Advanced Technology and Systems of Cross Point Memory

Albert Fazio
Intel Corporation, Santa Clara, CA, USA, email: Al.Fazio@intel.com

Abstract **-** Cross point memories are ideally suited to fill computer memory hierarchy gaps of memory capacity-cost and storage performance. System innovations exploiting the capability of 3D XPoint based cross point memory challenge historical notions of separate semantics of memory and storage, significantly boosting system performance.

I. Introduction

A decade ago, NAND began to fill the performance gap between DRAM and HDDs in computing memory hierarchy. Today, NAND SSDs are a fundamental tier in the memory hierarchy of a modern computing system. In the intervening decade, DRAM capacity and cost scaling has significantly slowed (Fig. 1) and NAND's performance growth has lagged its capacity growth (Fig. 2). Meanwhile, higher CPU core counts are requiring higher memory and storage capacity and bandwidth. As a result, two gaps have grown in computing memory hierarchy: load/store memory capacity-cost gap and storage performance gap. Commercialization of 3D XPoint [1], a non-volatile cross point memory with high capacity, low latency, and byte write capability is demonstrating ability to fill these gaps. New computing system architectures exploiting 3D XPoint are altering the notion of separate worlds between memory and storage semantics, significantly improving system performance. This paper reviews 3D XPoint memory technology, its emergence in computing systems, and discusses future directions of each.

II. Cross Point Memories

The industry researched 3D cross point memories [2-4] to simultaneously meet favorable performance and cell size characteristics. The challenges in realization has been the memory material system of choice and limited options for thin film selectors with appropriate Ion/Ioff characteristics. Memory material systems of metal oxides relying on oxygen vacancies [5], or electrochemical cells [6] are relatively easy to process but are subject to filamentary stochastic variation [7], limiting scaling. Choice of memory element material system capable of bulk switching and mating electrical and chemical behavior of memory and selector has been key for realizing successful technology commercialization.

III. 3D XPoint Technology

3D XPoint is the only cross point technology, to date, that has achieved high-volume production and is discussed as an example of cross point technology. 3D XPoint, based on atomistic storage, is an entirely new technology category. 3D XPoint consists of three key technological elements: 3D memory in backend metallization, with CMOS under the array; two terminal thin film selector and memory layers; and bulk (3D) atomistic nonvolatile memory switching mechanism (Fig. 3, 4). Each memory deck is formed by two standard double patterning steps and subtractive etching, facilitating process simplicity and low cost. Wordlines and bitlines are formed in low resistivity conductors, enabling low RC for low latency. The selector is based on an ovonic threshold switch [8]. Symmetrical blocking of the amorphous alloy provides for robust isolation of individual memory cells in cross point arrays and large sub-array tiles for low cost. Carbon electrodes provide diffusion barriers between active layers and ease process integration.

While the 3D XPoint memory element shares a chalcogenide-based genealogy with phase change memory (PCM) [9], it does not rely on any sub-lithographic interfacial heating / switching structures, such as mushroom, μ-trench, pore, cross spacer, or ring shapes [10-14]. Additionally, in departure from chalcogenide shared genealogy, the entire bulk of the memory material of 3D XPoint undergoes bonding transformation between high (fully crystalline) and low (fully amorphous) conductivity states with no partial dome switching (Fig. 5). This ensures tight statistical control of states (Fig. 6). In PCM, the set operation follows a classic stochastics crystallization growth process initiated at amorphous crystalline interface [15]. Due to the entirety of the bulk of the memory element undergoing atomistic (phase) transformation, the interface of such phases does not exist in the final states of 3D XPoint. In addition, 3D XPoint technology exploits deterministic write algorithms; no program verification is required. These fundamental architectural advantages enable low latency, high bandwidth, and high endurance non-volatile memory subsystems.

First generation 3D XPoint is 128Gbit die of 20nm half pitch. The die consists of 8K tiles, each tile has 2K wordlines, 4K bitlines, and 2 deck cross point array (Fig. 7). Each tile selects a single bit for read or write with direct write capability. The cross point architecture and low latency materials and interconnects enable the attributes shown in Table 1 of high capacity, low latency, direct write, and non-volatility. Density per die are significantly larger than DRAM; more comparable to NAND. Access size of 16 bytes per bank and multi independent read/write banks per chip enable memory compatible cache line load/store operation and efficient die to configure into low latency high performant systems. While read and write latencies are asymmetric (non-volatile memories require write energy barriers to be overcome), both are orders of magnitude faster than NAND, approaching DRAM latencies.

