SNIA PERSISTENT MEMORY PM SUMMIT

JANUARY 23, 2020 | SANTA CLARA, CA

PM: Media, Attachment, and Usage

Dave Eggleston, Intuitive Cognition Consulting

I ❤ DRAM! (But...)

	2010	2020	units
Market	$40B (2010)	$100B (2018)	$B
Price	-37%	-8%	$/year
Capacity	>2x	<2x	Gb/chip/2 years
DIMM Bandwidth	+31%	+31%	GB/sec/year
Channels/CPU	2	8 (going to 16)	
DIMMs/Channel	4-8	2 (going to 0.5)	
Channels/DIMM	1	2	

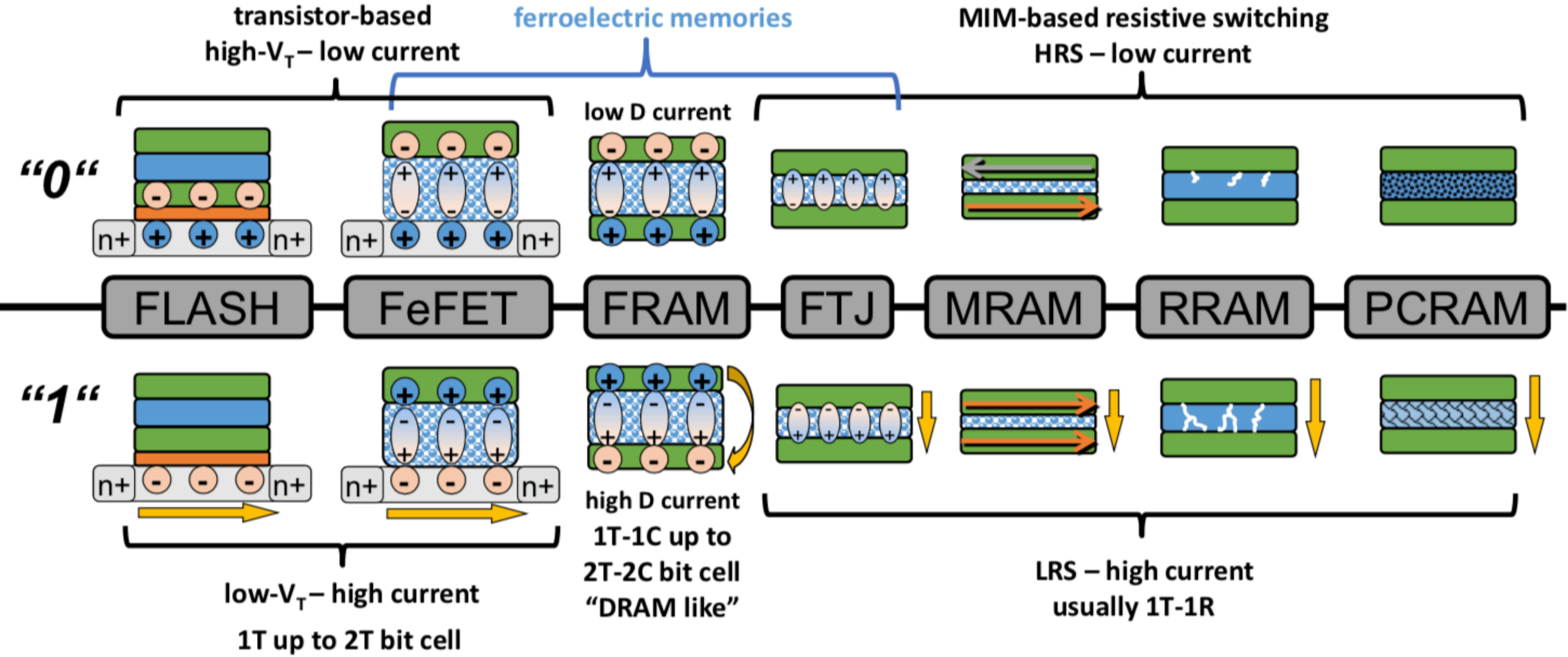
- DRAM tech and price scaling has slowed way down
- Running out of tricks to keep DDR bandwidth increasing
- Can PM do (part of) the DRAM job? More than that?

