RRAM Crossbar Arrays for Storage Class Memory Applications : Throughput and Density Considerations

A. Levisse[1], B. Giraud[2], J.-P. Noel[2], M. Moreau[3], J.-M. Portal[3]
[1]Embedded Systems Laboratory (ESL), Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland
[2]Univ. Grenoble Alpes, F-38000 Grenoble, France; CEA, LETI, MINATEC Campus, F-38054 Grenoble, France
[3]Aix-Marseille Université, IM2NP, CNRS UMR 7334, F-13453 Marseille, France
Email : alexandre.levisse@epfl.ch

***Abstract*—As more and more high density memories are required to satisfy the Internet of Things ecosystem, academics and industrials are looking for an intermediate solution to fill the gap between DRAM and Flash NAND in the memory hierarchy. The emergence of Resistive Switching Technologies (RRAM) proposes a potential solution to this demand for fast, low cost, high density and non-volatile memory. However, nowadays transistor-less RRAM-based architectures, such as Crosspoint, suffers of several issues such as sneakpath, IRdrop and periphery overhead. In this work, we propose to explore the positioning of RRAM crosspoint memories regarding DRAM and NAND in terms of density and write throughput. We present several design guidelines then show that for the optimal RRAM crosspoint architecture (2-layers with common bitline), massively multiple bank write is the solution to optimize density and write throughput to around 20-100Gbit/cm2 and 200-500MB/s respectively for 32 to 64 parallel access.**

Keywords*—Crossbar, Crosspoint, RRAM, IRdrop, Periphery Overhead, Write Throughput, Storage Class Memory.

I. Introduction

With the internet of things (IoT) emergence, more and more storage capabilities are required in all the levels and particularly in the servers which require huge quantities of high density, fast, low cost, and energy efficient memories. However, among nowadays available technologies, none is able to fulfill the aforementioned requirements. Flash NAND technologies are not scaling anymore in dimensions and while 3D stacked VNAND [1] can offer more and more density, their write throughput does not strongly increase due to an extremely slow write process relying on *Fowler-Nordheim* current. On the other hand, DRAM technologies density does not scale due to charge sharing issues during read operations, limiting the subarray size and the bitcell density. From that perspective, even if more and more capacity and throughput is enabled by 3D TSV stacked chips [2], it does not increase the memory density while increasing the static consumption as DRAM is a *Volatile Memory*. In this context there is no memory technology featuring both high density and high write throughput. This concept of intermediate density memory was materialized under the name of Storage Class Memories (SCM) [3]. It corresponds to a non-volatile memory of intermediate density and throughput which would fill the gap between DRAM and NAND Flash technologies in the memory hierarchy. However, there is still no clear answer to the SCM positioning or technology question and its use on the application level is still blurry.

Resistive Switching Memories Technologies (RRAM) [4] [5] [6] enabling low cost Back-End-of-Line (BEoL) integration seems to be a possible solution. Beyond the standard 1Transistor-1RRAM architecture (1T1R) featuring limited density, transistor-less architectures, such as crosspoint or Vertical RRAM (VRRAM), enabling $4F^2$ or more integration density are emerging. Among these, crosspoint is by far the most mature, and is currently under investigation by industrial for products [3] [7] [8].

However, these new technologies add new constraints such as *sneakpath* currents, *voltage drop* (IRdrop) and *periphery area overhead* constraints. These issues are currently investigated in the literature and some compensation techniques are currently explored to enable reliable write operations [9] [10] [11] [12]. Although IRdrop effect start to be widely explored lately [13] [14] [15], only few works considering the periphery design considerations and its effect on the memory density are reported [11] [16]. In this contribution, we propose to extend these works, by first, detailing the *memory decoder* (DEC) design, the effect of multiple bit-write per array on the *periphery/array area overhead* and some layout constraints related to peripheral circuits pitch-matching with the aggressive memory array pitch. Then, we explore various crosspoint memory organizations (*1-layer*, *CMOS under array*, *2-layer with common BitLine*), several writing strategies (single bit per array, multiple bit per array, parallel multiple bank) and determined that the natural positioning of crosspoint is in-between DRAM and Flash NAND technologies in terms of throughput and density. We also show that such performances can only be leveraged by massively parallel multiple-bank write (densities and write throughput from 20 to 100Gbit/cm2 and from 200 to 500Mbyte/sec respectively, can be achieved in 32 to 64 parallel multibank access configuration). These results support the current wave on RRAM crosspoint use for SCM memories.

The remainder of this paper is organized as follow. Section II presents the necessary background on this paper as well and some related previous works. Section III presents several design strategies which must be considered while designing a crosspoint memory peripheral circuitry. Section IV proposes to explore the positioning of crosspoint memories regarding DRAM and Flash NAND technologies. Finally, section V concludes the paper.

