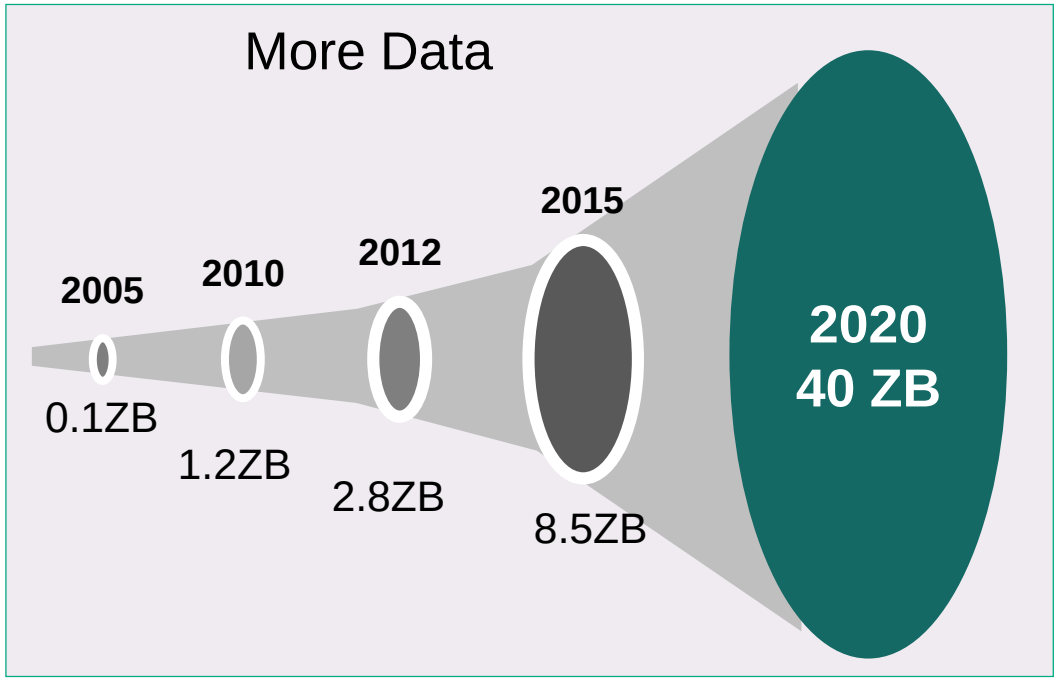
Hewlett Packard
Enterprise

Gen-Z
Memory-Driven Computing

Our vision for the future of computing

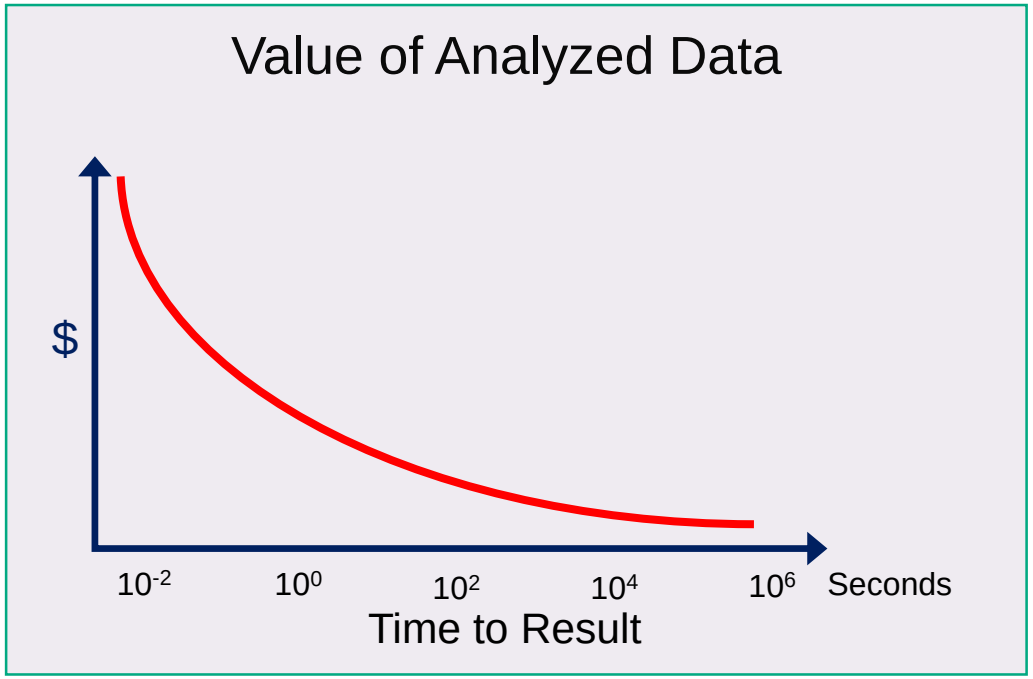
Patrick Demichel
Distinguished Technologist

Explosive growth of data



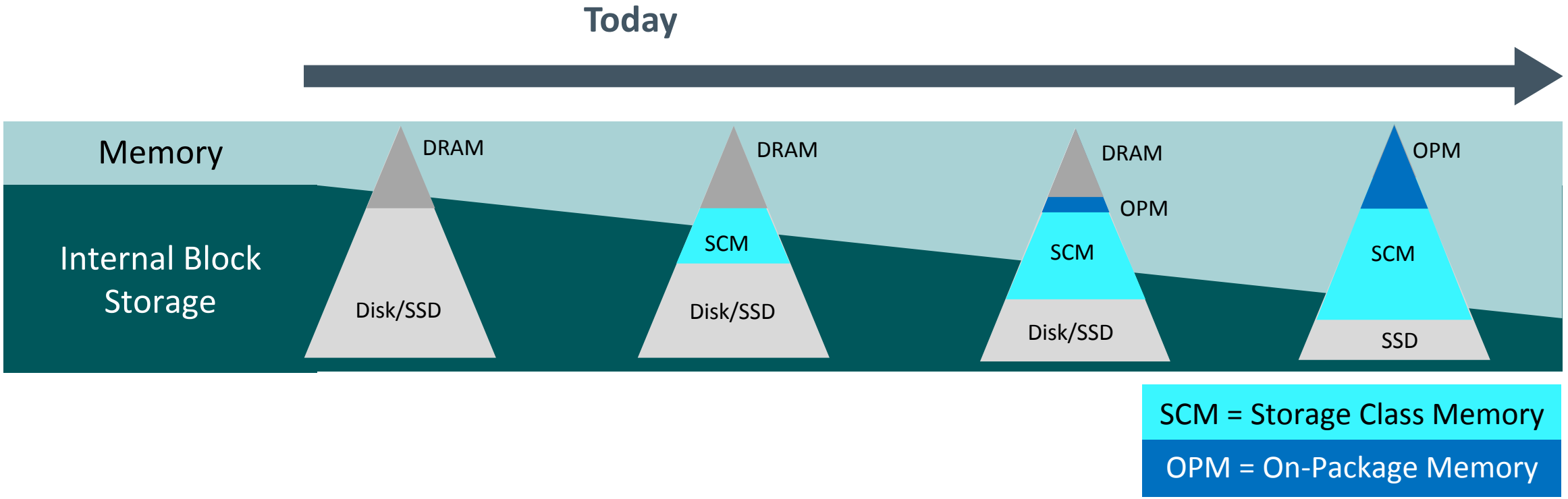
- More than 37% of total data generated in 2020 (40 ZB) will have Big Data value

