Architecting Phase Change Memory as a Scalable DRAM Alternative

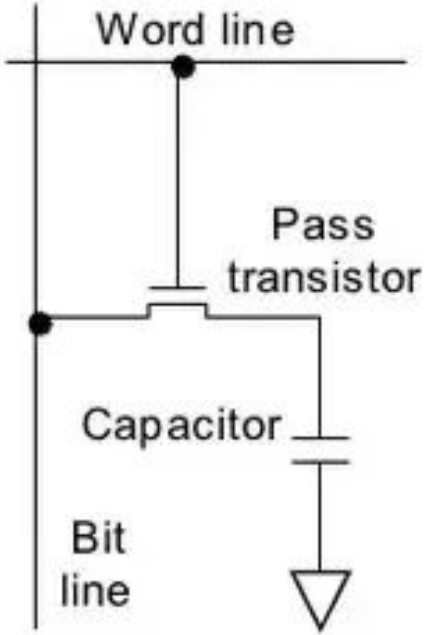
Benjamin C. Lee, Engin Ipek, Onur Mutlu, Doug Burger
ISCA 2009

Presented by Skanda Koppula
December 2018

Problem

Charge-based storage is hard to scale!

- Subthreshold charge leakage
- Neighbor interference and capacitive coupling (e.g. Rowhammer)
- *Capacitors*: large enough to store charge for reliable sensing
- *Access transistors*: large enough to exercise control over bit cell
- Manufacturability of <40nm DRAM was unknown

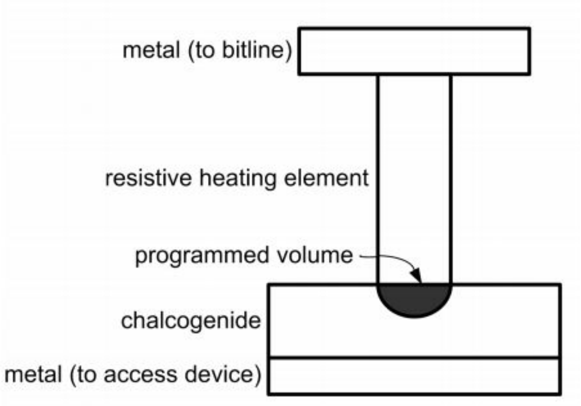


(a) DRAM cell

Image adapted from https://users.ece.cmu.edu/~omutlu/pub/understanding-latency-variation-in-DRAM-chips_sigmetrics16.pdf

PCM is an alternative memory technology

- Phase Change Memory (PCM) is a non-volatile resistive-based memory
- Each bit cell contains a deposit of chalcogenide glass (GST)
- GST is either a crystallized state (low resistance) or amorphous state (high resistance)



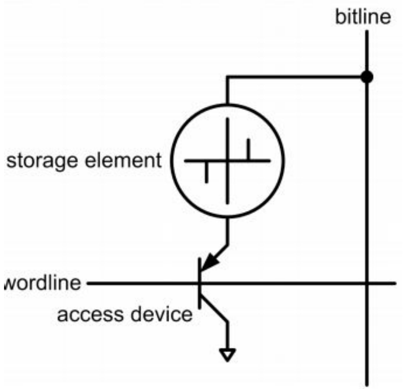


Image adapted from https://people.inf.ethz.ch/omutlu/pub/pcm_isca09.pdf

PCM is an alternative memory technology

- Phase Change Memory (PCM) is a non-volatile resistive-based memory
- Each bit cell contains a deposit of chalcogenide glass (GST)
- GST is either a crystallized state (low resistance) or amorphous state (high resistance)
- Application of current changes GST state

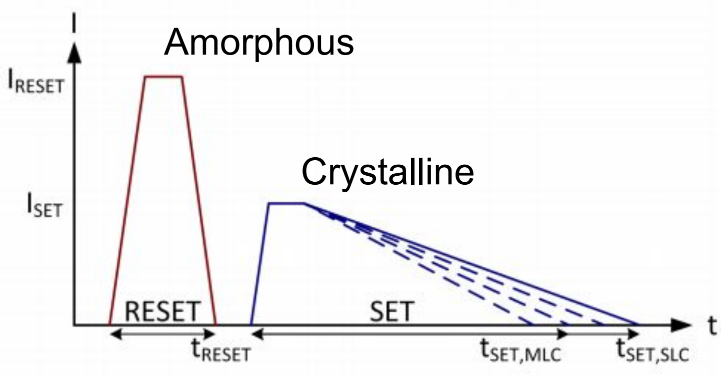


Image adapted from https://people.inf.ethz.ch/omutlu/pub/pcm_isca09.pdf

PCM is an alternative memory technology

- Phase Change Memory (PCM) is a non-volatile resistive-based memory
- Each bit cell contains a deposit of chalcogenide glass (GST)
- GST is either a crystallized state (low resistance) or amorphous state (high resistance)
- **Scalable with decreasing technology node**
- **But naive application of PCM consumes 2.2x more energy than DRAM and is 1.6x slower on SPEC benchmarks**

