Providing Storage at Memory Speed Using NVDIMMs Sponsored by the SNIA NVDIMM SIG

Open Server Summit
Session C-203
April 14, 2016

Description

- NVDIMMs are persistent memory modules that reside on the DDR DRAM channel, combining volatile DRAM and nonvolatile flash memory.
- Under normal power conditions, an NVDIMM operates just like DRAM with ultra low latency and essentially infinite endurance, but is still nonvolatile.
- During a power failure or system crash, the data in the DRAM is transferred to the flash and can be restored when normal conditions resume.
- Persistent memory enables applications to run faster due to greater I/O performance. Typical areas of interest include databases, Web 2.0, analytics, OLTP, and video and image processing.

NVDIMM SIG Panel

Session C-203

SNIA™ Solid State Storage Initiative

Mat Young

- VP of Marketing at netlist, brings 20 years of experience in the application and storage industry. He most recently served as the senior director of ESS technical marketing at the Data Propulsion Labs at SanDisk. Prior to SanDisk, he founded the European subsidiary of Fusion-io and helped build the Data Propulsion Labs as the technical marketing function of Fusion-io then Sandisk worldwide. Mr. Young began his career at Data General, and also held positions with Microsoft, Pillar Data Systems and 3PAR.

Bob Frey

- Bob Frey is Senior Director of Memory and Systems Engineering at SMART Modular Technologies. His group is responsible for developing DRAM-based products including Hybrid DIMM modules. He has 25 years of industry experience working at companies including Amdahl, Sony, SUN Microsystems, VERITAS, AdvanSys, and Maranti Networks. Bob has a system architecture background in servers, storage, and networking. He has contributed to the Linux kernel and participated in ANSI/INCITS and IEEE standards groups. He has a BA from Amherst College.

Bill Gervasi

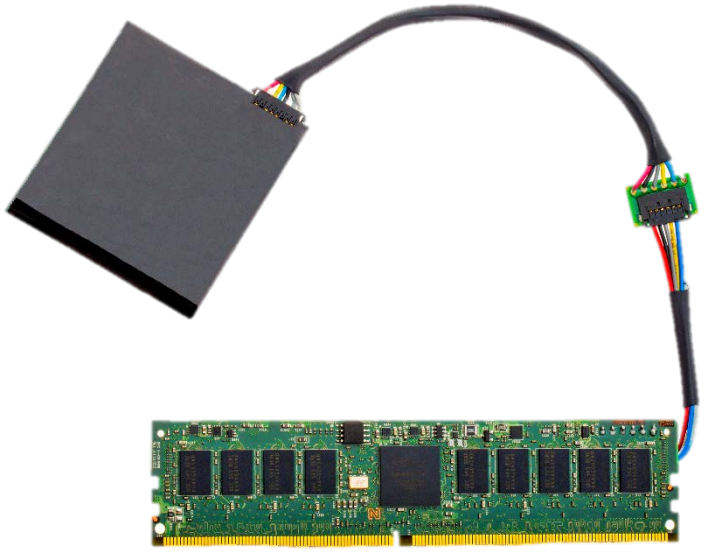
- Bill Gervasi is a well-known memory technologist and consultant. He has worked on the definition of DDR (double data rate) DRAM since its inception. He has introduced several DDR3 registered DIMMs into the JEDEC standardization process. He has also served on the JEDEC Board of Directors and chaired several JEDEC memory-oriented committees.
  He worked for Intel for 19 years as a system hardware designer, a software designer, and a field accounts manager. He has previous experience with S3, Transmeta, Netlist, and SimpleTech. He has provided expert testimony in court cases involving patent disputes. Bill also speaks often at technology trade shows and offers training in memory technology. He holds a patent for a high-density memory module.