Need answers … FAST!



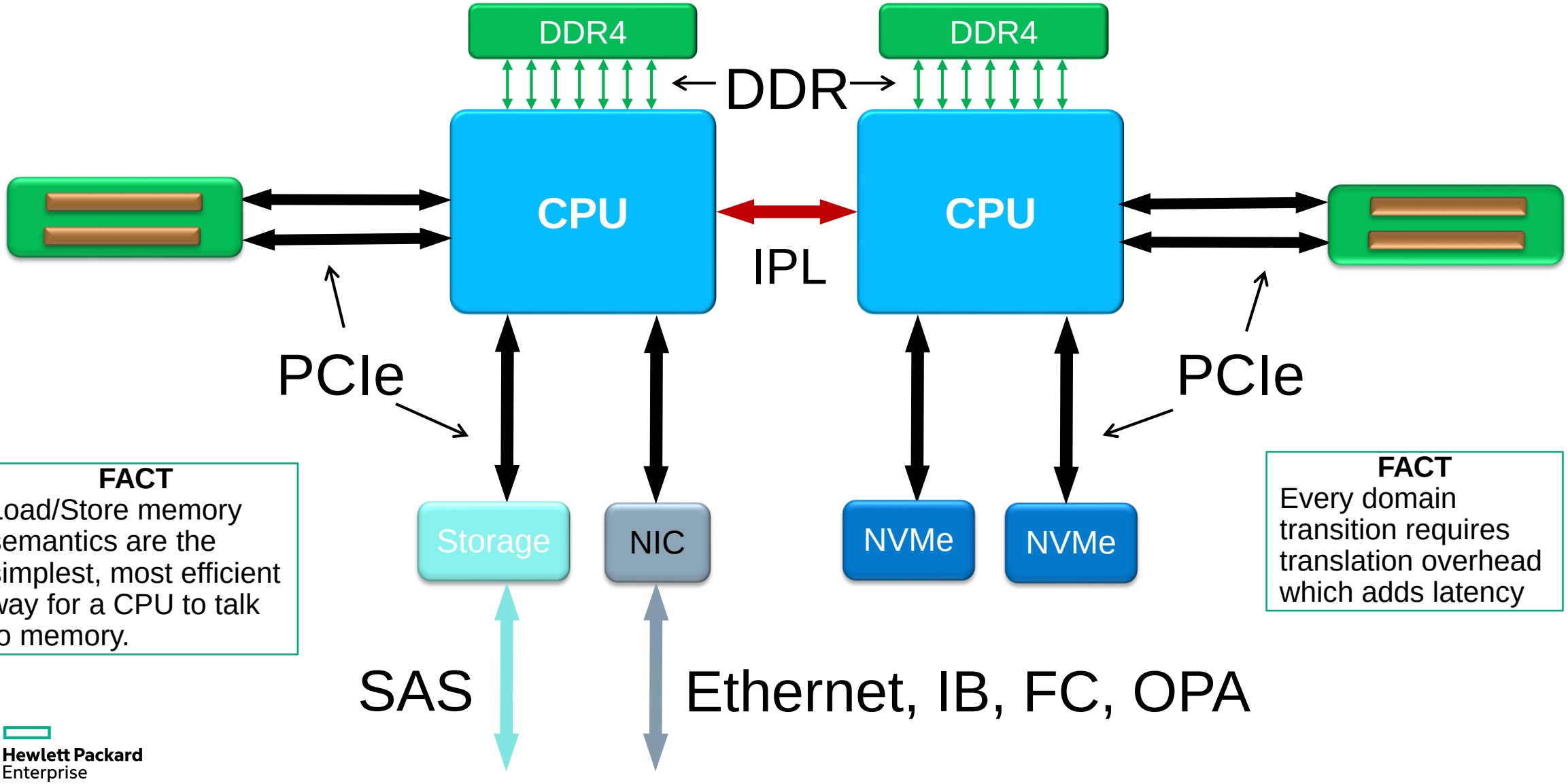
- Businesses demanding real-time insight
- Increasing amounts of data to be analyzed

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 server – 8 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