Data Source: J. Thomas Pawlowski, JTP, “Prospects for Memory”, SC’19 MCHPC

PM Media

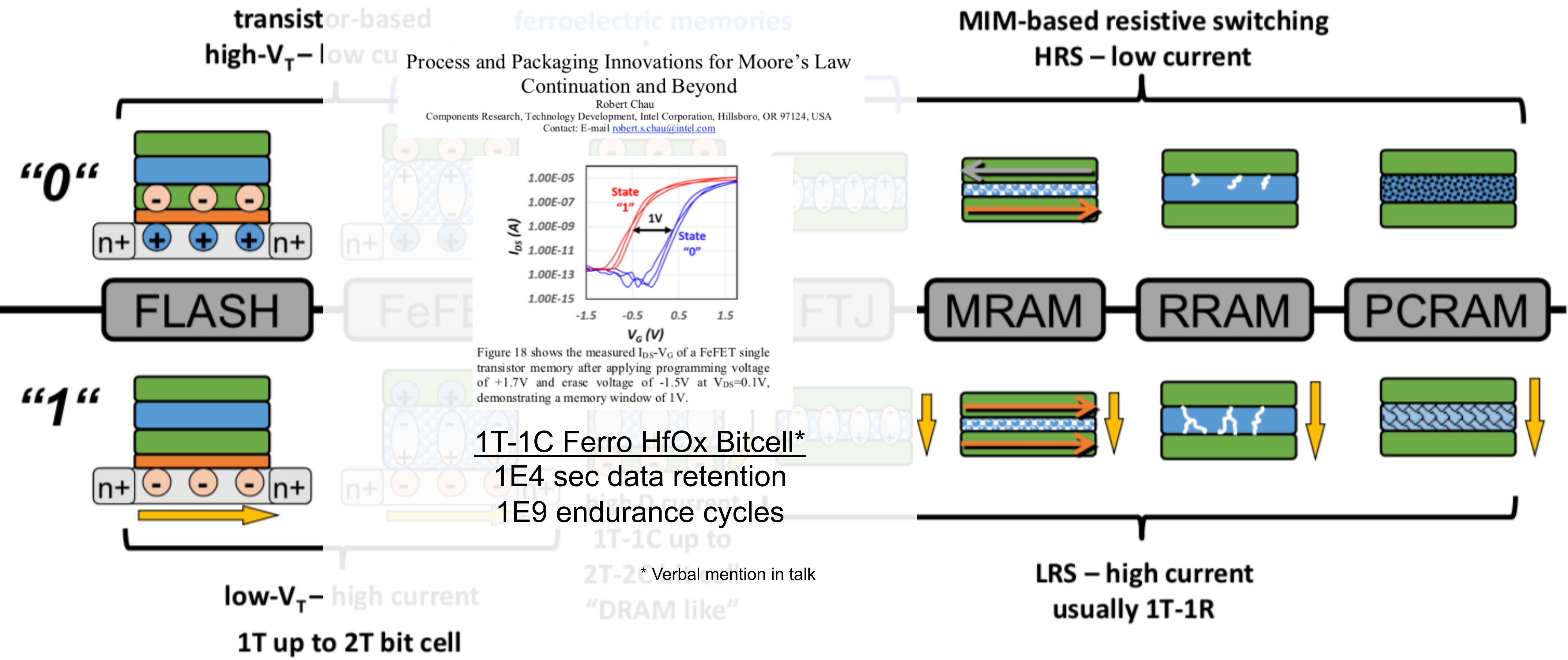
What, Why, and Who

PM Media: What Options



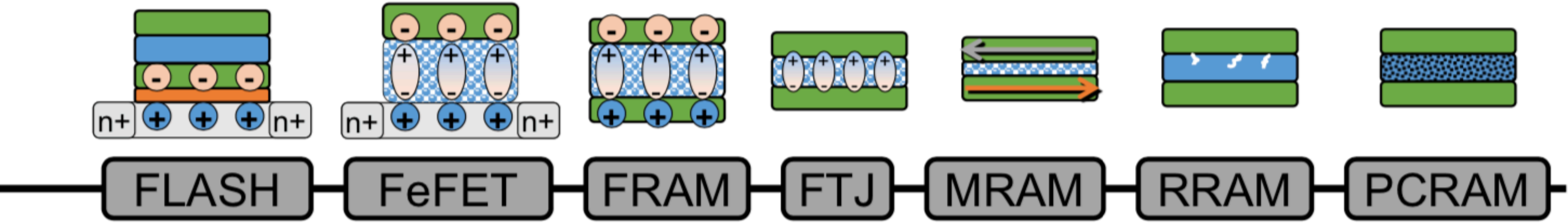
Source: J. Müller, GLOBALFOUNDRIES, "Ferroelectric Memories & Beyond", IEDM 2019

PM Media: Something New!



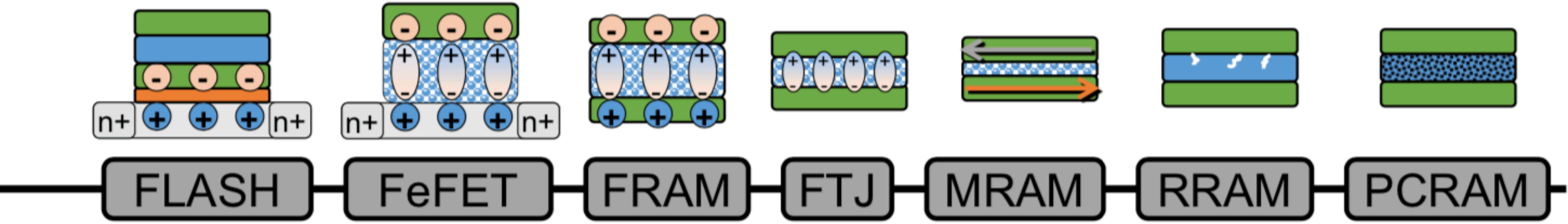
Source: R.Chau, Intel, "Innovations for Moore's Law Continuum", IEDM 2019

PM Media: Why we Love/Hate each



DENSE & CHEAP		FAST & CHEAP	FAST & FOUNDRY	CHEAP & ???	JUST RIGHT!
WAY TOO SLOW!		WAY TOO NEW!	WAY TOO $$$!	REL TOO ???!	HOW TO USE?

PM Media: Who is Doing What



FLASH
- SAMSUNG
- KIOXIA
- Micron
- SK hynix
- intel
- WD

FeFET
- intel
- Micron (rumored)
- SK hynix (rumored)
- GLOBALFOUNDRIES

MRAM
- tsmc
- GLOBALFOUNDRIES
- SAMSUNG
- UMC
- intel
- EVERSPIN TECHNOLOGIES

RRAM
- SONY

PCRAM
- intel
- Micron
- SK hynix
- SAMSUNG (rumored)