NVDIMM-N

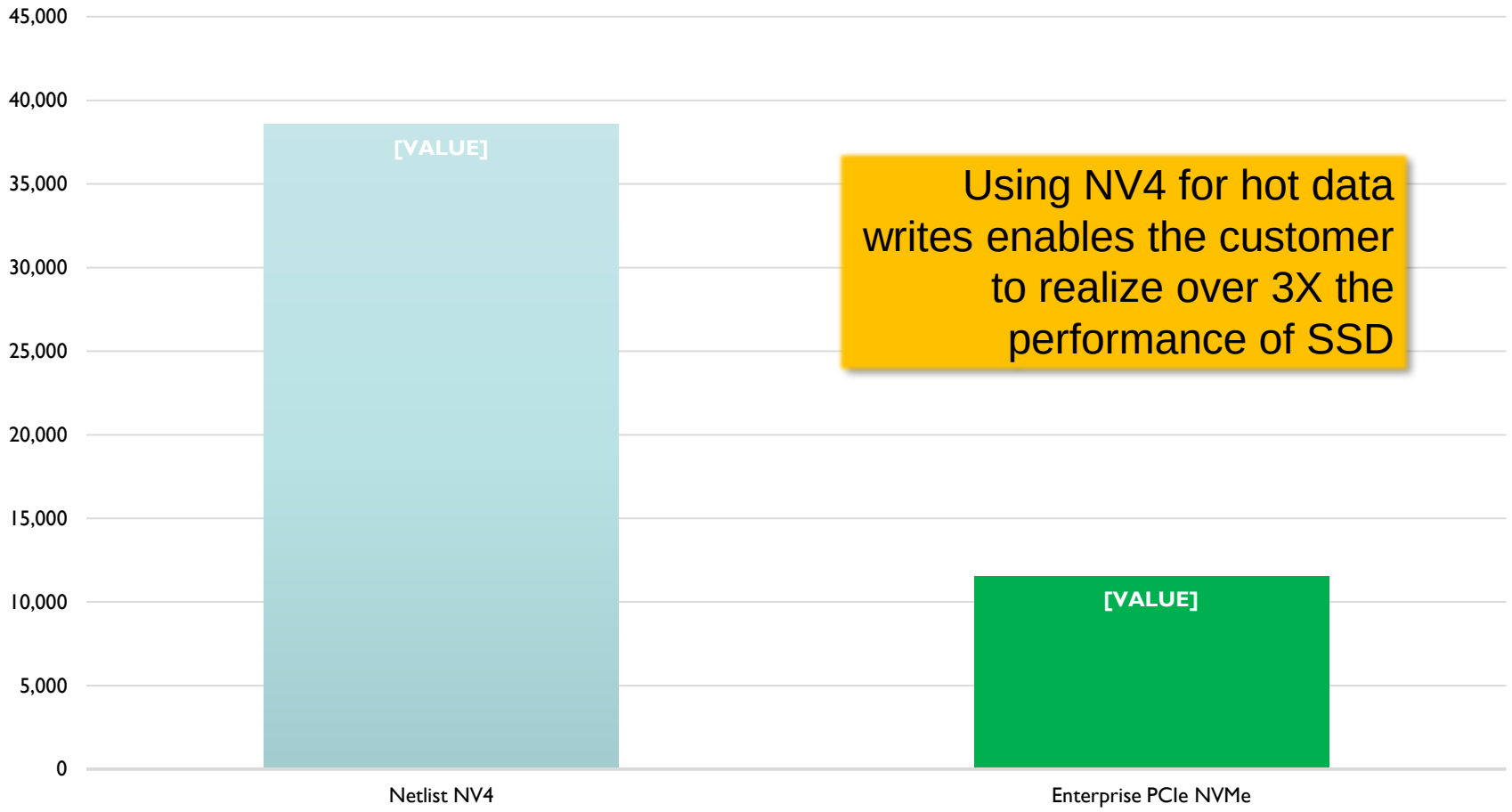
Mat Young

NVDIMM-N

- NVDIMM-N
- 8 GB and 16 GB Memory Capacities
- 2400 MT
- JEDEC Standard DDR4 RDIMM interface
- Fast backup and restore times < 1 min.
- Fast ‘reboot-ready’ recharge time < 4 min.

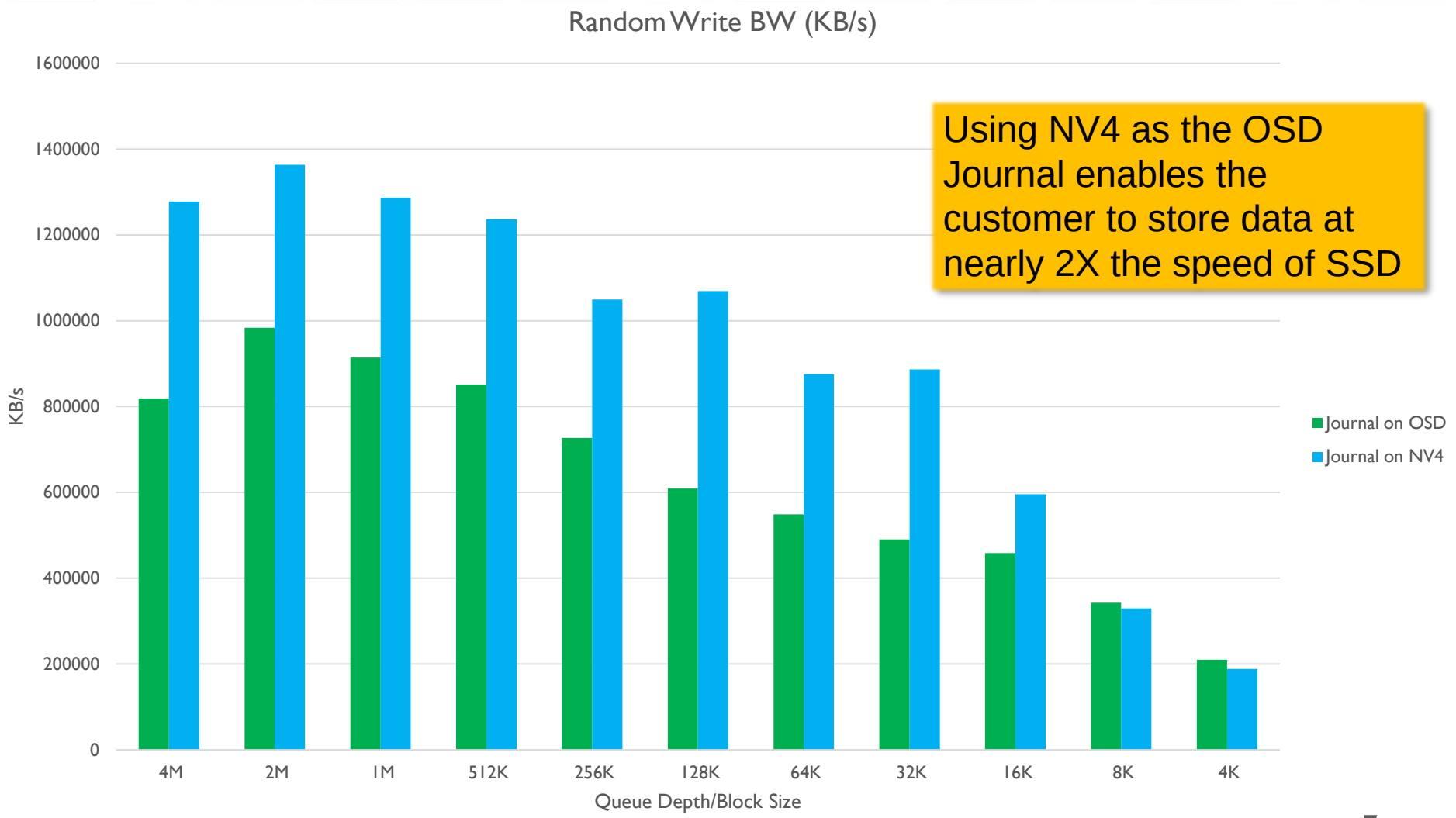


OLTP


Transactions Per Second
45,000
40,000
35,000
30,000
25,000
20,000
15,000
10,000
5,000
0
[VALUE]
[VALUE]
Netlist NV4
Enterprise PCIe NVMe
Using NV4 for hot data writes enables the customer to realize over 3X the performance of SSD

