Future High Density Memories for Computing Applications: Device Behavior and Modeling Challenges

Gianpaolo Spadini, Ilya V. Karpov, David L. Kencke*
NAND Solution Group
*Process Technology Modeling
INTEL Corporation
Santa Clara, CA 95054 USA
gianpaolo.spadini@intel.com

***Abstract*—Memory elements existing and under research are compared for their suitability in computer memory applications. Cross point arrays of phase change elements with matched isolation devices are found to be particularly attractive and the challenge to model them is analyzed.**

I. INTRODUCTION

Computer memory architecture is becoming more complicated due to the increased performance of processors and the limitation in cost / power / speed offered by existing I/O subsystems and their respective memory devices. Hard drives provide cheap storage with high latency and DRAM provides expensive and power hungry storage at low latency. A significant research effort is underway to find technologies that provide low latency data storage at high density with low cost and low power consumption.

II. COMPUTER MEMORY ARCHITECTURE

The traditional memory architecture of a general purpose computer includes:

Processor Registers and SRAM on board cache; Main Memory – (off package DRAM) and Storage (Hard Disk).

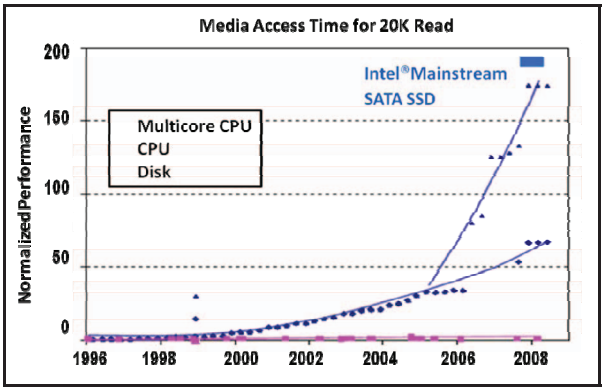

Figure 1. Historical performance improvement of CPU vs Memory and Storage

This architecture requires the latency of the Main Memory and the Storage to improve as the performance of the processor improves.

While the latency of the Main Memory and the Storage has improved slowly the performance of the CPU has improved much faster. The performance ratio of the processor to that of Storage has grown from 1.3 to 175 times in 13 years, as illustrated in Fig. 1 [1].

Recently the large gap in latency existing between the Main Memory and Storage (~ 100,000x) has been reduced by introducing SSD (Solid State Disks) made with well managed FLASH NAND devices.

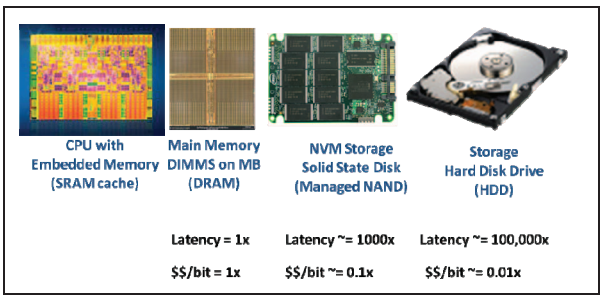

Figure 2. Memory+Storage architecture with SSD

The cost and performance of SSD makes the architecture shown in Fig. 2 very powerful and cost effective, maximizing today's processor utilization. The Storage may include a hard disk or the SSD may be the only storage unit.

In order for NAND to become a viable media in computing applications required several innovations: from the development of efficient wear leveling algorithms at the system level to allow the use of limited endurance media to scaling of the NAND array with multilevel storage (two bits per cell) to get the cost of SSD in the right range.

978-1-4244-7700-5/10/$26.00 ©2010 IEEE

The performance of FLASH NAND is about 1000 times worse than DRAM (with latency in the order of tens of microseconds vs. tens of nanoseconds) and no NAND performance improvements are expected. On the contrary, scaling is expected to increase NAND read latency.

Another possible improvement in the memory hierarchy is enabled by the large performance gap that exist between the NAND based SSD and the DRAM in the Main Memory (~1000x). The challenge here is that the cost gap is not very large (only 0.1x). Since the size of each memory cache must increase roughly 10x as we move out from the processor, a successful NVM system that fill this space must be made with devices with the cost of NAND and the performance in between NAND and DRAM. This justifies the search for a fast and low cost (small and scalable) NVM device that is happening in the last few years.

III. CROSS POINT MEMORIES

The research for a new NVM device has focused around the development of new storage elements that make use of thin film devices of materials not typically used in the semiconductor fabrication.

A few characteristics are common to the most interesting novel memory elements:

- Devices have two terminals, unlike the FLASH cell that is a three-terminal device, and behaves like a resistor with two or more programmable values.
- Devices are made in the back end (above CMOS metal interconnect) and require low temperatures to be manufactured.