II. BACKGROUND

A. RRAM technologies

RRAM technologies are today seen as the future of *Non-Volatile Memories* (NVM), from the connected objects, where it is expected to replace embedded flash memories, up to the high end server applications, where it has long been accepted as the straightforward replacement candidate for flash NAND technologies. The main interest of RRAM technologies compared to flash technologies lies in their low cost Back-End-of-Line fabrication process featuring fabrication facilities friendly materials and easy manufacturing process (few additional masks) [17].

There are three main RRAM technologies, which are the Filamentary Resistive Switching Technologies (ReRAM) [4], the Phase Change Memories (PCM) [5] and the Magnetic Memories (MRAM) [6]. Each technology features global pros and con. In general, these technologies have a wide range of programming conditions that makes it highly versatile [18]. Overall, whatever is the RRAM technology, two major criterions must be satisfied while programming it: (i) the programming current I_{prog} has to be injected in the device and (ii) the programming voltage V_{prog} has to be applied across the device. If one of those conditions is not satisfied, a reliable programming operation cannot be ensured.

B. High density RRAM architectures

Beyond the RRAM technology, choosing the array organization brings some other issues. Because of the two constraints aforementioned (I_{prog} and V_{prog}), transistor reliability might be reduced if the required V_{prog} exceed its range of operation and its area cannot be reduced because of the required I_{prog}. This lead to low density 1 Transistor-1 RRAM (1T1R) bitcells with a minimum bitcell area of around $12F^2$ [19] (where F is the minimum size metal half pitch) for a 28nm CMOS technology node as shown *Figure 1*-a. On the other hand, suppressing the selection transistor and replacing it by a BEoL selector [20] enables stackable $4F^2$ density bitcells as shown *Figure 1*-b. This architecture, named crosspoint (or crossbar) [7], suffers of new constraints such as the sneakpath currents and programming current control. As the selection transistor has been suppressed, the leakage current through unselected bitcells (sneakpath) is less controlled and leads to lower read margins [13] and distorted programming current I_{prog} [10].

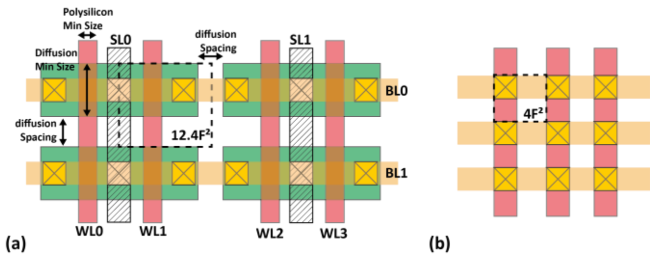

Figure 1 : (a) Layout view of a 1T1R memory array featuring minimum size transistors with a $12.4F^2$ density. (b) crosspoint memory array enabling a fully BEoL $4F^2$ density.

C. IRDrop Effects

The major issue of deeply scaled metallization levels is the effect of serial resistance increase while the current that must go through it doesn't scale down. The simple application of ohm law ends up in an increased voltage drop across the metal lines (namely IRdrop). This effect was explored in several works [11] [14] [15] [16] [21] and is known to be the source of limited read margins and write failure due to a reduced voltage across the accessed bitcell.

For one bit written in a crosspoint array, the equivalent circuit can be approximated as shown in Figure 2 for a single in-array bit-write. In this graph, the worst case bitcell is accessed (farthest away from the WL and BL drivers). Each non-selected bitcell is leaking a sneakpath current contribution (I_{sp}) while the programmed bitcell is consuming a programming current contribution (I_{prog}). The inset presents the evolution of the voltage along the selected WL and BL. The voltage first drops through the SWL parasitic resistance, then through the selected bitcell and finally through the SBL parasitic resistance.

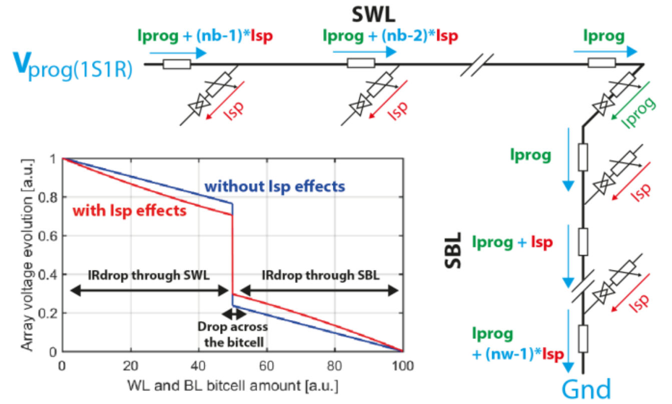

Figure 2: SWL and SBL of a crosspoint memory with detailed currents and voltages. Insert : Voltage in SWL, SBL and across the selected bitcell versus the in-array position.

