

High-performance, cost-effective 2z nm two-deck cross-point memory integrated by self-align scheme for 128 Gb SCM

Taehoon Kim, Hyejung Choi, Myoungsub Kim, Jaeyun Yi, Donghoon Kim, Sunglae Cho, Hyunmin Lee,
Changyoun Hwang, Eung-Rim Hwang, Jeongho Song, Sujin Chae, Yunseok Chun, Jin-Kook Kim
R&D Division, SK-Hynix, Icheon, Republic of Korea, email: taehoon12.kim@sk.com

Abstract— We demonstrate a high-performance and cost-effective cross-point memory (CPM) technology for two-deck 128 Gb storage class memory (SCM). The unit MAT size is 16 Mb consisting of a 2z nm 1S1M (one selector one memory) structure that is patterned by only two ArF-i steps per deck for a low cost per bit. The formidable task of self-align etch is enabled by the use of state-of-the-art etching and integration technology, which otherwise easily leads to hard fail or poor cell characteristics and reliabilities. New phase change materials (N-PCMs) are developed to have a large V_t window and a uniform V_t distribution for a sufficient read window margin (RWM) and a corresponding low raw bit error rate (RBER). New chalcogenide selectors (NCSs) are also developed to provide low V_t instability and very low leakage current. The new CPM is able to provide a sufficient RWM for 16 Mb MATs with very low latencies of write (set ≤ 300 ns) and read (≤ 100 ns). We also demonstrate its decent write disturbance and high reliabilities such as endurance and thermal retention.

I. INTRODUCTION

Ever-increasing demand for high-capacity and high-performance SCM for data centers and cloud computing systems has led to the pursuit of new types of SCM for both memory and storage applications [1]. The features desired of such memory are low latencies, byte-addressability, long endurance, and persistency at low cost. These requirements cannot be satisfied by the current NAND and DRAM technology [2]. In this paper, we introduce a high-performance, cost-effective two-deck CPM technology for 128 Gb SCM that is quite close to commercialization. We must emphasize that the data provided here does not represent the optimal single-cell performance but rather a typical 16 Mb arrays' performance for 128 Gb SCM. Therefore, this paper will provide the major issues and factors that must be considered at each development stage.

II. RESULTS AND DISCUSSION

A. Structure and Integration

Fig. 1 shows a cross section of the two-deck cell array with a peri under cell (PUC) structure. A Cu multi-layer was used for high-speed data transmission between the cell array and the PUC. Fig. 2 shows the floor plan of the 128 Gb die. The unit MAT size is 16 Mb, consisting of an array of 2z nm pillar cells. Each pillar consists of a memory, a selector, and electrodes separating each material physically and optimizing

its electrical properties, as shown in Fig. 3. To minimize the process cost per bit, it was necessary to perform integration with a self-align etch. Unfortunately, typical chalcogenide alloys are vulnerable to the conventional etching and cleaning processes. These technical difficulties have been directing PCM research towards the chemical vapor deposition (CVD) and atomic layer deposition (ALD) processes aiming damascene process for the last two decades. However, these processes are very slow and expensive, and the resulting film quality is poor both physically and electrically because of the porosity of the resulting structure, particularly at the sub-50 nm scale [3].

To overcome this, we used physical vapor deposition (PVD) for the cell stack deposition and developed new recipes for dry etching as well as cleaning for cross-point patterning. Integration schemes were also developed to minimize the process and integration damage. The self-align processes and integration schemes are shown in Fig. 3. The word lines (WLs) were patterned using a single ArF-i mask step after cell-material stack deposition. Both inter-layer dielectric (ILD) deposition and chemical mechanical polishing (CMP) were performed for bit line (BL) metal deposition followed by BL-patterning performed by another single ArF-i mask step.

B. Development of New PCM

Fig. 4(a) shows the IV behavior of a typical single cell, while Fig. 4(b) shows the V_t distributions of the array after the set and reset operations. A large read window margin (RWM) existed, even below -4.5σ , because of the N-PCM and NCS. Thus, a large RWM is a key enabler to satisfy the entire array's operations with a very low RBER and high production yield. To achieve this, the ΔV_t value between set and reset must be large enough, and the V_t distribution slope must be steep enough for the given array size, which can be expressed as

$$\text{RWM} = \Delta V_t - \sigma_{\text{array}} \times (\sigma_{\text{Set}} + \sigma_{\text{Reset}}), \quad (1)$$

where ΔV_t is the median V_t gap between reset and set, σ_{array} is the array size by normal quantile plot for the target BER, and σ_{Set} and σ_{Reset} are the V_t distribution slopes (standard deviations) of set and reset, respectively. A relatively easy way to increase ΔV_t is to increase the bandgap of the material by adding dopant. However, doing so will result in a longer t_{set} and a poor set distribution, as shown in Fig. 5(a). On the other hand, the N-PCM we developed exhibits a large ΔV_t without the disadvantage of t_{set} , as shown in Fig. 5(b). Furthermore, σ_{Set} and σ_{Reset} are also strong functions of the

selector's V_t instability, because each cell's distribution can be expressed by the combination of PCM and NCS as shown in (2) and (3).