SCM on the PCIe Bus

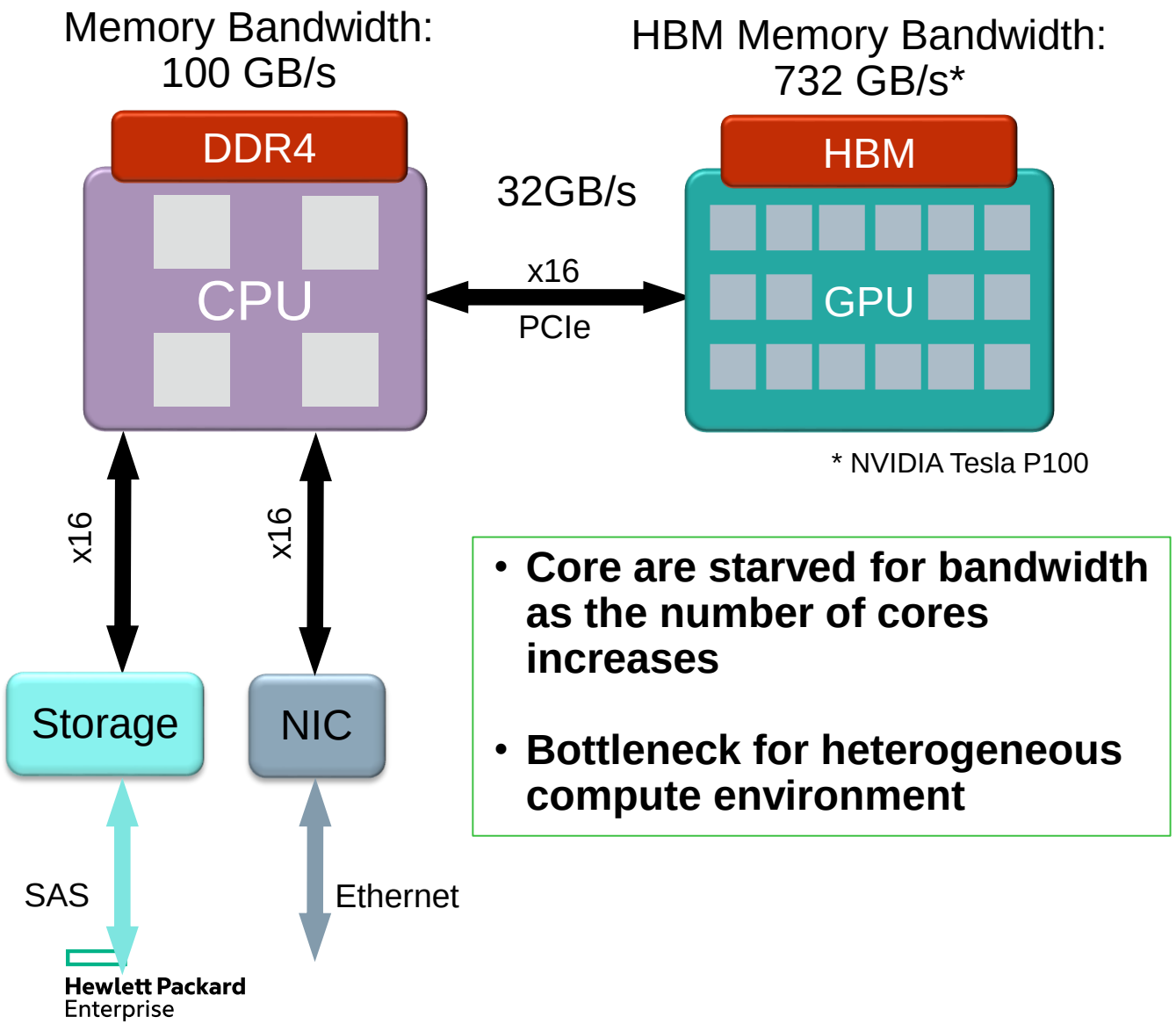
Pros

- Good enough bandwidth/latency
- Standard mechanical format
- Standard electrical interface
- Programming interface well known
- Environmental conditions are well known

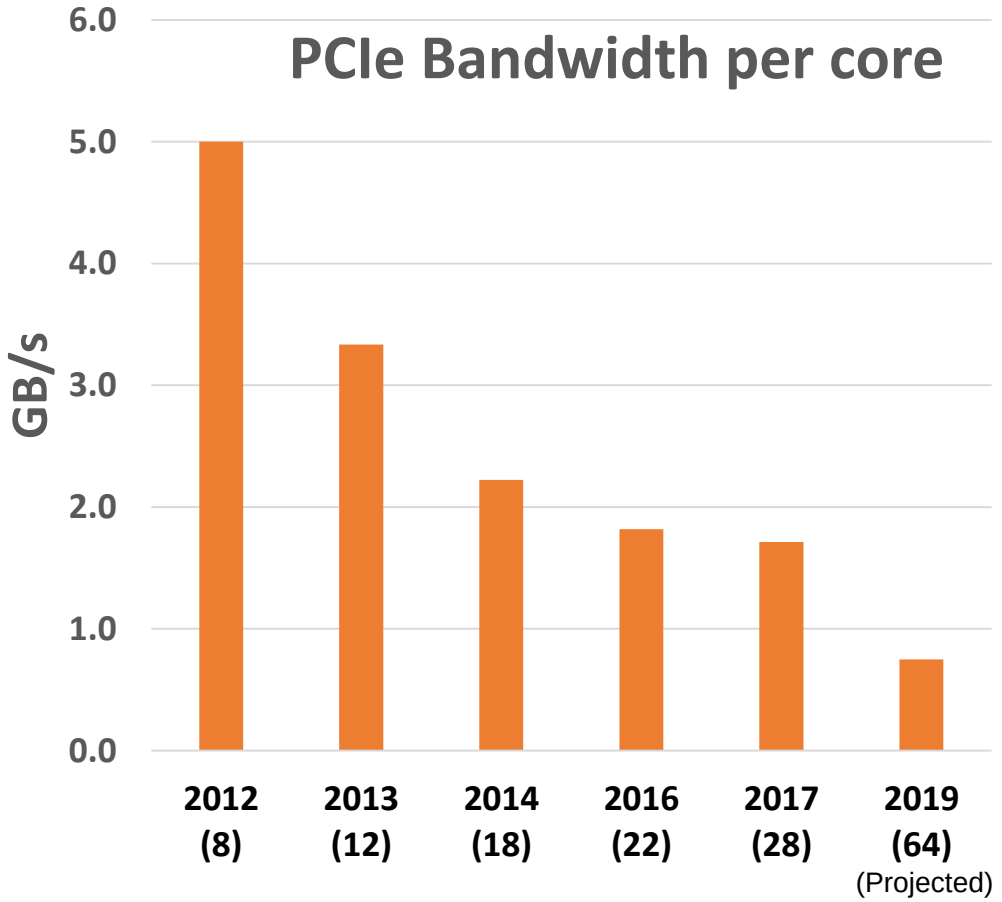
Cons

- Performance not keeping pace with CPU
  - 4GT/s, 8GT, 16GT/s (2003 – 2017, 14 years)
  - Bandwidth: PCIe x16 ~ < 1 memory channel (16GB/s)
- Asymmetric architecture (1x root, Nx endpoints)
- Architecture does not scale beyond the server
- Limited to 256 components per fabric
- No multi-path support