ceph - Improved Performance

SNIA™ Solid State Storage Initiative



Using NV4 as the OSD Journal enables the customer to store data at nearly 2X the speed of SSD

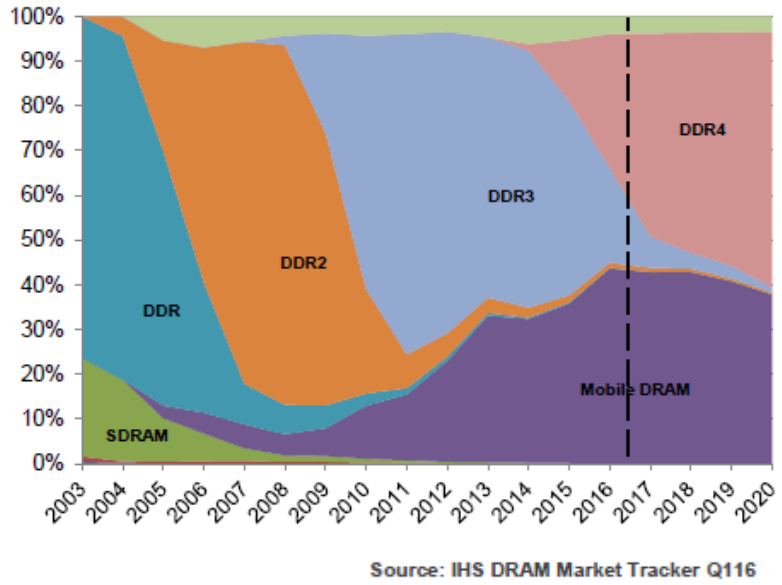
NVDIMM-F

Bob Frey

DDR3/4 NVDIMM(-N) Standardization

SNIA™ Solid State Storage Initiative

- DDR3 (2009-2014) ad hoc standardization
  - TEST/SAVE_n: pin 167
  - Intel MRC/BIOS – Vendor Specific
  - E820 Reserved
  - SM X9DRH-IF-NV
  - Linux NVDIMM Patch/ Windows Driver Kit



Source: IHS DRAM Market Tracker Q116

- DDR4 (2013-2019) electrical-mechanical (288-pin)
  - SAVE_n: pin 230
  - 12V: pin 1, 145
  - EVENT_n: pin 78
  - Byte Addressable I2C interface (JESD245)
  - Intel MRC/BIOS – JEDEC, ACPI 6.0 Type 12, NFIT, DSM
  - Linux 4.3/Windows Driver Kit

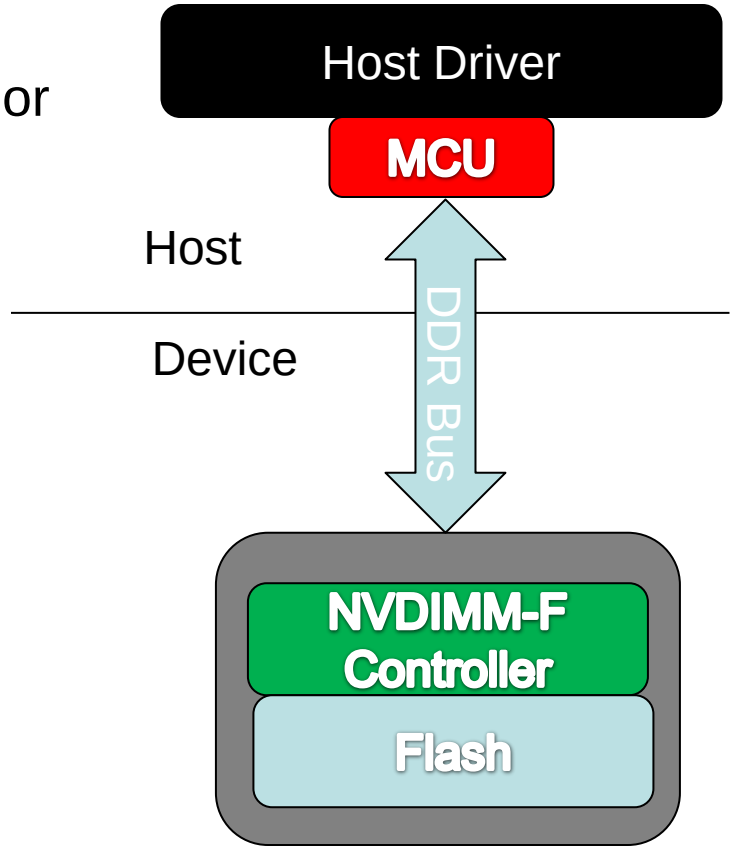
NVDIMM Definitions

NVDIMM Functions:
When used, these names infer a general function without specifying the interface type or form factor.