As presented in [11] and as visible in Figure 2, the dependency between the sneakpath current I_{sp} and the programming current I_{prog} regarding the IRdrop follows a different trend. While I_{prog} is injected one time for a single bit-write, I_{sp} is consumed non-linearly along the SWL and SBL. This result in a non-linear IRdrop effect for the sneakpath, amplifying the effect of Isp while it is linear for the I_{prog} current, as shown Figure 3-a and b. When multiple bits (n) are accessed in the memory array, the worst case current correspond to $n*I_{prog}$ while these contributions are removed from the Isp current sum.

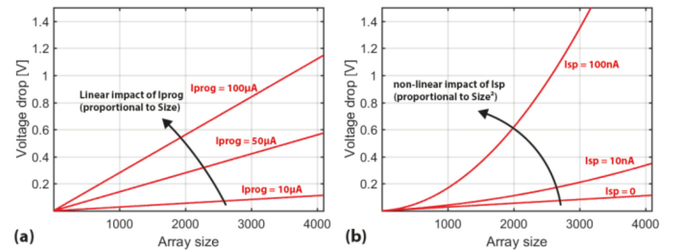

Figure 3: : Impact of the (a) programming (I_{prog}) and (b) sneakpath (I_{sp}) currents on the voltage drop across metal lines.

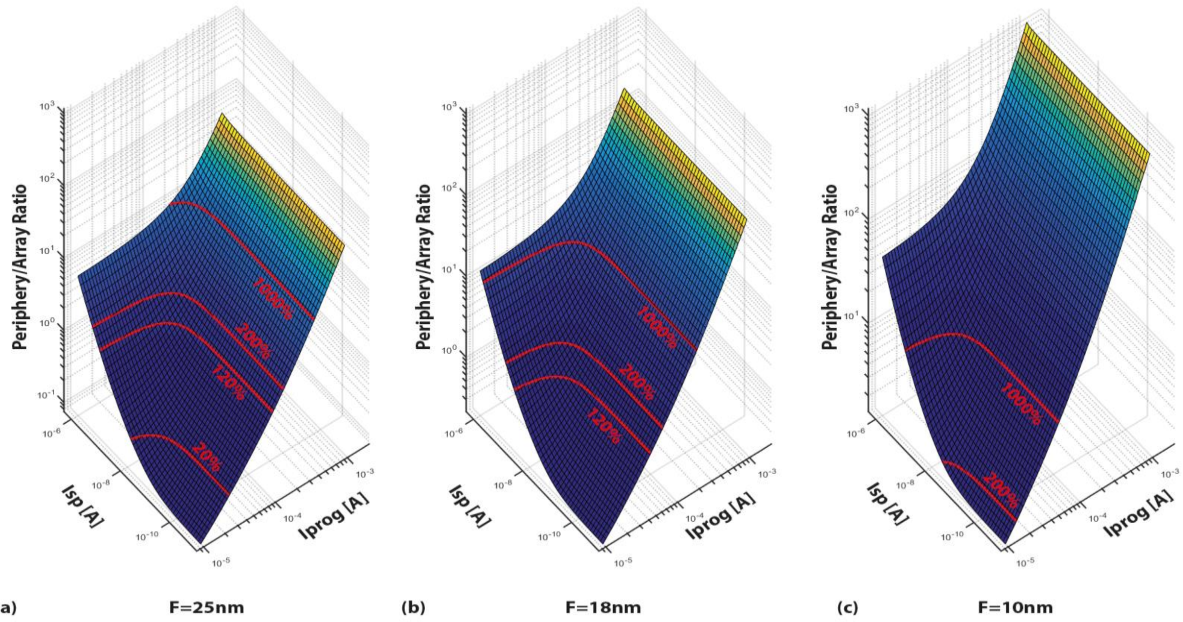

Figure 4 : 3D plot of the area overhead in a crosspoint memory array for several technology nodes: 25nm (a), 18nm (b) and 10nm (c). The area overhead increases with the reduction of the metal half pitch F.

III. Memory Periphery Design Considerations

While the crosspoint memory array itself is widely explored, there are only few works focusing on the scaling of peripheral circuitry. Particularly, due to the current requirement of 1S1R arrays (I_{sp} and I_{prog}), small, high current drive and high voltage transistors are required for WL and BL control. However, such devices do not exist. In other words, *middle voltage* management (2 to 5 volts) forbid the use of thin oxide gate transistors leading to wider transistors, at equivalent current drive. In the end, this increases the memory array periphery area.

