Hewlett Packard Enterprise

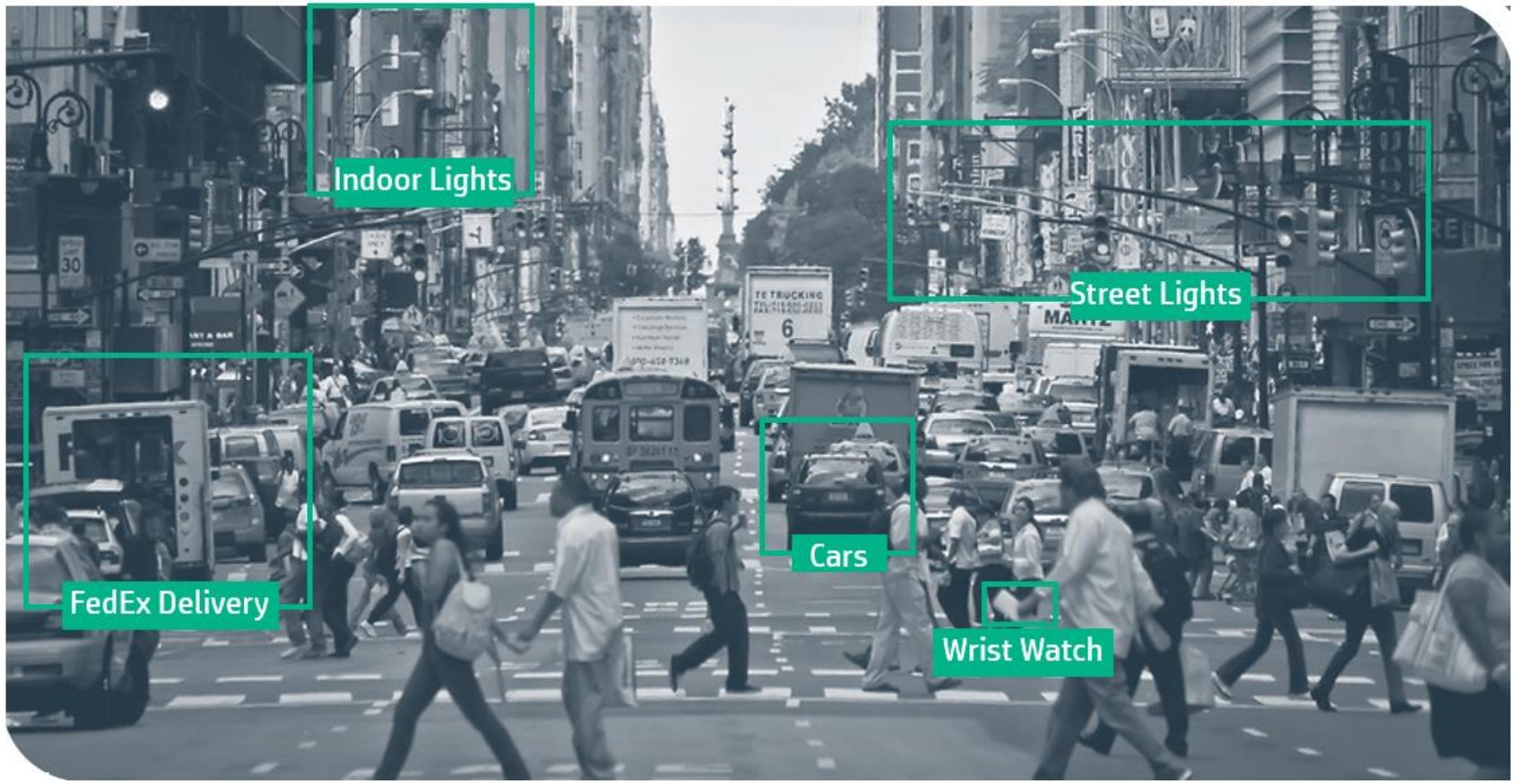
The role of NVM in the future of HPC workloads

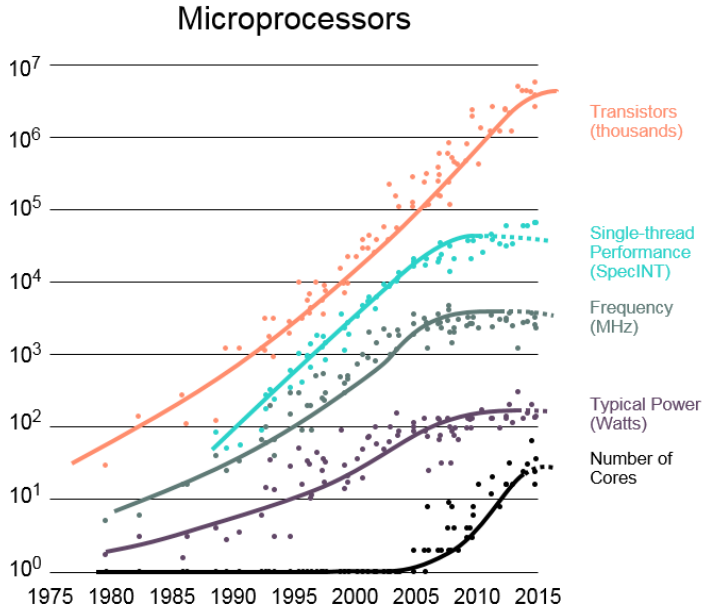
29 05 2017

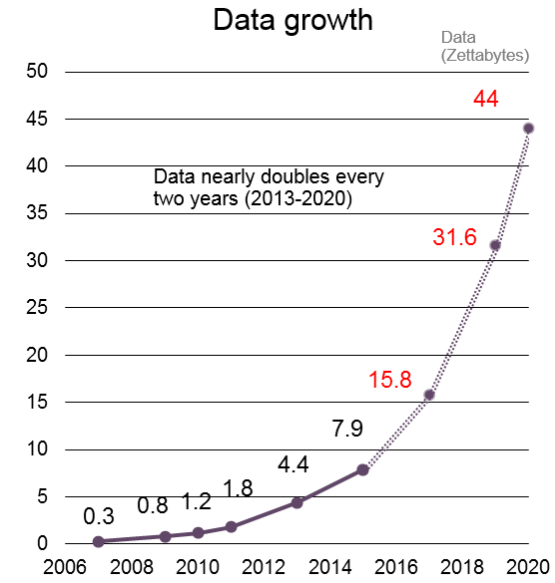
patrick.demichel@HPE.com

Distinguished Technologist EG EMEA

The New Normal: Compute is not keeping up with data explosion



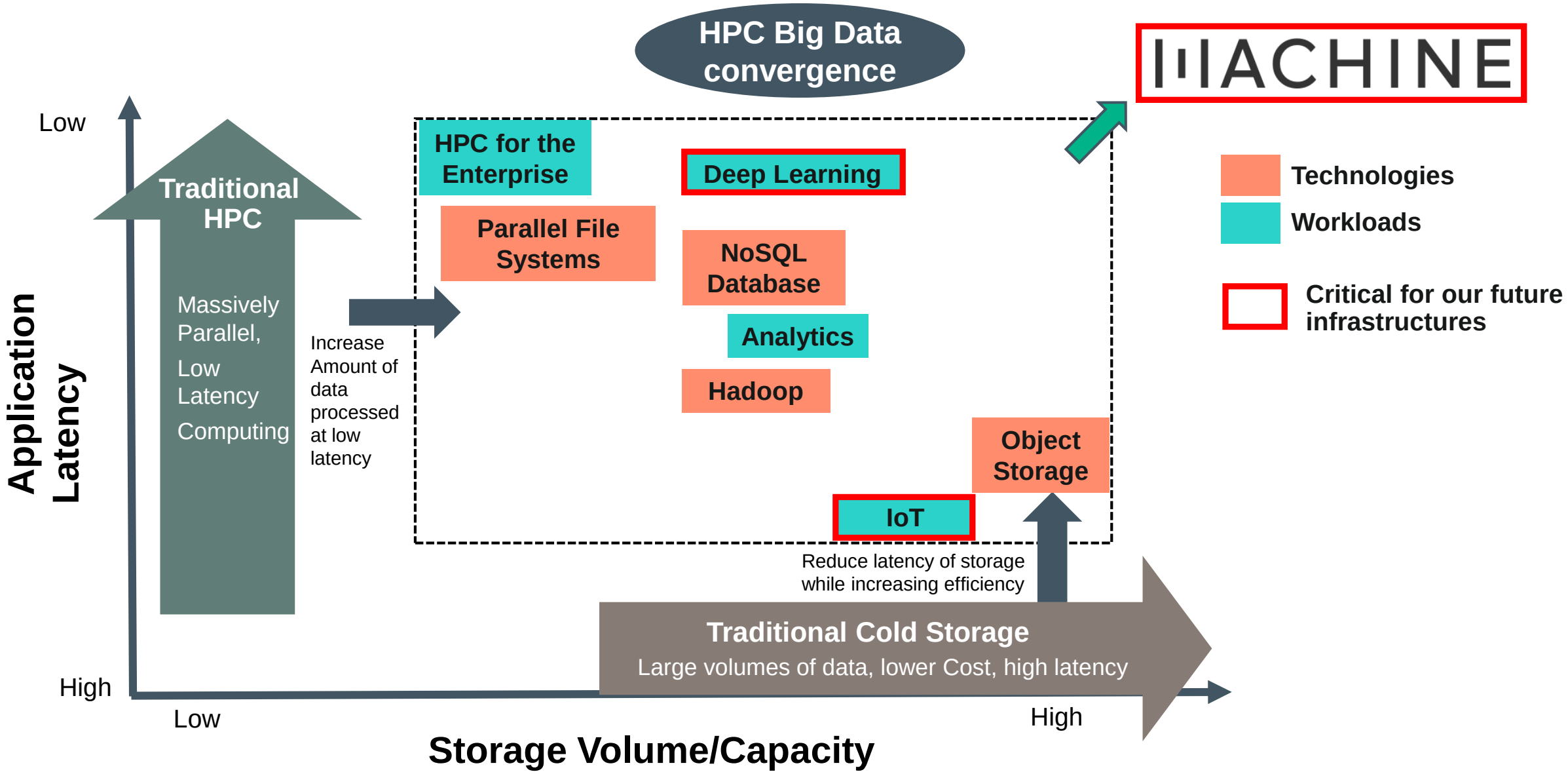




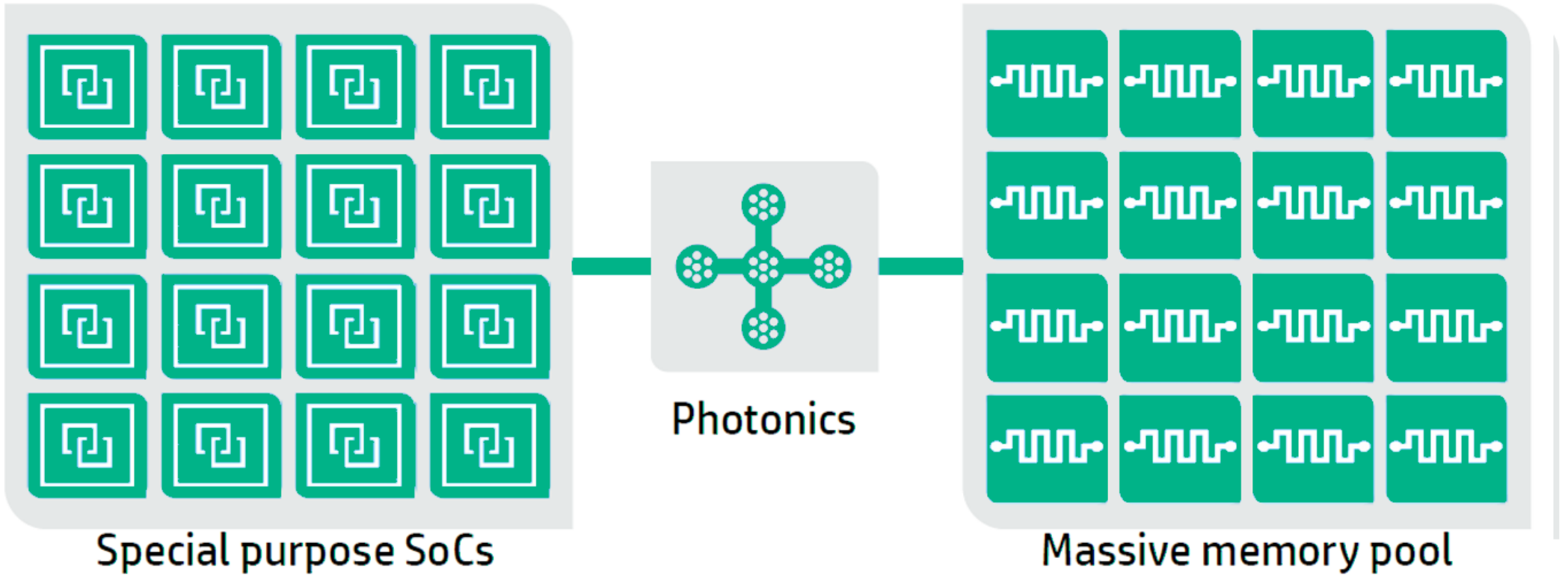
The end of scaling at just the wrong time …