Rumored Activity

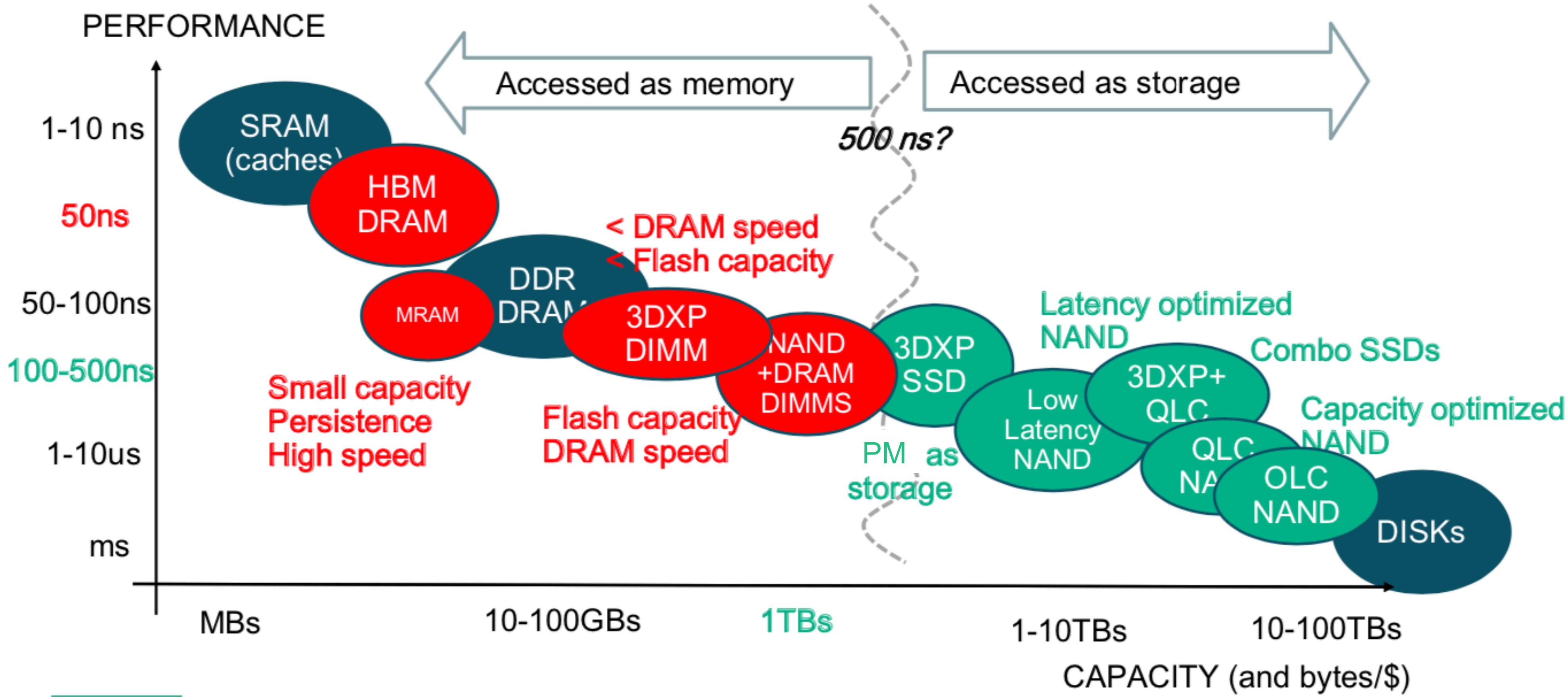
Source: Industry announcements & scuttlebutt

PM Attach

Options, DIMMs, Fabrics, Pooling, and Who

PM Attach: Options aplenty

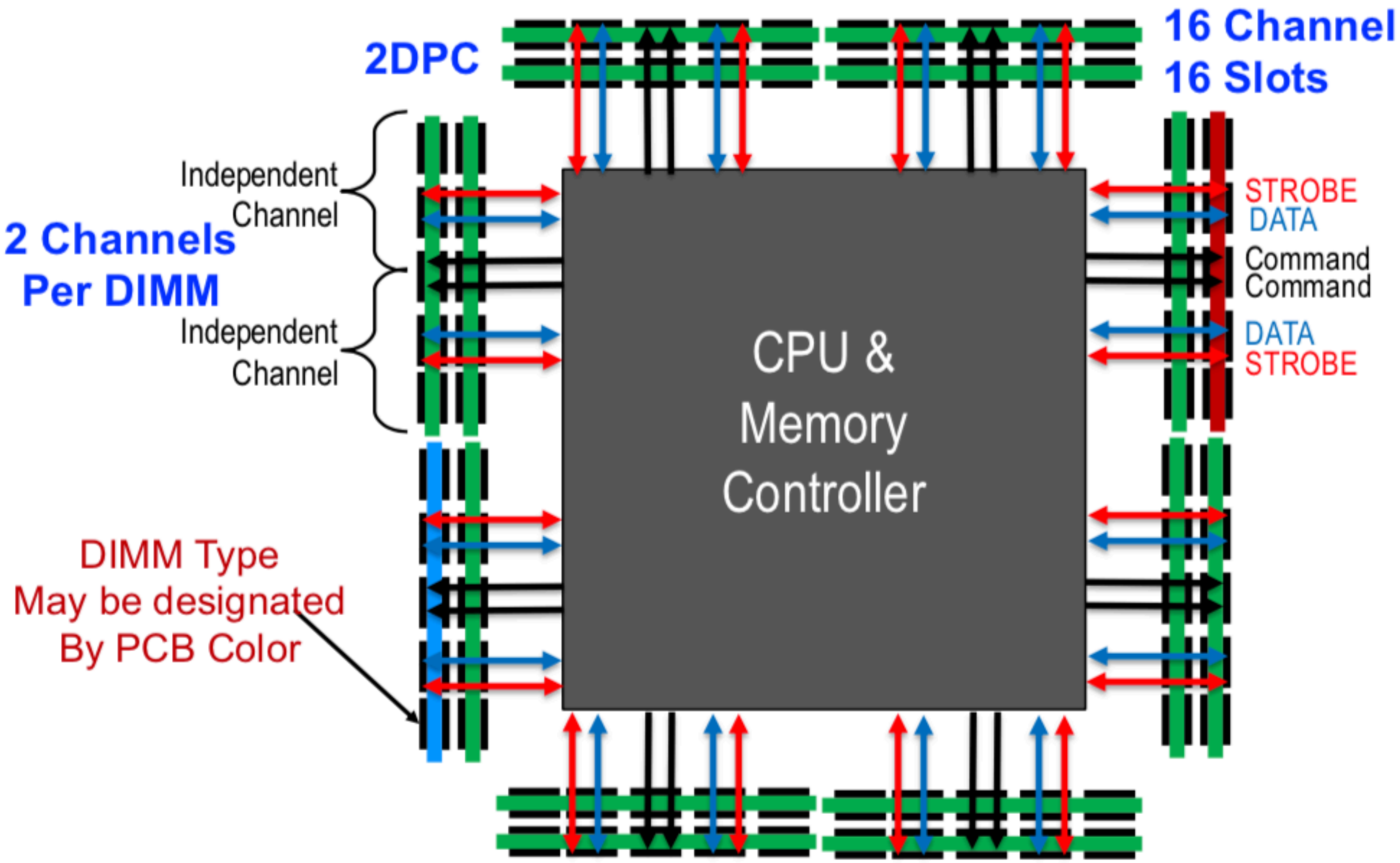
DATA ACCESS HIERARCHY IN 2019 – GETTING WORSE!



