support, and Politecnico di Milano for the long term scientific collaboration.

References

[1] A. Ghetti *et al.*, *IEEE Trans. on Nanotechnologies*, vol. 6, p. 659, 2007.
[2] A. Zaka *et al.*, in *IWCE*, p. 1, 2010.
[3] A. Mauri *et al.*, *Solid-State Electronics*, vol. 56, p. 23, 2011.
[4] A. Maconi *et al.*, *Solid-State Electronics*, vol. 74, p. 64, 2012.
[5] W.-s. Cho *et al.*, in *VLSI Technology Symposium*, p. 173, 2010.
[6] B.-I. Choe *et al.*, *IEEE Electron Dev. Letters*, 2014.
[7] H. Yoo *et al.*, in *Int. Memory Workshop (IMW)*, p. 147, 2013.
[8] A. Maconi *et al.*, in *IEDM*, p. 29, 2012.
[9] M. Toledano-Luque *et al.*, in *IEDM*, p. 203, 2012.
[10] G. Torrente *et al.*, in *SISPAD Conf. Proc.*, p. 21, 2013.
[11] G. Servalli, in *IEDM*, p. 113, 2009.
[12] A. Redaelli *et al.*, in *SISPAD Conf. Proc.*, p. 280, 2005.
[13] A. Redaelli *et al.*, *Journal of Applied Physics*, vol. 103, p. 111101, 2008.
[14] G. Novielli *et al.*, in *IEDM*, p. 589, 2013.
[15] A. Redaelli *et al.*, in *IEDM*, p. 750, 2013.
[16] I. Valov *et al.*, *Nanotechnology*, vol. 22, p. 254003, 2011.
[17] D. Ielmini, *IEEE Trans. on Electron Devices*, vol. 58, p. 4309, 2011.
[18] R. Degraeve *et al.*, in *Integrated Reliability Workshop (IRW)*, p. 3, 2012.
[19] R. Waser, *Journal of nanoscience and nanotech.*, vol. 12, p. 7628, 2012.
[20] L. Vandelli *et al.*, in *Int. Memory Workshop (IMW)*, p. 1, 2011.
[21] S. Sills *et al.*, in *VLSI Technology Symposium*, 2014.
[22] F. O. Heinz *et al.*, in *SISPAD Conf. Proc.*, p. 127, 2013.