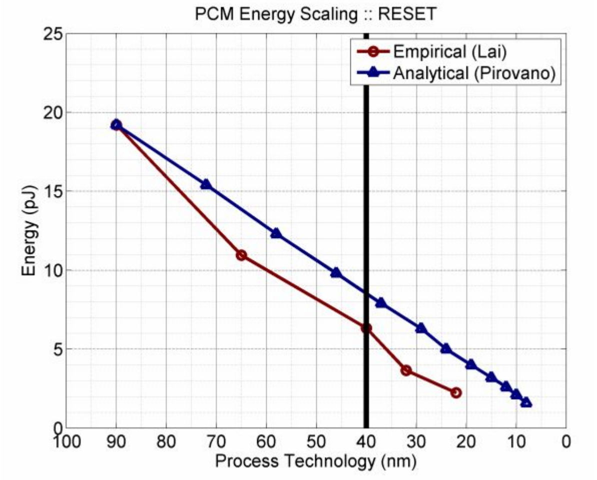


Image adapted from https://people.inf.ethz.ch/omutlu/pub/pcm_isca09.pdf

Key Ideas:

Phase Change Memory can be a competitive main memory solution if the system is architected with a PCM-optimized memory design

This paper contributes:

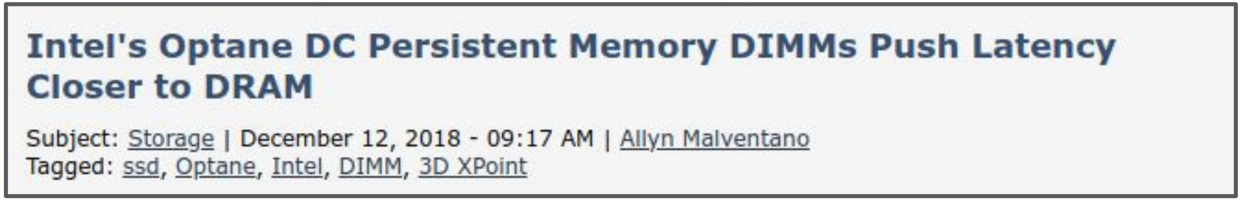
1. A study of row buffer reorganizations demonstrating:

   Narrow row buffers mitigate high energy writes, making PCM power comparable to DRAM

   Multiple row buffers allow write coalescing, reducing the PCM slowdown from 1.6x to 1.2x

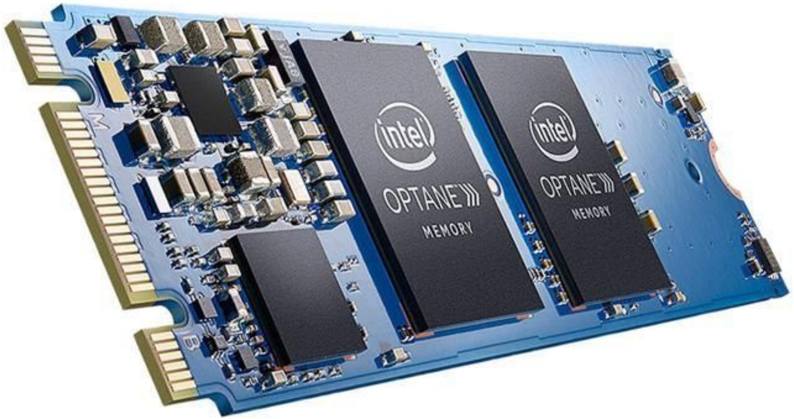
2. A proposal to track data modifications, and execute partial row writes
   - Extends PCM lifetime by four orders of magnitude

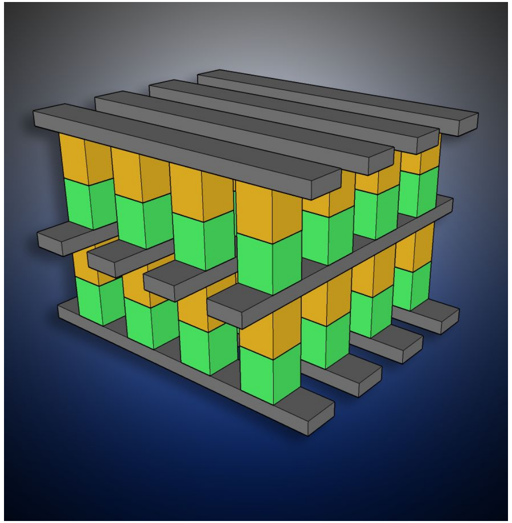
This is now!



