Flash Memory Summit

A Persistent CXL Memory Module with DRAM Performance

Bill Gervasi

Principal Systems Architect

bilge@nantero.com

35 YEARS OF NAND FLASH MEMORY INNOVATION

AGENDA

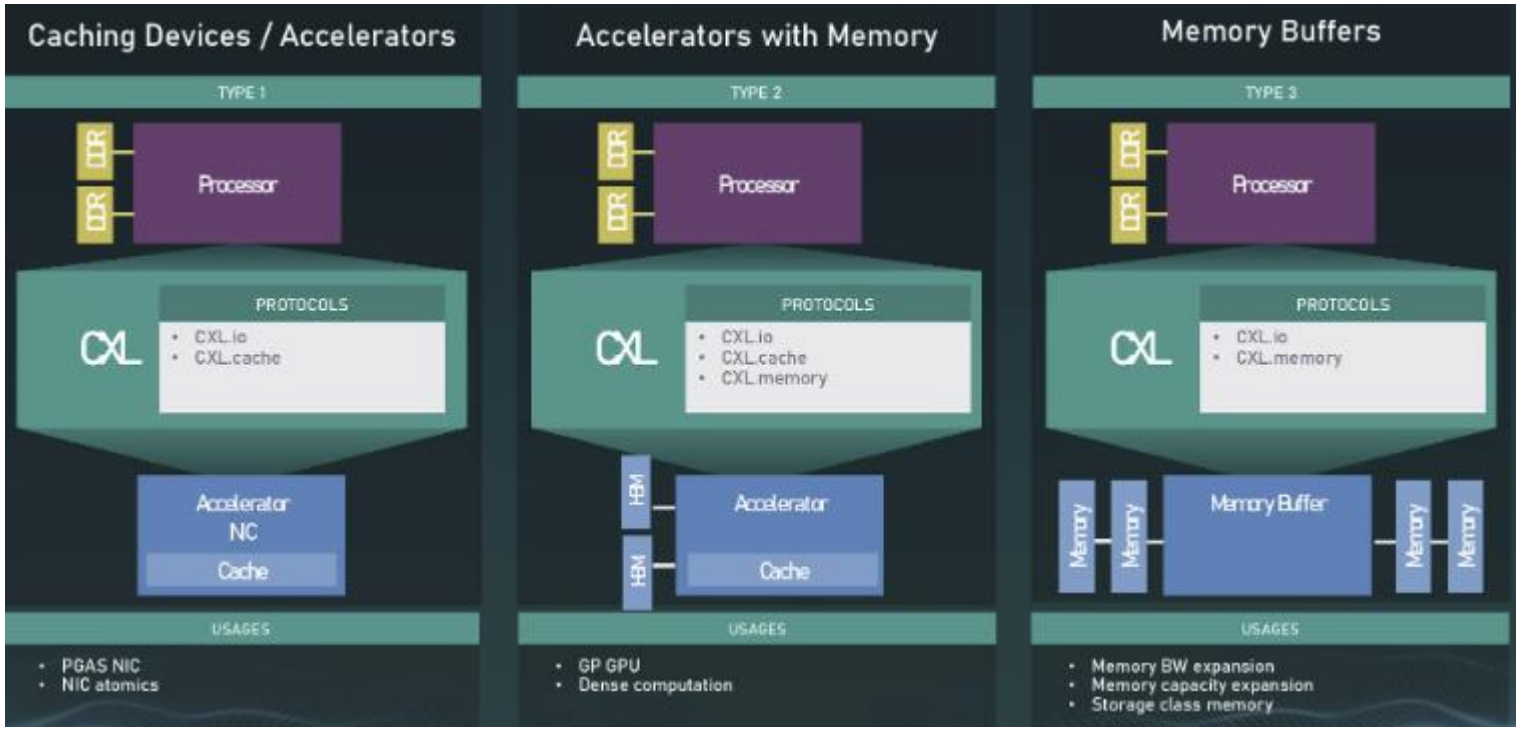
- CXL Expansion Memory is a growing trend
- CXL Memory faces some unique resistance
- System architecture adoption of persistent memory
- Nantero NRAM® persistent main memory
- How persistent memory addresses CXL concerns

