Persistent Memories: Markets and Applications 2019

Mark Webb

MKW Ventures Consulting, LLC

www.mkwventures.com

mark@mkwventures.com

Flash Memory Summit

Contents

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

Mark Webb, MKW Ventures

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

Mark Webb, MKW Ventures

How is PM Accessed

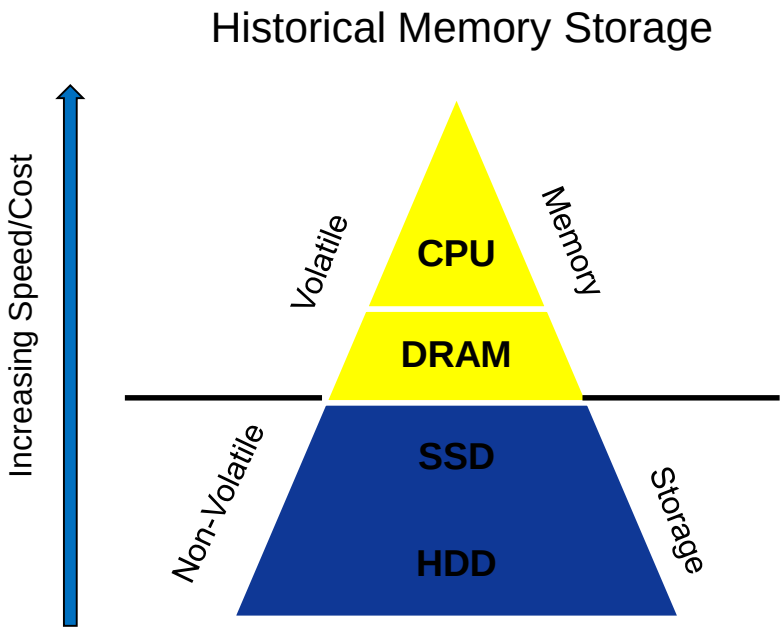
Interesting:

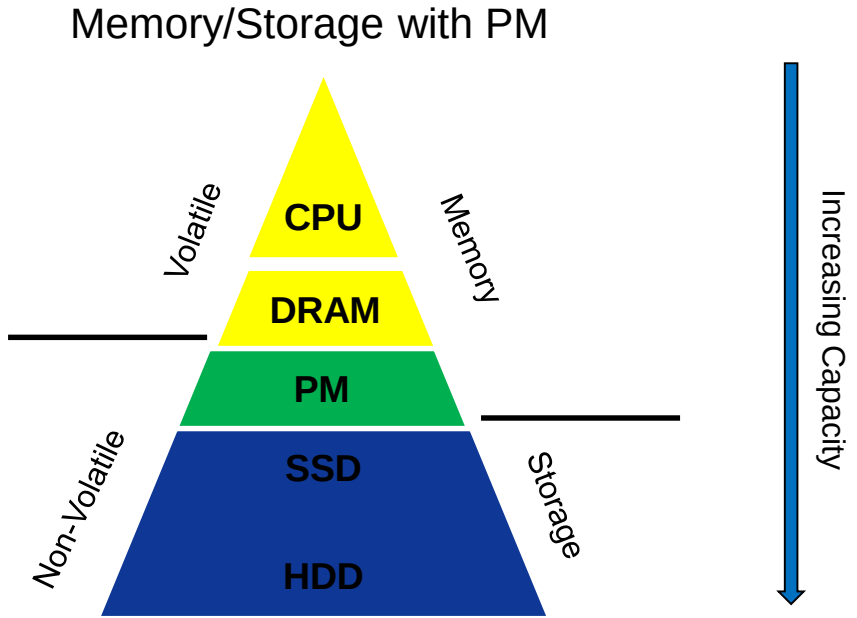
- Like DRAM: DDR bus. Parallel memory slots on server/PC board (Today). NVDIMM-N, NVDIMM-P or non-standard DDR4
- On New Bus: GenZ, OpenCAPI, CCIX, CXL
- Intel Optane PM DIMMS App Direct mode.

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

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)
  - Goal is 50% of servers by 2024

Mark Webb, MKW Ventures

Flash Memory Summit

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

Intel DC Persistent Memory

- After many launches, it is finally available.
- 128, 256, 512GByte DIMMS
- System 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

Mark Webb, MKW Ventures

High Density DIMM Applications

- 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)
  - AI: Low latency needed for fast look up and processing
  - VMs that are currently memory limited (10x more VMs/Server)
  - Video/entertainment/Animation (Large dataset). In Memory processing
  - Log files, In memory commit
  - Caching for Storage…. Faster and limits wearout of SSD.
- Simplified: All applications that ran PCIe/NVMe 3 years ago.

Mark Webb, MKW Ventures

Persistent Memory Applications (MORE)

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

- For Many apps, 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.

- 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

Mark Webb, MKW Ventures

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

Mark Webb, MKW Ventures

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
- PCM/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

Mark Webb, MKW Ventures

Flash Memory Summit

2019 Challenges

- While we now have Optane and NVDIMM-N, We have challenges to reach dream of Persistent Memory
  - Persistent Memory integration is hard. Delays in Optane DIMMs show challenges when compute system is controlled by Chipset owner.
  - New memory, slower RAM, balancing PM and DRAM
  - NVDIMM-P will hopefully do this on broader scale which is even more difficult. Implementation is not soon.
  - New memories show promise in low volume and samples but have yet to ramp (MRAM, ReRAM, Other PCM)
  - Other memories in research are 5+ years from revenue

Mark Webb, MKW Ventures

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.
  - If delayed, Alternatives like CXL would need to emerge
- Combination of memories is hard but provdes tradeoffs to prevent a niche market

Mark Webb, MKW Ventures

Predictions for Market

- NVDIMM-N 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 Webb, MKW Ventures

Persistent Memory

Revenue Growth "Guess-timate"

Year	Revenue Middle	Revenue High	Requirements to meet Middle
2020	$1.6B	2.4B	Optane, NVDIMM must takeoff ASAP
2022	$3.4B	$5.5B	Persistent memory multiple areas. NVDIMM-P Shipping, Multiple bus options evaluated
2024	$5B	$7B	Multiple new memories allow utilization in mobile, server, PCs

NOTES:
Numbers down 15-20% from 2018 due to Optane PM delays and NVDIMM-P delays
NVDIMM+SCM/NVRAM standalone memory only. Virtual memory on storage bus not included
NVDIMM could be DRAM+NAND, Fast NAND, SCM
Optane and NVDIMM-N dominate. "other options" are >500M 2024

Mark Webb, MKW Ventures

Flash Memory Summit

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