NVDIMM-N
Persistent DRAM using NAND Flash

NVDIMM-F
NAND Flash accessed as a block oriented device

NVDIMM-P
Combined persistent DRAM and block accessed NAND Flash



NVDIMM-F Motivation/Challenges

SNIA™ Solid State Storage Initiative

Storage Access Latency Hierarchy

- SRAM ~1X (10x faster)
- DRAM ~10X (10x faster)
- SCM ~100X (1000x faster)
- NAND ~100,000X (100x faster)
- HDD ~10,000,000X

Motivation

- Moving NAND to memory channel eliminates traditional HDD/SSD SAS/PCIe link transfer, driver, and software overhead. As storage latency decreases these factors dwarf the storage access percentage of an read/write.
- NAND low cost/high availability.

Challenges

- NAND 10,000x slower than DRAM. Attachment to memory channel must not interfere with DRAM performance.
- NAND block access vs. DRAM byte access.

NVDIMM-F: SPD Detection and Enumeration

SNIA™ Solid State Storage Initiative

SPD (Serial Presence Detect) EEPROM Module Type Key Byte 3

- 0x9M, NVDIMM
- Where M refers to the base memory architecture

Bytes 201~202 (0x0C9~0x0CA) (NVDIMM): Hybrid Module Media Types

These bytes define a media bit mask for all media types on the module. A setting of 1 in each bit position indicates the presence of that memory type on the module. The SPD is not considered as a media type in this context.

Byte 201		Byte 202	
Bit	**Media Type**	**Bit**	**Media Type**
0	Unknown/undefined	0	Reserved
1	SDRAM	1	Reserved
2	NAND Flash	2	Reserved
3	Reserved	3	Reserved
4	Reserved	4	Reserved
5	Reserved	5	Reserved
6	Reserved	6	Reserved
7	Reserved	7	Reserved

Examples:

Module Type	Media Types	Byte 201	Byte 202
NVDIMM-N	SDRAM, NAND Flash	0000 0110	0000 0000
NVDIMM-F	NAND Flash	0000 0100	0000 0000
NVDIMM-P	SDRAM, NAND Flash	0000 0110	0000 0000

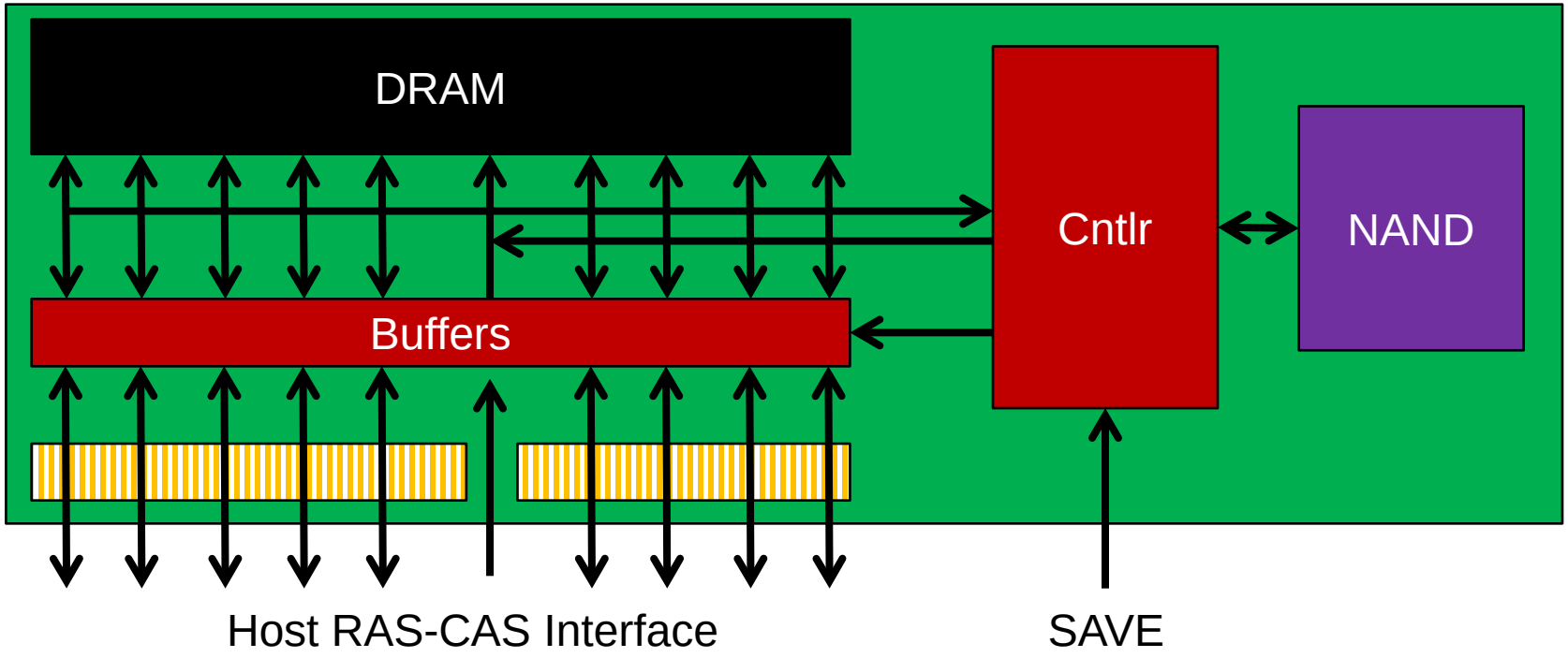
NVDIMM-F Looking Forward

- Standardized memory channel access protocol adopted by Memory Controller implementations.
- BIOS/MRC driver and MB support
- Performance validation
- Consult NVDIMM-F vendors

