Flash Memory Summit

Persistent Memories: Markets and Applications 2019

Mark Webb

MKW Ventures Consulting, LLC

mark@mkwventures.com

Contents

- Persistent Memory Definitions
- Applications and what is shipping today
- Technologies and memory configurations
- Revenue projections and forecasts

Flash Memory Summit

What is Persistent Memory

- It's a universal Non-Volatile Memory Technology (Device Geeks)
  - PCM, ReRAM, MRAM, NRAM, FRAM, Memristor, NVRAM
- It's a storage/memory concept (Storage Experts)
  - What if we wrote to address and didn't have to worry about data loss or storage later?
- Its BIG DATA Memory (End users)
  - I want to look at all my TBs of data like hot data

A Persistent Memory Definition (Updated)

- It's persistent … No need to worry about loss
- It's accessed like memory on memory bus
  - "Byte addressable" …. Not block mode
  - Anything can be virtual memory… but this is less interesting
- Speed: system less than 1us latency (I can do storage at 6-10us)
  - Raw memory read latency on order of 100ns
- Endurance "good enough" to meet needs required by application
  - ALL NVM have endurance issues. NONE can be cycled 10^8 times in real world
- Used for data being worked on and addressed by programs. Not primarily used as cold or warm storage

How is PM Accessed

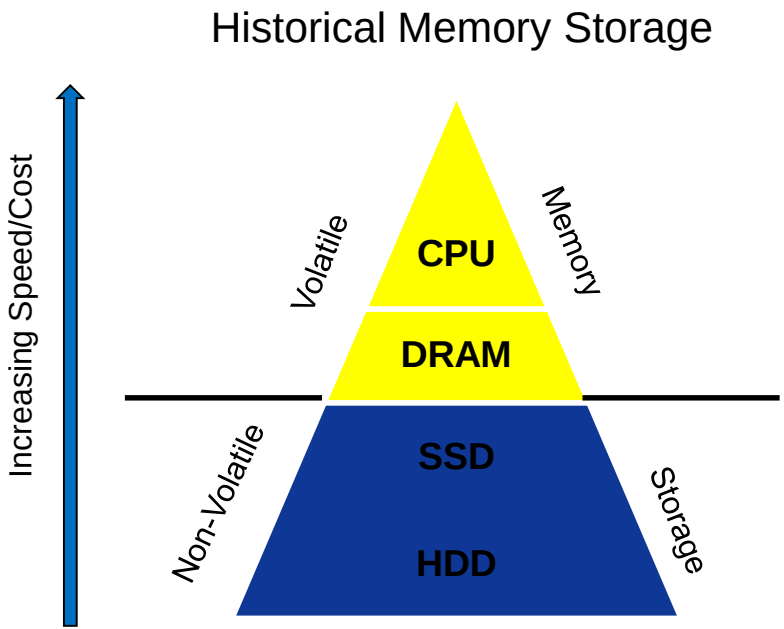
Interesting:

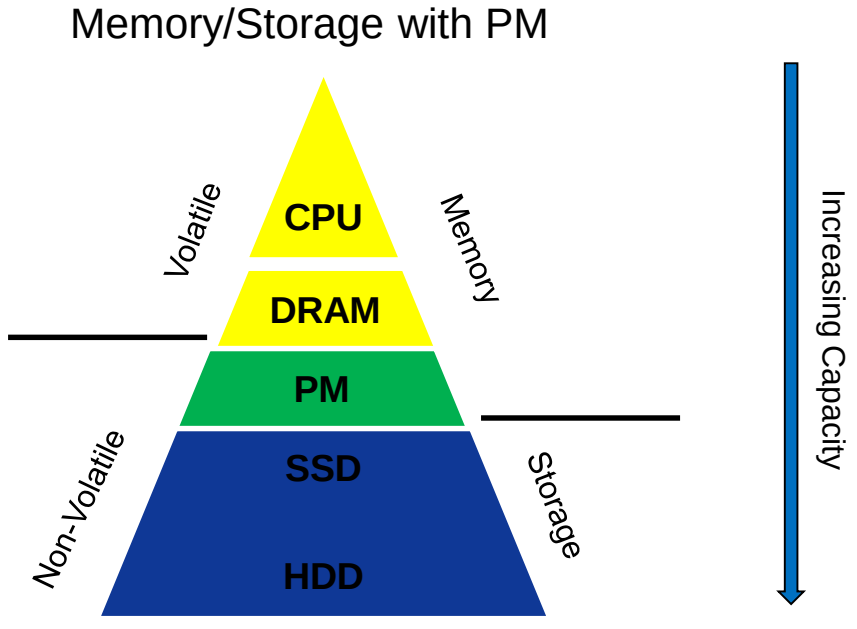
- Like DRAM: DDR bus. Parallel memory slots on server/PC board (Today). NVDIMM-N, NVDIMM-P or non-standard DDR4
- On New Bus: GenZ, OpenCAPI, CCIX, CXL
- Intel Optane PM DIMMS App Direct mode.