(**8B** people + **20B** mobile devices + **100B** social infrastructure + **1T** apps)

HPC and Big Data technology context and The Machine



3 disruptive technologies to the rescue

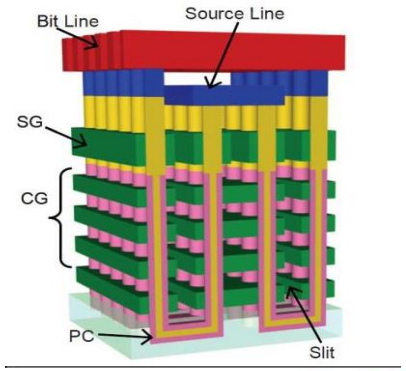


Electrons **Photons** **Ions**

Hewlett Packard Enterprise

Disruption #1: Non-volatile memories

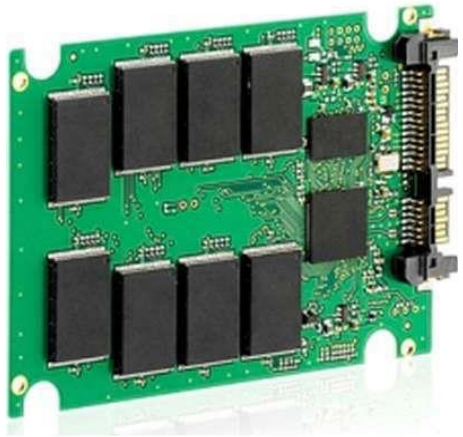
NVM and high speed memories are critical for extreme computing

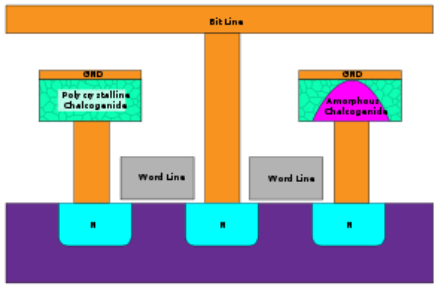


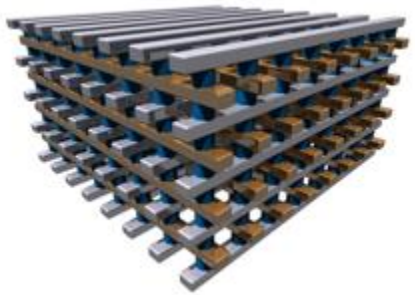
Reaching the physical limits of charge storage

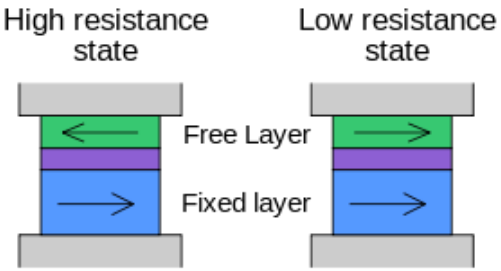
Non-Volatile memories – forms of memristor (Type 4)

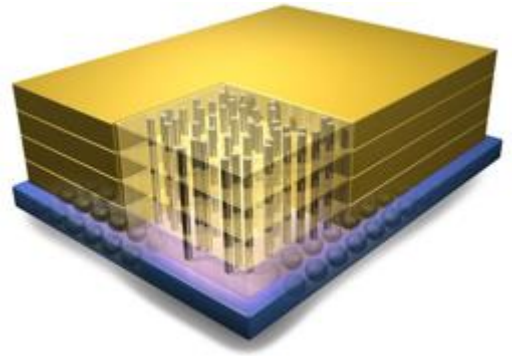
PCRAM **RRAM** **STTRAM** **HMC**











Technology	Density (μm²/bit)	Bandwidth (GB/s)	Latency Read (ns)	Latency Write (ns)	Energy Read (pJ/b)	Energy Write (pJ/b)
Hard Disk	N/A	0.5	3,000,000	3,000,000	2500	2500
Flash SSD [3] [6]	0.0021	1.0	25,000	200,000	250	250
DRAM [6] [30]	0.0038	51.2	55	55	24	24
PCRAM (22nm) [30]	0.0058	variable	48	150	2	19.2
Memristor (22nm) [8]	0.0048	variable	100	100	1-3	1-3