PCIe Architectural Limitations

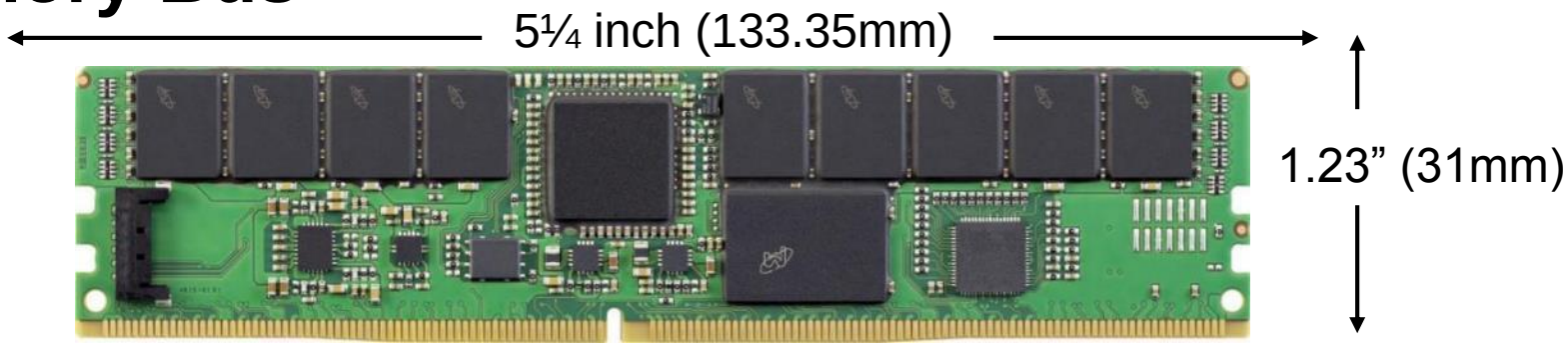


- **Core are starved for bandwidth as the number of cores increases**
- **Bottleneck for heterogeneous compute environment**



Hewlett Packard Enterprise

SCM on the Memory Bus



Pros

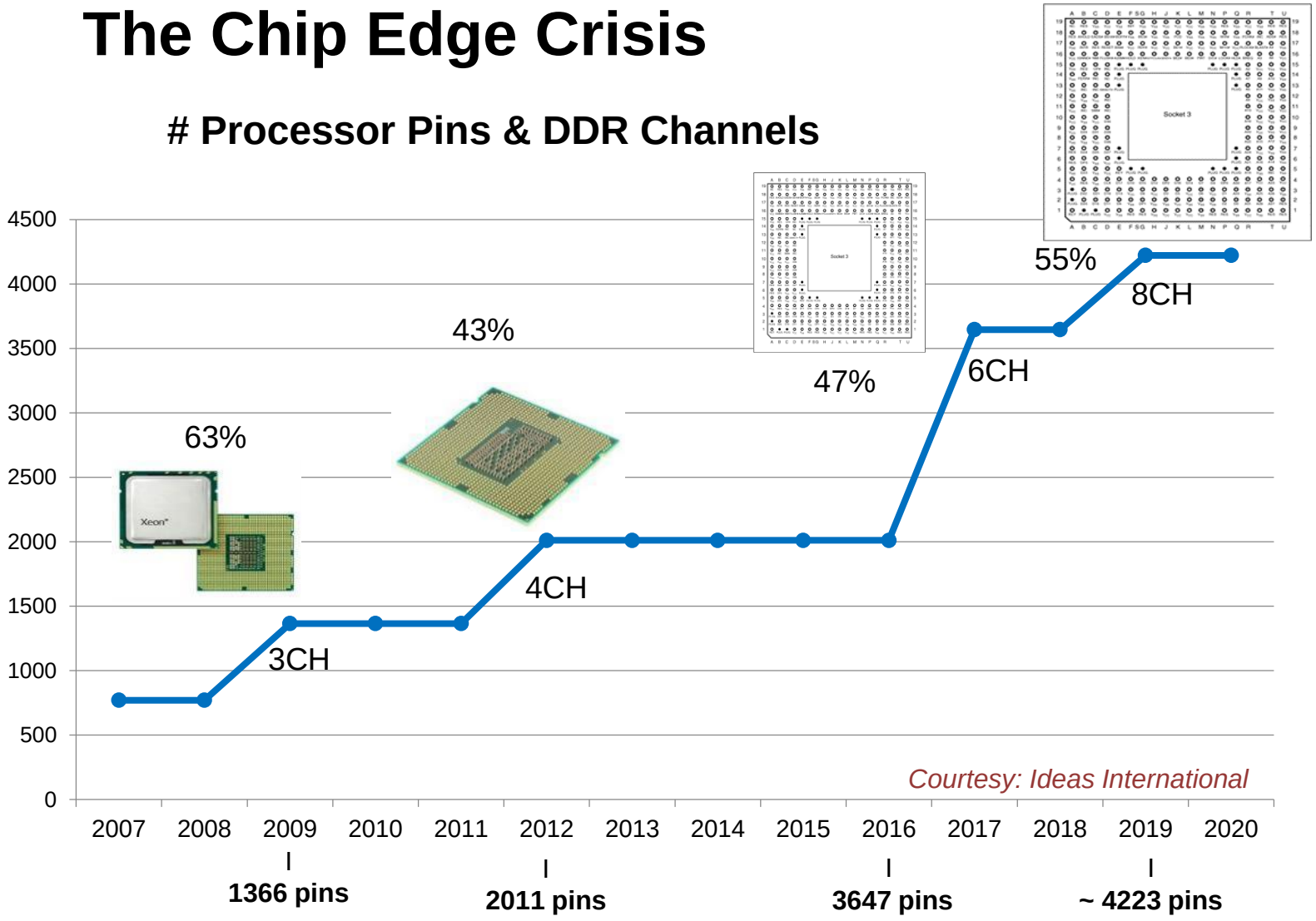
- Low latency, high bandwidth
- Memory semantic interface (Load/Store)
- Standard mechanical format (DIMM)
- Standard electrical interface
- Environmental conditions are well known