CXL Rising, but…

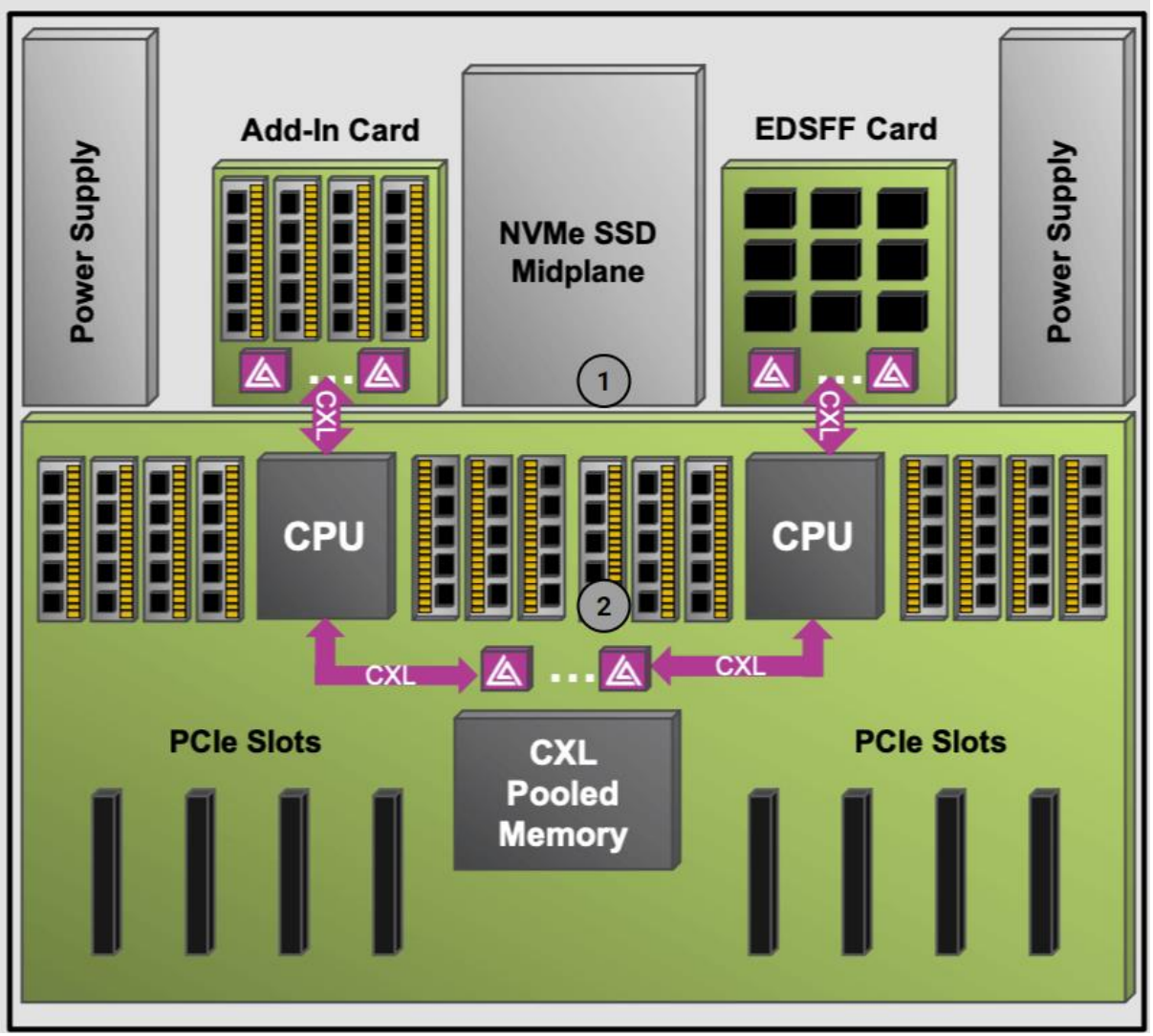
I'm sure we'll see dozens of presentations at FMS on the emergence of CXL as a revolutionary change in systems architecture

A Bluetooth for fabrics, so to speak

However, there is still a lot of work ahead of us to enable its potential



Focus on Memory Expansion



ASSUMPTIONS

CXL Expansion Memory does not replace the direct attached DIMM

CXL Expansion Memory is volatile, so local SSDs for checkpointing are still needed

CXL Expansion Memory is shared by many multi-core processors (very random access)