Source: P. Faraboschi, HPE, "The Data Access Continuum", SC'19 MCHPC

PM Attach: DIMMs getting complicated

Possible DDR5 Server Subsystem

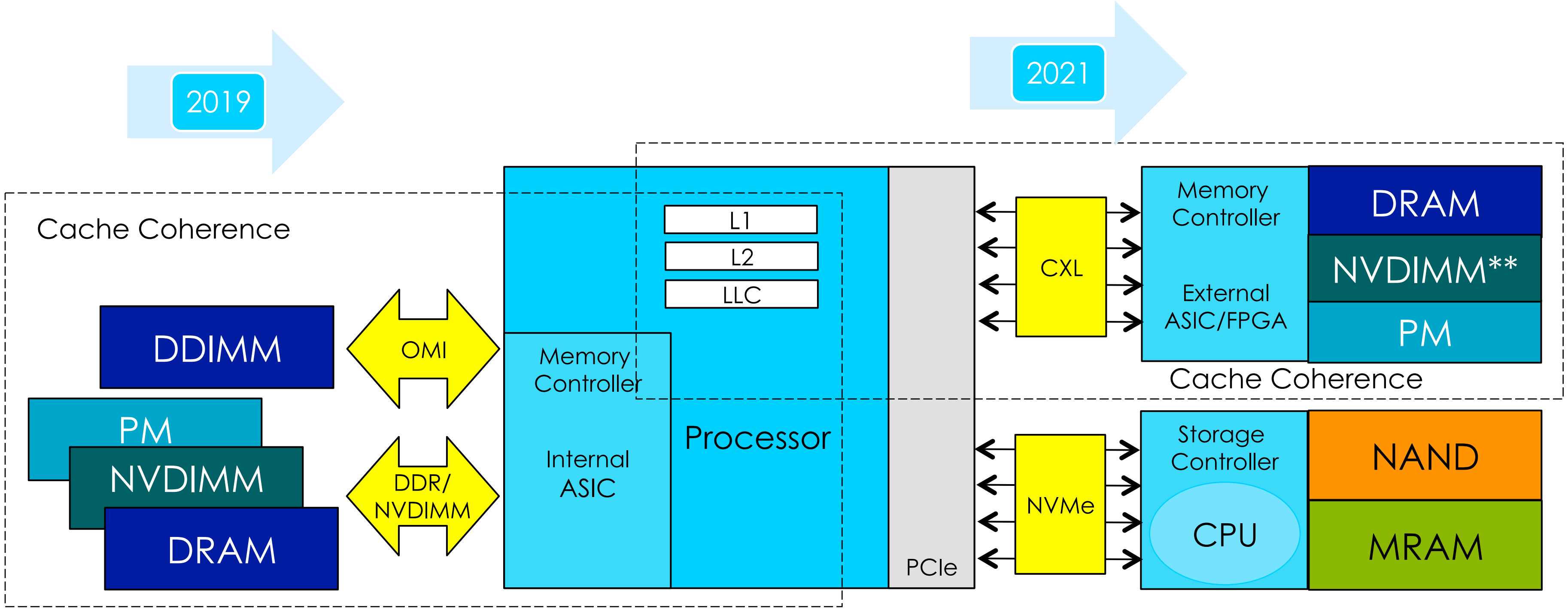


Different types of DIMMs **will** be common in DDR5

- RDIMMs
- LRDIMMs
- UDIMMs
- DDIMMs
- Optane DIMMs
- NVDIMM-N
- NVDIMM-P
- Etc…

Source: D. Rhoden, Montage, "Server Memory Evolution", JEDEC Training Class Oct 2019