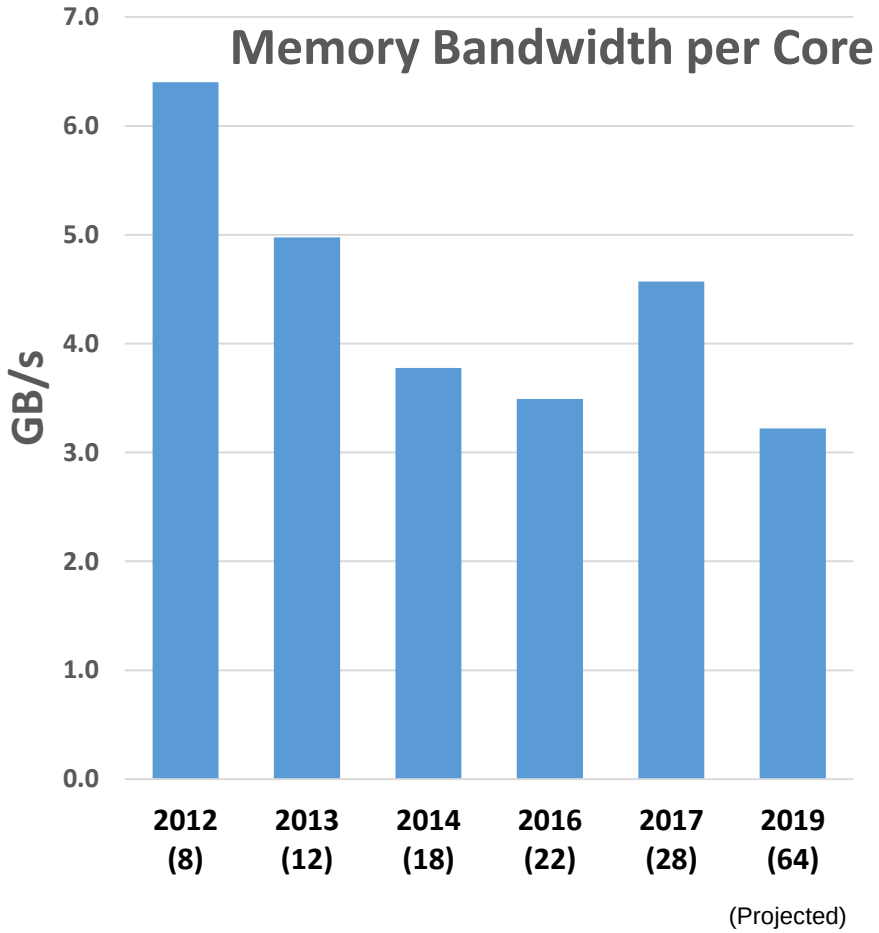
Cons

- Each NVDIMM consumes a Memory Slot
- DIMM format constrains capacity
- Constrained electrical interface
  - Single ended, pin intensive, fixed time budget
- Thermal & placement challenges
- Trapped persistence

The Chip Edge Crisis

Processor Pins & DDR Channels





Hewlett Packard Enterprise

What is required?

- Eliminate architectural bottlenecks by simplifying HW/SW interfaces
- Speak common language as CPUs and extend memory semantics to all devices
- Deliver the highest bandwidth, lowest latency to all devices
- Flexible architecture that comprehends technology trends
  - (i.e. convergence of storage and memory, photonics, heterogeneous compute)
- A framework that will keep up with the speed of innovation
- Reduce acquisition and operational costs, enable open ecosystem

An open, flexible fabric architecture that comprehends technology trends at the speed of innovation