The above makes it possible to consider their use in a cross point array: an array of two orthogonally overlapping metal lines with memory devices at each overlapping area. This also requires a non-linear device (Selector) to break the sneak paths currents into the cells in low resistance state in parallel to the one selected for either read or write (Fig.3).

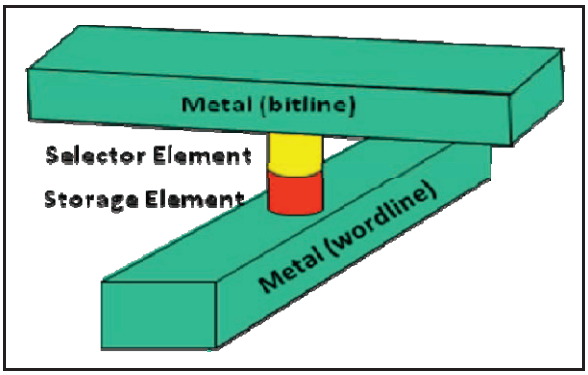

Figure 3. Cross point cell concept

Memory element and Selector should be considered together in the design of the cell for a cross point array.

The Selector must accommodate the requirements of the memory elements (bipolar or unipolar, on current) and must have sufficiently low leakage in the inhibit mode (considering it in series with the memory cell in both states) to allow for the cross point array to function.

TABLE I. RESISTIVE MEMORY ELEMENTS

Family	Example	Mechanism	Selector
Phase Change	Chalcogenide alloys e.g.: GST 225	Energy (heat) converts material between crystalline (low resistance) and amorphous (high resistance) phases	Uni-polar
Magnetic Tunnel Junction	Spin Torque Transfer RAM	Switching of magnetic resistive layer by spin-polarized electrons	Bipolar
Electro-chemical	e.g.: CuSiO2	Formation / dissolution of "nano-bridge" by electrochemistry	Bipolar
Binary Oxide (filaments)	e.g.: NiO	Reversible filament formation by Oxidation-Reduction	Uni-polar
Interfacial Resistance	Memristors e.g. Doped STO, PCMO	Oxygen vacancy drift diffusion induced barrier modulation	Bipolar

The key requirements for the use of cross point array based devices in computing applications are:

- Cost (cell size and fabrication cost)
- Future cost reduction path (scalability)
- Low programming energy (allow high throughput)
- Fast programming (low write latency)
- High read current (low latency) and low program current (tradeoff: read must not disturb)
- Adequate retention and endurance

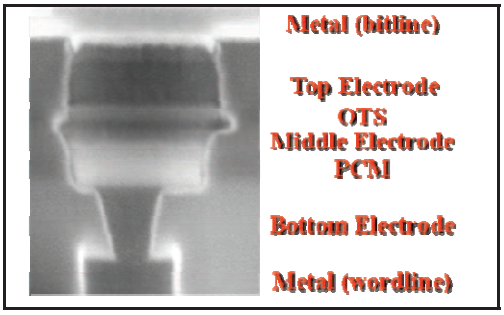

Figure 4. PCMS Cell

No device scores the best in all of the categories but one that shows a great deal of promise is the 3D PCMS (Fig. 4) cross point cell [2].

IV. PCMS MODELING

The PCMS cell consists of a PCM (Phase Change Memory) element stacked with an OTS (Ovonic Threshold

Switch) device. The OTS is a two terminal device made of amorphous chalcogenides whose composition is chosen to remain amorphous (Fig. 4) [3].

Simulations and modeling of the PCMS cell has proved useful in assessing its ability to scale. In particular the programming transitions are important: Set-to-Reset (low to high resistance) and Reset-to-Set.

The simulation of the Set-to-Reset transient of a PCM cell has been done with standard TCAD multi-dimensional simulation tools solving drift-diffusion transport coupled with heat generation/transport [4]. Crystalline and amorphous chalcogenide bandgap and traps are modeled [5]. Additional models produce conductivity in resistive metals and relate thermal to electrical conductivity by the Wiedemann-Franz relation. For this simulation the important features are obtained at high electrical current, when the PCM material is close to the melting point. The PCM material during this simulation starts in the crystalline state and its properties can be approximated by the use of a degenerate semiconductor. The key in this case is the choice of the correct parameters for the materials used in the cell (especially the thermal conductivity that is not well known close to the melting point).