PM Attach: Think Beyond the DIMM

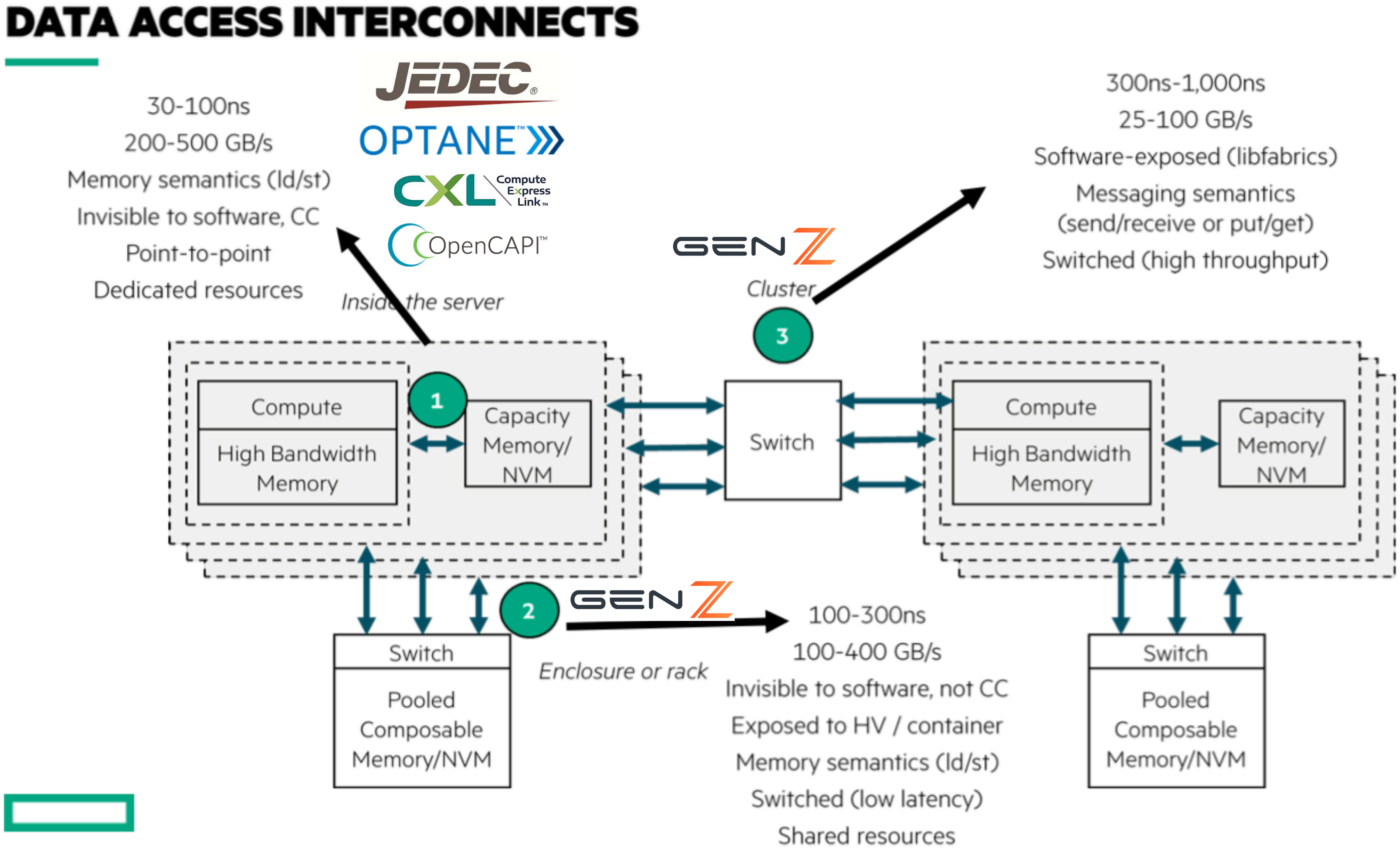


** Functionally equivalent to NVDIMM-N, not in DIMM form factor

- Today memory is direct-attached to the CPU
- New emerging interfaces will add high-speed differential CPU-attach options
- Systems will be aware of what type of memory or storage is available and how it is connected
- ***Lots of new types of memory, persistent memory and storage products are possible!***