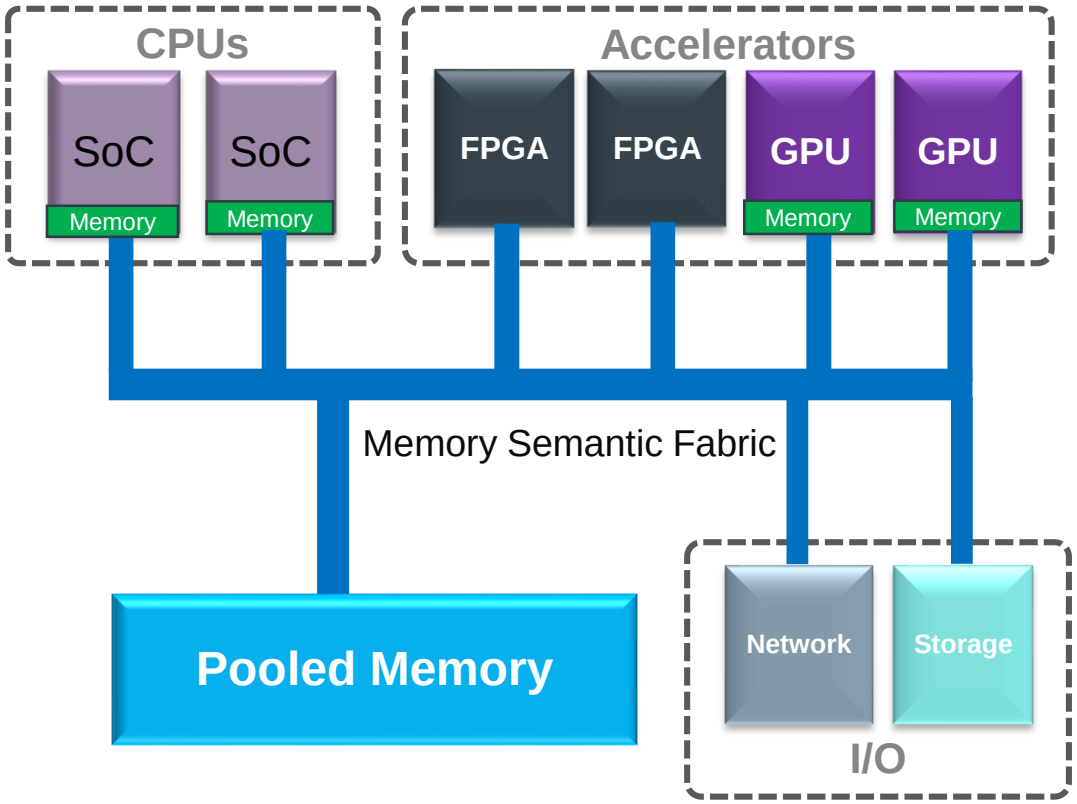
Memory Semantic Fabric

Communication at the speed of memory

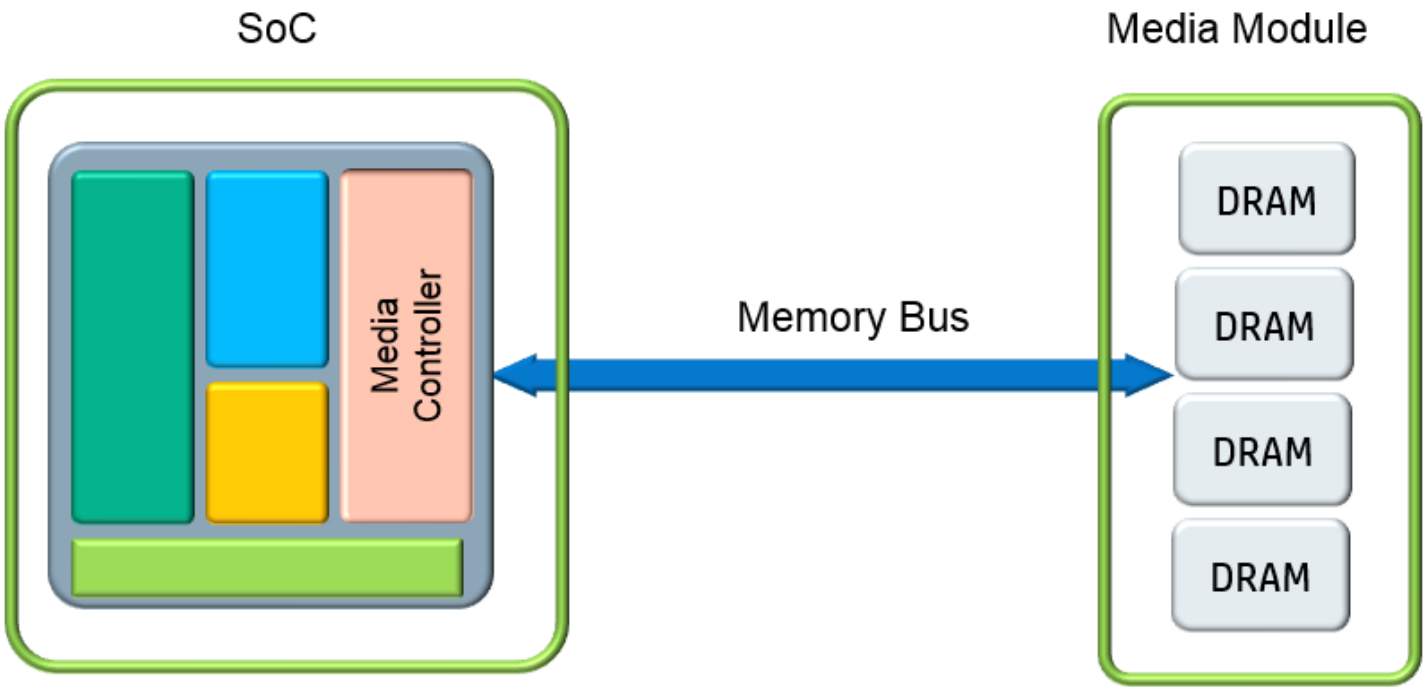
What is a Memory Semantic Fabric?

- A communication protocol that speaks the same language the CPU speaks: load/store, put/get, and read/write