IV. System Level View of Cross Point Technology

Unique attributes of 3D XPoint enable the creation of new classes of computing subsystems. Intel® Optane™ SSDs [16]

and Intel® Optane™ persistent memory modules (PMEM) [17] are described, as examples, to illustrate how capabilities of 3D XPoint are exploited to fill load/store memory capacity and storage performance gaps in computing (Fig. 8). Optane SSDs deliver performance per capacity, latency, and responsiveness under load not available with NAND SSDs. PMEM introduces an entirely new form of memory to computing, addressable like DRAM, but with larger capacities and persistence through power removal. Much like the memory on which they are built, these new subsystems blur the line between memory and storage.

Optane SSDs make 3D XPoint media available as operating system managed storage, connected as standard PCIe/NVMe SSDs. The SSDs are architected to deliver the value of 3D XPoint. The SSD controller features hardware only read/write path and spreads each 4kB data unit across multiple 3D XPoint die and across multiple banks within a die allowing reads to be returned in approximately 10 microseconds (idle latency) through the PCIe link. 3D XPoint direct write capability allows writes to complete quickly, in-place (no block erase), and at the native size of the operation (multiples of 4kBytes) eliminating the need for garbage collection and enabling a highly truncated latency tail. Latencies are more than an order of magnitude faster than NAND based SSDs (Fig. 9). With much higher endurance media and no write amplification, 3D XPoint SSDs deliver endurance much greater than NAND SSDs. Used as data caches or data tier, 3D XPoint absorb writes, enabling NAND SSDs to hold the bulk data, saving cost. Finally, Optane SSDs enable fast paging systems, extending main memory in a way NAND cannot.

PMEM makes 3D XPoint visible as system memory, providing the long dreamt of persistent, byte addressable, capacious memory. Connected as system memory with low latency cache coherent processor interconnects, these modules contain a controller featuring a low latency hardware read/write path to multiple 3D XPoint dies, providing average idle latency of 340ns (Fig. 10). First generation PMEM with capacities up to 512GB per module, more than double the maximum density of current DRAM DIMMs, enable 4.5TB per CPU socket. At the system level, PMEM can be managed either as a memory cached in DRAM, or as a persistent memory in its own address space distinct from DRAM. As a very large memory it enables more virtual machines to be hosted per physical CPU, avoiding time consuming swapping in and out of storage. As large persistent memory it enables very large databases to be held in memory and used immediately at bootup without time consuming storage load.

PMEM is the subject of intensive experimentation and development by many software developers, resulting in new topologies of computing systems with PMEM. Distributed Asynchronous Object Store (DAOS) is optimized for object store with PMEM [18]. In DAOS, the metadata storage pool (metadata, low-latency I/Os & indexing/query) are held in PMEM, while bulk data object store is held in NVMe SSDs. DAOS demonstrates significant performance increases over non-PMEM storage systems, setting world records in the IO500 Challenge [19]. In Oracle Exadata X8M, the storage file system is bypassed by remote direct memory access to PMEM on the storage server, achieving 10x latency reduction and >2.4x IOPs gain vs. X8 without PMEM [20].

V. Future Technology and System Directions

3D XPoint is beginning its scaling journey, in both capacity and performance. In 2nd generation, number of decks increases from two to four (Fig. 11), increasing capacity by 2x to 256Gbit per die, and increases bank count by 2x per die improving bandwidth. Future lateral scaling and deck count increases will continue to increase capacity, decreasing cost per bit. Material tuning and circuit design optimization will decrease latencies and increase bandwidth. A sign of a truly disruptive technology success is when it begets new industry architectures and solutions, such as what we are seeing with system innovations with PMEM noted above. The 3D XPoint capacity, cost, and performance scaling capability outlined ensures those new system innovations are just beginning.

VI. Conclusions

Cross point memories, as evidenced by 3D XPoint and Optane SSDs and PMEM, are proving ideal at filling the memory capacity-cost and storage performance gaps. System innovations exploiting 3D XPoint capabilities, blurring historical notions of separate semantics of memory and storage, are significantly boosting system performance.