With the above approximations and the proper choice of parameters to fit the available experimental data we find that scaling of the cell will reduce the reset current by a power law with an exponent *n* close to 1.7 for PCMS cell type B (Fig. 5).

$$\text{Ireset} \propto (\text{CellSize})^{n}$$

A key parameter in the Set-to-Reset simulation is the electrical and thermal contact resistance between the electrode and the PCM material. Including the correct value for the thermal end electrical contact resistance is necessary to properly fit experimental Current-vs.-Voltage device behavior (Fig. 6) and reset current (Fig. 5) [4].

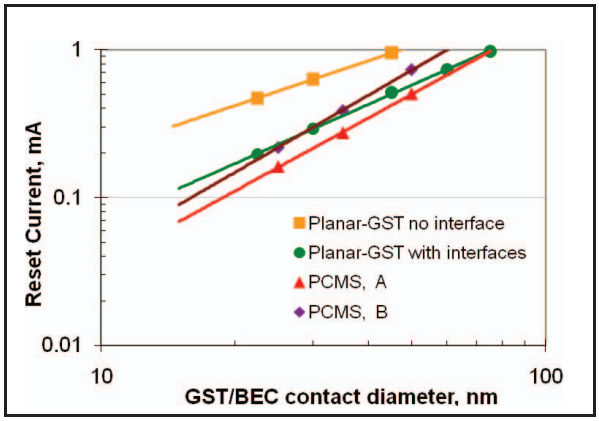

Figure 5. PCM Cell Reset Current Scaling

Adding the OTS to the cell does not change drastically the cell behavior since at high current the OTS will be in the on state and can be treated also as a degenerate semiconductor or a metal.

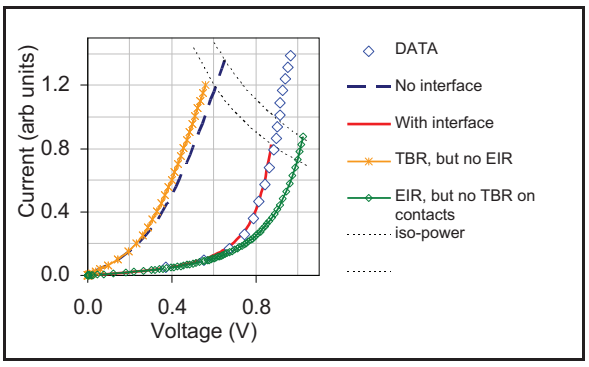

Figure 6. PCM Cell Current-vs-Voltage with different electrodes

The Set-to-Reset simulation can also help to predict the adjacent cell disturb due to thermal coupling. This simulation for an isotropic scaling shows that adjacent cell thermal disturb is not a problem [5, 6] and even if the thickness of the PCM material is not scaled this is manageable to about 10nm (Fig. 7).

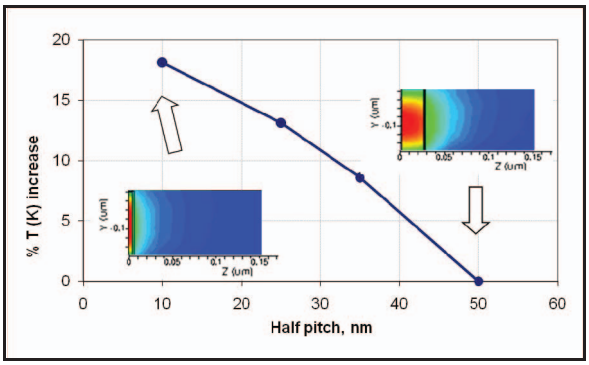

Figure 7. Adjacent Cell thermal disturb (half cell shown)

Transient Set-to-Reset simulations show how programming speed change as the technology scales. For isotropic scaling as PCMS cell becomes smaller shorter pulses can be used to program the cell (Fig. 8). The effect is mostly due to the reduction in the thermal transient which can be modulated by the choice of materials used to fabricate the cell.

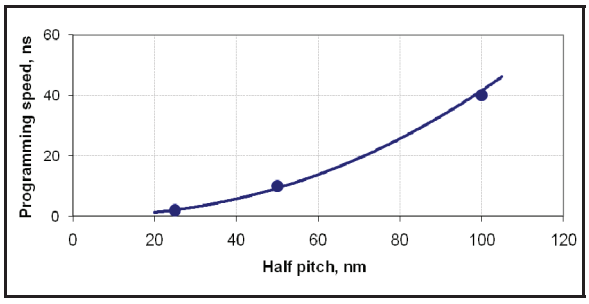

Figure 8. SET to RESET programming speed scaling

Numerical simulation of Reset-to-Set programming is a challenge due to the complexity and lack of understanding of

the effect known as threshold switching; this effect is also responsible for the Selector electrical behavior [3].

Threshold switching makes the Current-vs.-Voltage S-shaped and gives us an indication of a likely filamentation of the current path [7].