- Connectivity that extends beyond the server to the rack and data center

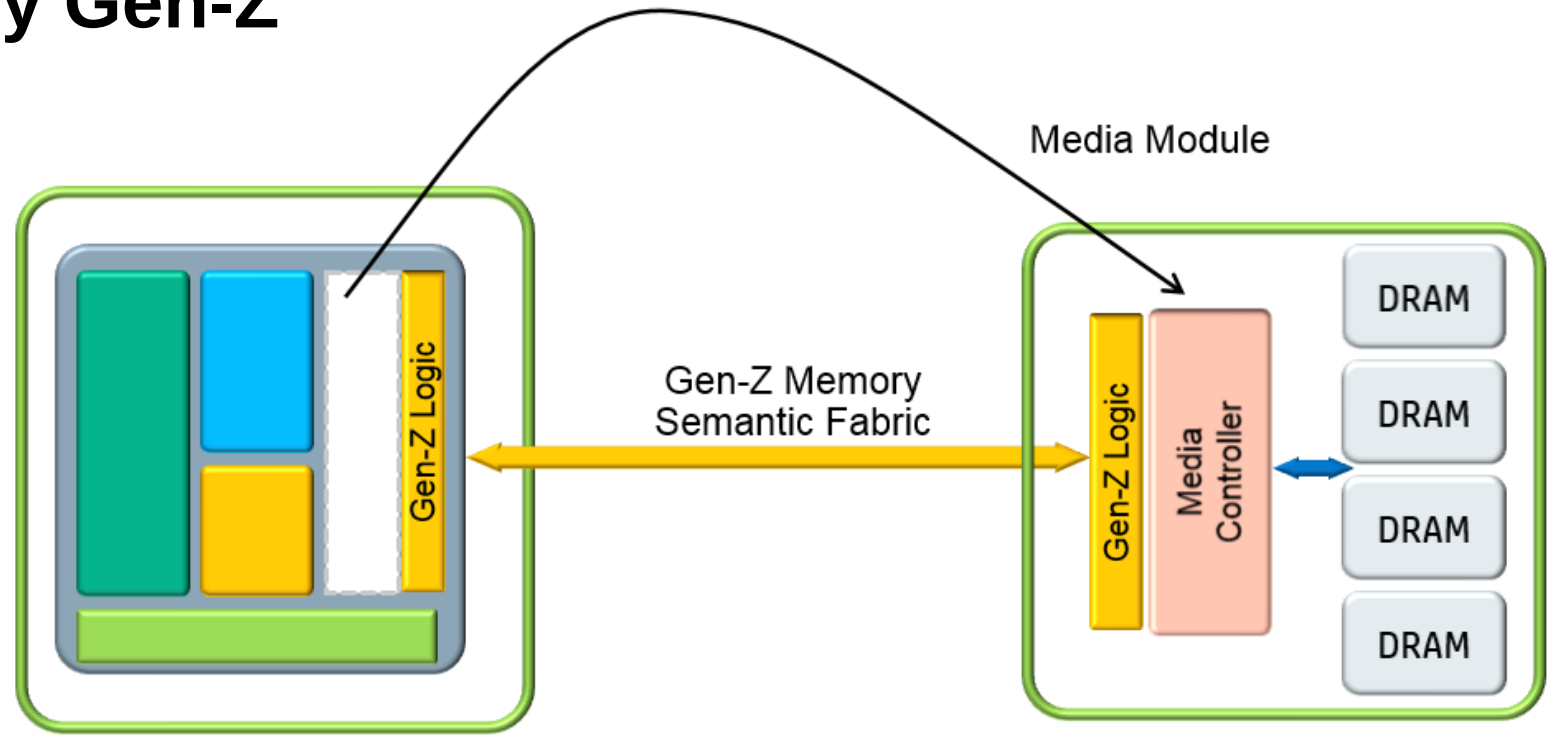


Current architectures



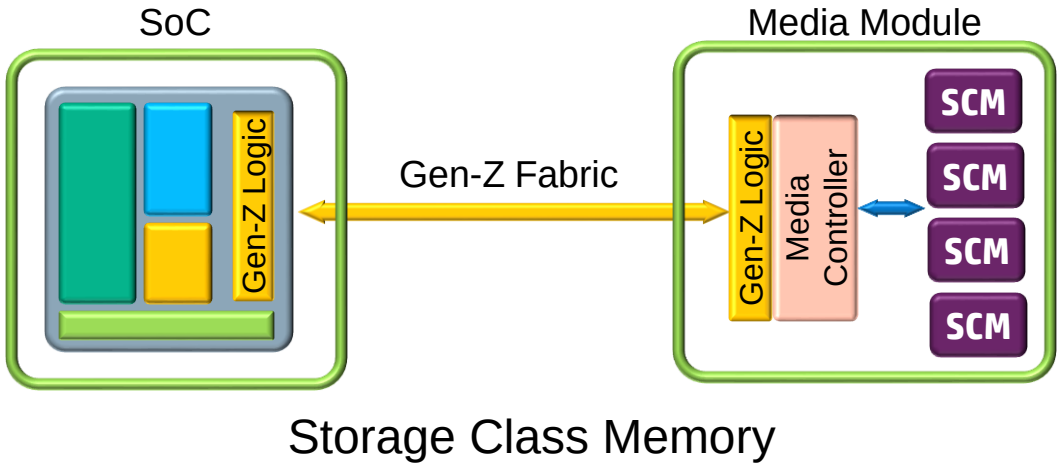
4-8 Memory Channels
17-25 GB/s / Channel
288pins / DIMM
Synchronous Interface

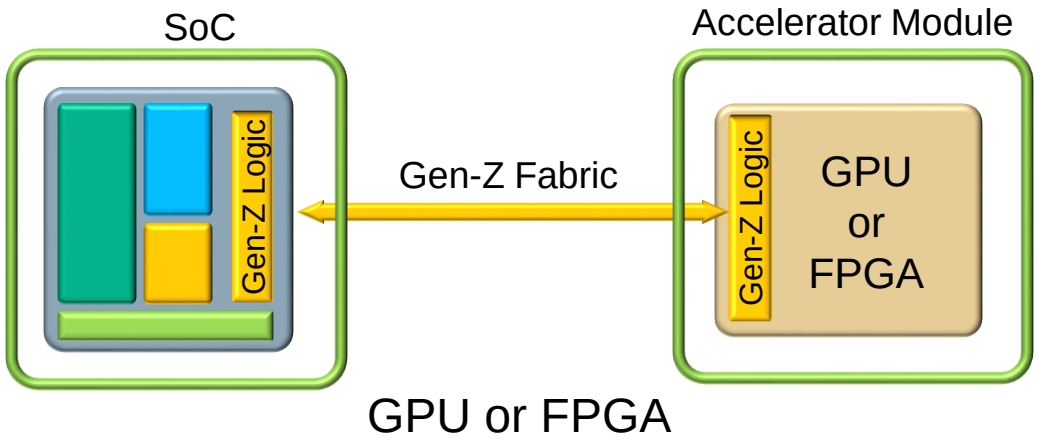
Replaced by Gen-Z

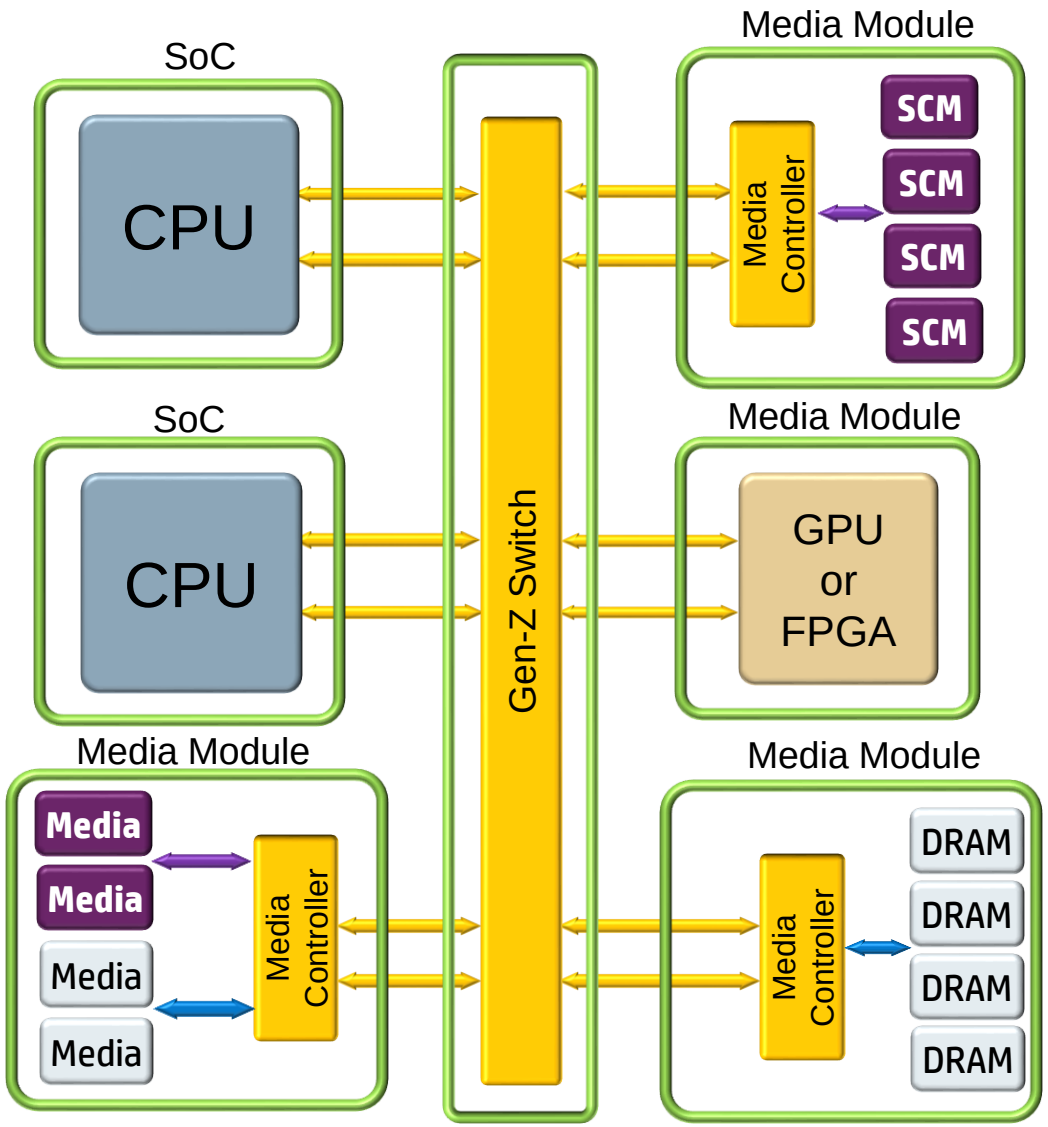


2-8 High-speed Serial Links
Low Latency, High performance
Split Memory Controller
Asynchronous Interface
SoC is media agnostic