References:

[1] "3D XPoint™: A Breakthrough in Non-Volatile Memory Technology", viewed 28 August 2020, https://www.intel.com/content/www/us/en/architecture-and-technology/intel-micron-3d-xpoint-webcast.html

[2] M. Lee, et. al., "2-stack 1D-1R Cross-point Structure with Oxide Diodes as Switch Elements for High Density Resistance RAM Applications", IEEE IEDM, 2007

[3] D. Kau, et. al., "A stackable cross point phase change memory", IEEE IEDM, 2009

[4] T. Kim, et. al., "High-performance, cost-effective 2znm two-deck cross-point memory integrated by self-align scheme for 128 Gb SCM", IEEE IEDM, 2018

[5] S Kim, et. al., "Breakthrough of Selector Technology for cross-point 25-nm ReRAM", IEEE IEDM, 2017

[6] S. Yasuda, et. al., "A Cross Point Cu-ReRAM with a Novel OTS Selector for Storage Class Memory Applications", IEEE VLSI, 2017

[7] S. Ambrogio, et. al., "Understanding switching variability and random telegraph noise in resistive RAM", IEEE IEDM, 2013

[8] D. Adler, et. al., "The mechanism of threshold switching in amorphous alloys", Review of Modern Physics, Vol. 50, No.2, 1978.

[9] H.-S. Philip Wong, et. al., "Phase Change Memory", Vol. 98, No. 12, Proc. of IEEE

[10] Y.H Ha, et. al., "An Edge Contact Type Cell for Phase Change RAM Featuring Very Low Power Consumption", IEEE VLSI, 2003

[11] F. Pellizzer, et. al., "Novel μTrench Phase-Change Memory Cell for Embedded and Stand-Alone Non-Volatile Memory Applications", IEEE VLSI, 2004

[12] M. Breitwisch, et. al., "Novel Lithography-Independent Pore Phase Change Memory", IEEE VLSI, 2007

[13] W.S. Chen, et. al., "A Novel Cross-Spacer Phase Change Memory with Ultra-Small Lithography Independent Contact Area", IEEE IEDM, 2007

[14] J.H. Oh, et. al., "Full Integration of Highly Manufacturable 512Mb PRAM based on 90nm Technology", IEEE IEDM, 2006

[15] A. Sebastian, et. al., "Crystal growth within a phase change memory cell", Nature Communications, 7 July 2014, DOI: 10.1038/ncomms5314

[16] F. Hady, et. al., "Platform Storage Performance With 3D XPoint Technology", Proceedings of the IEEE Vol. 105, No. 9, September 2017

[17] J. Izraelevitz, et. al., "Basic Performance Measurements of the Intel Optane DC Persistent Memory Module", arXiv:1903.05714 [cs.DC]

[18] Z. Liang, et. al., "DAOS: A Scale-Out High Performance Storage Stack for Storage Class Memory", Supercomputing Frontiers 6th Asian Conference, SCFA 2020, Singapore, Proceedings

[19] Virtual Institute for I/O, viewed 17 August 2020, https://www.vi4io.org/io500/start

[20] J. Shi, "Exadata with Persistent Memory: An Epic Journey", SNIA Persistent Memory Summit, Jan 2020

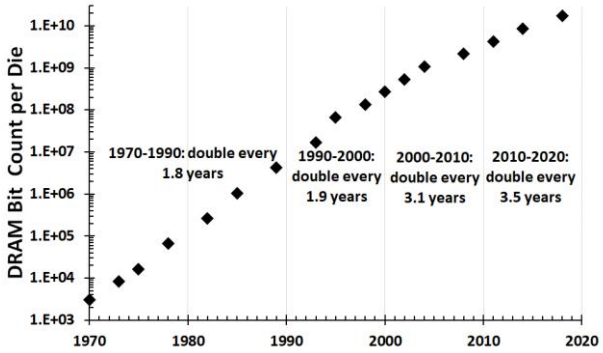

Figure 1: DRAM scaling over past 5 decades. Prior to 2000, DRAM scaling progressed at rate of doubling density in less than 2 years. Post 2000, DRAM scaling has slowed with rate in last decade slowing to doubling every 3.5 years.

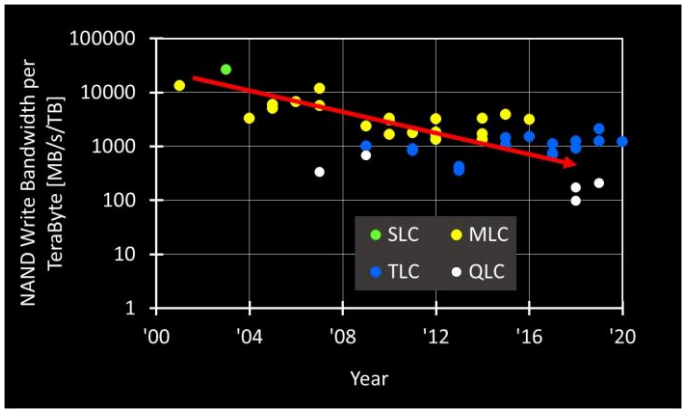

Figure 2: While capacity has grown for NAND, performance – write performance and in particular as more bits per cell are added – has not kept up with capacity growth. Data source: various IEEE, ISSCC, IEDM, and IMW papers.