$$\sigma_{\text{Set}} = (\sigma_{\text{Set_PCM}}^2 + \sigma_{\text{Set_NCS}}^2)^{1/2} \quad (2)$$

$$\sigma_{\text{Reset}} = (\sigma_{\text{Reset_PCM}}^2 + \sigma_{\text{Set_NCS}}^2)^{1/2} \quad (3)$$

C. Development of New Selector

It is well known that, in CPM, the off-current of the deselected selectors (I_{off}) should be small enough to both avoid read/write errors and provide a decent V_t distribution with limited IR drop. Fig. 6 shows the normalized IV behaviors of the various selector materials after integration without PCM. The selector's vertical leakage should be small enough in the subthreshold region to avoid read/write failures. Compared to the conventional ovonic threshold switch (C-OTS), NCS exhibits a much lower I_{off} value (<1 nA between 0.7 – 0.8 V_t), which is sufficient for 16-Mb arrays and higher.

Another important factor is the selector's V_t instability, including drift and random telegraph noise (RTN). Drift is an intrinsic property of amorphous materials in which defects are annihilated spontaneously [4]. The RTN phenomenon that exists in amorphous materials is thought to be caused by either the current or the corresponding voltage fluctuations that occur when the carriers' hopping paths vary [5]. Fig. 7 depicts the ways of RWM consumption caused by these two mechanisms. Fig. 8 shows the drifts of various materials. There are huge differences among the materials. In particular, NCS exhibits very little drift, even at 55°C , which guarantees several years with little RWM consumption. Fig. 9 shows the tradeoff between the drift and the RTN, which was defined using the standard deviation of 100 V_t reads. This result also indicates that a net improvement can be achieved by selecting a new material. However the tradeoff still exists. Therefore, these two parameters should be checked in determining the final RWM.

D. Read and Write Performances

Table 1 shows the overall performances of read and write, expressed via both the pulse widths and the latencies. The selector's quick turn-on and fast-charging characteristics make it possible to perform detections within 20–30 ns, which is a period short enough to guarantee a read latency shorter than 100 ns, including for both command and addressing. In the CPM structure, however, the set performance is more critical. Unlike the conventional line-type PCRAM in which crystal seeds surround the amorphous reset region, a self-align etched CPM has a confined structure that does not leave any crystal seeds behind after a full reset is performed, as shown in Fig. 10. Because of the nucleation step, particularly at the nano-scale, the set speed of the confined structure intrinsically becomes much slower than that of conventional PCRAM. Therefore, the set performance should be considered as carefully as the ΔV_t value. Fig. 11 shows that the set performance can easily be degraded by process damage, particularly in the low-probability region. In contrast, Fig. 12 shows that strategically integrated N-PCM can achieve a decent set distribution down to 300 ns without tail bits.

E. Disturbance and Reliability

The (thermal) write disturbance (WDT) is an unwanted reset to set transitioning of the adjacent (victim) cells during a reset write of the target (aggressor) cell. WDT is considered a major obstacle to PCRAM scaling. However, CPM's confined structure can assist in suppressing such disturbances. Decent integration for the uniform reset current distribution can also minimize the WDT by limiting the maximum reset current (I_{reset}). The benefit of such a solution is the elimination of the write-verify which means a huge advantage over traditional PCRAM on both write latency and power consumption. As explained, Fig. 13 shows no disturbed cell even at $1\text{E}5$ cycle.

With respect to reliabilities, Fig. 14 shows the write (set/reset) cycle endurance, which is one of the major advantages of SCM over NAND flash memory. Our device maintained a sufficient RWM even after $1\text{E}7$ cycle that is superior to NAND Flash at least more than three orders. The inset shows the variation of overall set/reset distribution including tail bits for each cycle. The RWM consumption by the write cycle is quite small. Finally, the thermal retention, which is determined by the crystallization of the PCM, was checked. N-PCM can maintain reset states for more than 10,000 hours at 85°C , including the tail bits, as shown in Fig. 15. Although there is a tradeoff between the set speed and the retention, different activation energies for the crystallization at low and high temperatures can improve the retention margin without degrading the set performance.