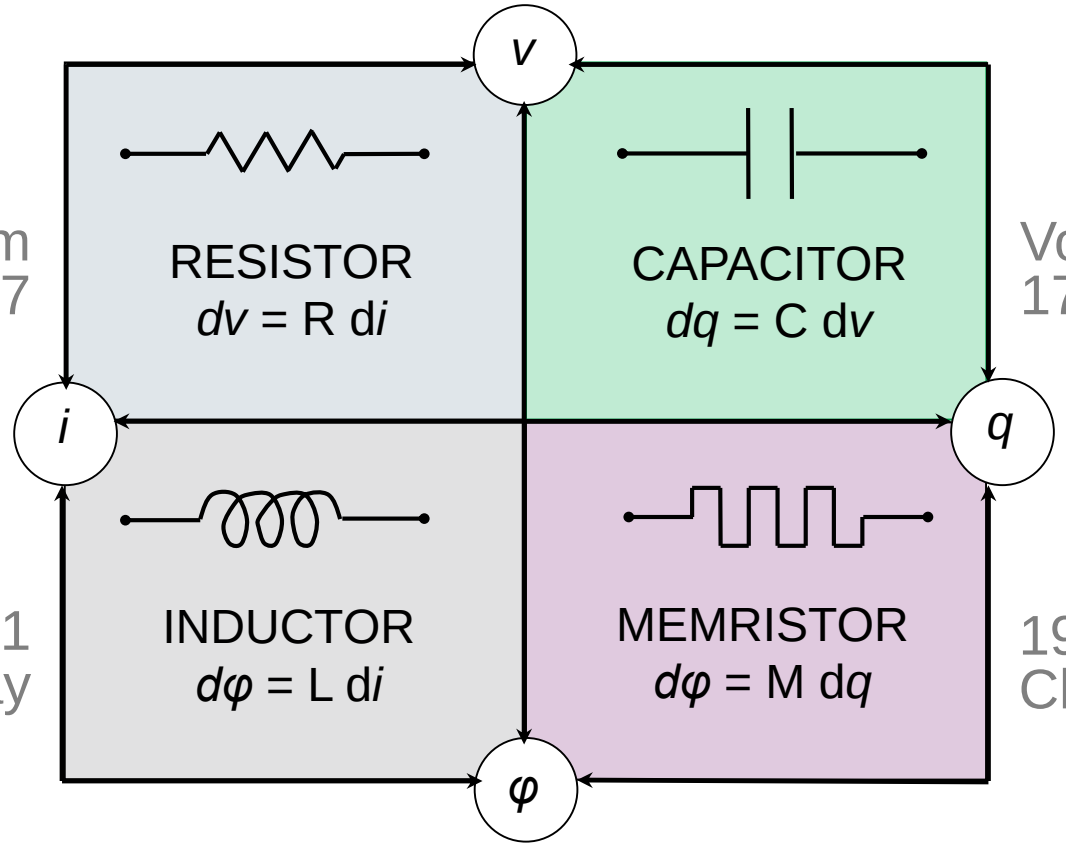
3D Flash

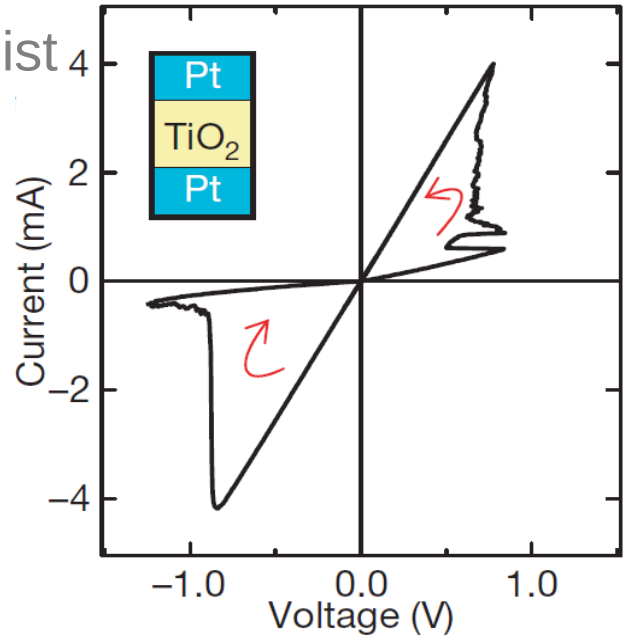
DRAM

The memristor : 4th fundamental cuircuit element

Predicted 1971
Leon Chua
U.C. Berkeley

Reduced to practice 2008
R. Stanley Williams
HP Laboratories





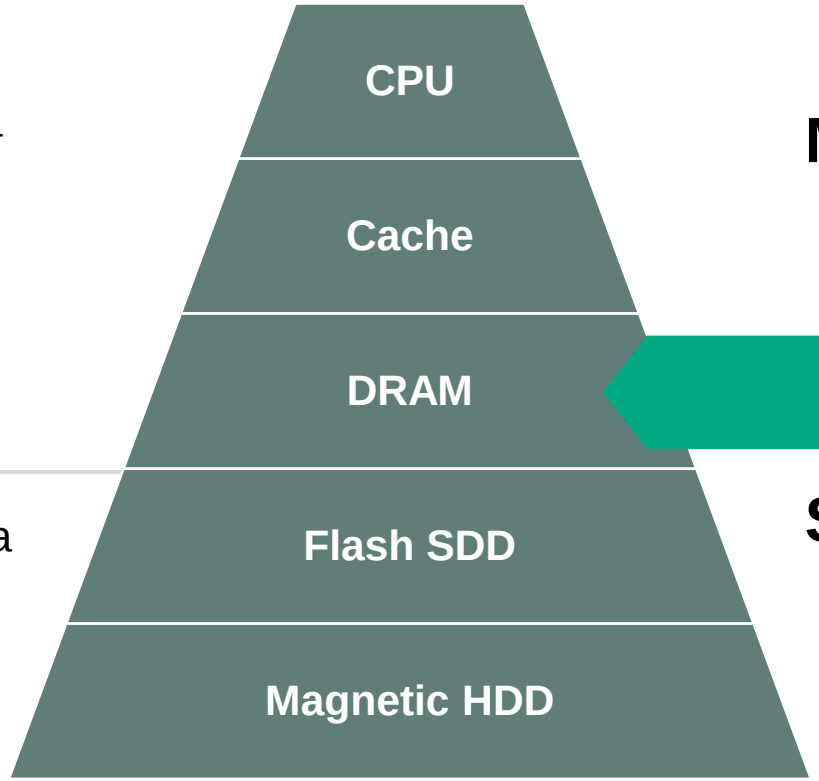
Convergence of memory and storage

Persistent Memory = The speed of memory with the persistence of storage

Load-Store Data

Low Latency

Volatile

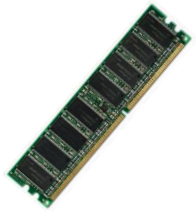


Block-Store Data

Long Latency

Non-Volatile

Memory



Persistent Memory

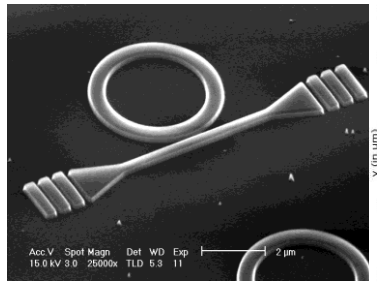
Storage

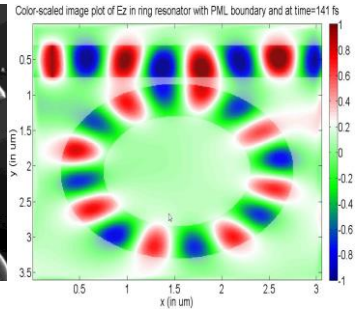
Speed

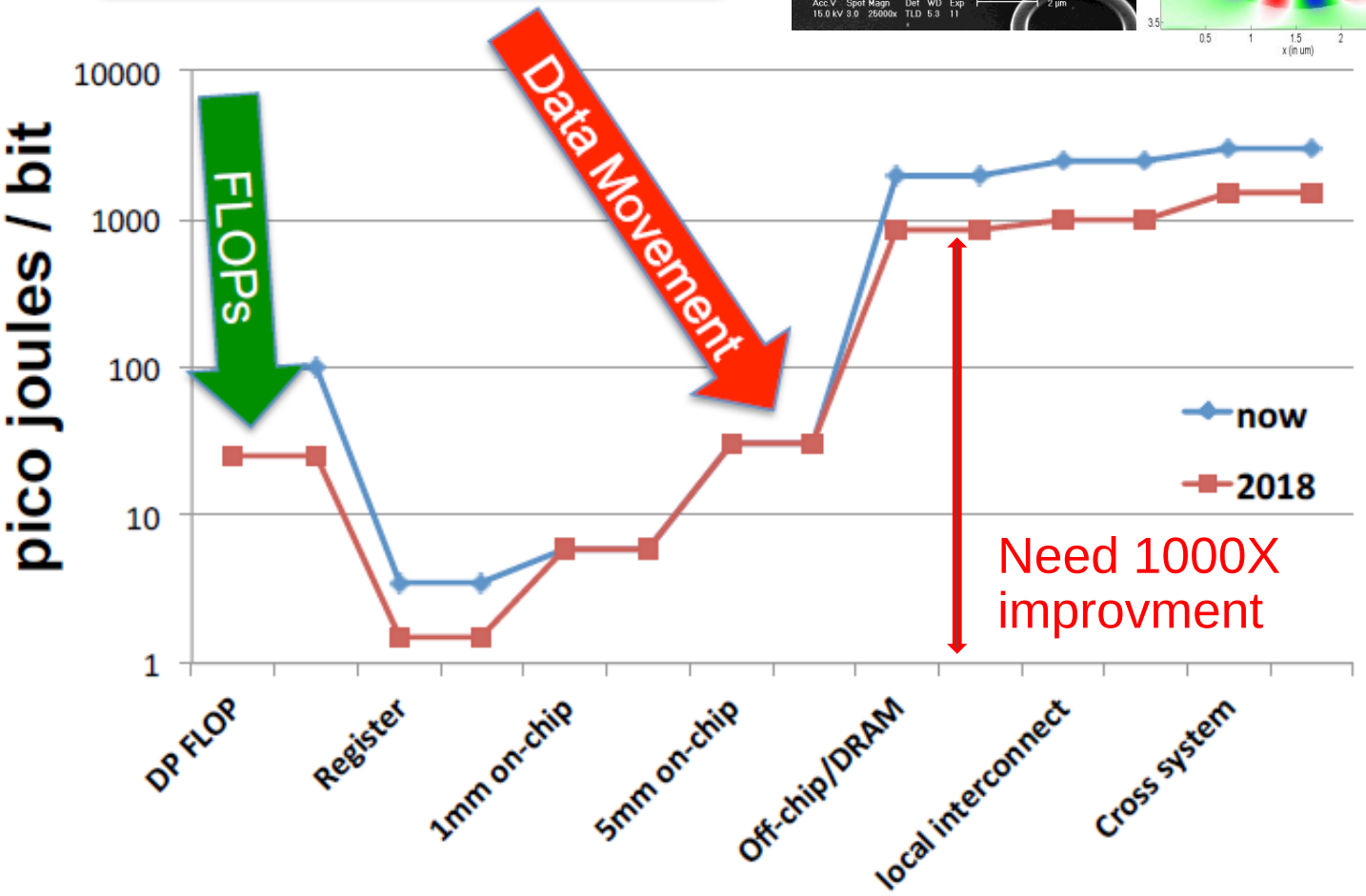
Persistence

Disruption #2 : Photonics

FLOPS will cost less power than on-chip data movement







Photonics

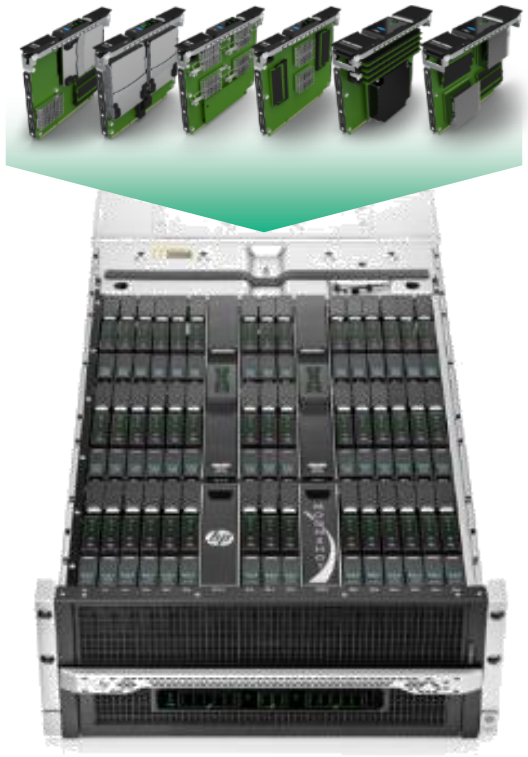
10^18 ops
*****
1Byte/ops
=
10^19bits
*****
1pj / bit
=
10MWatt

- ultra low energy
- low uniform latency
- high bandwidth
- low cost

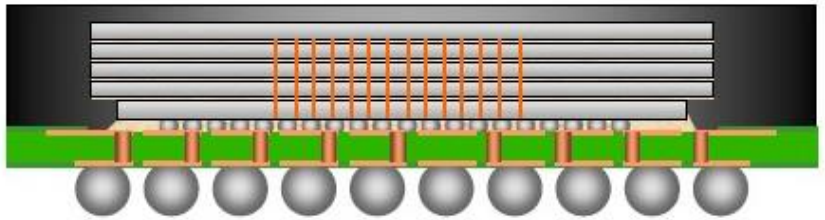
Katharine Schmidtke, Finisar

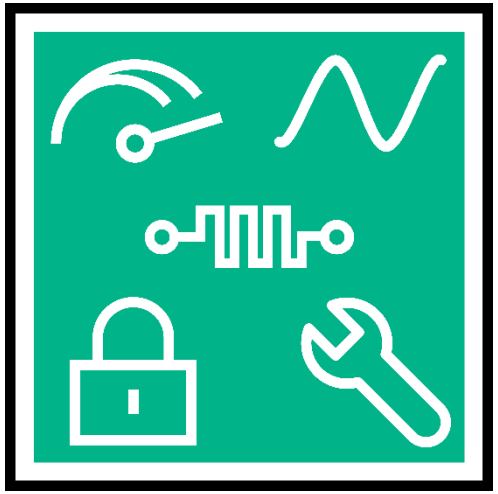
Disruption #3 : optimized architectures

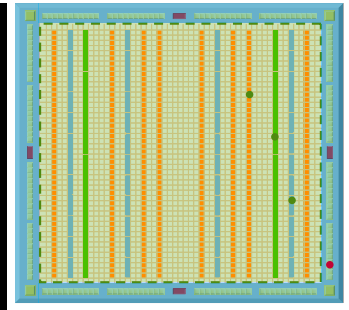
Special Purpose Cores

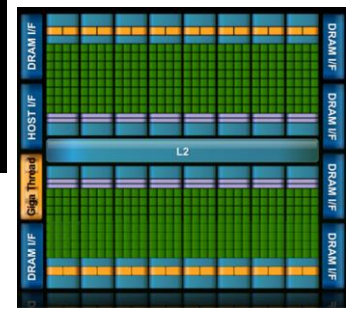


Reduced cost
Less energy
Less space
Less complex







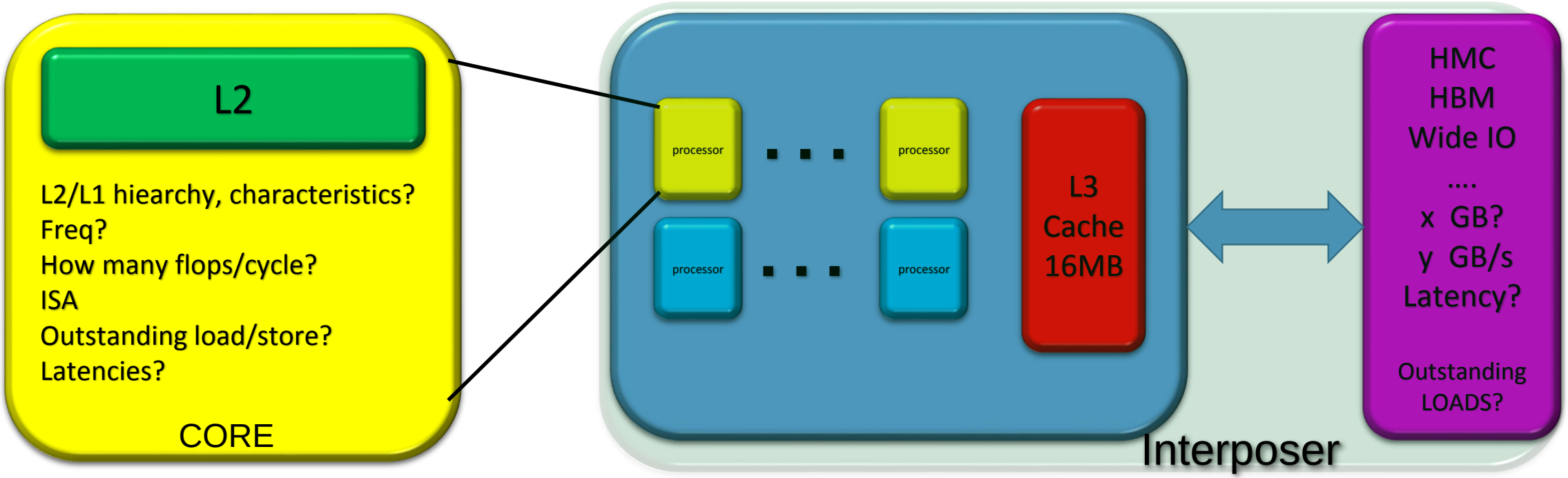


Integrated into a single package

Extreme scale compute requires ultra cheap and ultra efficient technologies

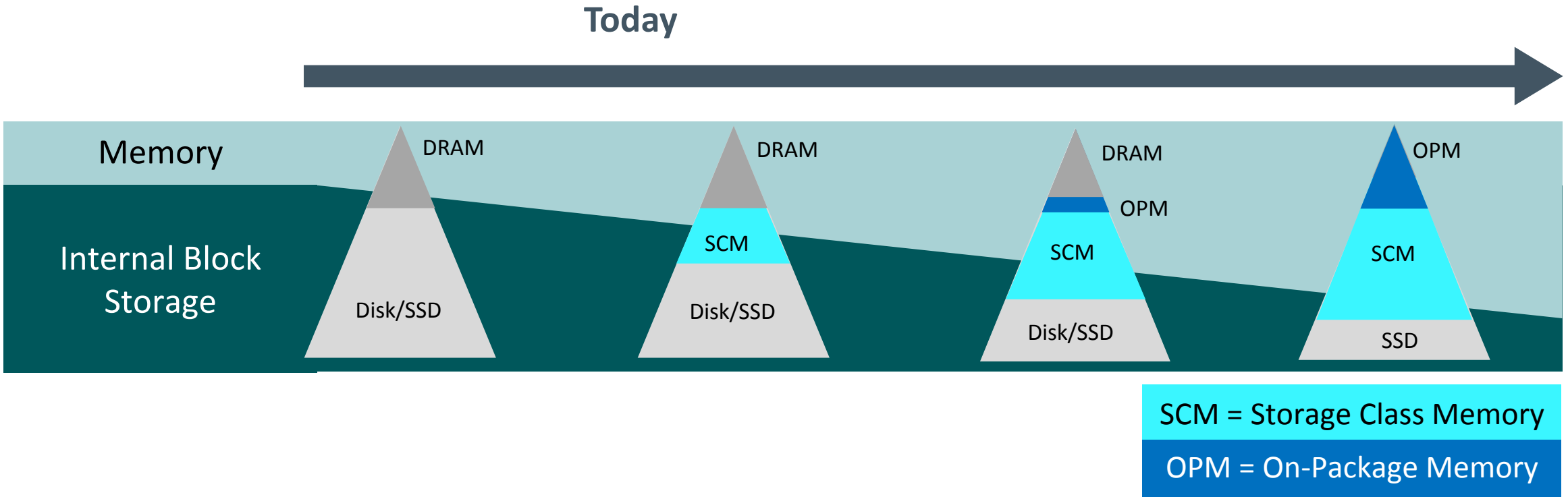
Challenges today : complexity to codesign hard+soft+integration in ecosystem