Intel's Optane DC Persistent Memory DIMMs Push Latency Closer to DRAM

Subject: Storage | December 12, 2018 - 09:17 AM | Allyn Malventano
Tagged: ssd, Optane, Intel, DIMM, 3D XPoint

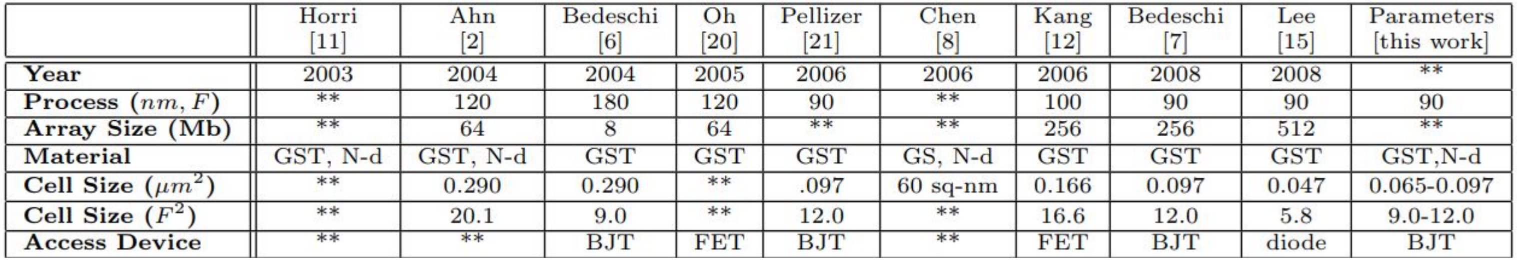




Images from https://www.pcper.com/news/Storage/Intels-Optane-DC-Persistent-Memory-DIMMs-Push-Latency-Closer-DRAM, https://en.wikipedia.org/wiki/3D_XPoint#/media/File:3D_XPoint.png

PCM Characteristics

- Survey of PCM Prototypes: 2003 - 2008

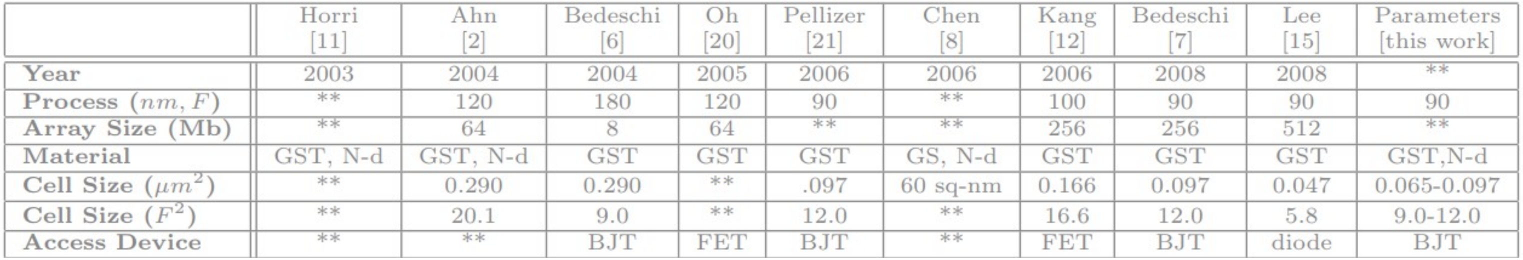


	Horri [11]	Ahn [2]	Bedeschi [6]	Oh [20]	Pellizer [21]	Chen [8]	Kang [12]	Bedeschi [7]	Lee [15]	Parameters [this work]
Year	2003	2004	2004	2005	2006	2006	2006	2008	2008	**
Process (nm, F)	**	120	180	120	90	**	100	90	90	90
Array Size (Mb)	**	64	8	64	**	**	256	256	512	**
Material	GST, N-d	GST, N-d	GST	GST	GST	GS, N-d	GST	GST	GST	GST,N-d
Cell Size (μm^2)	**	0.290	0.290	**	.097	60 sq-nm	0.166	0.097	0.047	0.065-0.097
Cell Size (F^2)	**	20.1	9.0	**	12.0	**	16.6	12.0	5.8	9.0-12.0
Access Device	**	**	BJT	FET	BJT	**	FET	BJT	diode	BJT

- Speculated future PCM settings in 2008:
    - GST bit cell material
    - BJT access device
    - 10^8 write cycles
    - 42ns read latency // 40uW read power
    - <100ns write latency // 480 uW write power
    - 9 - 12F^2 density using BJTs
    - 0.05W idle power