Less Interesting:

- *Intel Optane Persistent Memory in memory mode (Cached,not PM)*
- *Through NVMe/Storage bus: This is available today working with different memories but it is not my focus*
- *Block access (like an SSD on DRAM Bus)*

Historical Memory/Storage vs PM

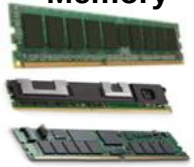




Mark Webb, MKW Ventures

How to work with 1TB of Data

OVERSIMPLIFIED!

196GB DRAM+8TB HDD	196G DRAM+1TB NVMe SSD+ 8TB HDD	1.5TB Persistent Memory+ 8TB HDD
Memory 	Memory 	Memory 
Storage 	Storage 	Storage 

Mark Webb, MKW Ventures

How to work with 1TB of Data

OVERSIMPLIFIED!

196GB DRAM+8TB HDD

- Data is on HDD
- Load part of it in to Memory
- Swap out blocks of data as needed until done
- Memory access times 30ns
- HDD access time mS
- Hope no power lost during work

196G DRAM+1TB NVMe SSD+ 8TB HDD

- Data is on HDD
- Load it all to SSD
- Load part of it in to Memory
- Swap out blocks of data with SSD until done
- Perhaps treat SSD as memory
- Memory access times 30ns
- SSD access time 10uS
- Hope no power lost during work

1.5TB Persistent Memory+ whatever

- Data is on HDD/SSD
- Load it all to persistent memory
- Complete work on data
- Leave it there or store it to SSD/HDD
- Memory access time is 30-350ns
- No SSD/HDD access needed
- If power lost, you are good

Mark Webb, MKW Ventures

Persistent Memory Applications … It's Here Today

- Server DIMMS/Main Memory for systems shipping NOW
- RAM requirements where max speed is needed and memory cannot be lost due to outage (NVDIMM-N)
- Log file, journaling, networks, fast restart requirements
- Applications with long processing times, Modeling
- Where quick recovery/reboot of server needed
- Financial transaction processing
- Still relatively low volume and penetration (<5% of servers)
  - Intel pushing VERY HARD to change this

What's Shipping Today

- NVDIMM-N is "classic" version of persistent memory DIMM
  - Addressed just like DRAM in a DIMM
  - Backed to NAND periodically or when power lost
  - Typical NVDIMM is 16G DRAM plus 32G of SLC NAND with control and capacitor/battery
  - Appears as 16GB of DRAM at DRAM speed
  - Downside: >2x the cost of DRAM. Limited to DRAM Density
- NEW: Intel Optane Memory (Finally!!)
  - App direct mode is the "classic" PM we want! Up to 512GB DIMMS
  - Memory mode not persistent so not focus, Same with block modes

Flash Memory Summit

Intel DC Persistent Memory

- After many launches, it is finally available.
- 128, 256, 512GByte DIMMS
- Speed is listed at 350ns Read Latency, Write is “higher”
  - Slower than DRAM, much faster than NAND
- No cycling limit in applications (a nice surprise)
- Ability to support TBs of addressable memory… and its persistent
- Requires Cascade Lake CPU, but is supported on majority of SKUs.
- MSRP* price is quite high… similar or even higher than DRAM
  - Cost is actually 60% of the cost of DRAM when in full productions (See presentation)
- TBs of addressable memory that meets our definition of PM shipping to people who want it. Big change from 2018

High Density Server DIMMs

- Future Apps: Large databases where loading and swapping portions is not efficient.
- Anything where faster loading, faster analysis provides monetary return to pay for it
- Examples:
  - Financial database/transaction processing ($/mS metrics available)
  - VMs that are currently memory limited (10x more VMs/Server)
  - Video/entertainment/Animation (Large dataset)
  - Log files, In memory commit
  - Caching for Storage…. Faster and limits wearout of SSD.
- Simplified: All applications that ran PCIe/NVMe 3 years ago.