Future processors



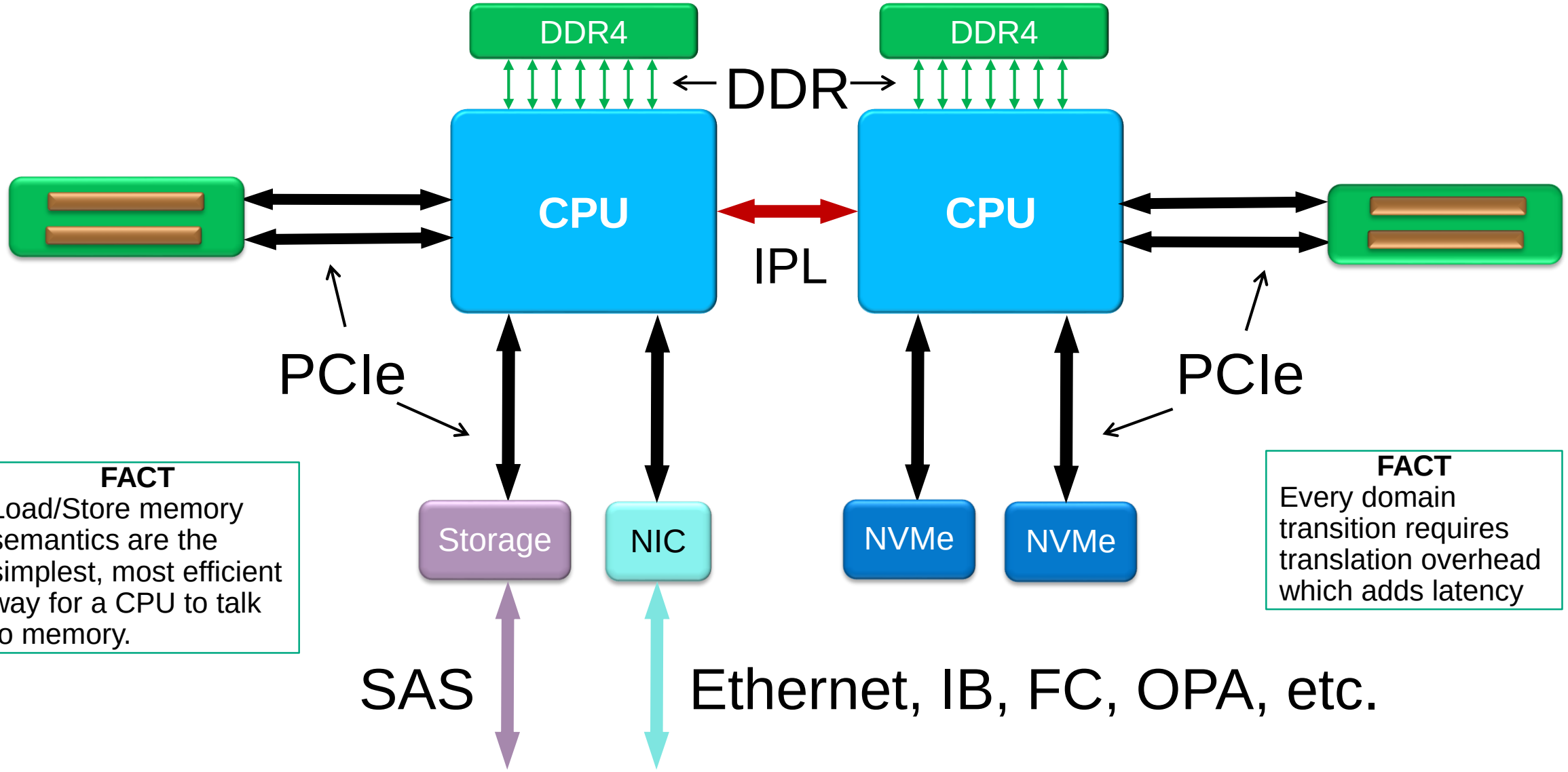
In future processors will have many cores and many heterogenous accelerators either GPU/FPGA/ASIC
We will have a high speed L4 cache of few GB
L3 shared and L2 private will have much smaller capacity in MB range
This High Speed Memory will be X times more rapid than higher levels , but will be limited to tens of GB
A very large pool of NVM will be attached and shared thru gen-Z
We can also have a very large local NVM pool to mitigate IO,latency,costs,... can be in TB

Memory/Storage Convergence: The Media Revolution



Memory Semantics is becoming pervasive in Volatile ***AND*** Non-Volatile Storage as these technologies continue to converge.

Typical 2 socket architecture – 8 possible interconnect types



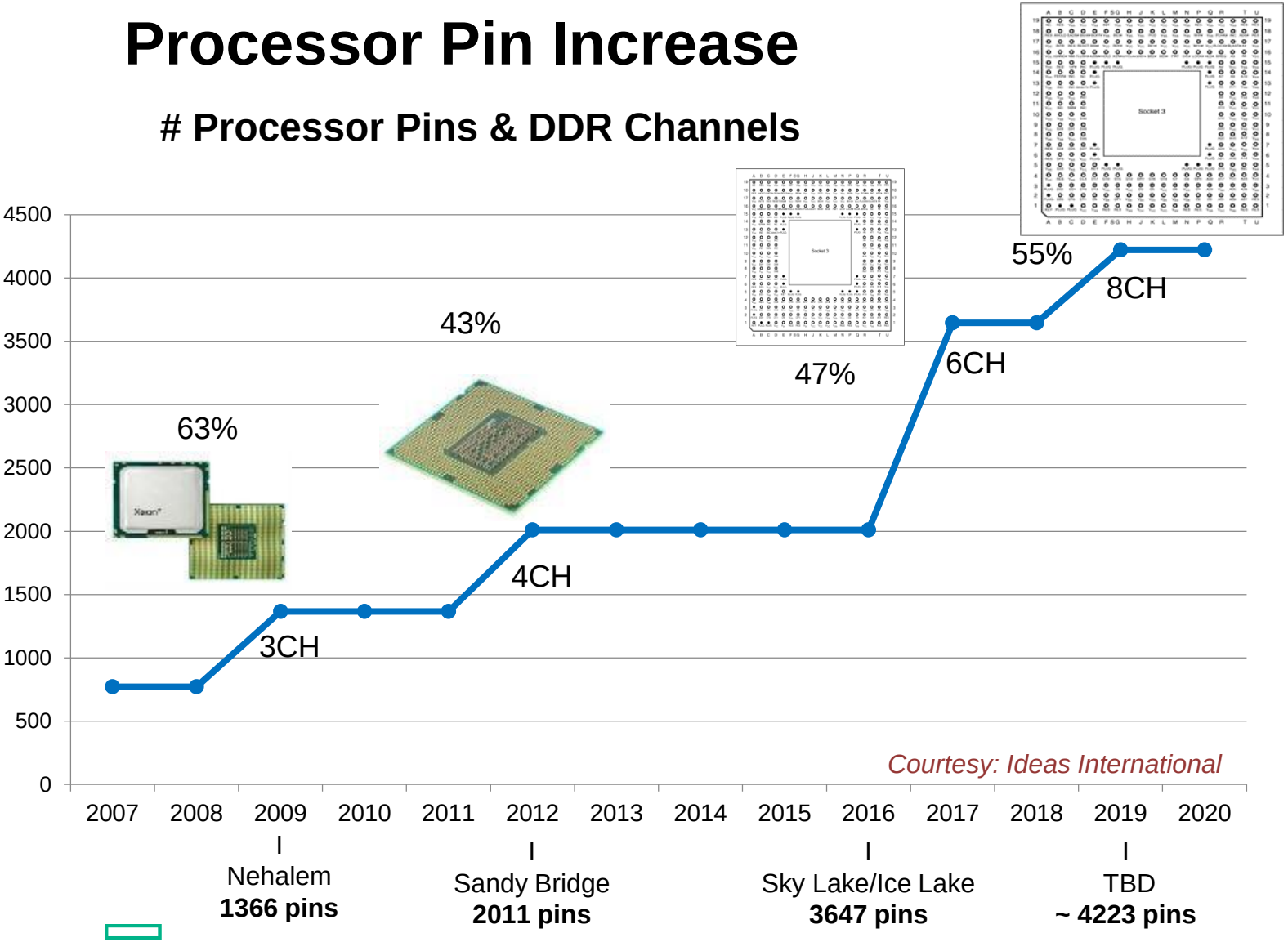
FACT
Load/Store memory semantics are the simplest, most efficient way for a CPU to talk to memory.

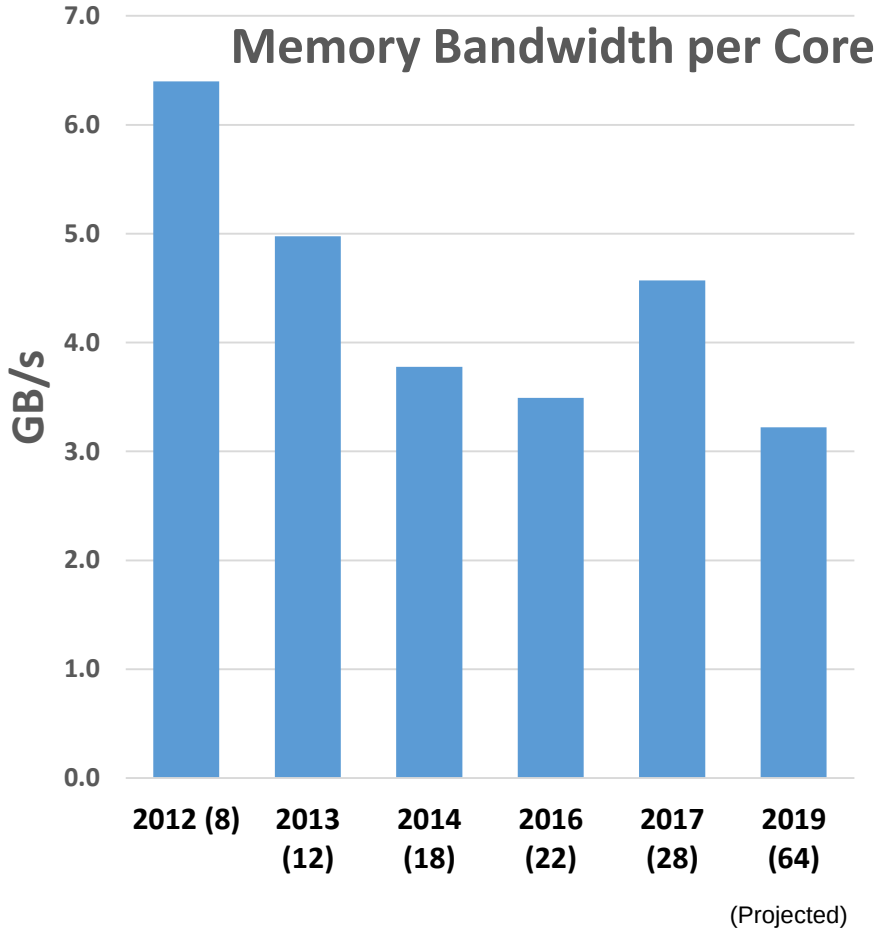
FACT
Every domain transition requires translation overhead which adds latency