In this section we propose to describe the main effects of such constraints on peripheral transistors. We first propose to study the limit between *low voltage* (i.e., thin oxide) and *middle voltage* (i.e., thick oxide) transistors. Then, we detail the evolution of the periphery/array area overhead for various BEoL technology nodes (metal half pitch F), for various I_{prog} and I_{sp} currents and for multiple-bits written in the memory array. Finally, we propose to discuss the pitch matching issues from a physical layout point of view and its effects on the IRdrop.

A. Decoding Logic Design Strategies

The two required near array periphery blocks are the *decoding logic* (DEC) and the *multiplexer* (MUX). We proposed optimized MUX and DEC architectures in previous works [11]. However, the question of voltage management has not been presented for the DEC block. In *Figure 5*, we present the DEC area evolution for two configurations. One with thin oxide DEC, *thick oxide* MUX and *Voltage Level Shifter* (LS – in the inset) in-between. The other one with thick oxide DEC and MUX while the addresses voltages are shifted before the DEC. Compared to flash technologies in which, due to huge *High Voltage* (HV) transistors, the DEC is designed in thin oxide transistors, in crosspoint RRAM technologies, middle voltage management enable innovative design strategies to optimize the periphery area. This way, a thick oxide DEC with LS blocks on the encoded addresses enables up to 1.5x of area reduction compared to the equivalent solution in thin oxide with LS blocks on the decoded addresses.

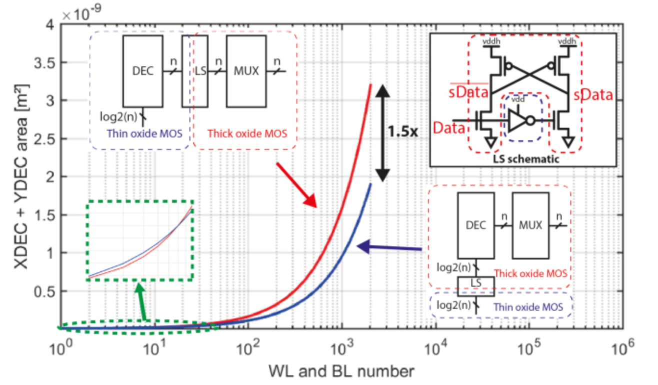

Figure 5: Comparison between two possible implementations of dynamic decoder. Blocks using thick oxide MOS consider a higher supply voltage. The implementation using thick oxide DEC block (blue) is more compact than the red one, thanks to a lower number of Level Shifters (LS) - $Log_2(n)$ *instead of* n.

B. Metal Scaling Effect on Area Overhead

Figure 4 presents 3D plots of the Periphery/Array area ratios for three different metal half-pitch versus I_{sp} and I_{prog} currents. With the reduction of the crosspoint metal lines pitch F (WLs and BLs), the IRdrop effect is increased by two different sources. The first one is the intrinsic metal resistivity which increases from around 6 to 9Ω.m-1 when F is scaled from 50nm to 10nm respectively [22]. As shown in Figure 4, the 20% area overhead can only be obtain for really low I_{prog} (lower than 20uA) for a F=25nm (a). When scaled down to F=18nm (b), 120% of area overhead line can be reached for I_{prog} lower than 40uA. This 120% area overhead corresponds to a 20% if the periphery can be entirely fitted under the memory array. Finally,

a F=10nm crosspoint memory array (c) doesn't give any acceptable area overhead values (1000% for a 50uA I_{prog}). This non-scaling is related to two effects on RRAM technologies when the devices dimensions are reduced: *(i)* The I_{prog} current doesn't scale. *(ii)* the Vprog voltage doesn't scale (but even worse, in STT-MRAM it tends to increase). Overall, for deeply scaled technology nodes, reducing the memory array size (by reducing F) does not reduce the periphery area.

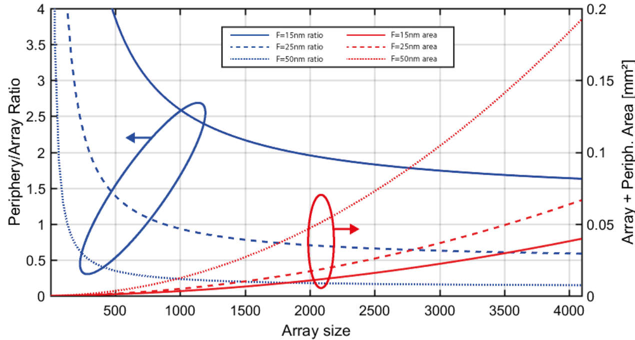

Figure 6: Impact of the array size on the Periphery/Array Ratio (blue) and Periphery+Array Area (red) for various metal half pitch.