PCM Characteristics: Then and Now

- Survey of PCM Prototypes: 2003 - 2008



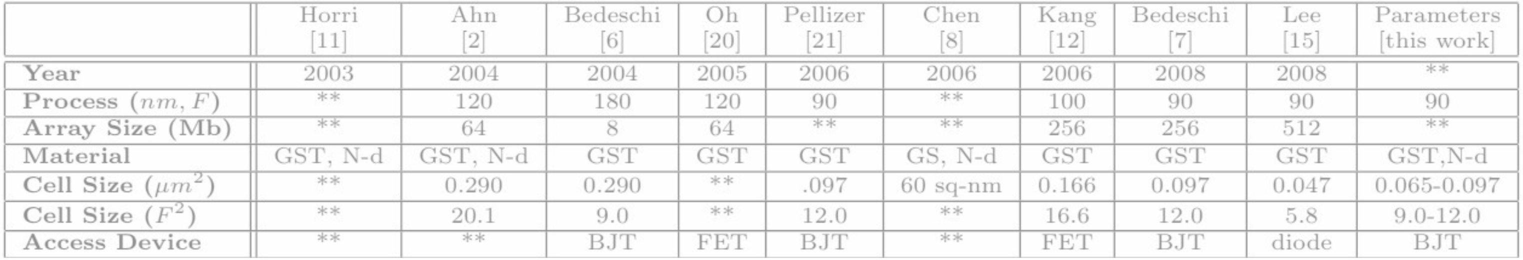
	Horri [11]	Ahn [2]	Bedeschi [6]	Oh [20]	Pellizer [21]	Chen [8]	Kang [12]	Bedeschi [7]	Lee [15]	Parameters [this work]
Year	2003	2004	2004	2005	2006	2006	2006	2008	2008	**
Process (nm, F)	**	120	180	120	90	**	100	90	90	90
Array Size (Mb)	**	64	8	64	**	**	256	256	512	**
Material	GST, N-d	GST, N-d	GST	GST	GST	GS, N-d	GST	GST	GST	GST,N-d
Cell Size (μm^2)	**	0.290	0.290	**	.097	60 sq-nm	0.166	0.097	0.047	0.065-0.097
Cell Size (F^2)	**	20.1	9.0	**	12.0	**	16.6	12.0	5.8	9.0-12.0
Access Device	**	**	BJT	FET	BJT	**	FET	BJT	diode	BJT

- Speculated future PCM settings in 2008:
  - GST bit cell material **(same as in 2017)**
  - BJT access device **(same as in 2017)**
  - 10^8 write cycles lifetime **(same as in 2017)**
  - 42ns read latency // 40uW read power **(same in 2017)**
  - <100ns write latency // 480 uW write power **(lower in 2017)**
  - 9 - $12F^2$ density using BJTs **(lower in 2017)**
  - 0.05W idle power **(same in 2017)**

Derived from Table 1 in Dynamic Adaptive Replacement Policy in Shared Last-Level Cache of DRAM/PCM Hybrid Memory for Big Data Storage

PCM Characteristics vs. DRAM

- Survey of PCM Prototypes: 2003 - 2008



	Horri [11]	Ahn [2]	Bedeschi [6]	Oh [20]	Pellizer [21]	Chen [8]	Kang [12]	Bedeschi [7]	Lee [15]	Parameters [this work]
Year	2003	2004	2004	2005	2006	2006	2006	2008	2008	**
Process (nm, F)	**	120	180	120	90	**	100	90	90	90
Array Size (Mb)	**	64	8	64	**	**	256	256	512	**
Material	GST, N-d	GST, N-d	GST	GST	GST	GS, N-d	GST	GST	GST	GST,N-d
Cell Size (μm^2)	**	0.290	0.290	**	.097	60 sq-nm	0.166	0.097	0.047	0.065-0.097
Cell Size (F^2)	**	20.1	9.0	**	12.0	**	16.6	12.0	5.8	9.0-12.0
Access Device	**	**	BJT	FET	BJT	**	FET	BJT	diode	BJT

- Speculated future PCM settings in 2008:
  - GST bit cell material (N/A in DRAM)
  - BJT access device (N/A in DRAM)
  - 10^8 write cycle lifetime **(10^8x lower than DRAM)**
  - <50ns read latency // 40uW read power **(5x higher than DRAM)**
  - <100ns write latency // 480 uW write power **(12x higher than DRAM)**
  - 9 - 12F^2 density using BJTs **(1.3x higher than DRAM)**
  - 0.05W idle power **(14x lower than DRAM)**

From https://people.inf.ethz.ch/omutlu/pub/lee_isca09_talk.pdf

Mechanisms

How do we fix we high read and write latencies?

How do we fix the high read and write power consumption?

Key PCM/DRAM difference:

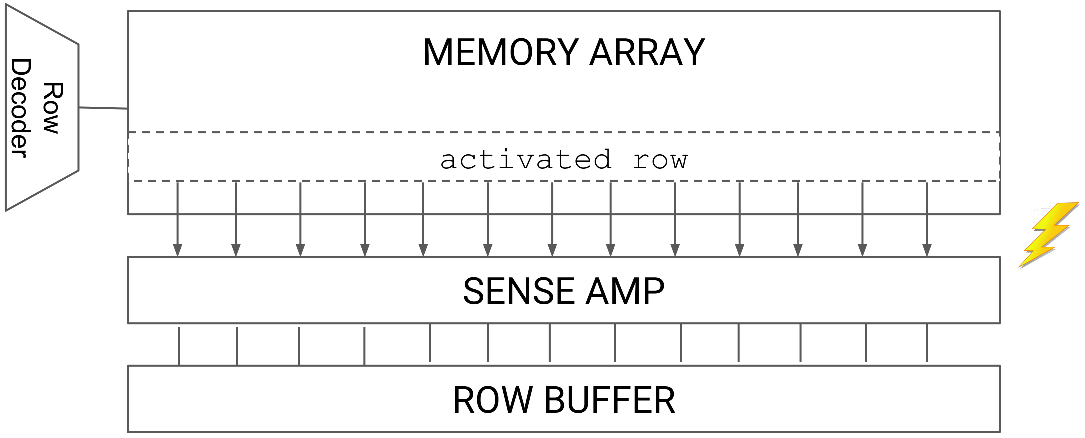
PCM row activation + read is non-destructive