Hewlett Packard Enterprise

Processor Pin Increase

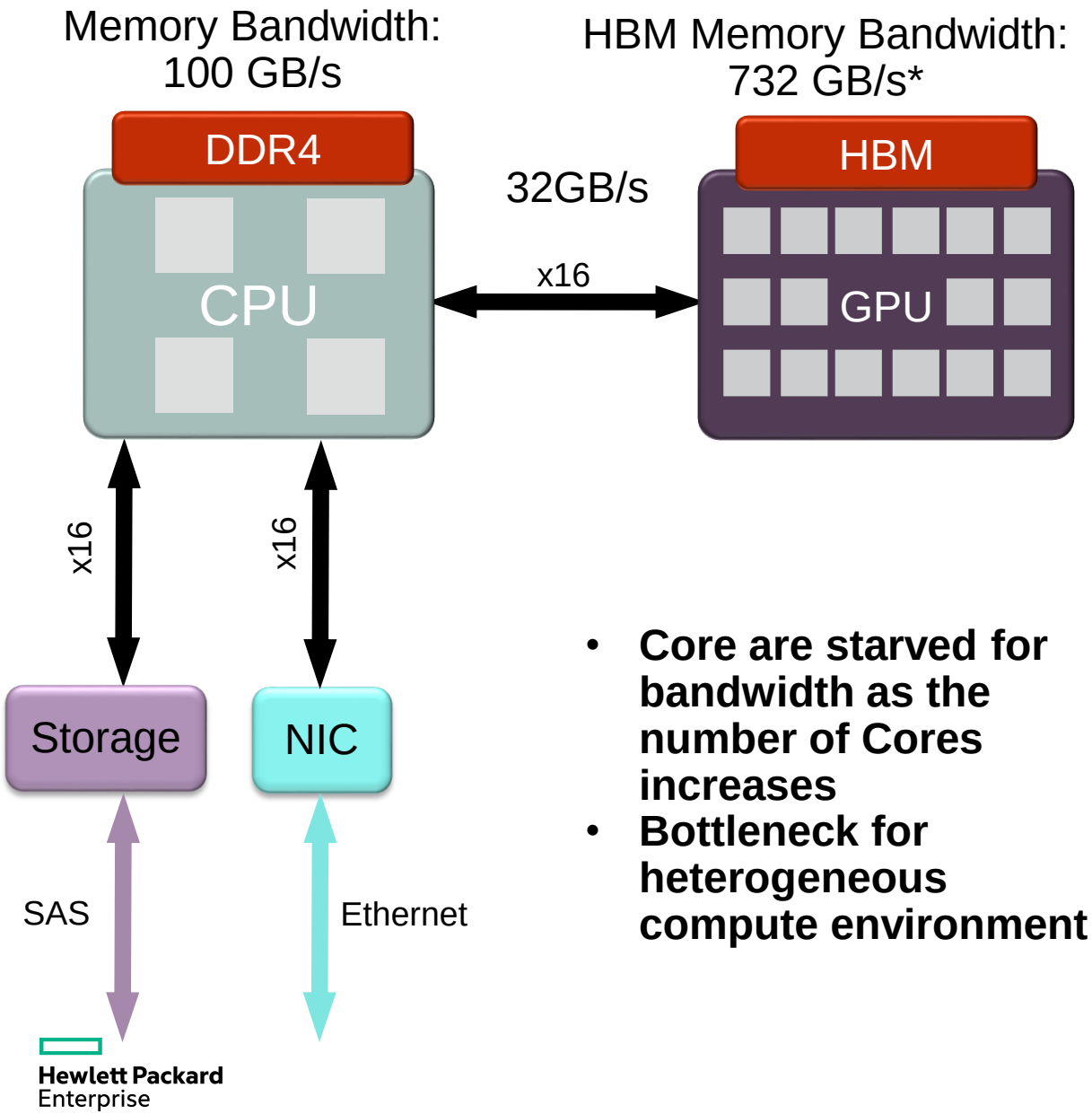
Processor Pins & DDR Channels



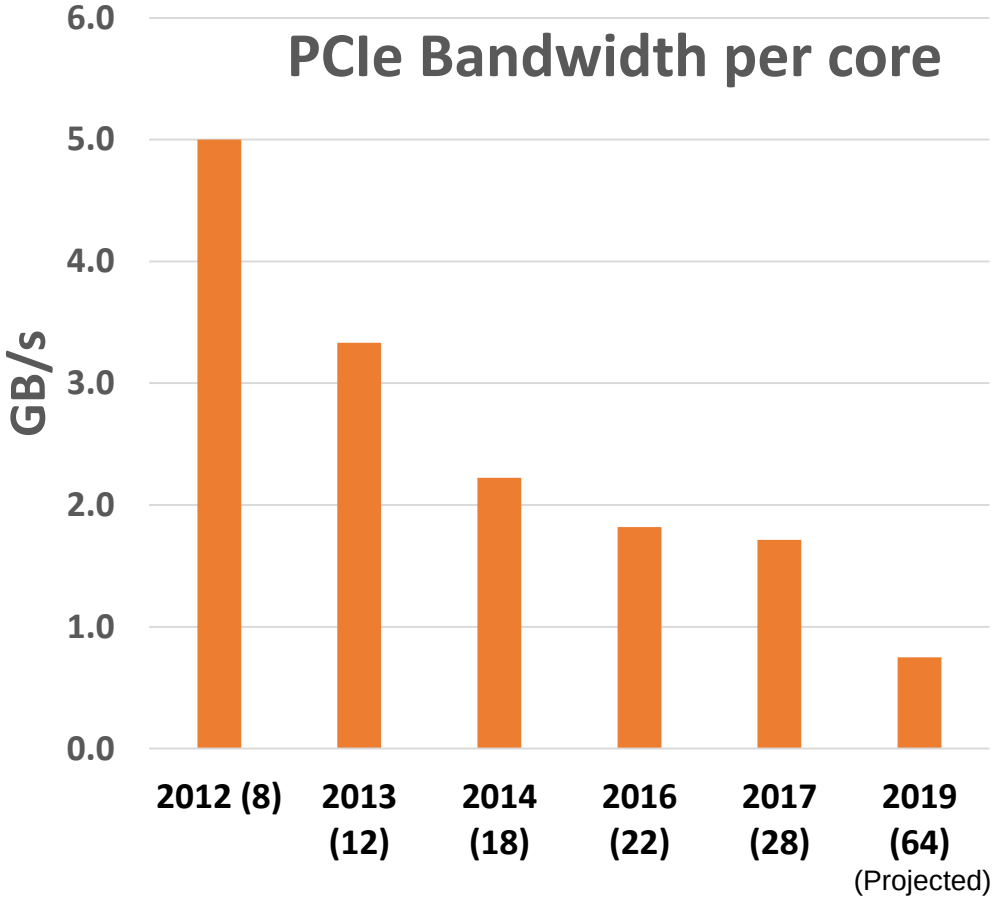


Hewlett Packard Enterprise

Architectural Limitations



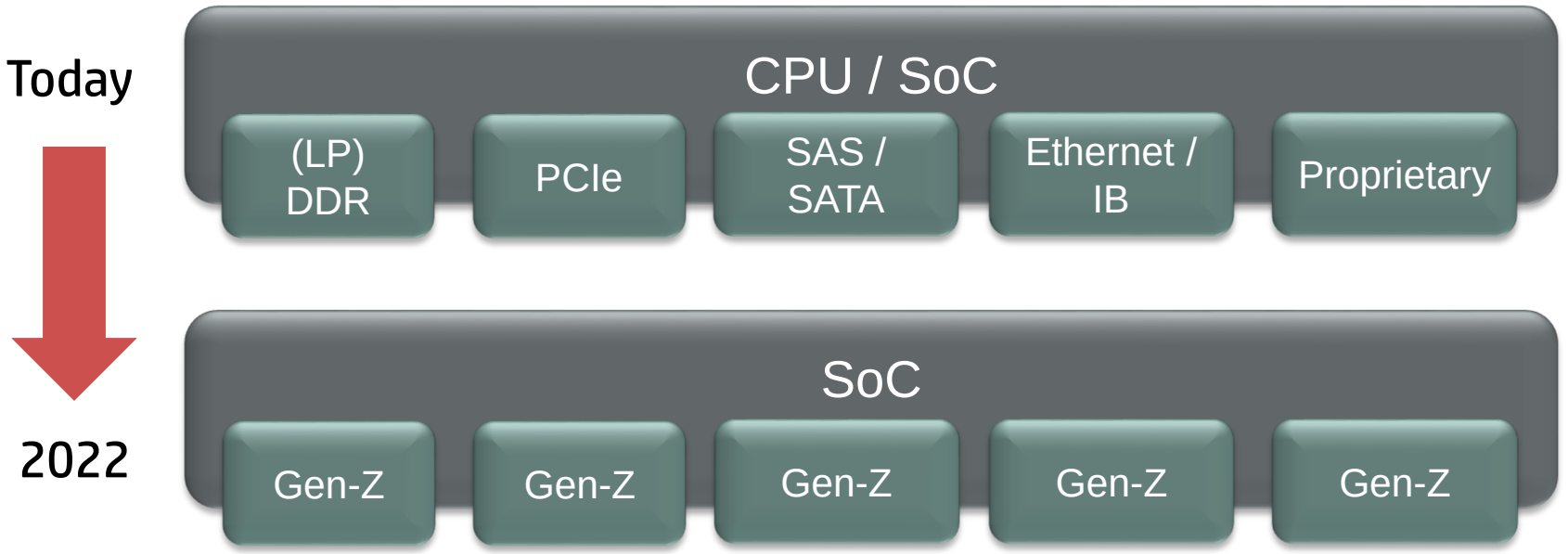
- **Core are starved for bandwidth as the number of Cores increases**
- **Bottleneck for heterogeneous compute environment**



* NVIDIA Tesla P100

Hewlett Packard Enterprise

Drive a New and Open Protocol: Gen-Z



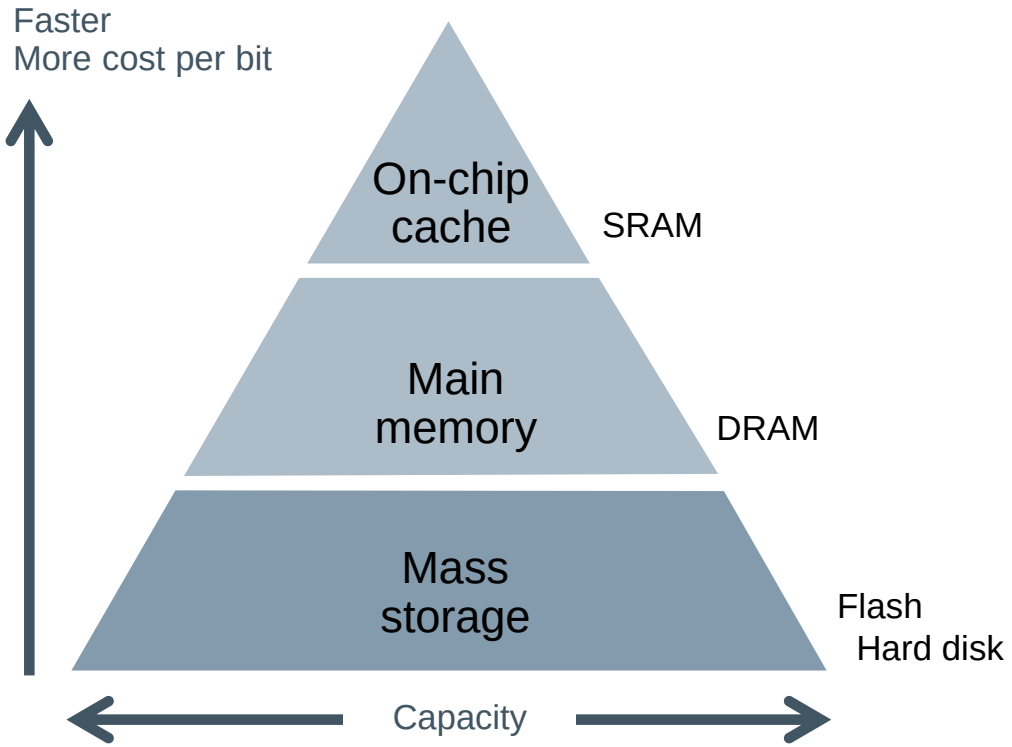
- Scalable, general-purpose interconnect and protocol
  - Replaces processor-local interconnects—(LP)DDR, PCIe, SAS/SATA, etc.
  - Replaces global fabrics for ultra-low-latency communications at scale
- Provide a flexible load-store domain for memory ops and message passing

Making the memory hierarchy obsolete

Massive Memory Pool

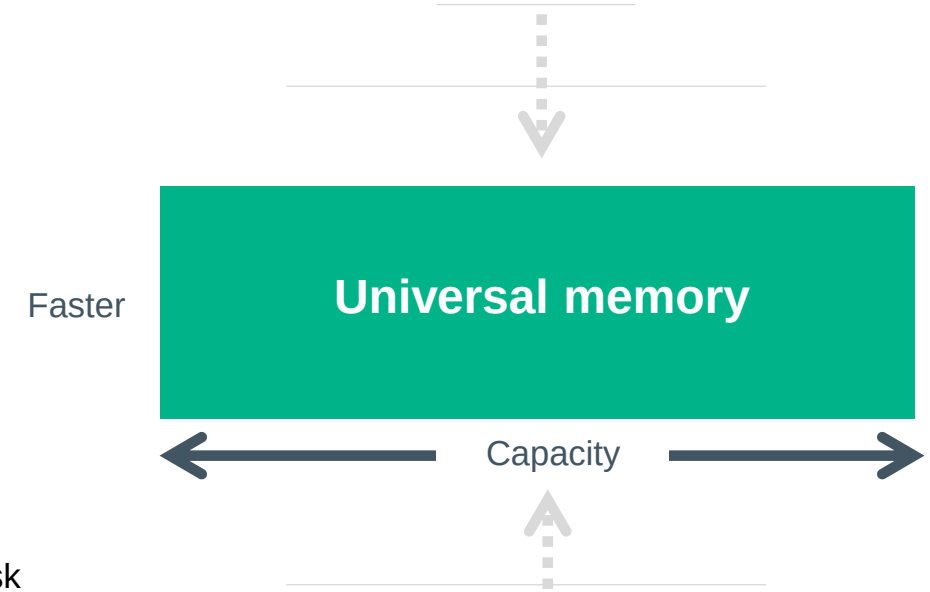
Today

Constant balance between cost and performance



The Machine

Enabling massive data sets



http://genzconsortium.org

Gen-Z: A New Data Access Technology

GEN Z

AMD ARM BROADCOM CAVIUM CRAY Lenovo Mellanox Micron Microsemi redhat

DELL EMC Hewlett Packard Enterprise HUAWEI IBM IDT SAMSUNG SEAGATE SK hynix Western Digital XILINX

Open Standard

High Bandwidth Low Latency

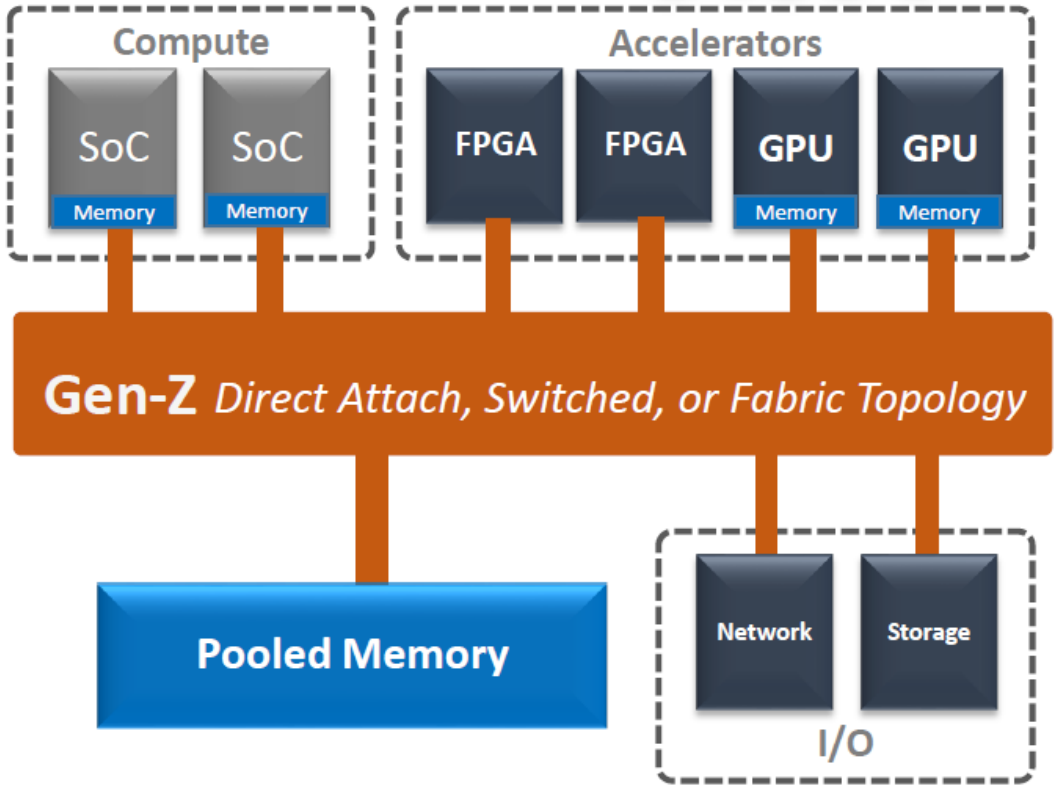
- Memory Semantics – simple Reads and Writes
- From tens to several hundred GB/s of bandwidth
- Sub-100 ns load-to-use memory latency