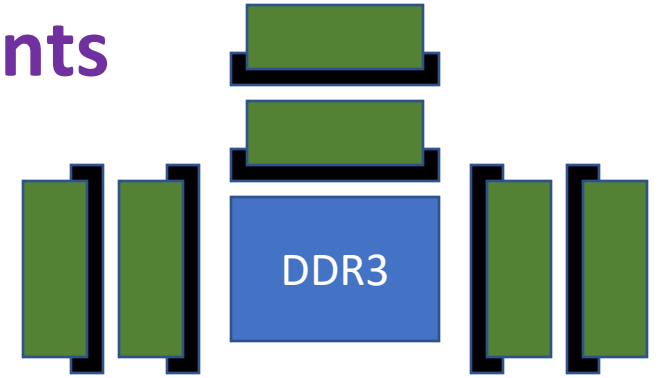
6X

1 CH, 4 DPC, 2R/D, 256Mb = 4 GB

2 CH, 3 DPC, 2R/D, 1 Gb = 24 GB

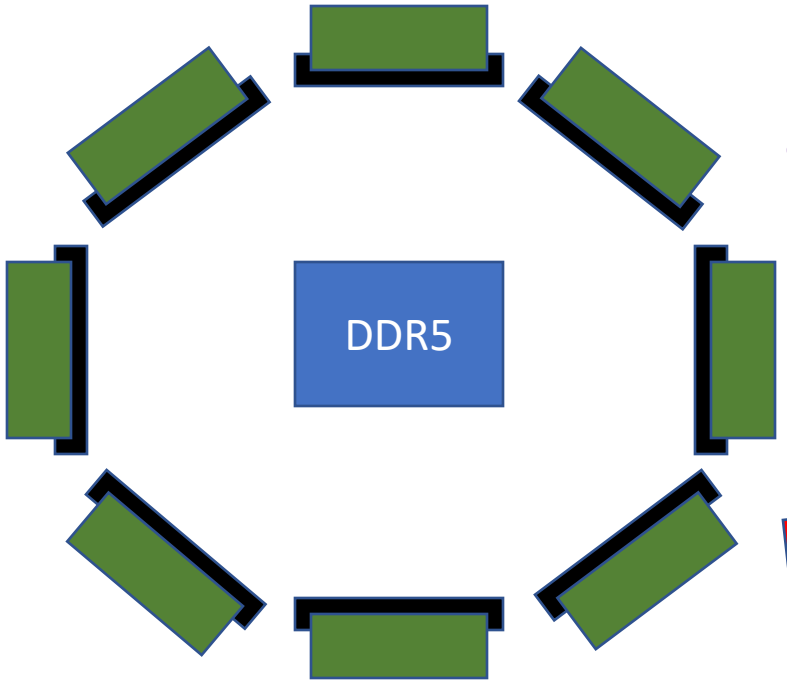
4X

Increasing frequency is slowing DIMM improvements



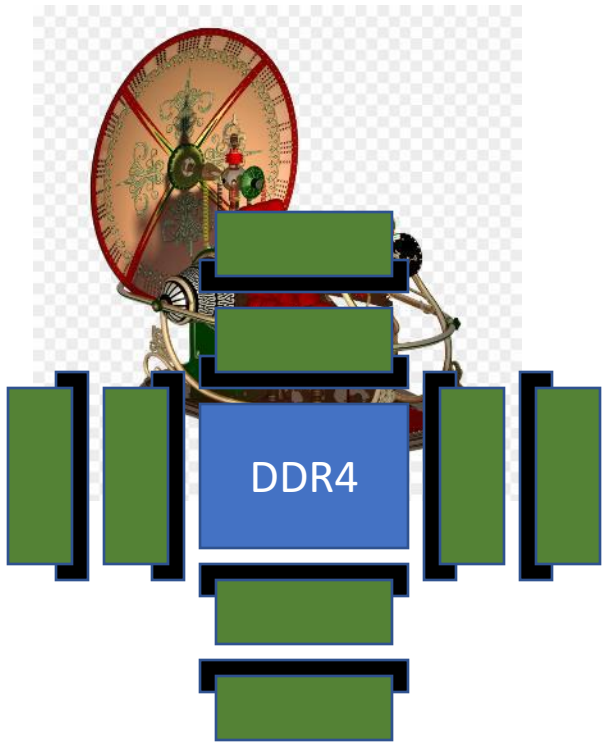
3 CH, 2 DPC, 2R/D, 4 Gb = 96 GB

5X



1-2X

8 CH, 1 DPC, 2R/D, 16 Gb = 512 GB
8 CH, 1 DPC, 2R/D, 32 Gb = 1 TB



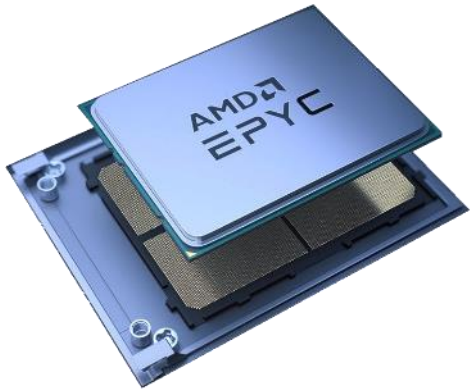
4 CH, 2 DPC, 2R/D, 16 Gb = 512 GB

CH = channel
DPC = DIMMs per channel
R/D = ranks per DIMM

Assumes no 3DS

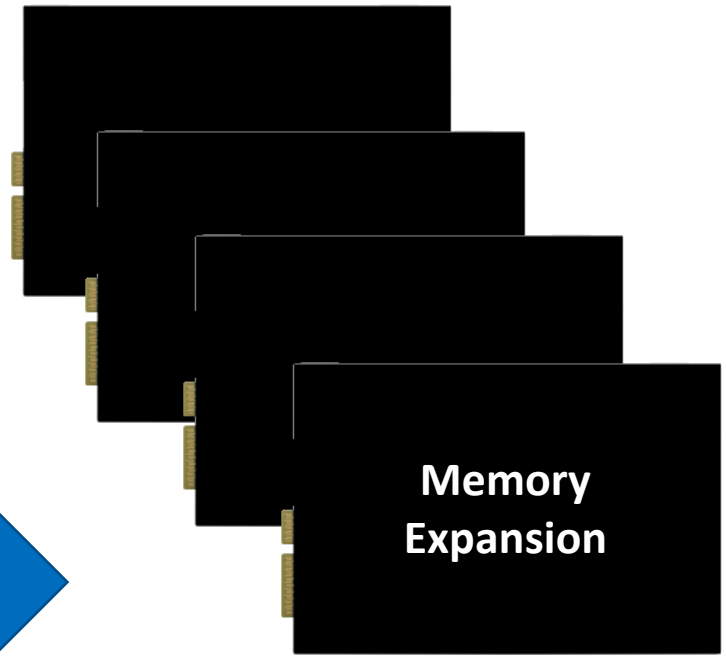
The slowdown of capacity expansion from DRAM on DIMM explains the momentum for CXL memory expansion

Memory expansion beyond DIMMs
End-user friendly pluggable modules

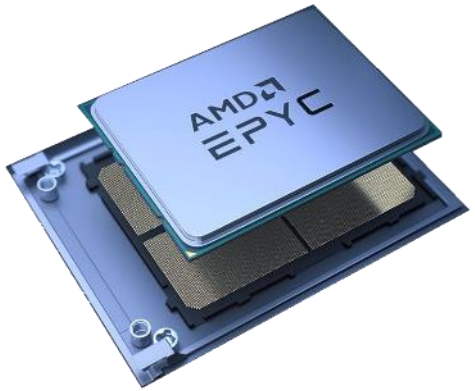


DDR

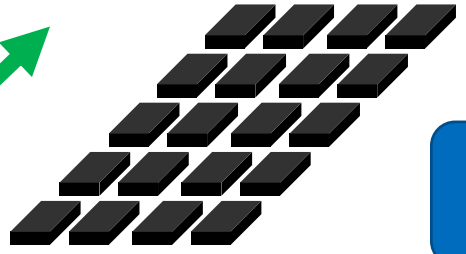
CXL



So, What's Not to Love About CXL?