Unified switching model [8] provides a framework to model numerically the formation and dynamics of the filament. Moreover, such model of filament formation can describe the switching process in both PCM and OTS elements of PCMS cell but much work is left to implement this capability in general purpose modeling packages.

One last aspect of a PCMS device behavior that is challenging to model is the temporal change in device parameters that follows a programming event. This "drift" exhibited by amorphous materials must be observed or its effect must be extrapolated for very long time spans to explain the memory array operation. For example, the low field resistance and the threshold voltage of a PCM device after a Set-to-Reset operation will increase vs. time in power law or logarithmically, respectively, as shown in Fig. 9.

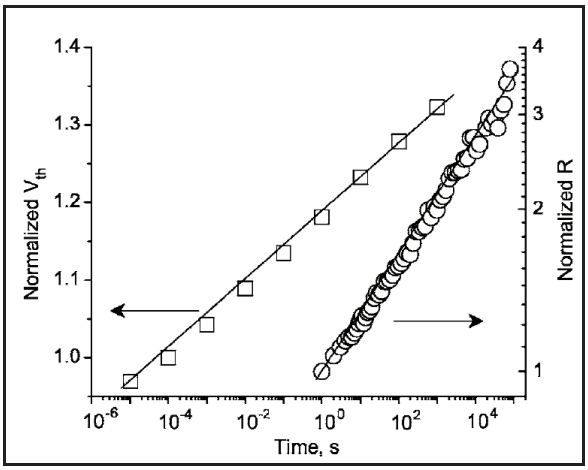

Figure 9. Low field resistance and threshold voltage change with time after programming

V. CONCLUSIONS

While there has been a large progress in capability to analyze phase change devices, some aspects of their operation are not yet known and cannot be included in device simulators.

Table II groups numerous physical processes involved in each of the main memory device operation. The nature of the interfaces, high-temperature material properties, and physical effects such as thermo-electric powers, latent heat, and thermal expansion all continue to be areas of interest. Simulation of the OTS device operation, especially including its transient behavior, requires further development and study. The many physical interactions within a PCMS cell provide a rich and important area of opportunities for modeling research and development.

TABLE II. PHYSICAL PROCESSES FOR SIMULATION OF KEY DEVICE OPERATION

PCMS Device Operation/ Physical process	**SET to RESET**	**RESET to SET**	**READ**
Threshold Switching	X	X	X
Nucleation and growth		X	X
Melt and quench		X	
Heat transport	X	X	X
Electrical and Thermal Contact resistance	X	X	
Mechanical Stress	X	X	X
Subthreshold Conduction		X	X

REFERENCES

[1] A. Fazio, "Future Directions of Non-Volatile Memory in Compute Applications," IEEE IEDM Proceedings, pp. 641-644, 2009.

[2] D. Kau, S. Tang, I. Karpov, R. Dodge, B. Klehn, J. Kalb, J. Strand, A. Diaz, N. Leung, J. Wu, S. Lee, T. Langtry, K. Chang, C. Papagianni, J. Lee, J. Hirst, S. Erra, E. Flores, N. Righos, H. Castro and G. Spadini, "A stackable cross point phase change memory," IEEE IEDM Proceedings, pp. 617-620, 2009.

[3] S. Ovshinsky, "Reversible Electrical Switching Phenomena in disordered structures," Physical Review Letters, Vol. 21, pp.1450-1453, 1968.

[4] D. Kencke, I. Karpov, B. Johnson, S. Lee, D. Kau, S. Hudgens, J. Reifenberg, S. Savransky, J. Zhang, M. Giles, G. Spadini, "The Role of Interfaces in Damascene Phase-Change Memory," IEEE IEDM Proceedings, pp. 323-326, 2007.

[5] A. Pirovano, A. L. Lacaita, A. Benvenuti, F. Pellizzer, S. Hudgens, and R. Bez "Scaling Analysis of Phase-Change Memory Technology," IEEE IEDM Proceedings, pp. 699–702, 2003.

[6] S. Kim, P. Wong, "Analysis of Temperature in Phase Change Memory Scaling," Electron Device Letter, V28, pp. 697-699, 2007.

[7] B. Ridley, "Specific Negative Resistance in Solids," Proc. Phys. Soc., V82, pp.954-966, 1963.

[8] M. Nardone, V. Karpov, D. Jackson, and I. Karpov, "A unified model of nucleation switching," Appl. Phys. Lett. 94, p.103509, 2009.

[9] I. Karpov, M. Mitra, D. Kau, G. Spadini, Y. Kryukov and V. Karpov, "Fundamental drift of parameters in chalcogenide phase change memory," J. of Appl. Phys. 102, p. 124503, 2007.