This effect is illustrated in Figure 6 which shows the periphery/array area overhead (blue) and the array+periphery area (red) versus the memory array size for F varying from 15nm to 50nm. A smaller F lead to higher area overhead for the same array size. This effect could be counterbalanced by increasing the array size, however, IRdrop effects limits the array size to smaller and smaller arrays when F is reduced.

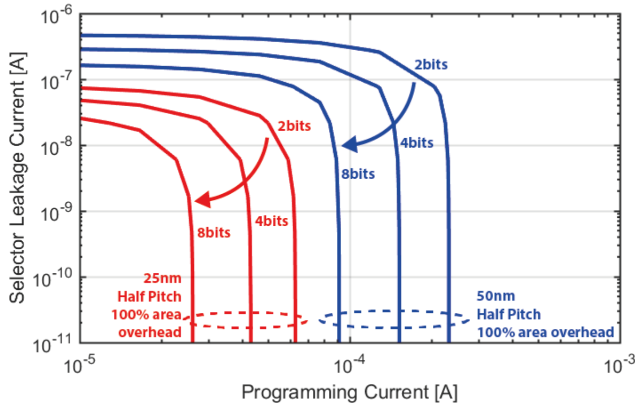

Figure 7: Area overhead of crosspoint arrays versus selector leakage current and RRAM programming current for 2 to 8 bits written per array for F=25nm (red) and 50nm (blue).

C. *Multiple Bit-Write Effect on Area Overhead*

When writing simultaneously *n* bits in a crosspoint memory array, the total current consumed on the array is $n*I_{prog}$ (i.e. sum of I_{prog}). This leads to two effects: *(i)* the IRdrop which directly depends on the I_{prog} current is strongly increased and *(ii)* the peripheral circuitry W sizing increases accordingly to compensate for the higher current consumption. *Figure 7* shows the evolution of the 100% area overhead line (cf. 3D graphs Figure 4) for F=50nm (blue) and F=25nm (red), when 2, 4 or 8 bits are written simultaneously in a single array, versus I_{sp} and I_{prog}. As expected, writing more bits in parallel increases the IRdrop effects and the periphery area requiring lower I_{prog} and I_{sp} to keep the area overhead ratio constant.

D. *Periphery-Array Pitch Matching Constraints*

Additionally, to periphery/array area overhead constraints which appears when scaling down the metal half pitch F. Fitting the MUX transistors on a 2*F pitch is a new challenge. Figure 8 shows the physical layout of a F=25nm memory array connected to a middle voltage transistors MUX [11]. In this example, each array line is connected to 3 transistors (2 p-type and 1 n-type) with a minimum length of 250nm (i.e. 10*F). Thereby, among 5 contiguous array WLs or BLs, only 1 will be directly connected to its MUX driver. For the other, an additional metal length, up to 5 times the MUX width, will be added. Such a configuration introduces a non-regular IRdrop effect among the memory array from one line to the other. When applied to the results from [14], a variable Bit-Error Rate is expected to be found along the memory array. One solution could be to double the metal access to the array to reduce the parasitic serial resistance.

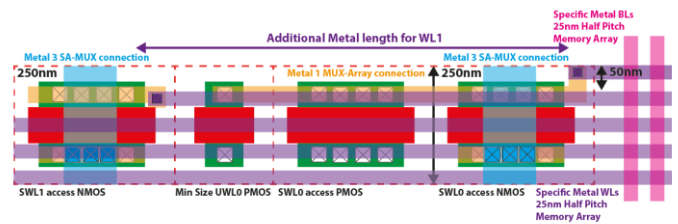

Figure 8: Connection between the MUX gates and decoder are not represented for the sack of clearness but they may be done using metal 1 and metal 2 horizontal routing. In this view, as in [7] [23], the crosspoint array is processed in a high BEoL metal level with aggressive pitch (F = 50nm).

IV. CROSSPOINT POSITIONNING

Once the previous considerations are taken into account. Considering optimized MUX and DEC blocks design [11], we propose to explore three architectures among the published crosspoint memories: (i) standard planar structure with no periphery under a single layer crosspoint array (i.e., standard FEoL-based memory architecture). (ii) single layer crosspoint array with *CMOS under Array* (CuA). (iii) common bitline 2-layer crosspoint array with CuA. It is important to note that any architecture that would stack more crosspoint array would not improve the density because near array periphery cannot be reduced more [16]. As a reference, all the industrial papers or products considering crosspoint memories consider the common bitline 2-layer architecture [3] [7] [8] [23].