Performance



Volatility

Capacity

CXL Interface

CXL DRAM Controller

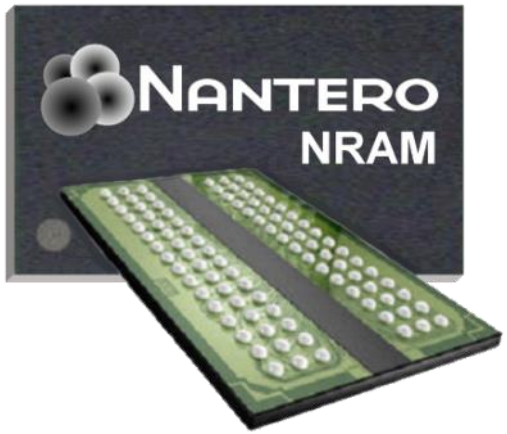
Latency

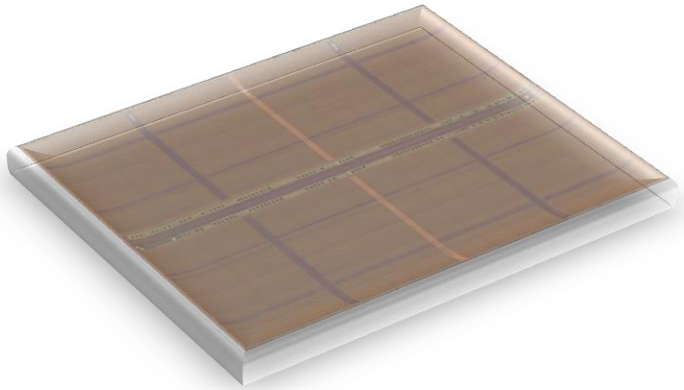
Power

PCIe Bus

These concerns may slow the rate of CXL adoption

Nantero DDR5 NRAM®





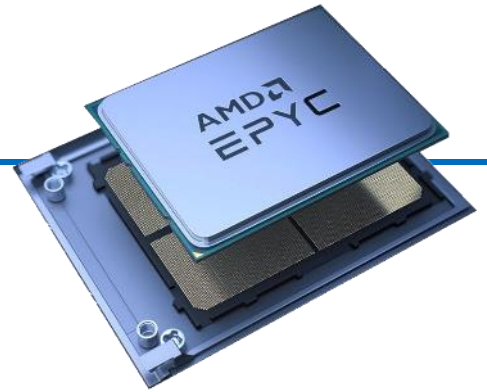
DDR5 SDRAM speed

Non-volatility

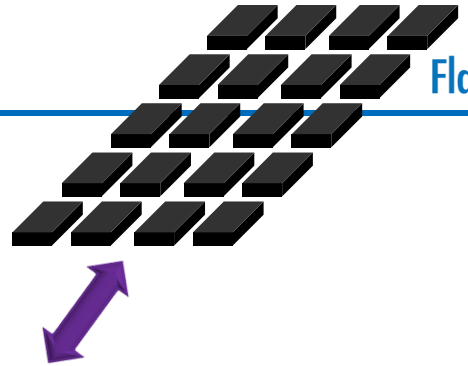
Scalable beyond DRAM

Lower power than DRAM

Let's explore how this affects user's concerns about CXL for memory expansion



Addressing Latency



CXL Interface

CXL DRAM Controller

Type 3: CXL.mem

Latency

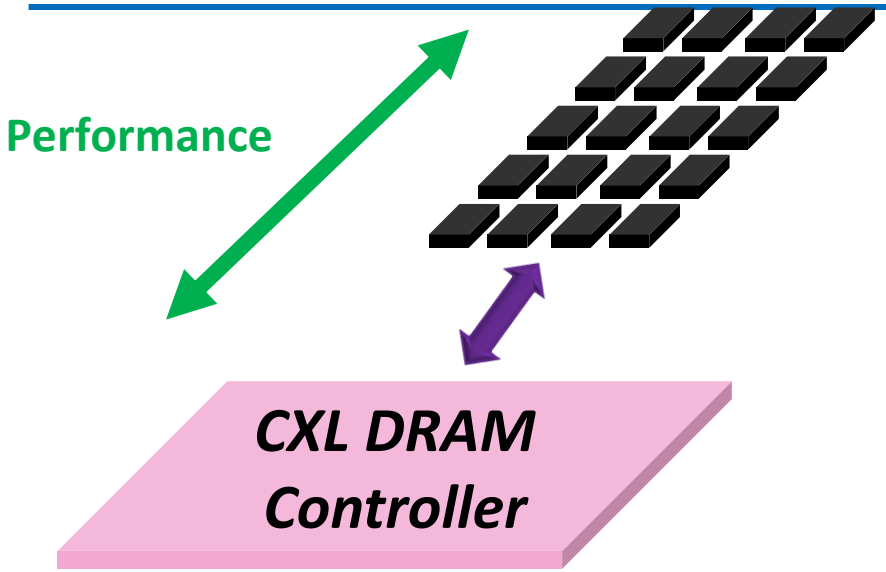
PCIe Bus

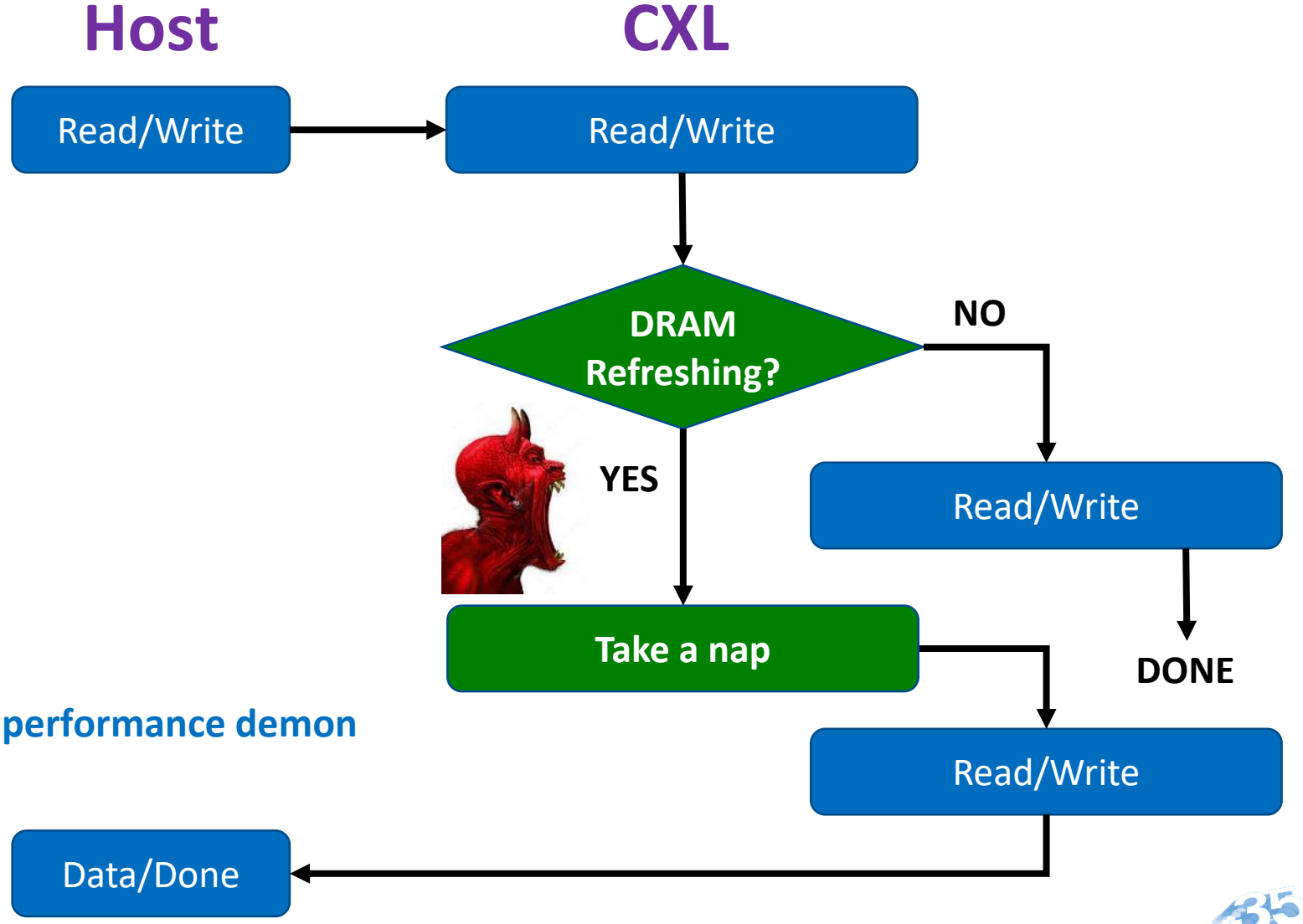
Latency adder is real: CXL, SERDES, media