⇒ **Narrow and multiple row buffers!**

Mechanisms

DRAM

Selected row's charge setting is lost and subsequently restored

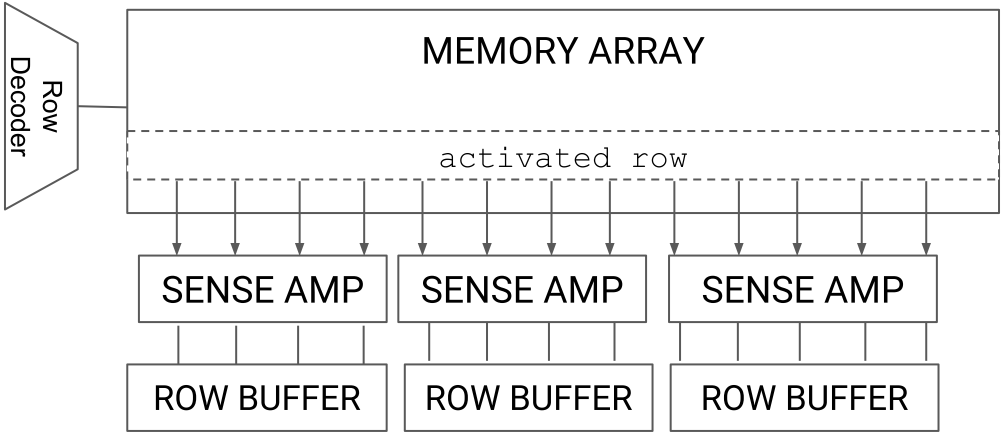


1-8KB

Mechanisms

PCM Proposal

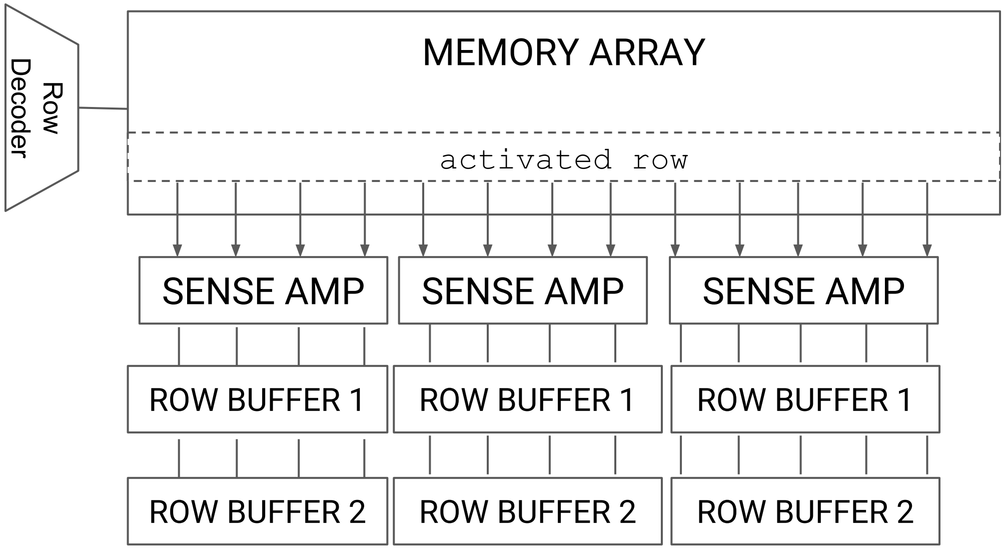
- No row restoration enables ‘Narrow Buffers’
    - Divide the length of the sense amp and row buffers by up to 32X to read/write only the pertinent row section