NVDIMM-P
A New Hybrid Architecture

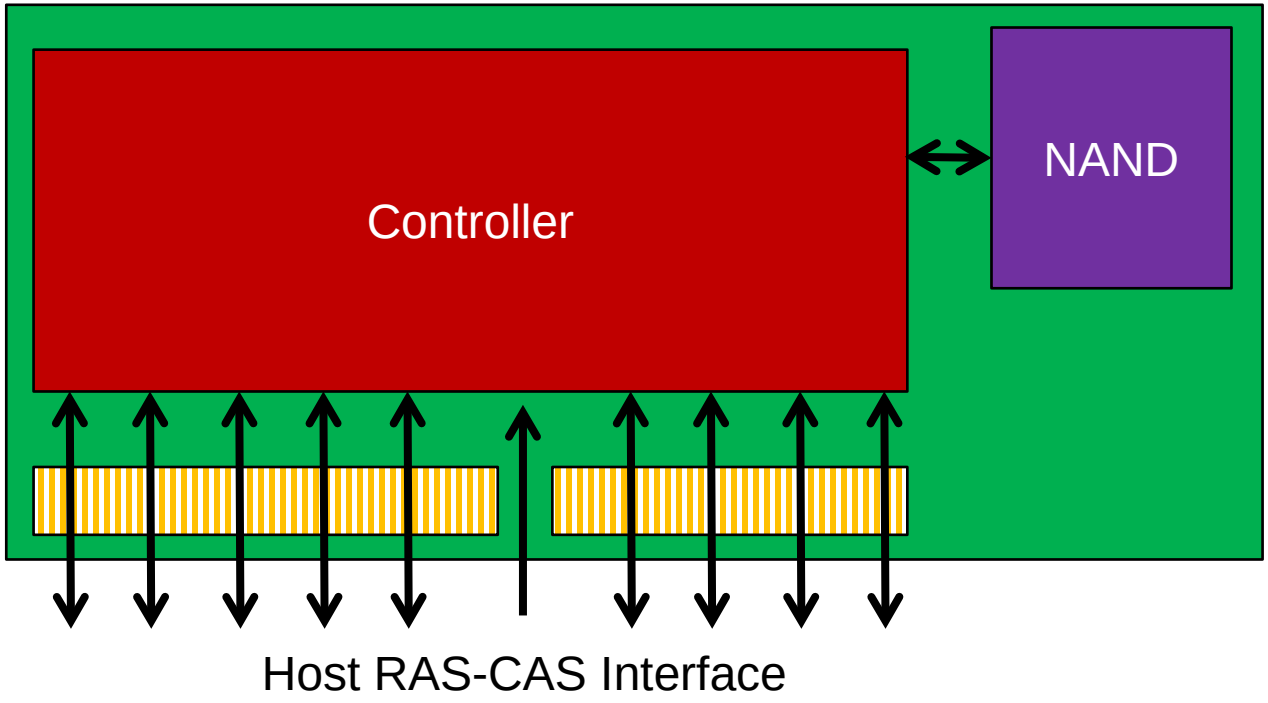
Bill Gervasi

NVDIMM-N: DRAM Persistence



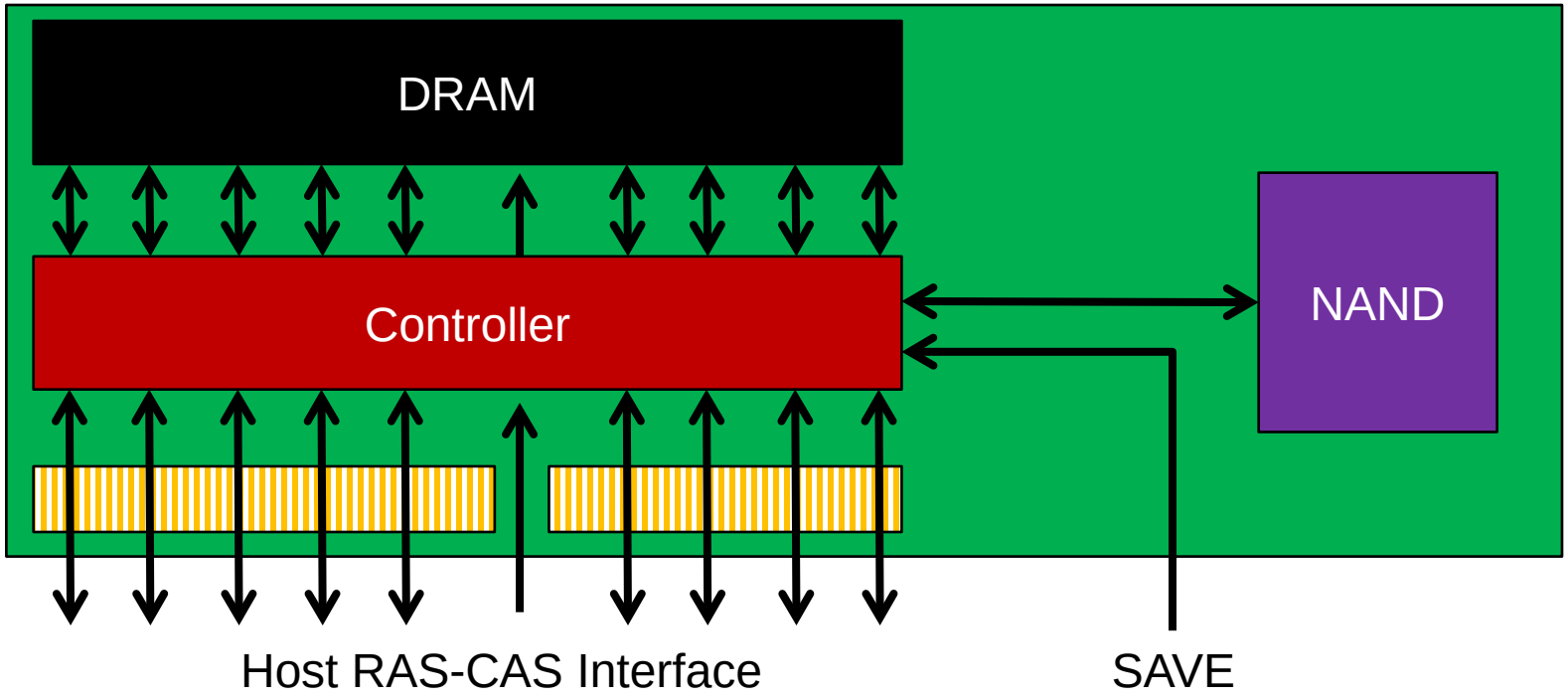
- DRAM accessed at DRAM speeds
- Contents saved to NAND on power fail
- Restored to DRAM when power resumes