Full duplex nature of CXL helps offset penalties

Read and write buffers in a CXL DRAM controller can reduce some penalties

Addressing Performance





Media timing is (mostly) fixed

Non-deterministic refresh is the performance demon

What if There Were No Refresh?

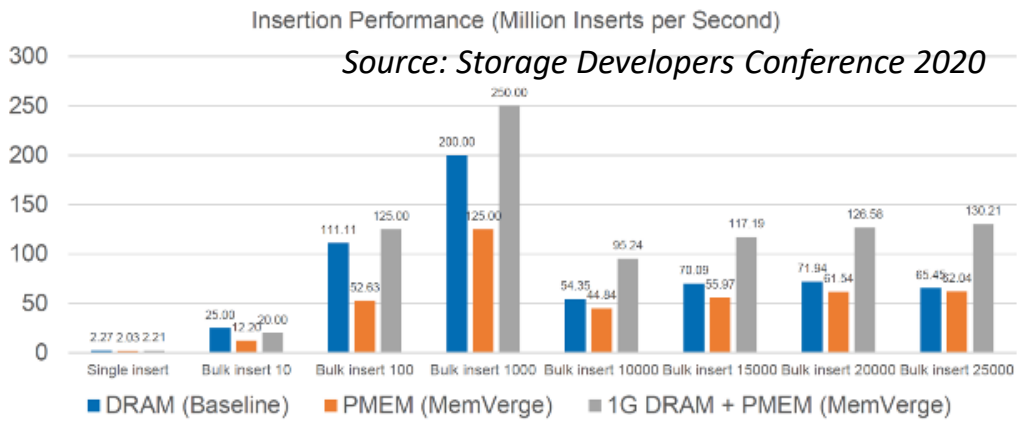
Flash Memory Summit

1. 111% DIMM-NRAM vs DIMM-SDRAM
2. 96% CXL-NRAM vs DIMM-SDRAM
3. 111% CXL-NRAM vs CXL-SDRAM
4. 165% CXL-NRAM vs CXL-Optane

NRAM's data persistence eliminates refresh, making CXL accesses more deterministic

Offsets some effects of latency penalty

	Ref	Refi	Avail	Norm	Norm	Norm	Norm
DRAM, 1X Ref	350	3900	91%	111%	96%	111%	164%
DRAM, 2X Ref	350	1950	82%	100%	87%	100%	148%
DRAM, 4X Ref	350	975	64%	78%	68%	78%	116%
Optane	0	0	100%	64%	58%	68%	100%
NRAM	0	0	100%	111%	96%	111%	165%

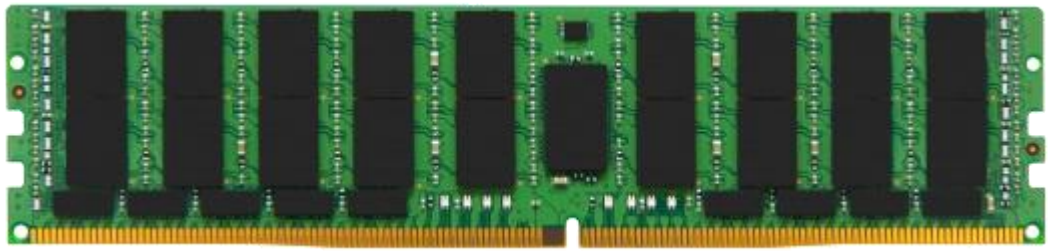


* Half-duplex assumed; full duplex model to be built; likely gets better

** Assumes pipelining of CXL requests offsets 85% of penalties

35 YEARS OF NAND FLASH MEMORY INNOVATION

Addressing Power



DDR5 → 6400 Mbps speed @ 25% increase in power
Register, data buffers consume at least 2W per DIMM
40 DRAMs per DIMM: 11 + 2 = 13W total DIMM power

CXL Controller

DDR5 → Same power issues as DIMM
CXL controller consumes at least 6W
Comes down to media capacity