F. Second-Deck Properties

The primary advantage provided by the two-deck structure is the doubling of the bit density while achieving the same number of net die per wafer. As a solution, a common BL structure was chosen to minimize the number of local/global BL-selection transistors. However, this resulted in a different polarity in the second deck, which caused an offset in the cell's characteristics. In addition, the differing thermal histories of the two decks can also cause an offset. This offset should be compensated for by modulating the integration between the two decks. Fig. 16 shows a comparison of the offsets in the set/reset V_t before and after the integration modulation, which clearly demonstrates that the offsets were successfully corrected for by the modulation.

III. CONCLUSION

In this letter, we demonstrated a cost-effective 128 Gb CPM technology that is nearly ready for commercialization. We developed a set of N-PCMs and NCSs for large RWM and V_t stability, which are both necessary factors for a low RBER and decent reliability margins. However, because of the intrinsic vulnerabilities of chalcogenide alloys, such characteristics can easily deteriorate or even be washed out after integration. To overcome this, robust materials, new etching, and cleaning recipes with novel integration schemes were developed.

In conclusion, functional 16 Mb MATs with sufficient RWM were successfully obtained. These MATs also exhibited great reliability in areas including write endurance, retention, and drift as well as little WDT.

REFERENCES

- [1] S. Nazari, "Using storage class memory in next generation designs" *Flash memory summit keynote 10* (2017)
- [2] G. W. Burr et. al, "Overview of candidate device technologies for storage-class memory," *IBM J. Res. Develop.*, vol. 52, no. 4, pp. 449-464, (2008)
- [3] W. Kim et. al, "ALD-based Confined PCM with a Metallic Liner toward Unlimited Endurance," *IEDM Tech. Dig.*, pp. 4.2.1-4.2.4, (2016)
- [4] D. Ielmini et. al, "Physical interpretation, modeling and impact on phase change memory (PCM) reliability of resistance drift due to chalcogenide structural relaxation," *IEDM Tech. Dig.*, pp. 939-942., (2007)
- [5] D. Dong et. al, "The Impact of RTN Signal on Array Level Resistance Fluctuation of Resistive Random Access Memory" *IEEE Electron Device Lett.*, Vol. 39, (2018)

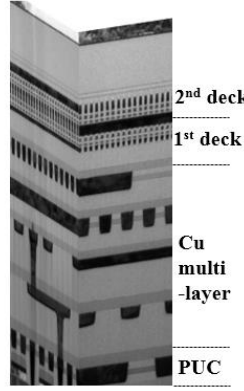


Fig. 1. TEM cross section of two-deck cell array with Cu multi-layer and PUC.

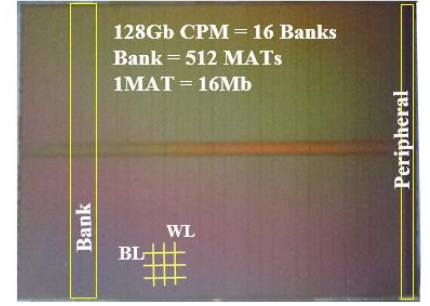


Fig. 2. Floor plan of 128 Gb die consisting of 16 banks.

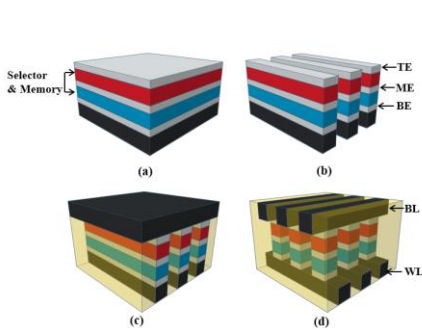


Fig. 3. Self-align process integration schemes: (a) cell stack material deposition, (b) after self-aligned WL patterning, (c) ILD deposition, CMP, and BL deposition, and (d) self-aligned BL patterning.