We compare these three architectures with commercial Flash NAND and DRAM products density and technology nodes. While Flash NAND requires HV transistors to operate, its programming current are kept low as all the programming operations are performed in *Fowler-Nordheim* mode. This way, more than 80% of area efficiency can be achieved (the considered densities are the following 76Gb/cm² for F=16nm [24], 56Gb/cm² for F=19nm [25] and 28Gb/cm² for F=32nm [26] and for 3D VNAND [1] more than 200Gb/cm^2 while considering a relaxed pitch higher than 40nm). On the other hand, due to charge sharing effect between the accessed cells and the BLs, DRAM density is limited (the considered bit

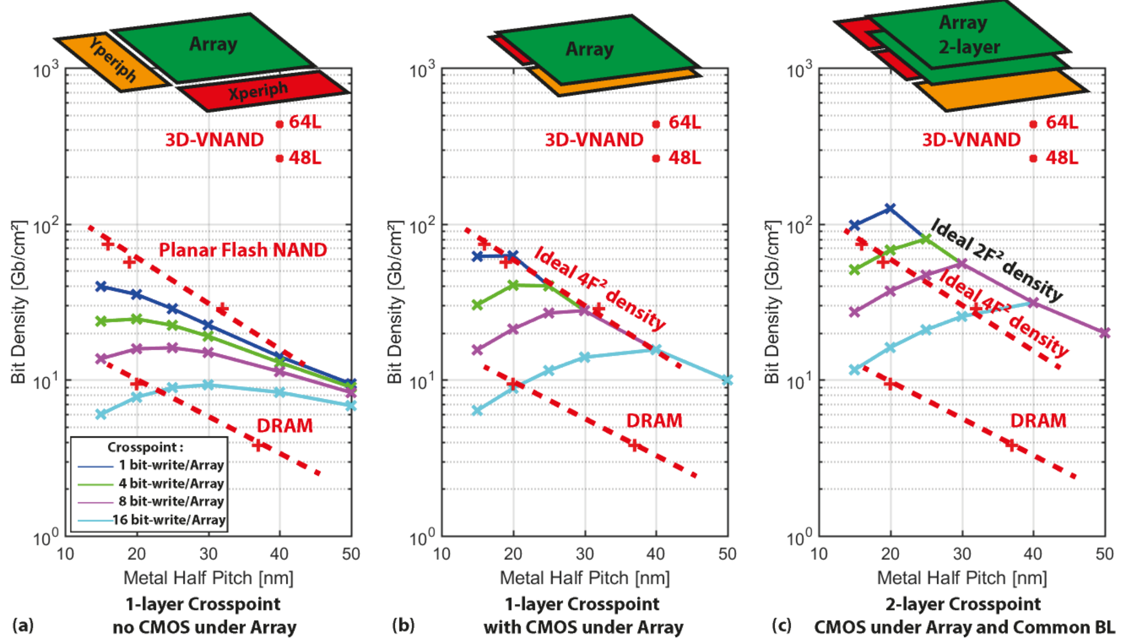

Figure 9: Bit density versus metal half pitch graph of crosspoint memory arrays with three memory array configuration for different number of parallel bit-write per array. (a) 1-layer crosspoint memory with no CMOS under array. (b) 1-layer crosspoint memory with CMOS under array. (c) 2-layers crosspoint memory with CMOS under array. For each configuration, bit densities are compared with DRAM and NAND flash memory technologies.

densities are 9.4Gb/cm² for F=20nm [27] and 3.8Gb/cm² for F=37nm [28]). When compared to Flash NAND memories, due to huge area overhead, crosspoint memories exhibit lower bit density although the bitcell area is the same (4F²). This gap increases as more bits are written in parallel. Figure 9 shows the bit density of a crosspoint memory using I_{prog}=30µA and Isp=10nA versus the metal half pitch (from 50nm down to 15nm) for different word-length written in parallel in a single array (1bit to 16bits), and compared to other memory technologies (planar and vertical 3D Flash NAND and DRAM). Various array configurations are considered. In (a), 1-layer planar crosspoint memory array. In (b), the memory array ratio is reduced of a 100% factor in order to simulate a CMOS-under-Array (CuA) integration of the peripheral circuitry. Finally, in (c), a 2-layer memory array with common BLs and CuA is considered. In this configuration, the BLs MUXs and DECs are in common. Each configuration is illustrated on the top by a schematic of the array and the periphery.