Source: A. Sainio, SMART, “PM Benefits AI/ML”, IMC Summit 2019

PM Attach: Think Pooling and Fabrics



Source: P. Faraboschi, HPE, “The Data Access Continuum”, SC’19 MCHPC

PM Attach: Who is Doing What

OPTANE™

DDR-T

intel

JEDEC®

NVDIMM-N

Micron

SMART® Modular Technologies

AGIGA TECH®

NVDIMM-P

SAMSUNG

Micron

SK hynix

CXL Compute Express Link™

intel

Micron

SAMSUNG

SK hynix

OpenCAPI™

SMART® Modular Technologies

GEN Z

SMART® Modular Technologies

SAMSUNG

SK hynix

Rumored Activity

Source: Industry announcements & scuttlebutt

PM Usage

Value Prop, Tiering, Pooling, Computational Memory

PM Usage: Value Propositions

1. PM = More Memory

   Pro: Existing applications want it

   Con: Compete directly vs. DRAM on price

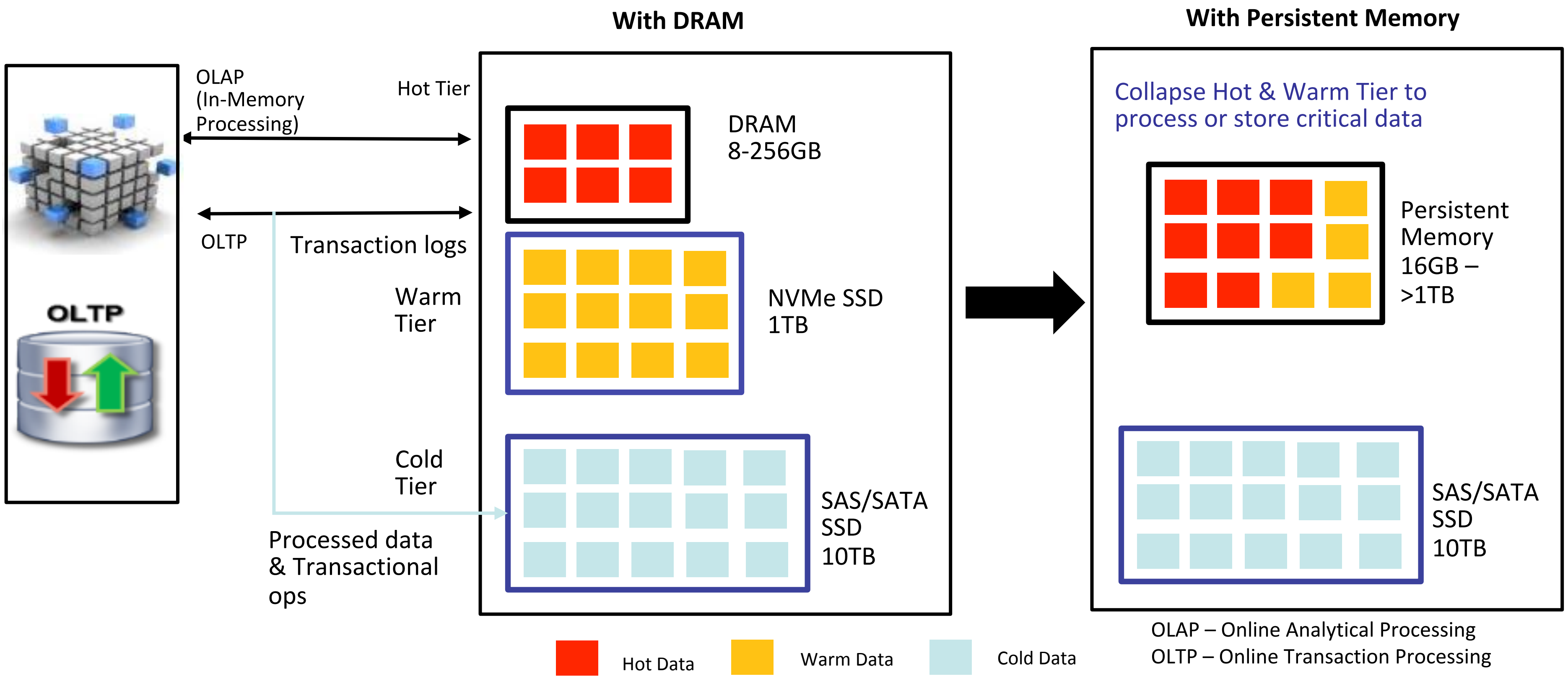
2. PM = Persistent & Coherent

   Pro: Take full advantage of features

   Con: Application needs to know

Source: D. Eggleston, Intuitive Cognition Consulting, “Overview of PM”, FMS 2019

PM Usage: Data Tiering



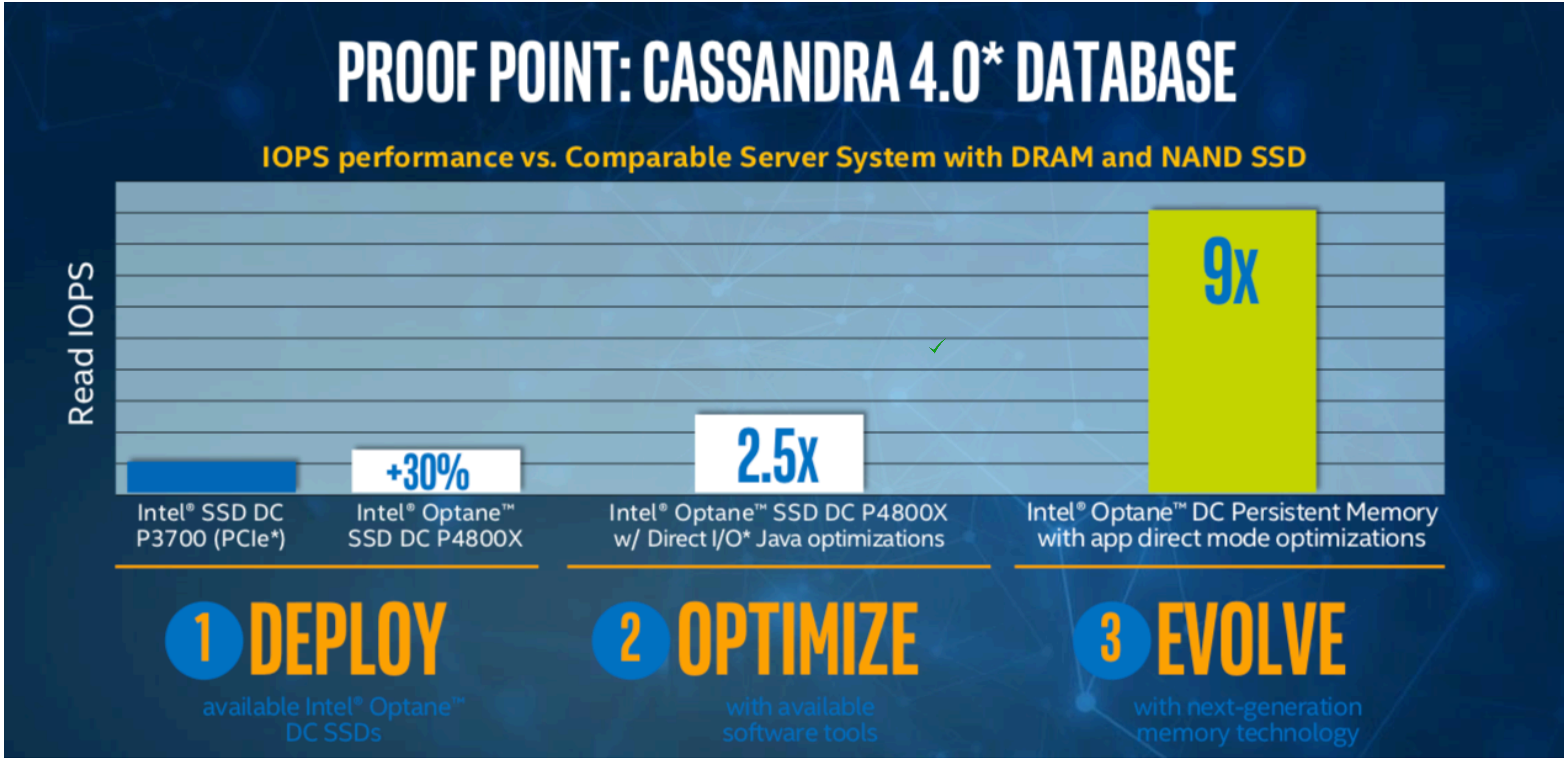
OLAP – Online Analytical Processing
OLTP – Online Transaction Processing

✓Persistent

✓Coherent

Source: A. Sainio, SMART, "PM Benefits AI/ML", IMC Summit 2019

PM Usage: Move SSD tasks into PM



✓Persistent

✓Coherent

9x the Read IOPS!