NVDIMM-F:
Block Accessed NAND Flash



- No DRAM
- Flash accessed in native block format

NVDIMM-P
Combines DRAM & Flash

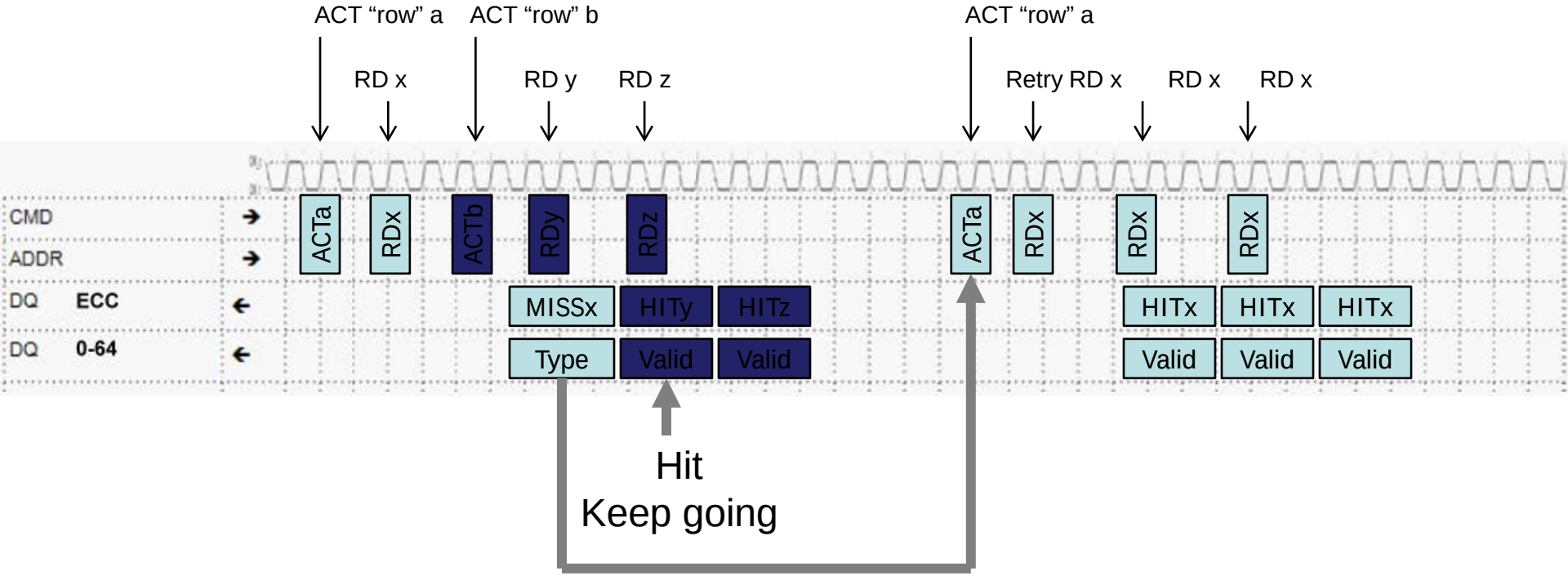


Variations Under Discussion

- NVDIMM-controlled cached interface
  - Host blindly requests access
  - On cache hit, request satisfied immediately
  - On cache miss, request postponed + retried
- Host-directed cache interface
  - Host memory tracks NVDIMM cache
  - Forces cache fill & replacement
  - No cache misses on DDR4/DDR5 bus