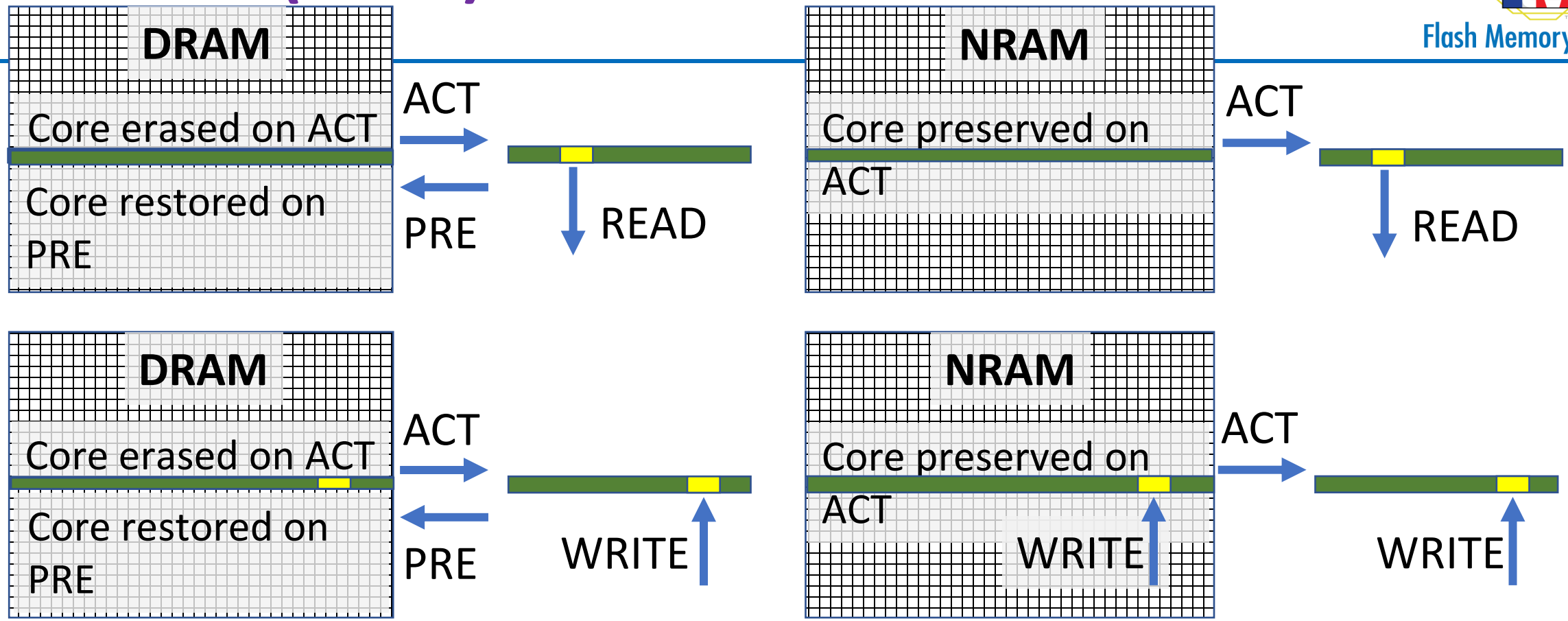
Compared to… what? DDR5 LRDIMM?

E1.S: 20 chips
GB/W: ½ DIMM capacity @ 11.5W
~ 56% efficient

E3.S: 40 chips
Better GB/W: 1X @ 17W
~ 77% efficient

E3.S: 80 chips
Even better GB/W: 2X @ 28W
~ 92% efficient

(Non-)Destructive Activation



Precharge operation required even if data is only read
8192+ECC bits always rewritten

No precharge operation required,
On writes, 64+ECC bits written to the array and page buffer

Non-Destructive Activation = Lower Power

Even with the **volume** cranked to **11***

(Data transfers when a DRAM would be refreshing)**

NRAM saves 21% power over DRAM
while delivering 11% more data
at the same clock rate

34% better throughput per watt than DRAM

Assume
70% data active
2/3 read
1/3 write
20% refresh
10% idle

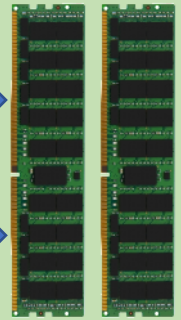
																						Totals	Ratio
																			**				
CMD	ACT		ACT		ACT		ACT		ACT		ACT		ACT		ACT		ACT			REF	REF	NOP	
DATA		RAP	RAP	RAP	RAP	RAP	RAP	RAP	RAP	RAP	RAP	RAP	RAP	WAP	WAP	WAP	WAP	WAP	WAP	RAP	WAP		
DRAM	65.9	139.7	205.6	139.7	205.6	139.7	205.6	139.7	205.6	139.7	205.6	139.7	205.6	115.5	181.4	115.5	181.4	115.5	115.5	139.7	139.7	53.1	3295	126%
NRAM	65.9	106.75	172.65	106.75	172.65	106.75	172.65	106.75	172.65	106.75	172.65	106.75	172.65	82.55	148.45	82.55	148.45	82.55	82.55	106.75	82.55	53.1	2612	79%