Source: A. Huffman, Intel, “Solid State Transformation of the Data Center”, SDC 2018

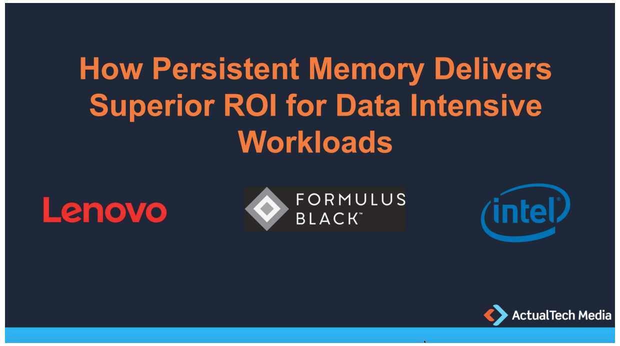
PM Usage: Move SSD tasks into PM

FORSA Enhances Performance and Usability of Persistent Memory:

FORMULUS BLACK™

	Intel NVMe SSD P4510	Intel Optane DCPMM Linux Native (Block)	Intel Optane DCPMM FORSA LEM (Block)
MySQL Transactions Per Second	3,824	9,211	10,299
MySQL 99th Pctl. Latency (ms)	84.5	21.5	17.6

Intel Cascade Lake - NVMe SSD	Intel Cascade Lake - DCPMM
2x 8160M CPU 24C, 2.1GHz	2x 8260M CPU 24C, 2.4GHz
24x DDR4 2666 MT/s (384GB)	12x DDR4 2666 MT/s (384GB)
n/a	12x Intel Optane DCPMM (1.5TB)
FORSA 3.0	FORSA 3.0
System Memory: 384GB	System Memory: 384GB
MySQL TPS: 3,824	MySQL TPS: 10,299
MySQL 99 Pctl. Latency: 85ms	MySQL 99 Pctl. Latency: 18ms
System Price: $22,879.80*	System Price: $29,614.95*
$USD / TPS: $5.98	$USD / TPS: $2.88



✓Persistent

✓Coherent

2.4x the transactions...

@ 50% lower $ per transaction!!!

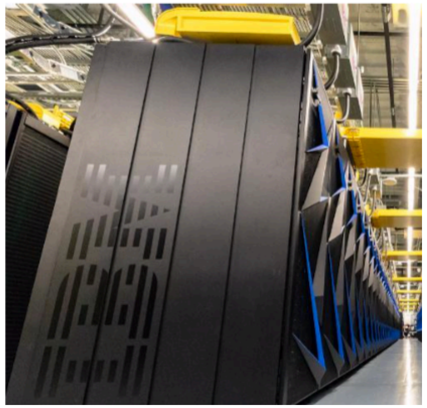
Source: J. Xie, Formulus Black, “PM Delivers Superior ROI”, Webinar 2020

PM Future: "Pooled Memory Appliance"

All Top Supercomputers Are Heterogeneous

Both within their Processors/Accelerators and their Memories

- World's Top: Summit 200 PetaFlops at 14MW
- 4,608 nodes
  - 9,216 IBM POWER9 22-core CPUs and 27,648 Nvidia Tesla GPUs
  - > 95% of Flops comes from GPUs
- Over 600 GiB of coherent memory per node
  - 6×16 = 96 GiB HBM2
  - 2×8×32 = 512 GiB DDR4 SDRAM
  - Addressable by all CPUs and GPUs
- 800 GB of non-volatile RAM that can be used as a burst buffer or as extended memory
- Challenging to program system with heterogeneity
  - Not all software can be ported



JTP

Opportunity of Memory Heterogeneity

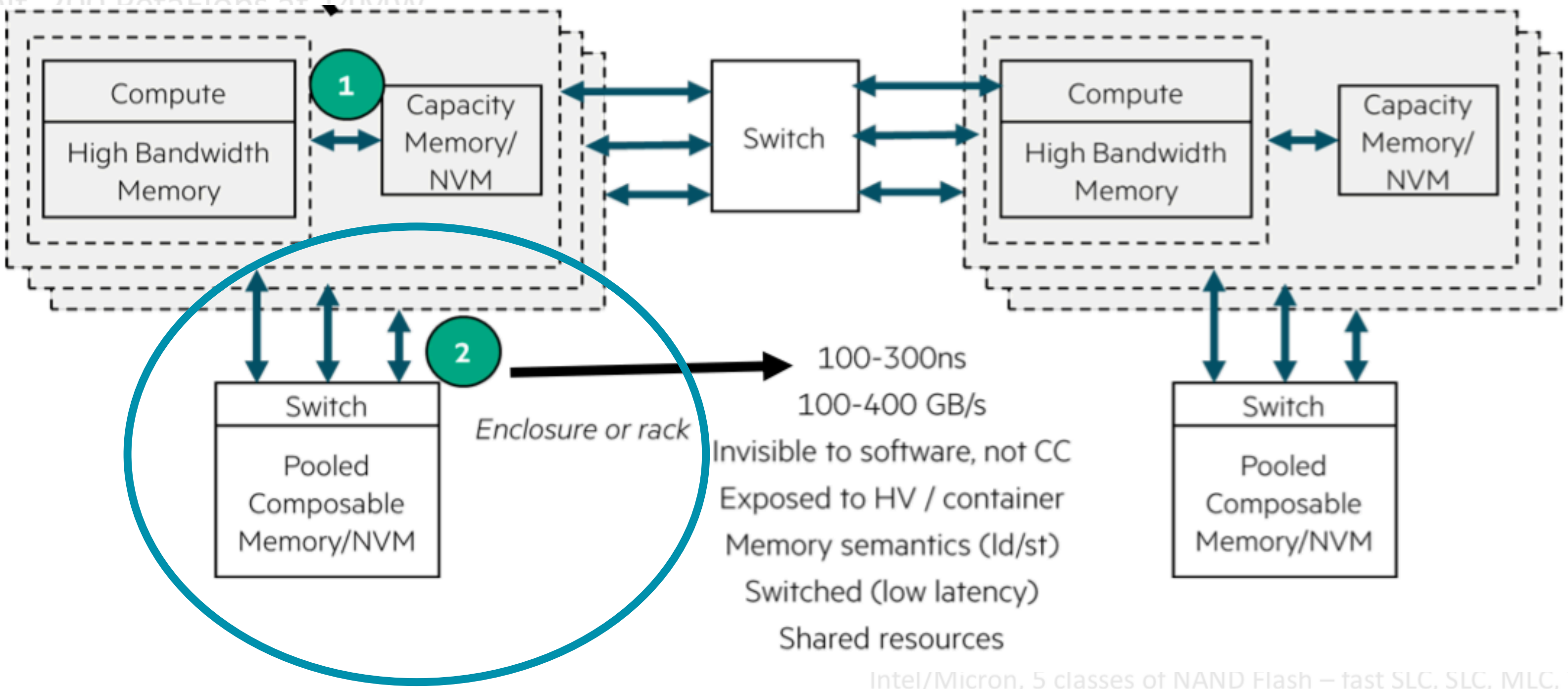
- Why have systems become heterogeneous?
  - Slowdown in frequency scaling, lithography scaling, interconnect scaling
  - Increase in specialization, problems that are embarrassingly parallel, higher intensity compute
  - Mixed compute requirements - variable access patterns, some very regular, others very irregular such as pointer chasing, indirection
  - High energy and latency of data movement - finer grain accelerators, data transformations
  - Advent of emerging memory technologies
  - Advances in packaging
- Challenges
  - Compute and Software - offload and communications overhead, variable performance, determinism, programming complexity, mapping increasingly varied workloads
  - Emerging memory explosion
    - Recent introduction of STT-MRAM, ReRAM into foundries, introduction of 3DXpoint by Intel/Micron, 5 classes of NAND Flash – fast SLC, SLC, MLC, TLC, QLC, with varied maturity and performance
  - Attaching the memory into hardware
  - Additional memory tiers presents challenge to app's software and operating system
  - Data movement between memory types
  - Widened variety of data access patterns