NVDIMM-P Cache Control – READ

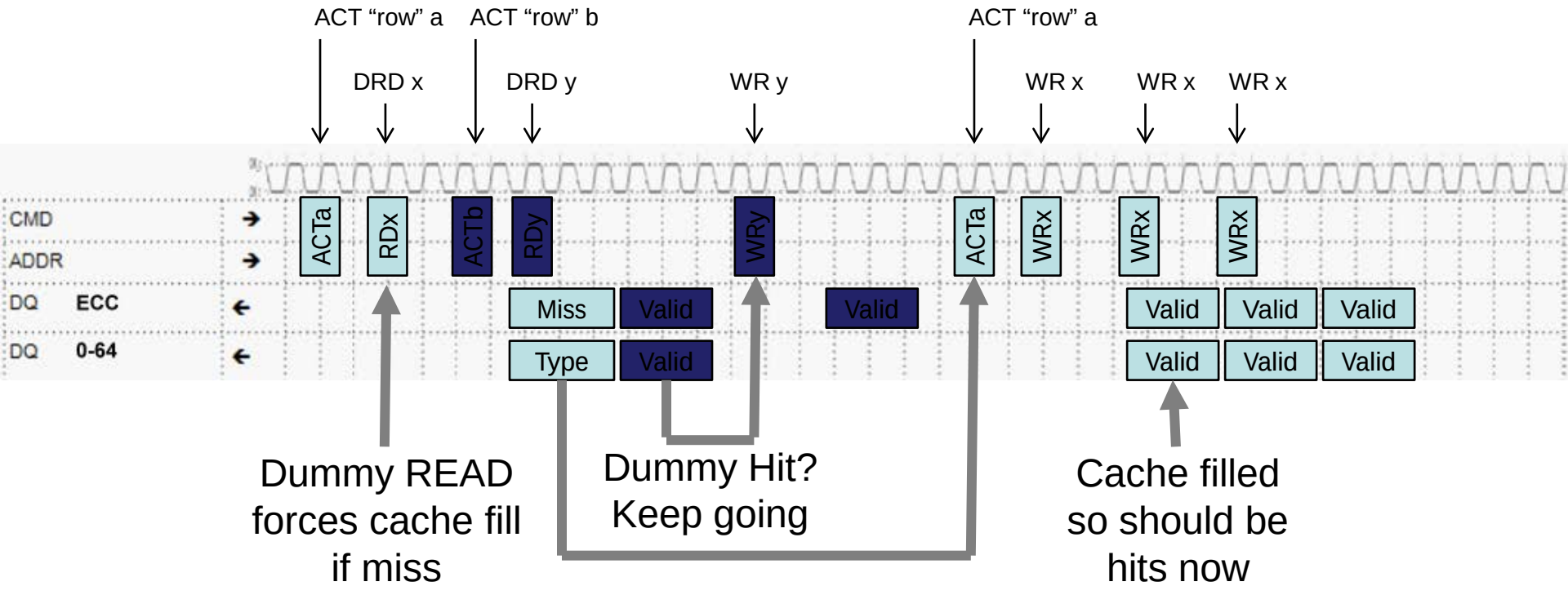
- Data response for READ includes cache hit/miss in ECC
- Retry only needed on miss
- O-o-O transactions allowed by simple try-retry



- Miss – Retry
- "Miss Type" can include hint how long to wait for retry

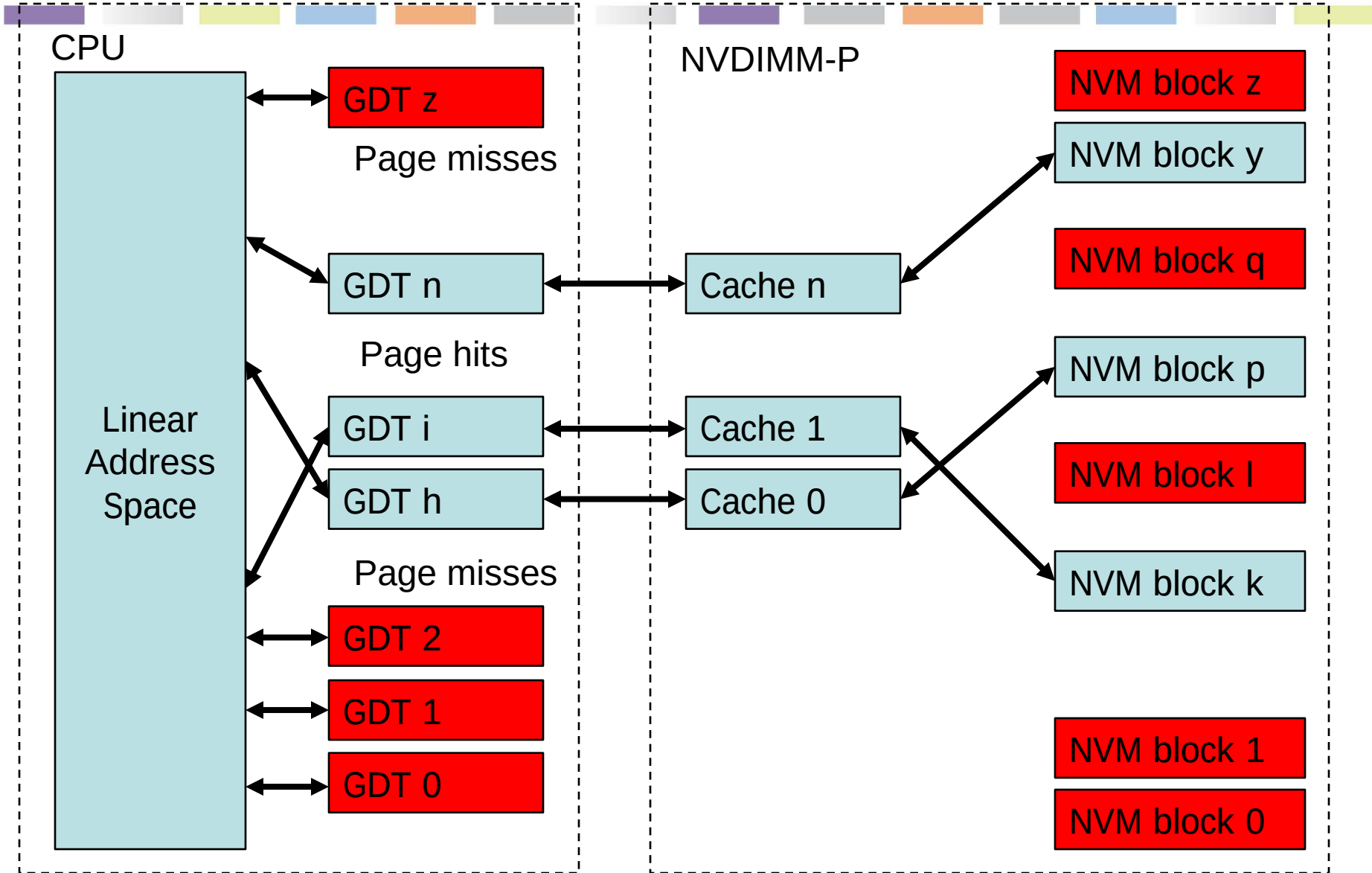
NVDIMM-P Cache Control – WRITE

- Dummy READ to check cache status, force cache fill
- Dummy READ gets valid reply? Cache hit, go ahead
- Subsequent WRITEs do not require ACK