Persistent Memory Applications (MORE)

CE/Mobile Devices (Not NVDIMMS…. Fast NVM)

- Smaller density replacing Capacitor/battery backed DRAM, replacing SRAM/DRAM/Flash. CE device optimization
- For cost-speed reasons, these applications often optimize NAND and DRAM and HDD in gaming/CE systems
- Potential to create a memory system that is fast enough and allows less chips, faster overall speed, better reliability.

- For Many apps, lower density is OK enabling more media (memory types) options
  - 16M SRAM+1G DRAM+8G NAND could use MRAM for aspects.
  - 2G DRAM+16G NAND could go to ReRAM/PCM-3D Xpoint

Flash Memory Summit

Memory Types/Media

	Latency	Density	Cost	HVM ready
DRAM	*****	***	***	*****
NAND	*	*****	*****	*****
MRAM	*****	*	*	***
3DXP	***	****	****	****
ReRAM	***	****	****	**
Other	***	**	**	*

Combined Today

Alone or Combined In future

Notes: NOR/SRAM and low density Not in Included (Small), Low density FeRAM not included

Coming Persistent Memory/ SCM Technologies

- NVDIMM-N meets the specs but is very expensive and density <=DRAM
  - “classic” PM, 16GB-64GB, 2-3x the cost of pure DRAM
  - Shipping today, Market is estimated at $750M and growing 30-50% CAGR
- Optane Persistent Memory Media (Don’t call it 3D Xpoint)
  - We expect multiple competitors using crosspoint PCM technology
- ZNAND/Fast NAND: slower than DRAM, cycling limitations (good for SSDs)
- MRAM: Much more expensive than DRAM (But close on speed)
- ReRAM: Slower than DRAM, Cycling limitations (much like Optane)
- Nothing is replacing DRAM. Its combined with DRAM or used in applications where being 5-7x slower than DRAM is OK

Example of Cost Challenges

- 2019 estimated Cost (not price) per Bit (DRAM RDIMM=1x)
  - MRAM: 5x
  - NVDIMM-N: 1.5x
  - ReRam (today): 0.9x
  - 3D Xpoint (today): 0.6x
  - Fast SLC NAND (today): 0.15x
- DRAM+ReRAM/Optane/NAND is lower cost/bit, more capacity at “similar” performance

Mark Webb, MKW Ventures

Flash Memory Summit

DRAM/NVM Combinations

- Coming solutions are some DRAM merged with lots of NVM.
  - Lower cost, near DRAM performance, managed endurance
- 3D-Xpoint persistent memory combines DRAM DIMMs and 3D Xpoint DIMMs with processor/memory controller managing data
- Z-NAND and solutions from All NAND and NVDIMM vendors will use similar architecture
  - Cheaper than DRAM, Lots of memory, Managed endurance
- NVDIMM-P SHOULD Provide standard to allow all of these.
  - Delays are changing the story on what will dominate

Flash Memory Summit

Predictions for Market

- NVDIMM is established and very useful for high speed, low density (<64GB) DIMM applications at higher cost.
  - Bit growth is strong, revenue growth dependent on pricing
- Long delayed Intel Optane PM is shipping today with widespread Intel Support. Soon to be a Billion dollar market
- Last Years Comment: “”If we are having “what’s possible” discussions at end of 2019, Market will be much, much lower than middle revenue”
  - Thanks to NVDIMM-N and Intel, we are not having this discussion!!

Mark's Summary

- Persistent memory is here in NVDIMM-N, Optane DIMMS
- To grow, we need to be cost effective.
  - DRAM replacement by expensive tech won't work broadly
  - Memory that is too slow won't work broadly
  - Neither DRAM nor NAND are getting replaced in next 5 years
- DRAM + NVM will be the PM future (like NVDIMM-P)
  - Includes Optane Persistent Memory which requires DRAM
- Revenue could grow 35% CAGR if technologies deliver
  - New bus, NVDIMM-P or other alternate to Intel, NVDIMM cost reduction

Mark Webb, MKW Ventures