Advanced Workloads & Technologies

- Real time analytics
- Enables data centric and hybrid computing
- Scalable memory pools for in memory applications
- Abstracts media interface from SoC to unlock new media innovation

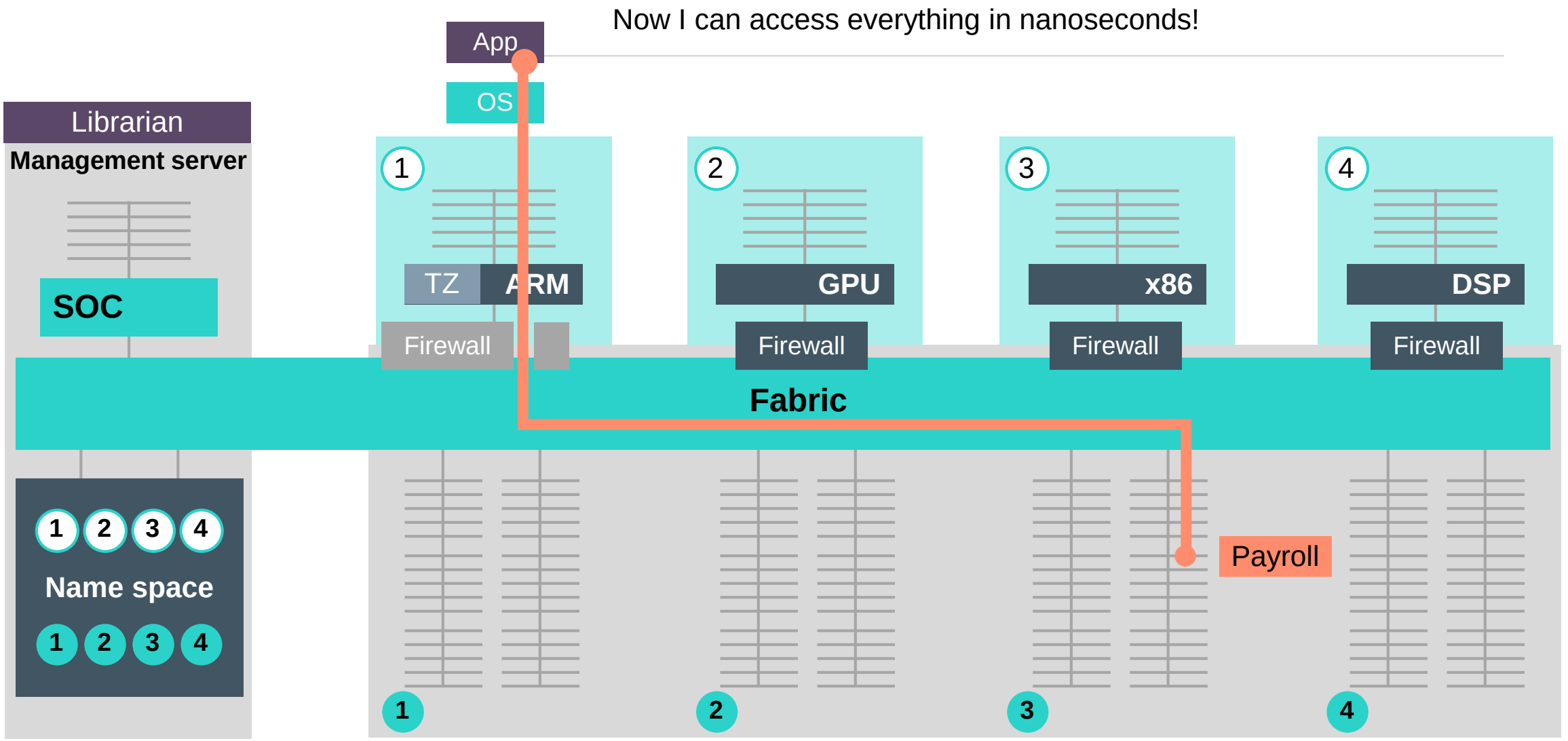
Secure Compatible Economical

- Provides end-to-end secure connectivity from node level to rack scale
- Supports unmodified OS for SW compatibility
- Graduated implementation from simple, low cost to highly capable and robust
- Leverages high-volume IEEE physical layers and broad, deep industry ecosystem



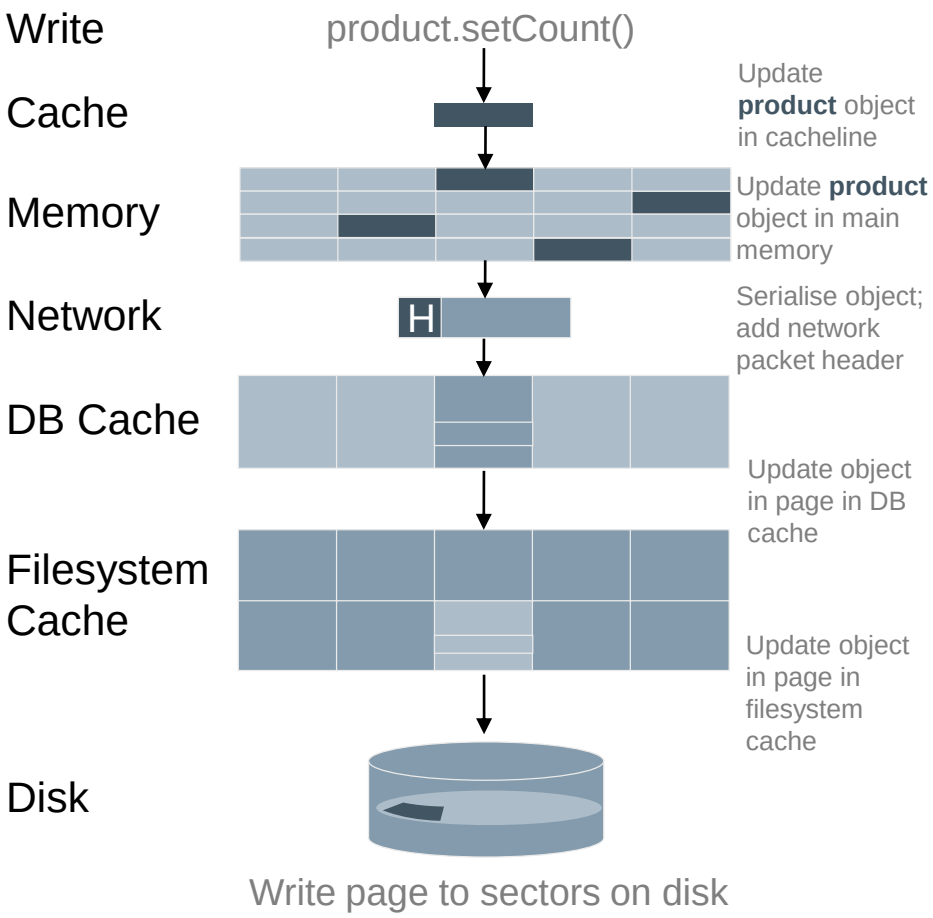
Rack Scale

Semantic of access : load/store gen-Z protocol



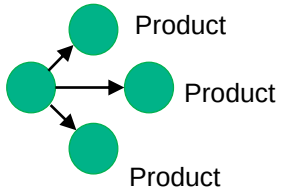
Simplicity: Fewer data layers

Conventional Data Formats



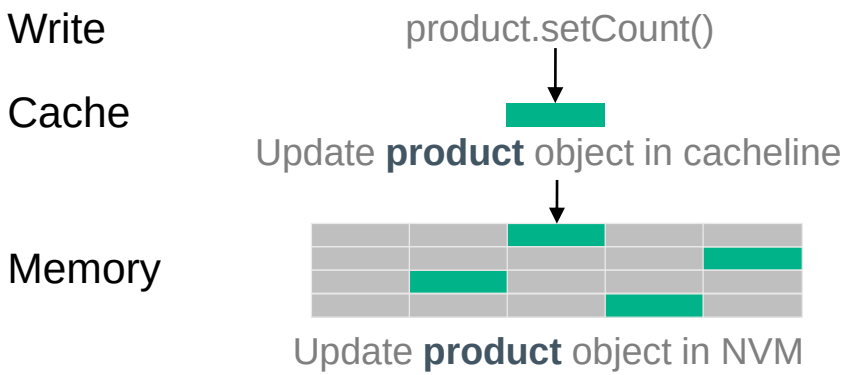
"Track the **products** my company sells."