* The number 11 is a registered trademark of Spinal Tap

DDR5 on DIMM

DDR

DDR



2x40 pkgs

DRAM power: 2 x 11 W
Buffer power: 2 x 2 W

100%

DDR5 on CXL

CXL

80 pkgs

DRAM power: 22 W
Buffer power: 6 W

108%

NRAM on CXL

CXL

80 pkgs

NRAM power: 14 W
Buffer power: 6 W

79%

NRAM CXL lower power than two DDR5 DIMMs or DDR5 on CXL

Additional power saving if SSDs can be eliminated

Addressing Volatility

Flash Memory Summit

Let your data die on power fail
Checkpoint to SSD for recovery

Add supercapacitors
to NVDIMMs, CXL

Use non-volatile NRAM

35 YEARS OF NAND FLASH MEMORY INNOVATION

Addressing Capacity

Flash Memory Summit

DDR5 NRAM®

16 Gb

32 Gb

DDR5 SDRAM compatible

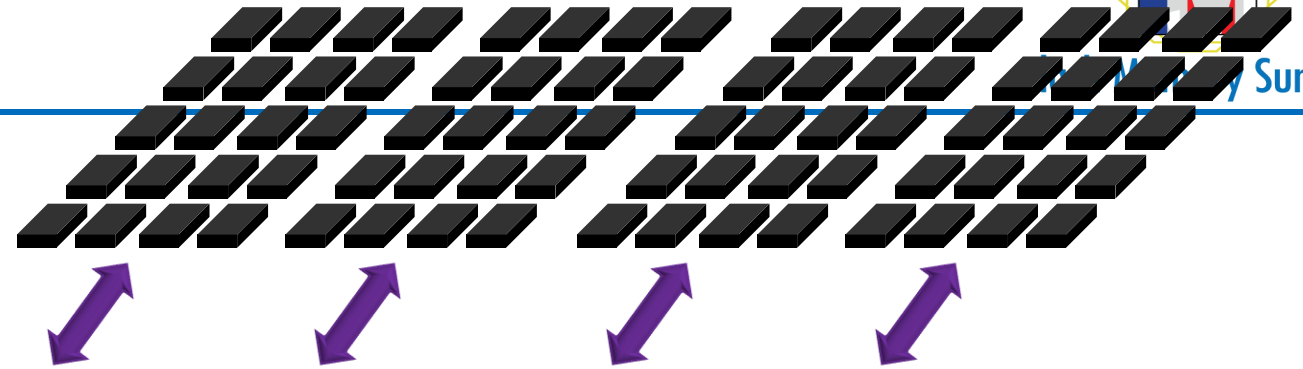
DDR5 NVRAM extensions

35 YEARS OF NAND FLASH MEMORY INNOVATION

Massive Memory Expansion!

64 GB to 2 TB per module

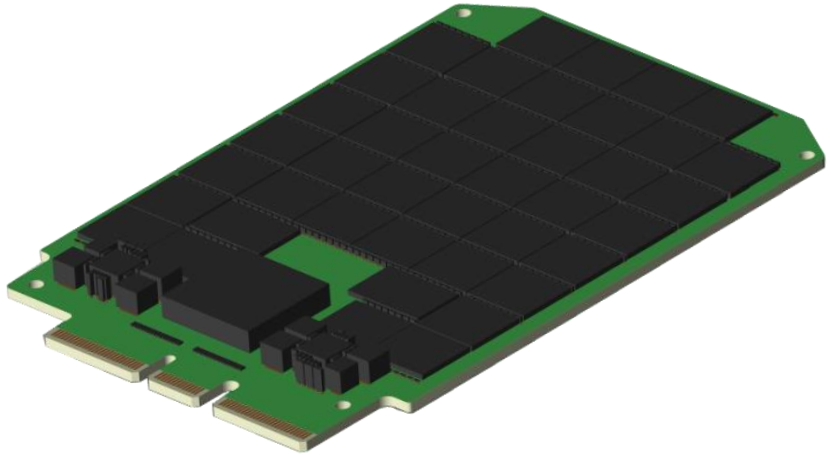
CXL E3.S



CXL NRAM Controller

DDR5-6400 NRAM
2 channels x
2 subchannels x
2 ranks
@ 25.6 GB/s

CXL 2.0
PCIe 5.0
8 lanes
@ 32 GB/s
Full duplex

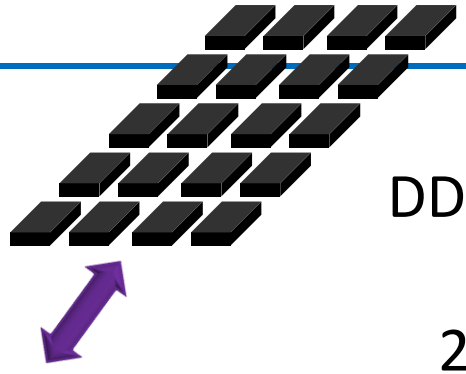


CXL E1.S

NANTERO

Commodity Memory Expansion

32 GB to 256 GB per module

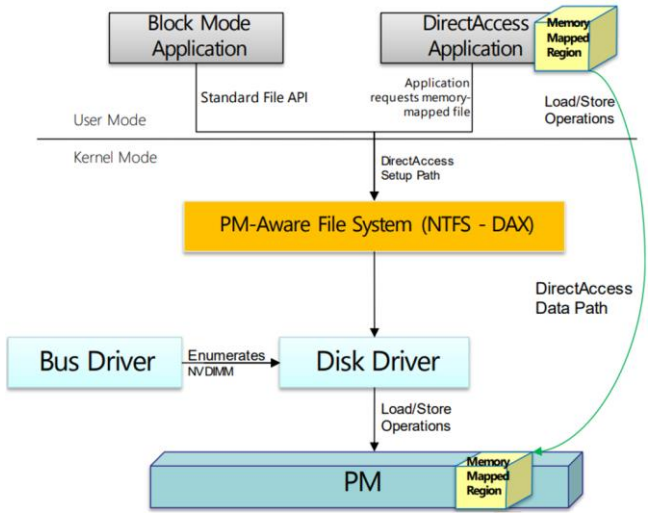


DDR5-6400 NVRAM
1 channels x
2 subchannels x
1 ranks
@ 25.6 GB/s

CXL NRAM Controller

CXL 2.0
PCIe 5.0
8 lanes
@ 32 GB/s
Full duplex

So Where Do We Go From Here?