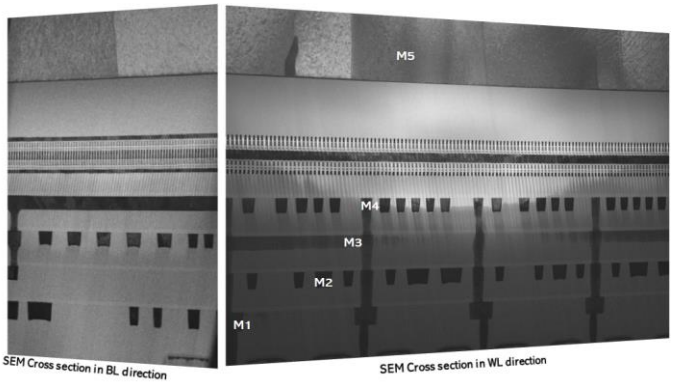

Figure 3: 2-deck 3D Xpoint tight pitch array residing above 4-level metal CMOS, where CMOS under the array contains cross point memory wordline / bitline decoders, sense amplifiers, control logic, etc.

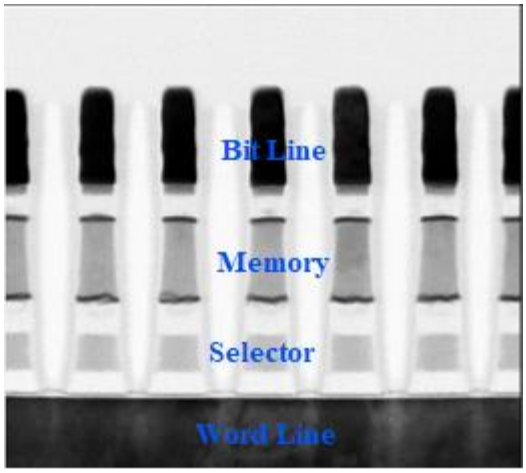

Figure 4: Single deck cell cross-section, with wordline and bitline, selector and memory materials, as noted. Each are separated by carbon electrodes (white colored region).

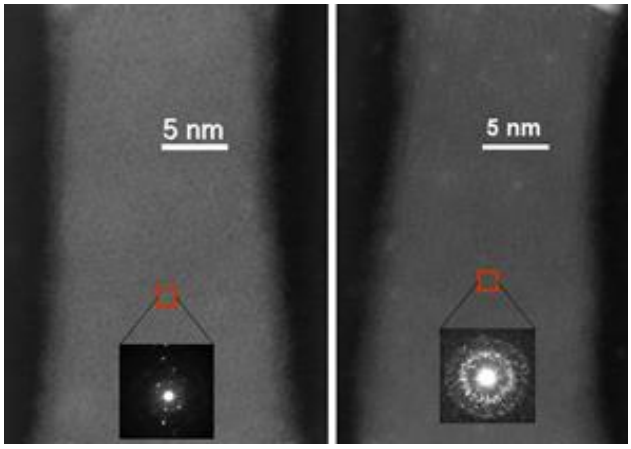

Figure 5: TEM cross section showing entire cell in fully crystalline and fully amorphous in SET (left), RESET (right) states.

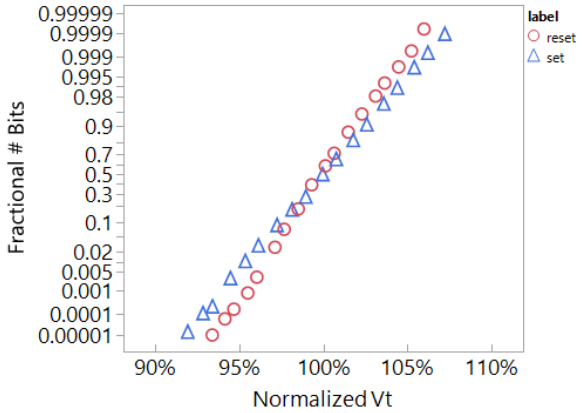

Figure 6: Distribution of threshold voltages of 3D Xpoint SET and RESET states. Tight distributions results from 3D volumetric switching between fully crystalline and fully amorphous states.