Due to their area hungry peripheral circuitry, not stacked crosspoint memories (without CuA) shows lower density (40Gb/cm² at F=15nm) than planar NAND memories. This effect becomes worse when multiple bits are written in parallel in the array. Visible density optimums are due to the fact that when *F* is scaled, the memory array area reduces but the peripheral circuitry does not. The optimal configuration consists in stacking 2 layers of crosspoint with the BLs in common and the periphery underneath the array (Figure 9-c). In this configuration, the BLs MUX and DEC are in common. This way, the periphery area is reduced. Thus, the density follows the ideal 2F² density (2 stacked layers of 4F²) and drops when the periphery overflows. It is interesting to note that the density drops faster than for the 1-layer configuration due to smaller array area (cf. Figure 9-a). Additionally, better densities can be obtained using unipolar memories tanks to a simpler peripheral circuitry.

Next, we consider the memory write throughput in the estimation, in order to accurately place the memory architecture in the memory hierarchy. While DRAM exhibit at max 20GBytes/cm^2 [2], its write throughput exceed the GBytes/sec. On the other hand, the slow writing time of Flash NAND technologies (writing is performed with first, a block erase of few MBytes followed by write pages of few kBytes each taking around the millisecond) limits their write throughput to less than 20MBytes/sec.

Table 1 : Considered write and read conditions

Operation	Time	Current	Voltage
Read	50ns		
Set	100ns	30uA	2.5V
reset	100ns	30uA	2.5V

Crosspoint memory write throughput is determined at the array level. Oppositely to NAND flash technologies, a block erase is not required, each single bit can be written either to 0 or 1 (i.e., *set* or *reset*). In order to avoid overwrite phenomenon and to optimize the power consumption, a read operation is performed before each write. Then the read output and the input word to be written are compared. Finally, only the bits in a different state are written in two steps. One *set* and one *reset* step. Table 1 summarizes the considered conditions for *read*, *set* and *reset* in a crosspoint memory array. We consider an overall full programming cycle time of 250ns.

While it has been shown that the time distribution can strongly move with cycle to cycle or device to device variability, we assume that a smart circuit such as [29] is considered. This

way, it enables reliable write operation while limiting the programming time. We are aware that the time, voltage and current values presented Table 1 might seems optimistic, but device engineering is still ongoing. Additionally, we assume that in SCM positioning, retention time is less critical, relaxing the programming current constraints.

Figure 10 presents the evolution of the write throughput versus the memory bit density of commercial DRAM and flash NAND chips compared to a common BL 2-layer crosspoint memory in CuA configuration. Write operations are considered performed sequentially in a single crosspoint memory array (dark blue curve). Writing more bits in the same memory array leads to lower density but multiplies the write throughput. However, in such configuration, crosspoint cannot even compete with Flash NAND write performances. Thereby, as introduced in [11], the solution consists in parallel multiple bank access to increase the write throughput while not decreasing the bit density (cyan curve). We neglect here the density effect of far periphery and controller additional complexity. Finally, crosspoint memory appears in an intermediate positioning between Flash NAND and DRAM in term of write throughput and bit density.

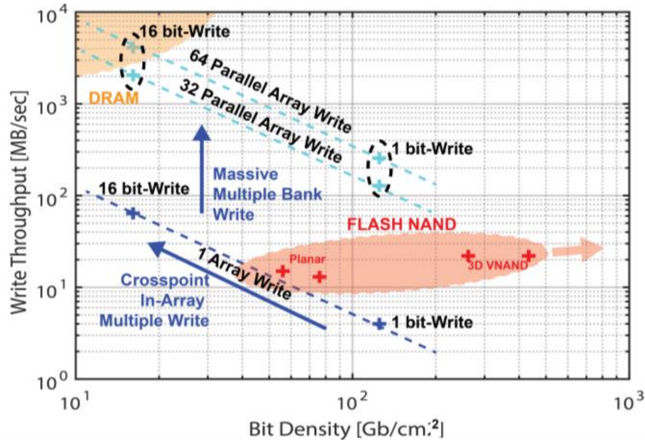

Figure 10: Write throughput versus the bit density for crosspoint (blue), Flash NAND (red) and DRAM (orange). Massive multi-bank write in crosspoint places it in an intermediate positioning.

V. Conclusion

In this work, we proposed an overview of the design constraints related to RRAM crosspoint memories design. Thereby, based on our previously published IRdrop and periphery overhead models, we showed the effect of multiple bit-write on periphery area and IRdrop effect. We also show that stacking more than 2-layers of crosspoint array doesn't bring any density improvement as the periphery cannot be shared more and pitch matching issues become critical. Finally, we explored the positioning of crosspoint in the memory hierarchy by comparing several crosspoint architectures in term of density and write throughput. We showed that crosspoint memory must be written in massively multiple bank approach (32 to 64 parallel access) to be positioned in-between Flash NAND and DRAM technologies as a SCM.

Acknowledgment

This work has been partially funded by the ERC Consolidator Grant COMPUSAPIEN (Agreement No. 725657) and by European project ECSEL-PANACHE.