JTP

41

Source: J. Thomas Pawlowski, JTP, "Prospects for Memory", SC'19 MCHPC

PM Future: "Pooled Memory Appliance"



✓ Persistent

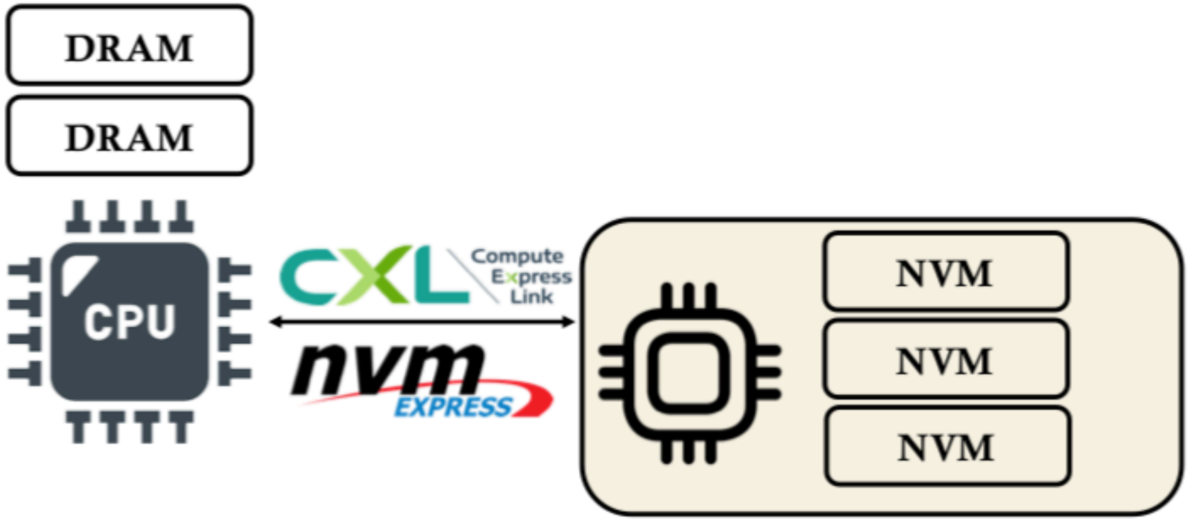
✓ Coherent

Source: P. Faraboschi, HPE, "The Data Access Continuum", SC'19 MCHPC

PM Future: "Computational Memory"

EIDETICOM

Use Case: CXL-Based NVDIMM + Accelerator



CXL-based NVDIMM+Accelerator:

- Controller chip includes compute functions (e.g. AI, search, graph database)
- Controller chip can be programmed via PCIe driver (e.g. NVMe).
- NVM can still be exposed to host and accessed via CXL.mem (volatile or persistent)

✓Persistent

✓Coherent

A new initiative?

SNIA®

Source: S. Bates, Eideticom, "Successfully Deploying PM and Acceleration", PIRL 2019

Conclusions

4 Things I think I think

4 Things I Think I Think

PERSISTENT MEMORY SNIA PM SUMMIT
JANUARY 23, 2020 | SANTA CLARA, CA

1. I ♥ DRAM! (But…)
   - Watch: Tech and Price scaling stall gets worse
2. 3DXP/PCM still the top PM Media for the job
   - Watch: Ferro HfOx 1T-1C research advances
3. DIMM slots too precious for PM
   - Watch: CXL becomes primary PM attachment; big bandwidth increase
4. Best PM usage values Persistence and Coherence
   - Watch: Pooled Memory Appliances and “Computational Memory”

Thank you

Please visit www.snia.org/pmsummit for presentations

Dave Eggleston
Intuitive Cognition Consulting
Technology & Business Strategy

Email: dave@in-cog.com
Twitter: @NVM_DaveE
LinkedIn: linkedin.com/in/deggleston/

Abstract

The commercialization of Persistent Memory is well underway, fundamentally changing the way processors load and store information. The characteristics of the non-volatile memory media (3DXP, PCM, MRAM, RRAM, others) determine the basic performance, reliability, and cost that Persistent Memory can deliver to the system. Brand new methods of processor attachment (DDR5, DDR-T, NVDIMM, CXL, CCIX, Gen-Z, OpenCAPI) for Persistent Memory promise to broaden its usage far beyond "DRAM replacement", opening up vast resources enabling memory and power hungry applications.

In this talk, the speaker will:

- Discuss the basic characteristics of Persistent Memory Media
- Articulate how Persistent Memory enhances server architectures
- Identify, compare, contrast the new methods for Persistent Memory attachment
- Propose innovative ways applications will utilize Persistent Memory to overcome memory and power limitations
- Discuss who is doing what within the Persistent Memory ecosystem