Product Manager



Application Developer

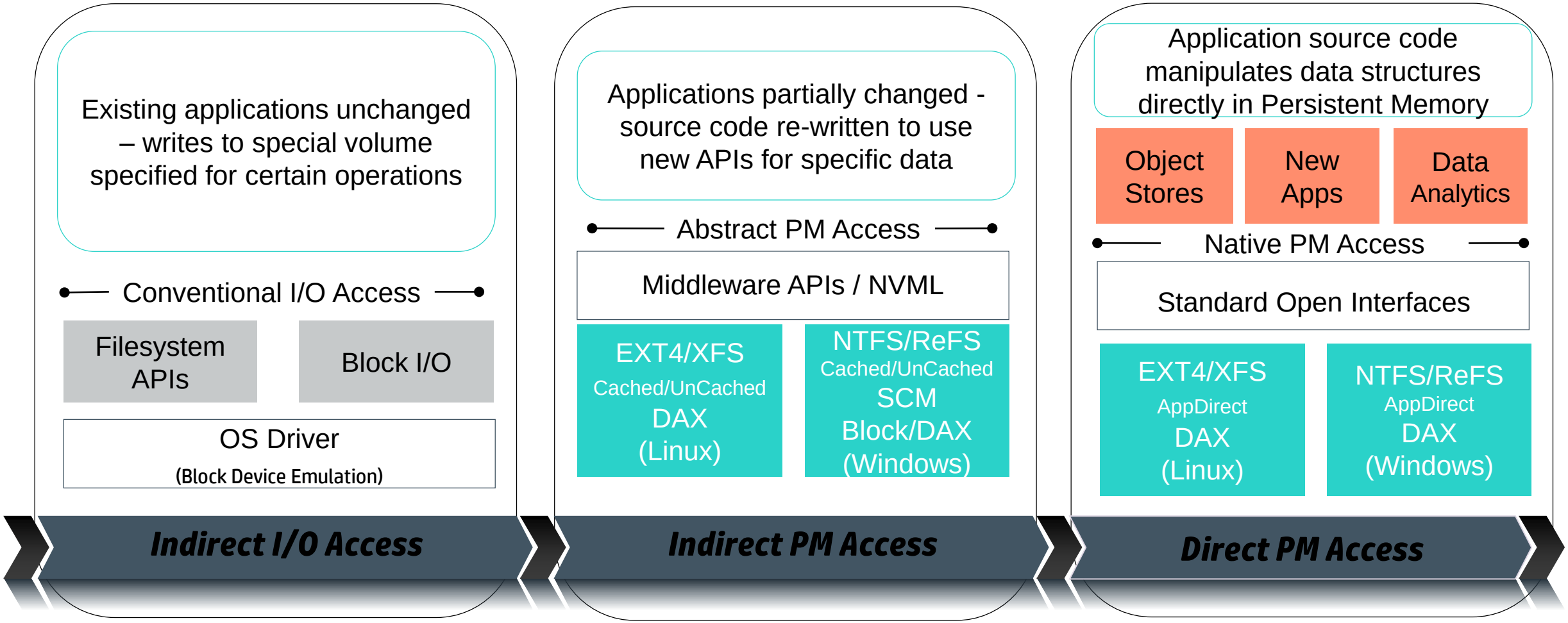
MDS Data Formats



Update now persists in shared non-volatile memory

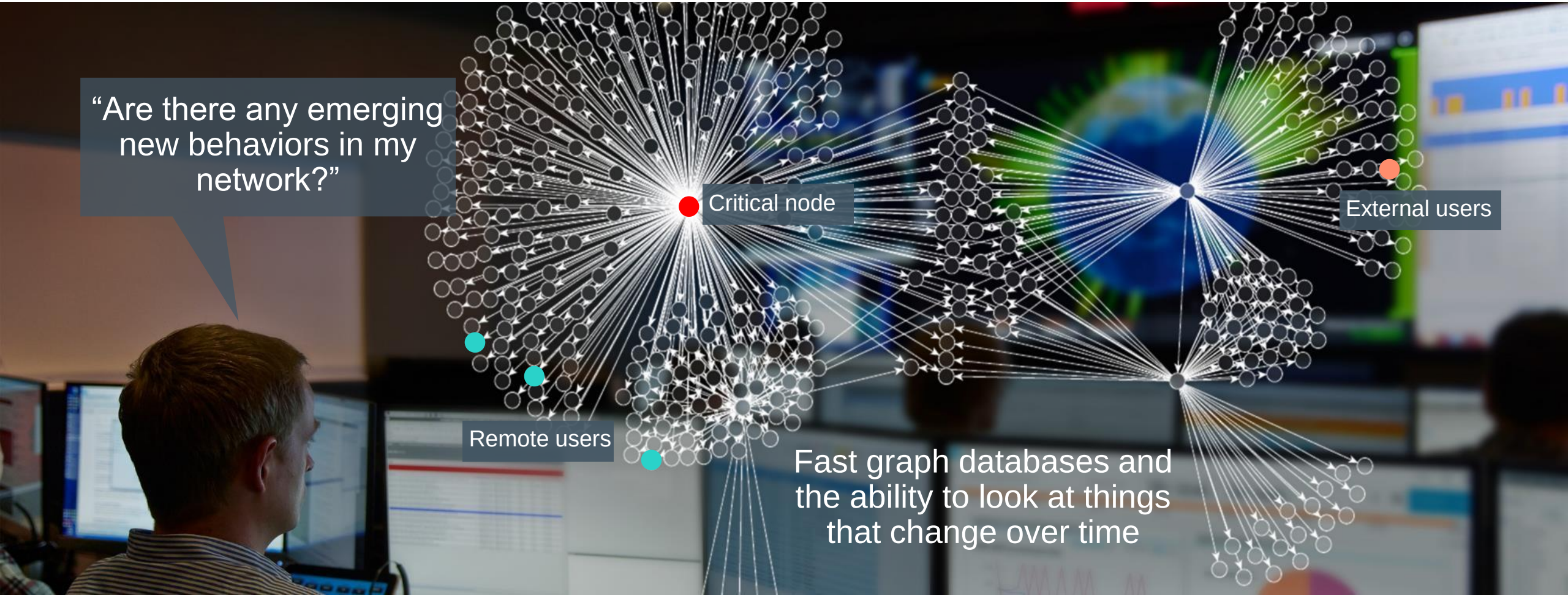
Hewlett Packard Enterprise

Application Programming Models to Persistent Memory



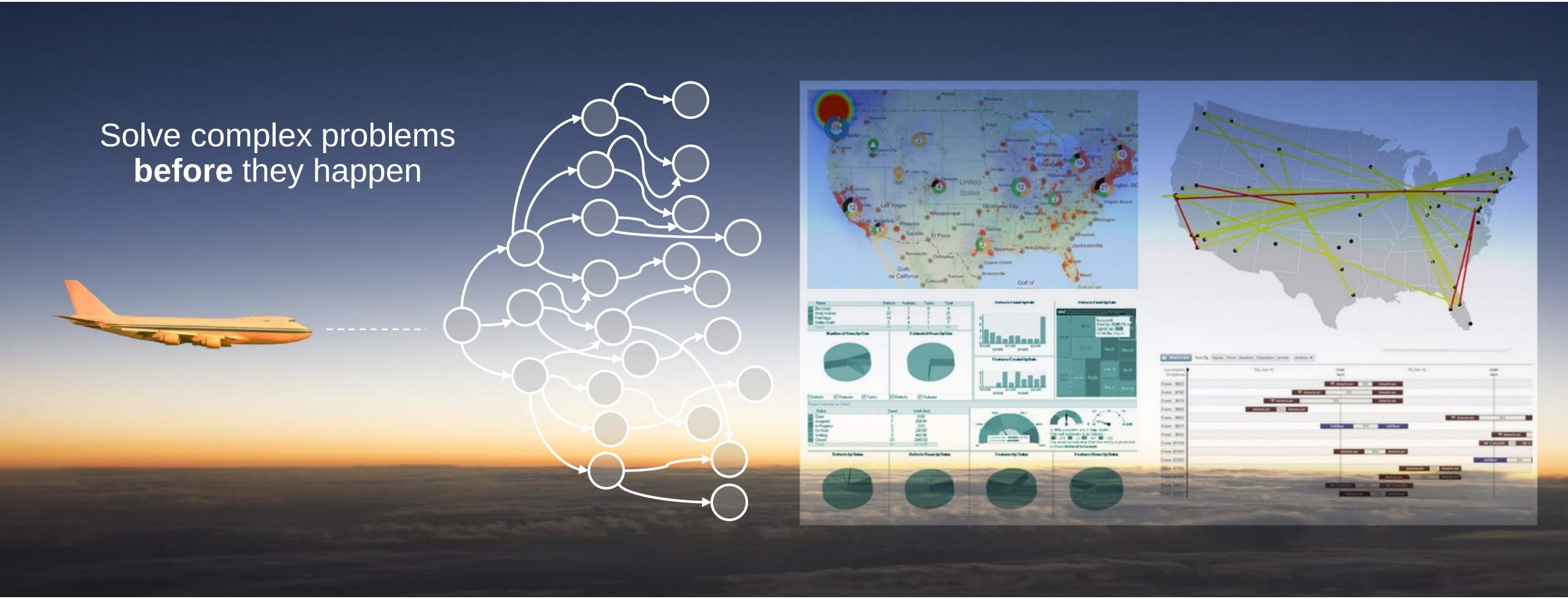
Graph analytics time machine

Massive memory and fast fabrics to ingest all data



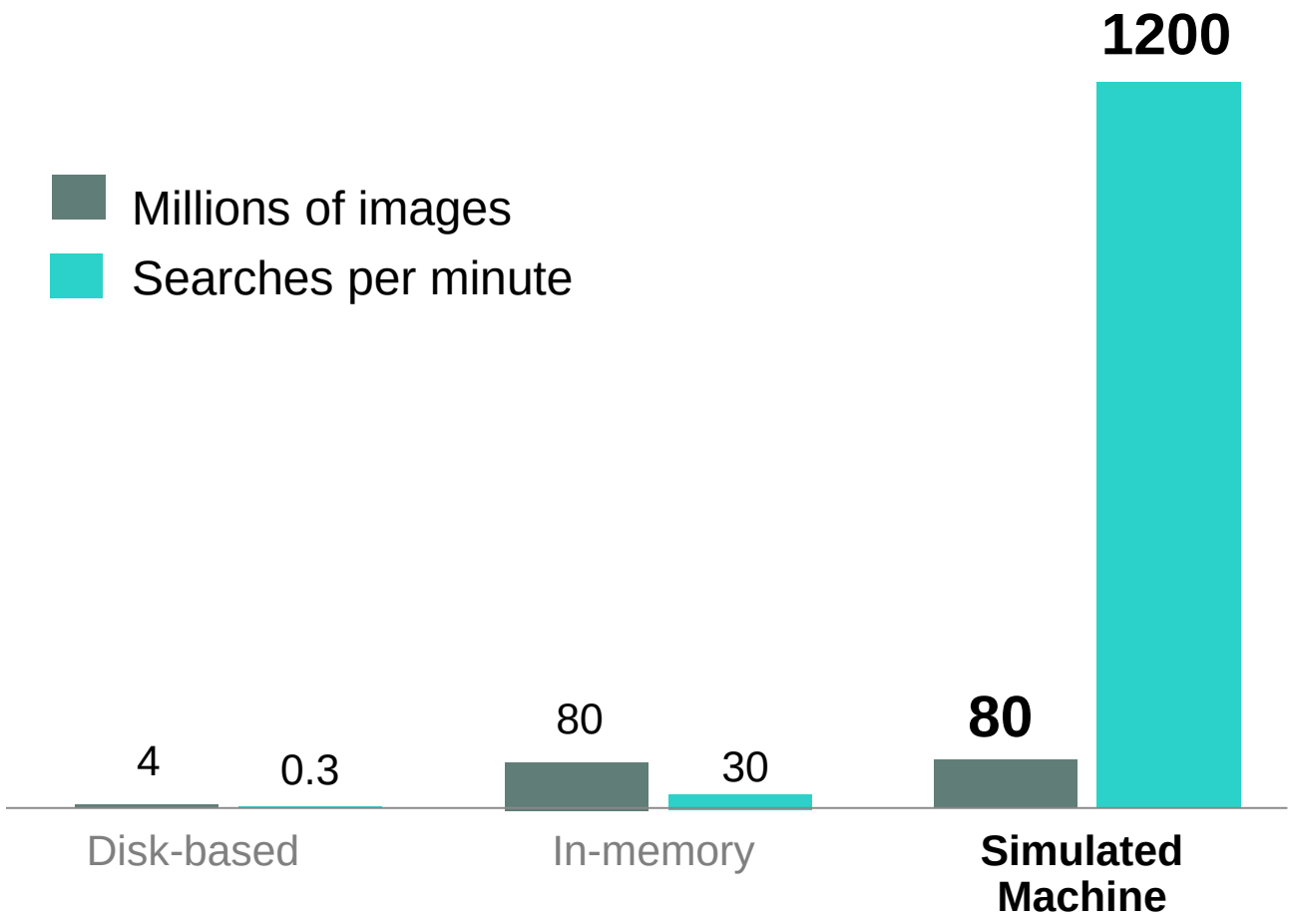
What if we could pre-compute an almost infinite set of “what ifs”?

Optimization over a large search space in real time becomes realistic



Performance demonstration – similarity search

From offline to decision time



Use cases:

Content-based image/video retrieval

Near-duplicate web page detection

Similar document retrieval

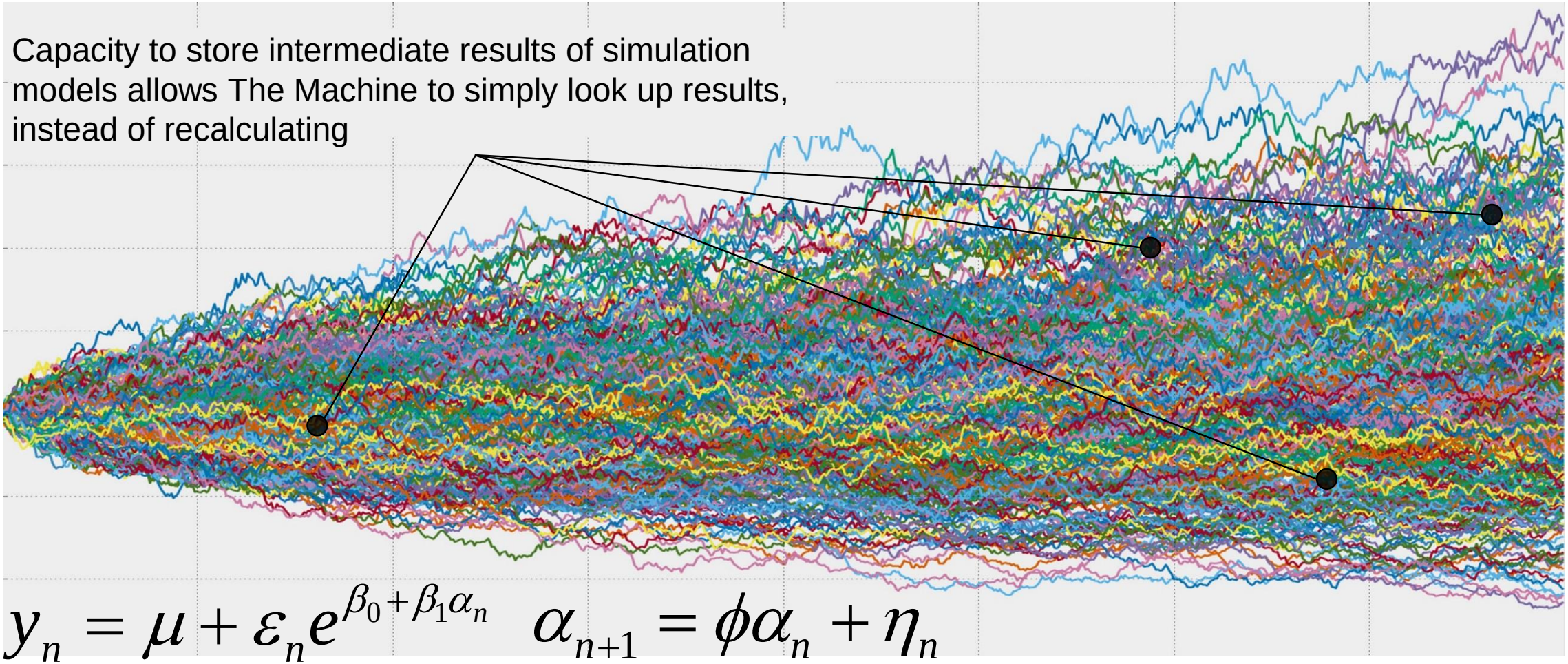
Outlier detection for e-commerce fraud mitigation