References

[1] K.-T. Park et al., "Three-Dimensional 128 Gb MLC Vertical nand Flash Memory With 24-WL Stacked Layers and 50 MB/s High-Speed Programming," *IEEE JSSC,* 2015.

[2] L. D. Uk, "A 1.2V 8Gb 8-channel 128GB/s high-bandwidth memory (HBM) stacked DRAM with effective microbump I/O test methods using 29nm process and TSV," *IEEE ISSCC,* 2014.

[3] C. Paolo, "Non Volatile Memory Evolution and Revolution," *IEEE IEDM,* 2015.

[4] H.-S. P. Wong et al., "Metal-Oxide RRAM," *Proc. of the IEEE,* 2012.

[5] H.-S. P. Wong et al., "Phase Change Memory," *Proc. of the IEEE,* 2010.

[6] D. Apalkov et al., "Magnetoresistive Random Access Memory," *Proc. of the IEEE,* 2016.

[7] T.-y. Liu et al., "A 130.7-mm 2-Layer 32-Gb ReRAM Memory Device in 24-nm Technology," *IEEE JSSC,* 2014.

[8] A. Kawahara et al., "An 8 Mb Multi-Layered Cross-Point ReRAM Macro With 443 MB/s Write Throughput," in *IEEE JSSC*, 2013.

[9] A. Levisse et al., "Capacitor based SneakPath compensation circuit for transistor-less ReRAM architectures," *IEEE Nanoarch,* 2016.

[10] A. Levisse et al., "SneakPath Compensation Circuit for Programming and Read Operations in RRAM-based CrossPoint Architectures," *IEEE NVMTS,* 2015.

[11] A. Levisse et al., "Architecture, Design and Technology Guidelines for Crosspoint Memories," *IEEE Nanoarch,* 2017.

[12] B. Giraud et al., "Advanced memory solutions for emerging circuits and systems," *IEEE IEDM,* 2017.

[13] A. Chen, "A Comprehensive Crossbar Array Model With Solutions for Line Resistance and Nonlinear Device Characteristics," *IEEE TED,* 2013.

[14] M. Manqing et al., "Design an Analysis of Energy-Efficient and Reliable 3-D ReRAM Cross-Point Array System," *IEEE TVLSI,* 2018.

[15] A. Ciprut et al., "Energy-Efficient Write Scheme for Nonvolatile Resistive Crossbar Arrays With Selectors," *IEEE TVLSI,* 2018.

[16] A. Levisse et al., "High Density Emerging Resistive Memories: What are the Limits," *IEEE LASCAS,* 2017.

[17] E. Vianello et al., "Resistive Memories for Ultra-Low-Power embedded computing design," *IEEE IEDM,* 2014.

[18] A. Grossi et al., "Experimental Investigation of 4-kb RRAM Arrays Programming Conditions Suitable for TCAM," *IEEE TVLSI,* 2018.

[19] W. C. Shen et al., "High-K metal gate contact RRAM (CRRAM) in pure 28nm CMOS logic process," *IEEE IEDM,* 2012.

[20] R. Aluguri et al., "Overview of Selector Devices for 3-D Stackable Cross Point RRAM Arrays," *JEDS,* 2016.

[21] G. Piccolboni et al., "Vertical CBRAM (V-CBRAM): from experimental data to design perspectives," *IEEE IMW,* 2016.

[22] [Online]. Available: http://www.itrs2.net/.

[23] "Techinsights - 3Dxpoint analysis," [Online]. Available: http://www.techinsights.com/about-techinsights/overview/blog/intel-3D-xpoint-memory-die-removed-from-intel-optane-pcm/.

[24] M. Helmet al., "A 128Gb MLC NAND-Flash device using 16nm planar cell," *ISSCC,* 2014.

[25] N. Shibata et al., "A 19nm 112.8mm2 64Gb multi-level flash memory with 400Mb/s/pin 1.8V Toggle Mode interface," *IEEE ISSCC,* 2012.

[26] B. T. Park et al., "32nm 3-bit 32Gb NAND Flash Memory with DPT (double patterning technology) process for mass production," *IEEE VLSIT,* 2010.

[27] M. Brox et al., "An 8Gb 12Gb/s/pin GDDR5X DRAM for cost-effective high-performance applications," *ISSCC,* 2017.

[28] T.-K.-J. Ting et al., "An 8-channel 4.5Gb 180GB/s 18ns-row-latency RAM for the last level cache," *IEEE ISSCC,* 2017.

[29] G. Sassine et al., "Sub-pJ Consumption and Short Latency Time in RRAM Arrays for High Endurance Applications," *IEEE IRPS,* 2018.