- Miss – Retry
- "Miss Type" can include hint how long to wait for retry

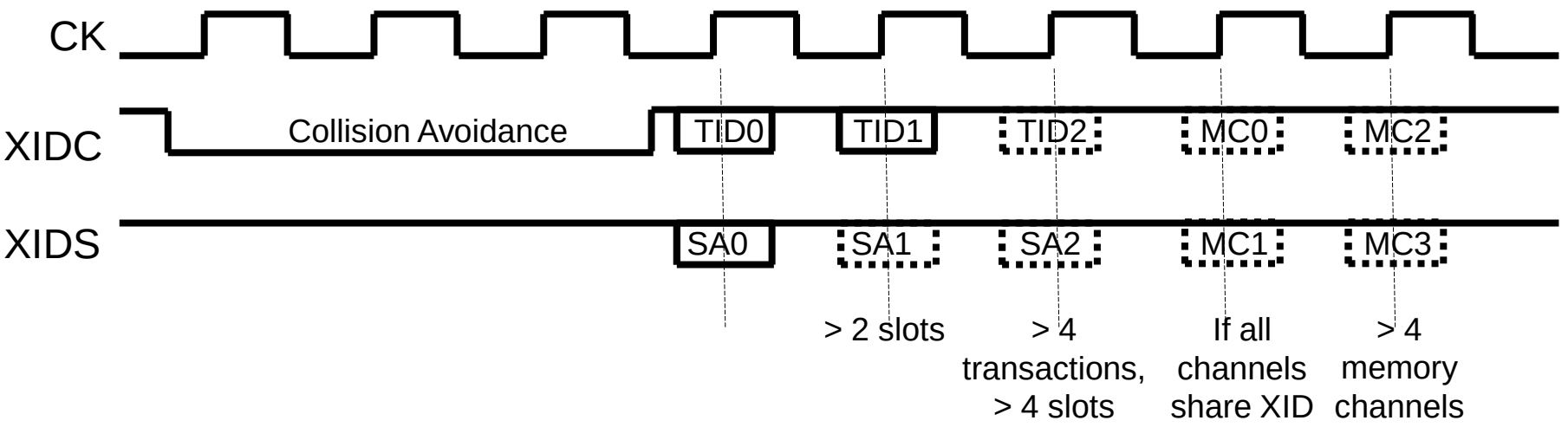
Host Directed Cache Control – R/W



Optimizing for NVDIMM

- Polling on cache misses consumes bus bandwidth
- NVM activities are non-deterministic
  - Cache replacement
  - Error scrubbing
  - Wear leveling
- Sideband signals may be used to speed up response to miss
- Support for out-of-order needed

Considering Sideband ID Bus



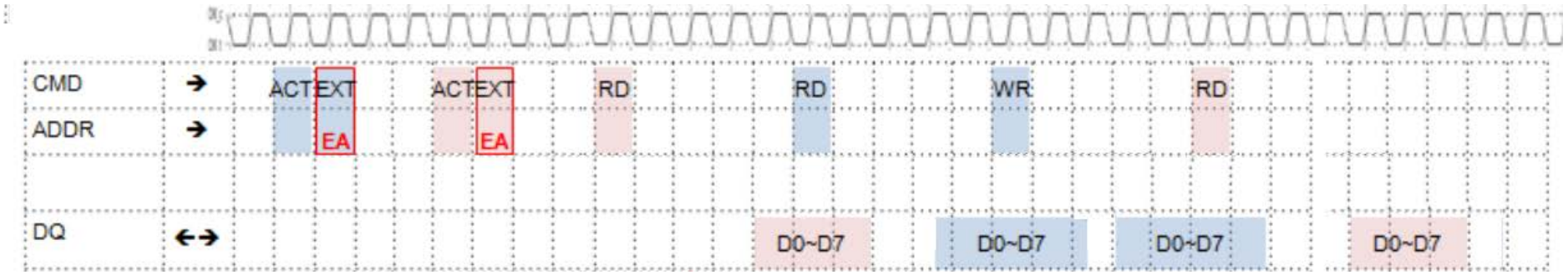
- Host waits for XID on cache miss
- XID bus indicates completed O-o-O transaction

Linear Addressing

- Current DDR4 bus limits DIMM capacity to 128GB of linear mapped memory
- Desire to put multiple TB on an NVDIMM
- Transaction IDs for O-o-O included

Considering Extended Addresses

Function	CKE Previous Cycle	CKE Current Cycle	CS_n	ACT_n	RAS_n / A16	CAS_n / A15	WE_n / A14	BG0 - BG1	BA0 - BA1	C2-C0	BC_n / A12	A17	A13	A11	A10 / AP	A9 - A0
Bank Activate [ACT]	H	H	L	L	RA			BG	BA	V	RA					
Bank Activate Extension [EXT]	**H**	**H**	**L**	**H**	**L**	**H**	**H**	V	V	V	V	V	V	V	H	EA



- Increases total memory capacity by up to 2^{15}

NVDIMM-P Summary

- NVDIMM-P definition in discussion
- Existing DDR4 protocol supported
- Extensions to protocol under consideration
  - Sideband signals for transaction ID bus
  - Extended address for large linear addresses

Thank You!

SNIA
Solid State Storage Initiative