Fingerprint matching

Scalable object recognition

Nearest-neighbor classification

Complex models converge in minutes not days

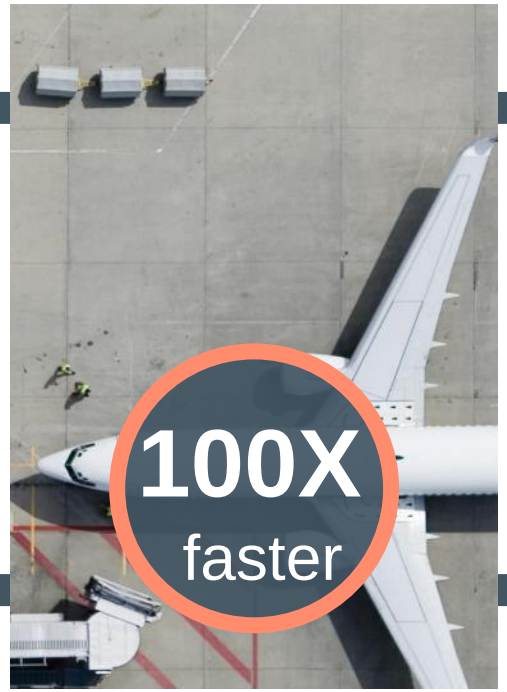


Transform performance with Memory-Driven programming

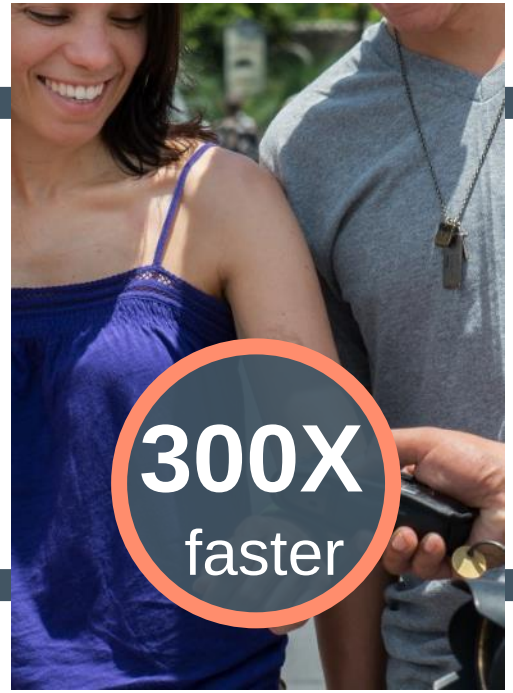
Modify existing frameworks

In-memory analytics

New algorithms



Large-scale graph inference

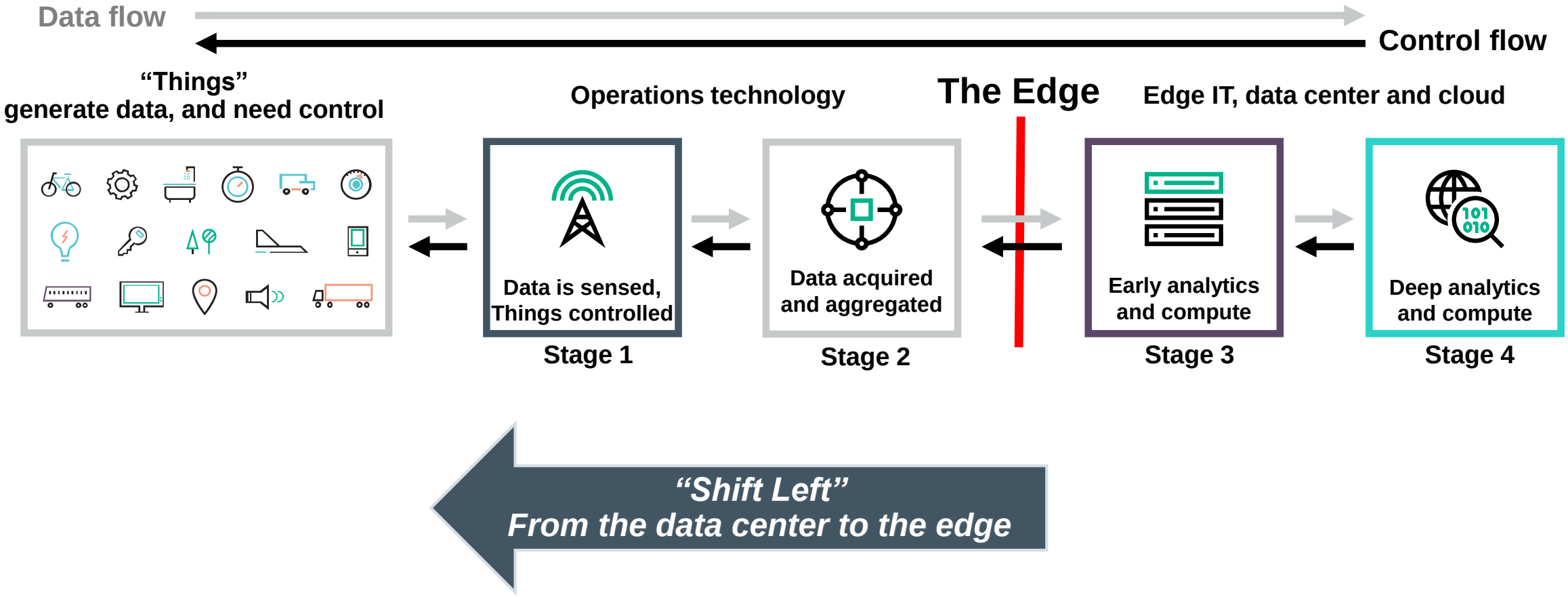


Similarity Search

Completely rethink

Financial models

NVM will play a major role in the IoT world



Too much data to move the data ; need to move the codes

Need also to store the data where created ; need to move only the metadata

Deep Learning and Edge Computing

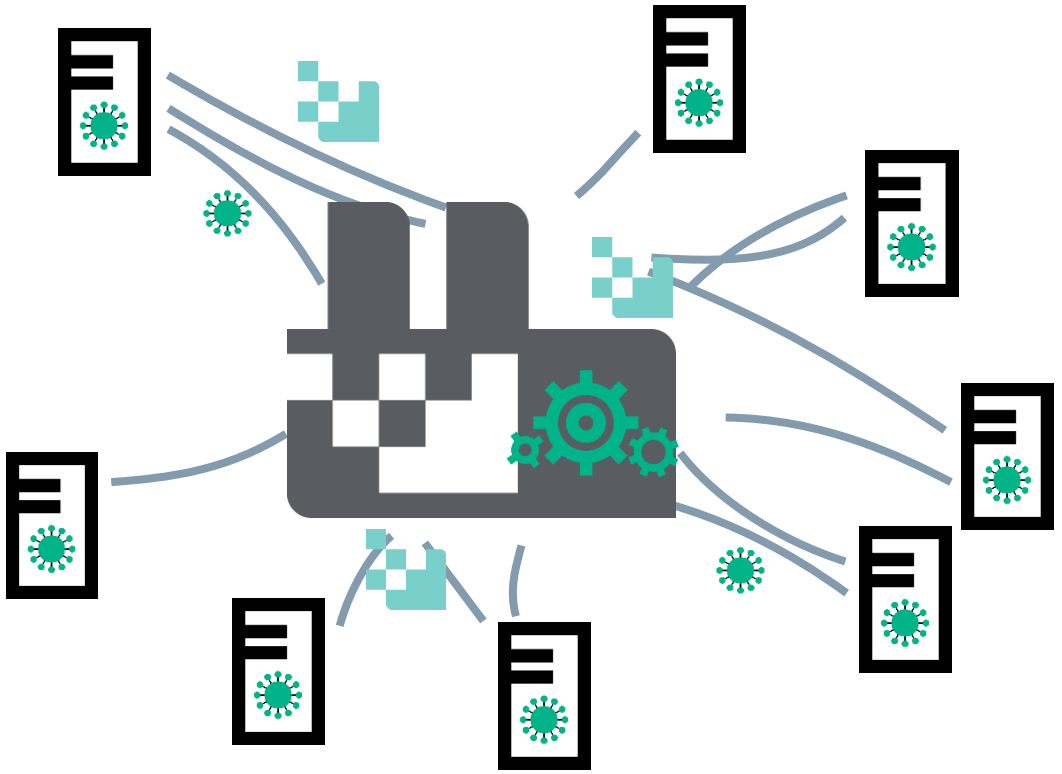
Center : training

Collects some data

Continously trains models

Sends models to edge nodes

Large scale simulations



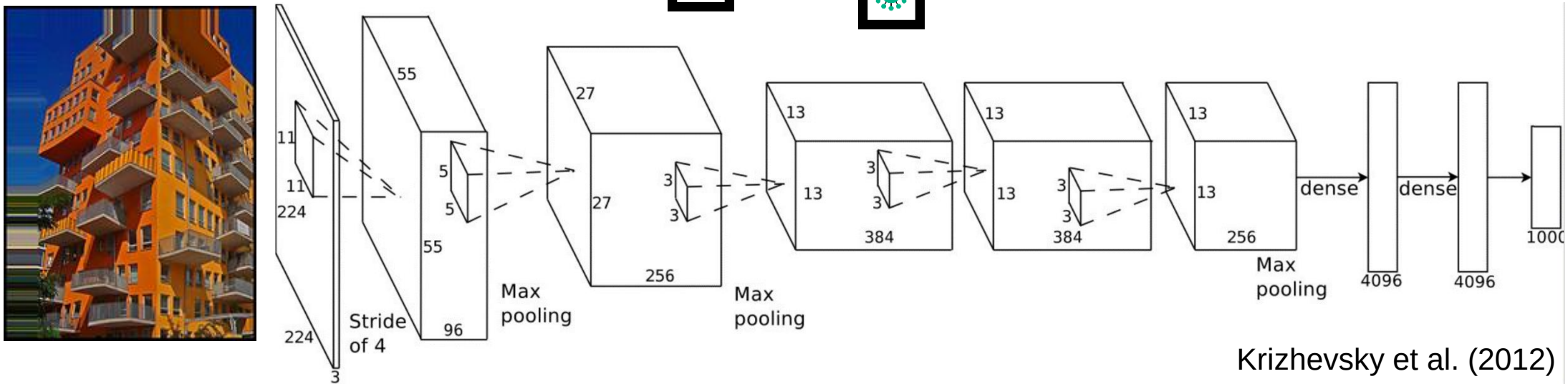
Edge Node : inference

Gets trained model

Uses the model in real-time

Collects data

Sends some data to center



"Building"

Krizhevsky et al. (2012)

Hewlett Packard
Enterprise

Thank you

More resources on The Machine

Industry articles, blogs, and social media outlets:

The Machine on Hewlett Packard Labs Webpage (http://labs.hpe.com/research/themachine/)
Videos: Story on The Machine (https://www.youtube.com/watch?v=NwWF1LSmBJY) and
The Machine: Future of Computing (https://www.youtube.com/watch?list=PL0_ubpZ6vGcAm1sLOSyQWYx_WTJ_u9zNr&v=NZ_rbeBy-ms)

IEEE

- Adapting to Thrive in a New Economy of Memory Abundance – Computer Magazine special article
  http://www.labs.hpe.com/pdf/IEEE_Adapting_to_Thrive_in_a_New_Economy_of_Memory_Abundance.pdf
- At IEEE's Rebooting the Computer Conference, A New Economy of Memory Abundance
  http://community.hpe.com/t5/Behind-the-scenes-Labs/At-IEEE-s-Rebooting-the-Computer-Conference-A-New-Economy-of/ba-p/6818400
- Blah, blah, technology, blah: Sharing the MDC Vision with the IEEE Conference
  http://community.hpe.com/t5/Behind-the-scenes-Labs/Blah-blah-technology-blah-Sharing-the-MDC-Vision-with-the-IEEE/ba-p/6875502
- Memory-Driven Computing – how will it impact the world?
  http://community.hpe.com/t5/Behind-the-scenes-Labs/Memory-Driven-Computing-How-will-it-impact-the-world/ba-p/6796925

Technical articles from TheNextPlatform

- Drilling Down Into The Machine From HPE
  http://www.nextplatform.com/2016/01/04/drilling-down-into-the-machine-from-hpe/
- The Intertwining Of Memory And Performance Of HPE's Machine
  http://www.nextplatform.com/2016/01/11/the-intertwining-of-memory-and-performance-of-hpes-machine/
- Weaving Together The Machine's Fabric Memory
  http://www.nextplatform.com/2016/01/18/weaving-together-the-machines-fabric-memory/
- The Bits And Bytes Of The Machine's Storage
  http://www.nextplatform.com/2016/01/25/the-bits-and-bytes-of-the-machines-storage/
- Non Volatile Heaps And Object Stores In The Machine
  http://www.nextplatform.com/2016/02/08/non-volatile-heaps-object-stores-machine/
- Operating Systems, Virtualization, And The Machine
  http://www.nextplatform.com/2016/02/01/operating-systems-virtualization-machine/
- Future Systems: How HP Will Adapt The Machine To HPC
  http://www.nextplatform.com/2015/08/17/future-systems-how-hp-will-adapt-the-machine-to-hpc/
- Spark on Superdome X Previews in-memory on The Machine
  http://www.nextplatform.com/2016/04/11/spark-superdome-x-previews-memory-machine/
- Programming for Persistent Memory takes Persistence
  http://www.nextplatform.com/2016/04/25/first-steps-program-model-persistent-memory/
- First Steps in the Program Model for Persistent Memory
  http://www.nextplatform.com/2016/04/25/first-steps-program-model-persistent-memory/