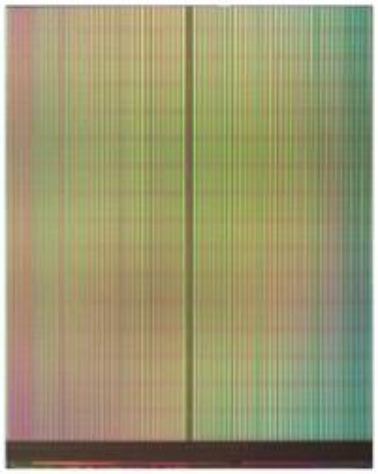

Figure 7: Die photograph of 128Gbit 3D XPoint

1st Generation 3D Xpoint Media Attributes	
Attribute	Value
Cell Feature Size (half pitch)	20nm
# Decks	2
Die Capacity	128 Gbits
Bank Access Size	16 Bytes
# Independent Banks	16
Read Latency per Bank	100nSec
Write Latency per Bank	500nSec
Write Bandwidth (MB/s) per TeraByte (cf. NAND, Fig. 2)	>35,000MByte/sec/TB

Table 1: Key attributes of 1st generation 3D XPoint media

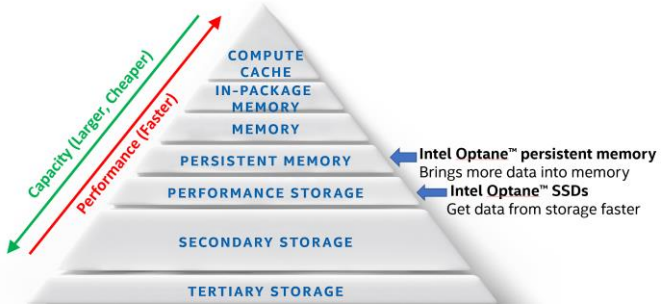

Figure 8: Memory hierarchy pyramid showing where Optane SSD and PMEM, both based on 3D Xpoint: a cross point memory technology, fits between DRAM memory and NAND secondary storage, with regards to performance (latency and bandwidth) and capacity (capacity and cost per bit).

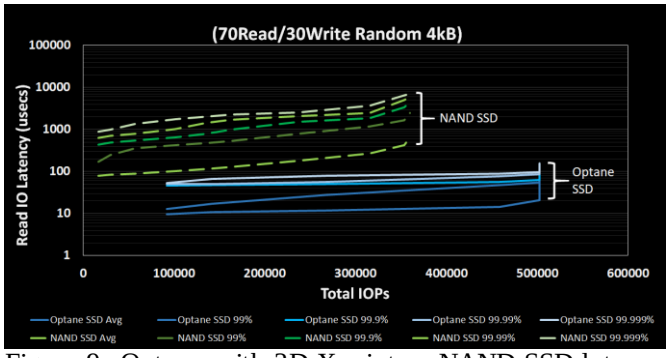

Figure 9: Optane with 3D Xpoint vs NAND SSD latency tails. 3D XPoint based SSDs achieve latency more than an order of magnitude faster than NAND SSD, with flat and tight quality of service response, from average through 99.999% latency, over full IOPs load range.

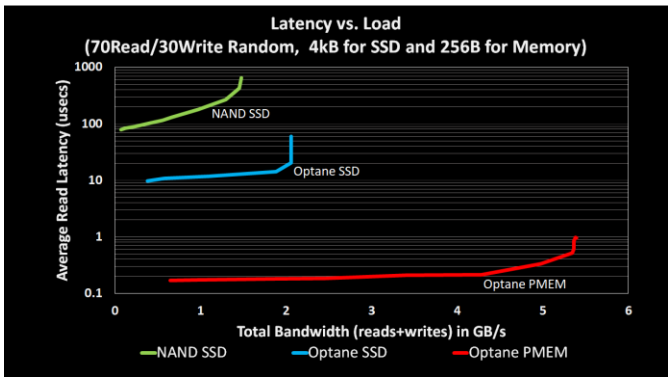

Figure 10: Latency at load comparisons: NAND SSD vs 1st generation Optane SSD and PMEM, both of which based on 1st generation 3D XPoint.

Figure 11: Cross Section of 2nd generation 3D XPoint with 4-deck array. Die capacity and bank count each doubling.