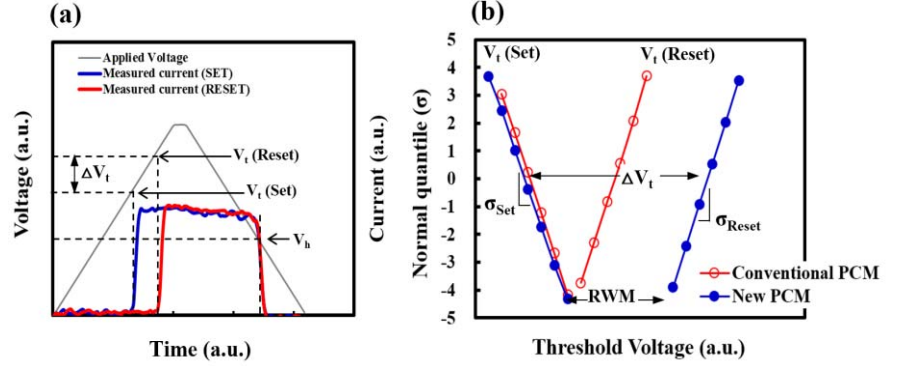


Fig. 4. (a) Voltage measurements by time for V_t and V_h (hold voltage) detection and (b) V_t distributions for set and reset. N-PCM exhibited much larger RWM than C-PCM.

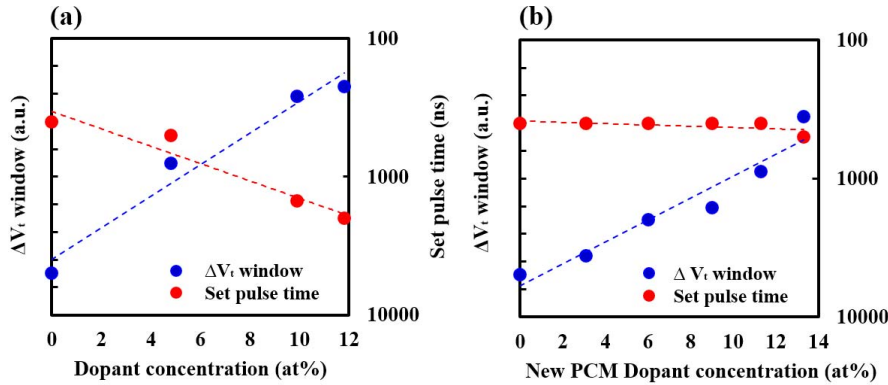


Fig. 5. Variations in ΔV_t by dopant concentration, which is proportional to bandgap: (a) conventional dopant results showing tradeoff between ΔV_t and t_{set} and (b) N-PCM dopant exhibiting increasing ΔV_t without t_{set} degradation.

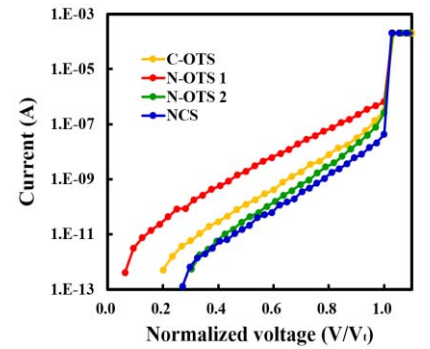


Fig. 6. Normalized I-V behaviors of various selector materials after integration without PCM.

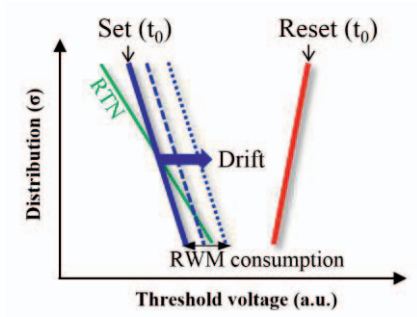


Fig. 7. Plot of set/reset V_t to explain RWM consumption by drift and RTN.

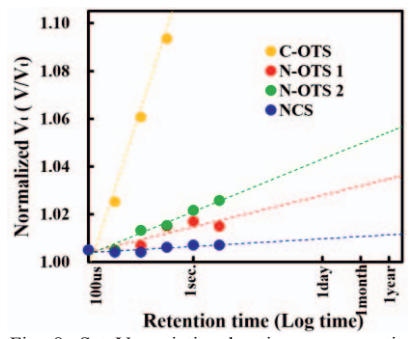


Fig. 8. Set V_t variation by time, representing drift from various selectors: C-OTS, N-OTS, and NCS.

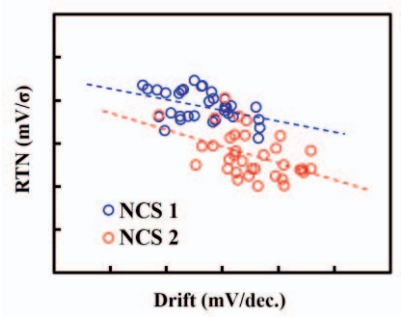


Fig. 9. Tradeoff between set drift and RTN at 55 °C, where RTN is defined as the standard deviation of 100 V_t reads.