Software support via DAX assists in moving…

from mounted drives…

…to RAM drive…

…to direct access mode

But the next step in exploiting persistent main memory requires

Boot from CXL

Power fail restart from CXL

Similar to BAEBI

JEDEC STANDARD

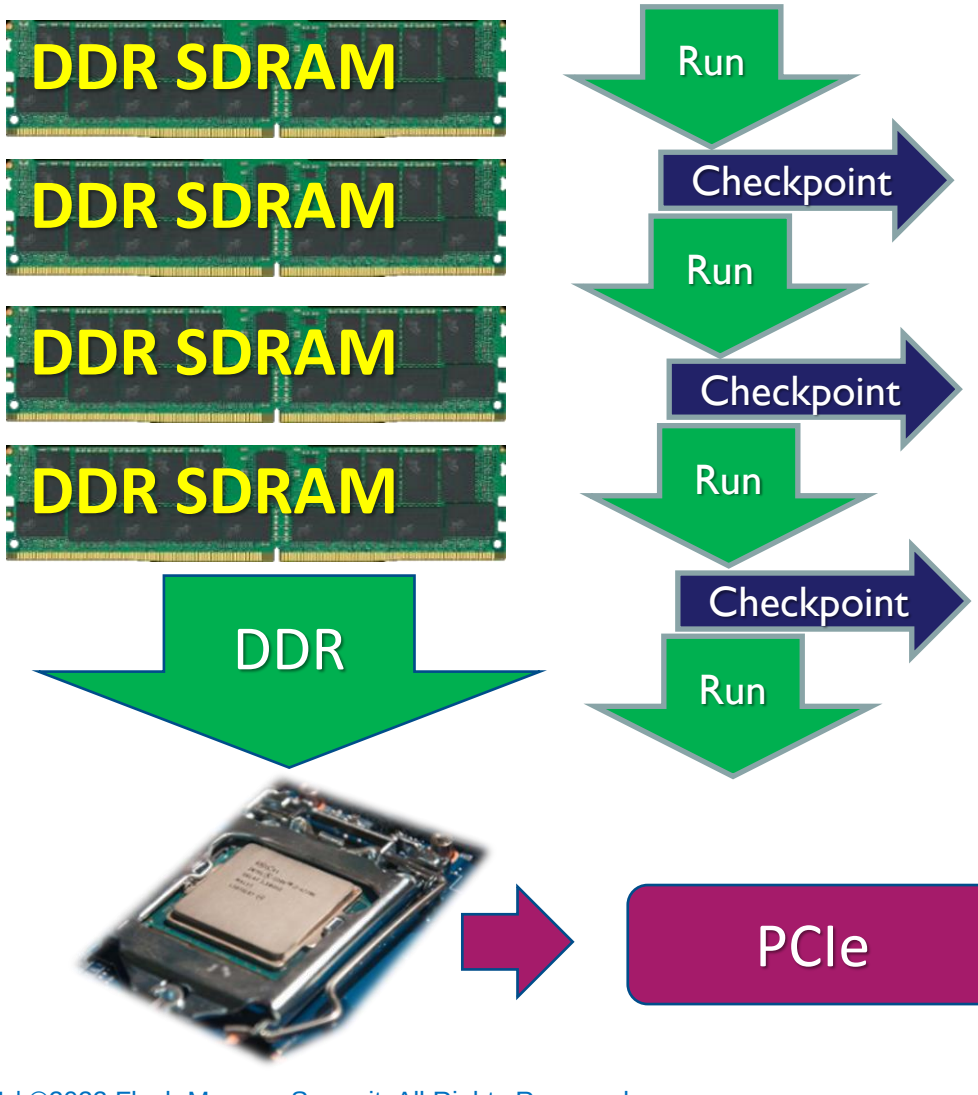
Byte Addressable Energy Backed Interface

JESD245E

(Revision of JESD245D, July 2020)

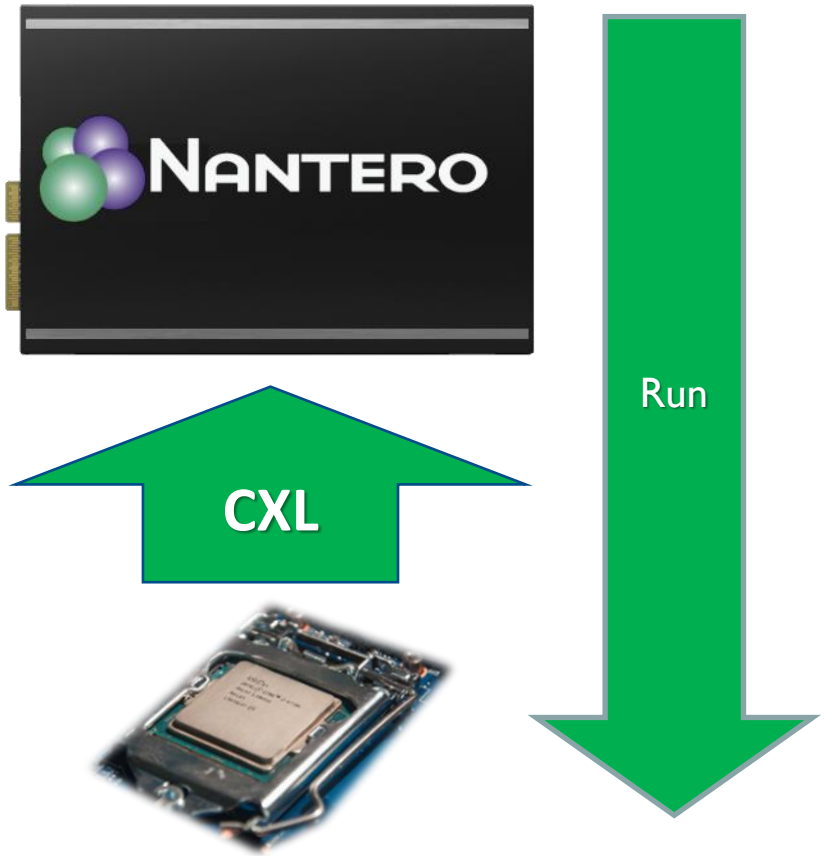
The old paradigm

"checkpointing"



The new paradigm

"leave it in place"



Flash Memory Summit

35 YEARS OF NAND FLASH MEMORY INNOVATION

The Holy Grail of computing systems is when **Instant On** is realized

Flash Memory Summit

Non-volatile memory media with DRAM performance makes **Instant On** achievable

Instant On

35 YEARS OF NAND FLASH MEMORY INNOVATION

SUMMARY

CXL for memory expansion has some valid concerns

Persistent memory on CXL addresses these concerns

The Holy Grail is end to end data persistence for Instant On

Nantero NRAM is a DDR5 class persistent memory

The DAX model helps existing systems move gracefully to PMEM

Shameless plug for BRDG Bridge to Connect

Education non-profit

Please visit the booth

Thank You

Bill Gervasi

bilge@nantero.com