Mechanisms

PCM Proposal

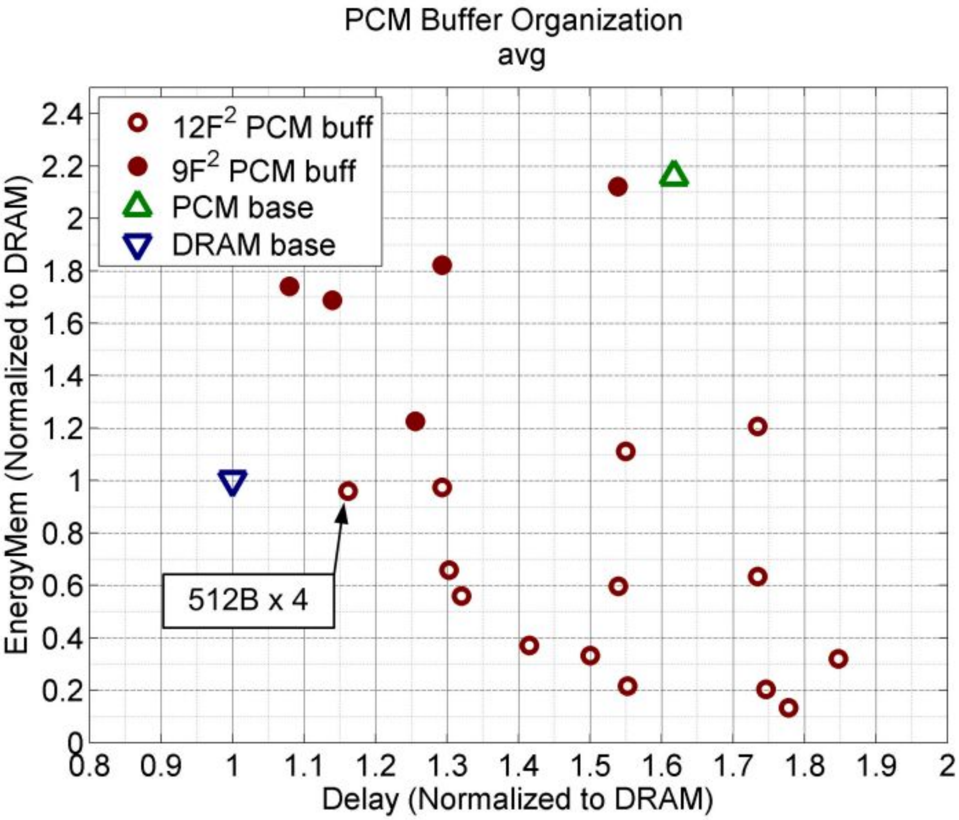
- No row restoration enables ‘Narrow Buffers’
- Multiple row buffers allow write coalescing (LRU eviction)
    - Reduce number of writes that hit the memory array: **better endurance!**



What does this buy us?

- Smaller sense-amps avoid fan-out area blow-up of wide sense-amps
- Partial row writes avoid large current/power waste for parts of row that are not modified
- Multiple row buffers increase write coalescing!

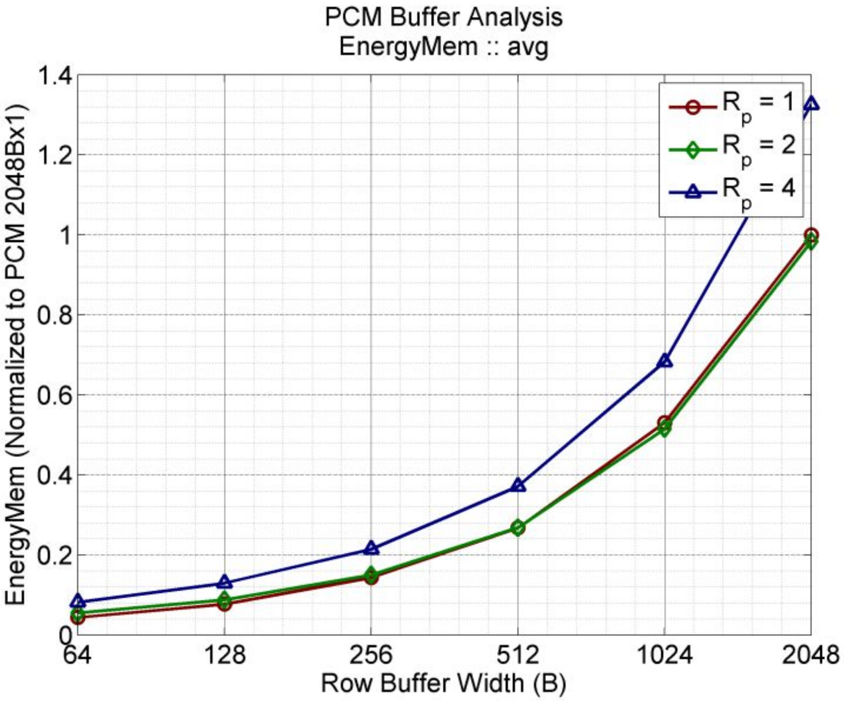
What does this buy us?