Process technology	2z nm
Cell size	4F2 / two deck
Cell selector	NCS
Cell memory	N-PCM
Organization	8Gb* 16 Banks =128Gb
Read latency	≤100ns
Write latency	≤ 30ns (reset) ≤300ns (set)

Table 1. Basic die information, including structure and pulse widths (latencies).

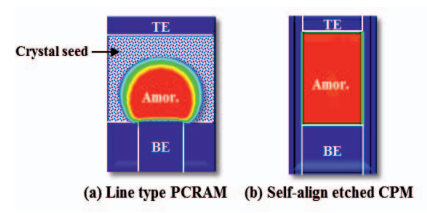


Fig. 10. Different environments for set operation (crystallization) in (a) line type and (b) confined structure. Note that there's no crystal seed in (b) after full amorphization, which requires additional nucleation step.

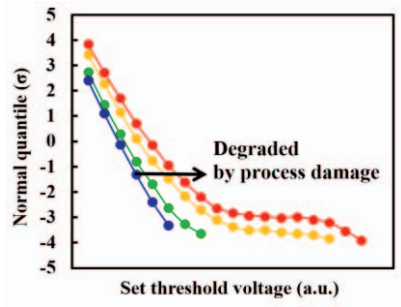


Fig. 11. Aggravating set tail by process damage from different etch recipe and integration.

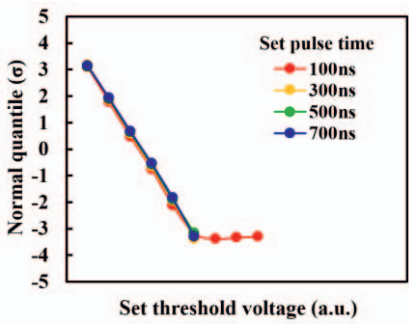


Fig. 12. Variations in set distribution by different set time. Note that tail starts to develop from 100 ns.

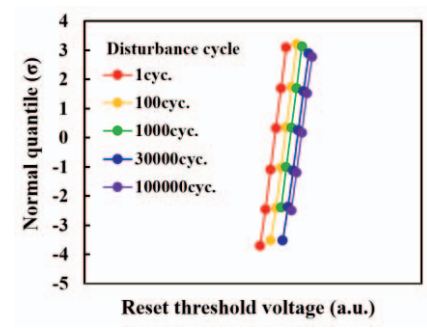


Fig. 13. Variations of victim cell's reset distribution by increasing aggressor's reset write cycle. For better visibility, X-axis offset is made intentionally.

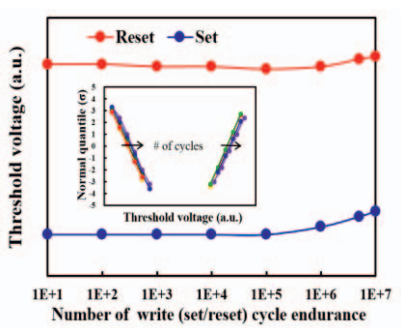


Fig. 14. Write cycle (set/reset) endurance of median cell (main) and entire array distribution, including tail bits (inset). Wide RWM is maintained even after 1E7 cycle.

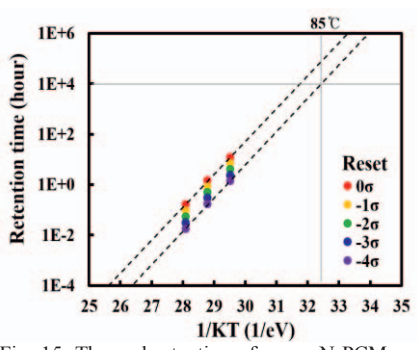


Fig. 15. Thermal retention of array. N-PCM can maintain reset state for more than 10,000 hours at 85 °C, even for the tail bits.

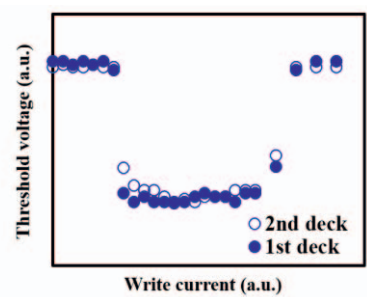
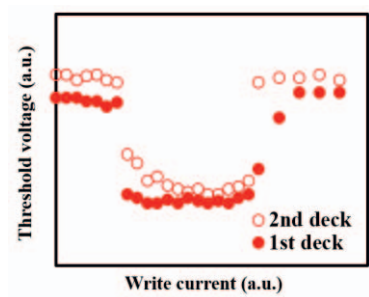


Fig. 16. Different V_t -I behaviors in first and second decks (a) before and (b) after integration modulation. Both decks exhibit similar behaviors after modulation.