Each dot is a different buffer width/number of buffers combination

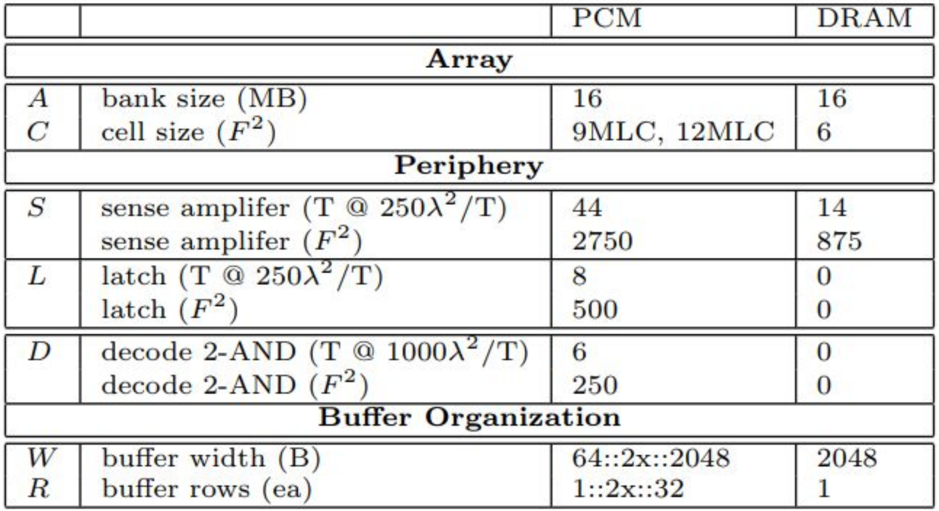
PCM base is (presumably) the undivided, nominal width.

What does this buy us?



What does this cost us?

- Need additional decoders to mux between multiple row buffers
- PCM uses current sense amp which consumes **more** area than the voltage sense amp of DRAM
- Latches to keep data in the multiple row buffers
- Narrow buffers reduce write-coalescing



		PCM	DRAM
	Array		
A	bank size (MB)	16	16
C	cell size (F^2)	9MLC, 12MLC	6
	Periphery		
S	sense amplifer (T @ $250\lambda^2$/T)	44	14
	sense amplifer (F^2)	2750	875
L	latch (T @ $250\lambda^2$/T)	8	0
	latch (F^2)	500	0
D	decode 2-AND (T @ $1000\lambda^2$/T)	6	0
	decode 2-AND (F^2)	250	0
	Buffer Organization		
W	buffer width (B)	64::2x::2048	2048
R	buffer rows (ea)	1::2x::32	1

Mechanism

- Buffer reorganization improves PCM energy consumption
- Buffer reorganization improves PCM application slowdown

- **Can we further improve the lifetime and write endurance problem?**
  - **Yes, if we reduce the number of writes even more!**

Mechanism: Partial Writes

Reduce number of writes to PCM array by **tracking dirty data from caches**

During eviction, **only dirty words are written back**

Trade-off:

‘Modest’ increase in cache state (3.1% extra cache bits) to reduce writes

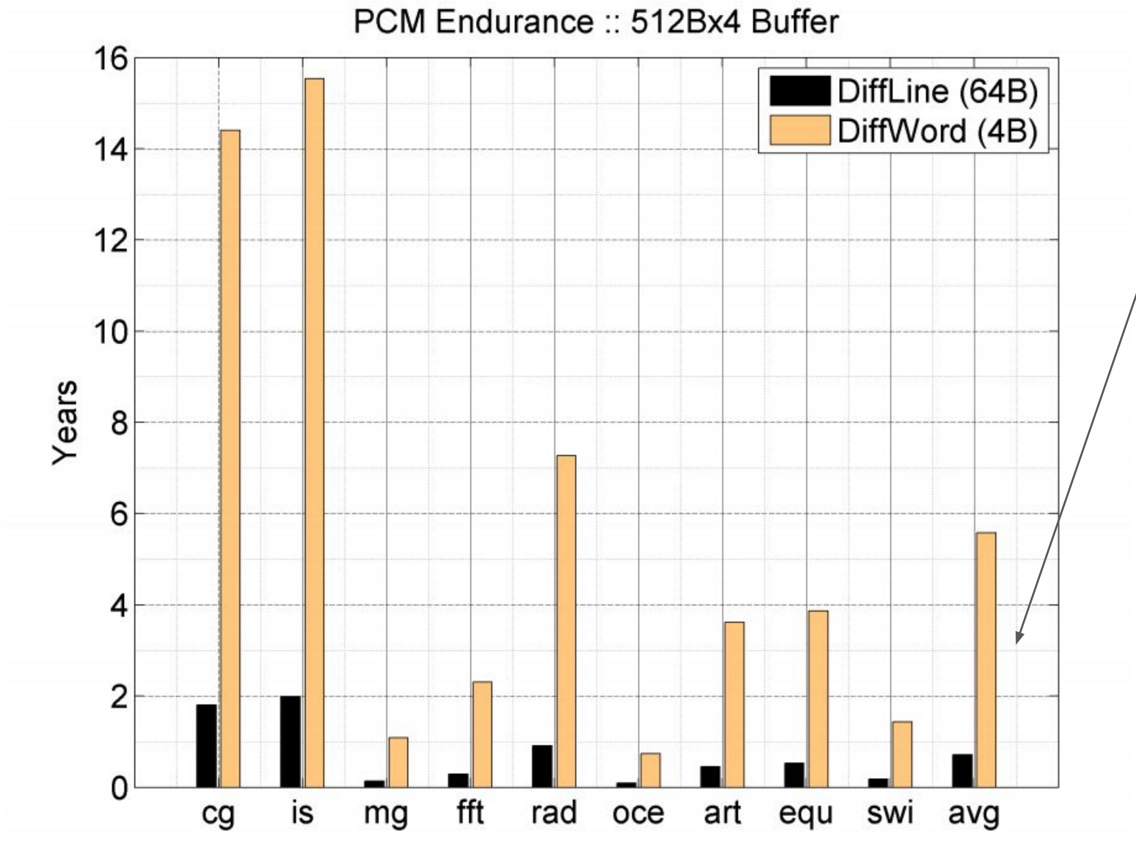
Evaluation: Partial Writes

Uses an analytical model to estimate number of PCM array writes. Factors in:

- Different application write intensity
- Buffer organizations
- Dirty-tracking granularities

Nominal lifetime running the benchmarks 24/7 is 525 hours ~ 0.05 years

Evaluation: Partial Writes



Improves endurance to:
0.7 years (cacheline gran.)
5.6 years (word granularity)

Reduced writes to PCM array by 41% (cacheline) and 92% (word granularity)

Strengths

- **Demonstrates that PCM has the potential to be a viable memory solution**
    - Highly prescient of future trends!
- Presents two very effective techniques to reduce energy, latency, and write wear of PCM memories: buffer reorganization and partial writes
- Does all of this in a time when physical PCM modules were not widespread or very available
- Good exposition of how PCM works

Weaknesses

- Baselines are weak
  - Uses DDR2-800 as a baseline (2008 paper)
  - No comparison with equivalently optimized DRAM (e.g. multiple row-buffers)
- DRAM energy analysis is analytical, based on MICRON datasheets and technical notes
  - VAMPIRE: “Don’t always trust the datasheets: actual consumption may be lower”
- Endurance estimates are based on self-created approximate analytical model, no evaluation on how this aligns with empirical data
- Exposition of core idea can be improved
  - Some graph axis mislabeled and other labels missing (e.g. Figure 8)
  - Details missing on evaluation experiments

Questions?
(before proposals and discussion starters)

Architecting Phase Change Memory as a Scalable DRAM Alternative

Benjamin C. Lee, Engin Ipek, Onur Mutlu, Doug Burger

Presented by Skanda Koppula
December 2018

Questions and Proposals

Can we extend the characterization?

- How does DDR4/LPDDR4/NAND Flash compare to PCM today?
- How valid are the energy simulation results and endurance models to commercially available PCM modules?

Questions and Proposals

Can we extend the characterization?

- How does DDR4/LPDDR4/NAND Flash compare to PCM today?
- How valid are the energy simulation results and endurance models to commercially available PCM modules?

Does PCM introduce new security threads?

- How sensitive is the read and write process in PCM to external temperature variation?
  - Are RowHammer like attacks possible in PCM by playing with external temperatures or targeted heating of chip area?
- What are current wear-leveling algorithms for PCM?
  - Do manage at fine enough granularity to prevent targeted attacks that induce early aging in specific memory sections?

Questions and Proposals

How can we combine memory architectures?

- Hybrid memory: how would we architect a PCM-DRAM-Flash-HBM-SRAM memory systems for current and future workloads?
    - What would the cache hierarchy look like? Buffer organization?
    - Do we have simulators for such systems?
    - What are the ideal partitioning and controller policies?

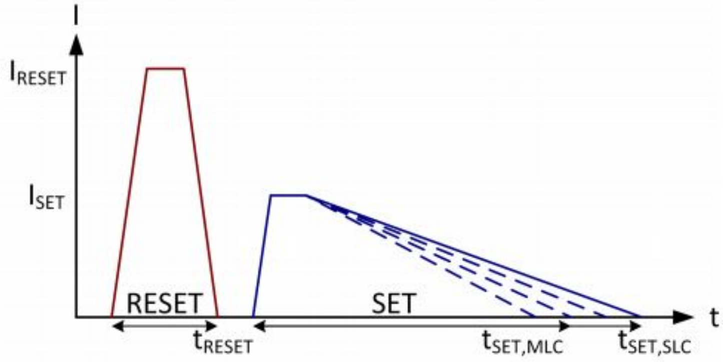
Questions and Proposals

How can we combine memory architectures?

- Hybrid memory: how would we architect a PCM-DRAM-Flash-HBM-SRAM memory systems for current and future workloads?
  - What would the cache hierarchy look like? Buffer organization?
  - Do we have simulators for such systems?
  - What are the ideal partitioning and controller policies?

Miscellaneous

- Is it possible to reduce the expensive SET/RESET voltages/latencies and compensate with ECC?


I
I_{RESET}
I_{SET}
RESET
SET
t
t_{RESET}
$t_{SET,MLC}$
$t_{SET,SLC}$