Gen-Z is a high speed memory semantic protocol







- Supports DRAM, Flash, SCM, NVM
- Decouples CPU/memory design, enabling independent innovation

Hewlett Packard Enterprise

Open: Broad Industry & Device Support

GEN Z Consortium Members

ALPHA DATA, AMD, Amphenol, ARM, BROADCOM, CAVIUM, CRAY, DELL EMC, FIT, Hewlett Packard Enterprise, HUAWEI, IBM, IDT, Lenovo, LOTES, Mellanox TECHNOLOGIES, Micron, Microsemi, NOKIA, PLDA GROUP, redhat, SAMSUNG, SEAGATE, SK hynix, Western Digital, XILINX

- **Components**
- **Intellectual Property**
- **Connectors**
- **Subsystems**
- **Systems**
- **Software**

GEN Z Component Categories


CPUs

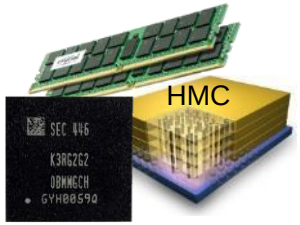

GPUs


Special Purpose Processors

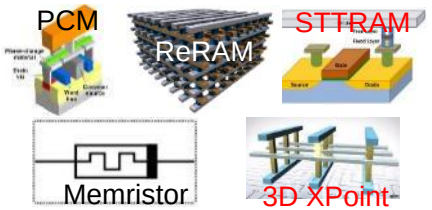

Storage Controllers


Switches/ Gateways


FPGAs



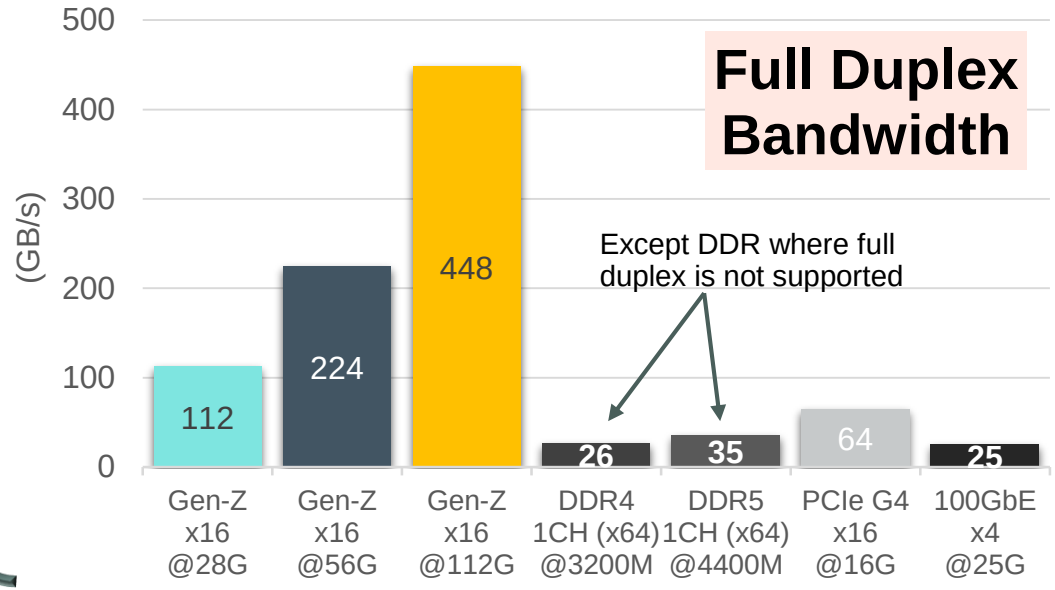
Volatile Memory

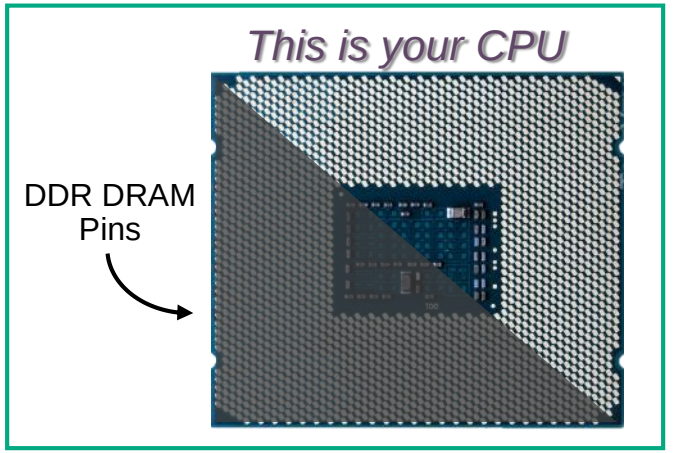


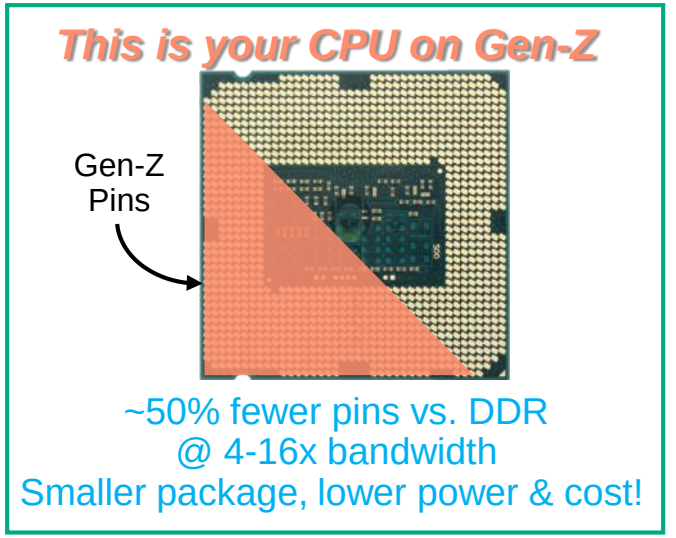
Storage Class Memory

Gen-Z Physical Layer

- Uses IEEE802.3 physical layer standards
  - Serial speeds of 16G to 112G & beyond
  - Component placement freedom (not next to CPU)
- Multi-lanes per link (x1 to x32 & beyond)

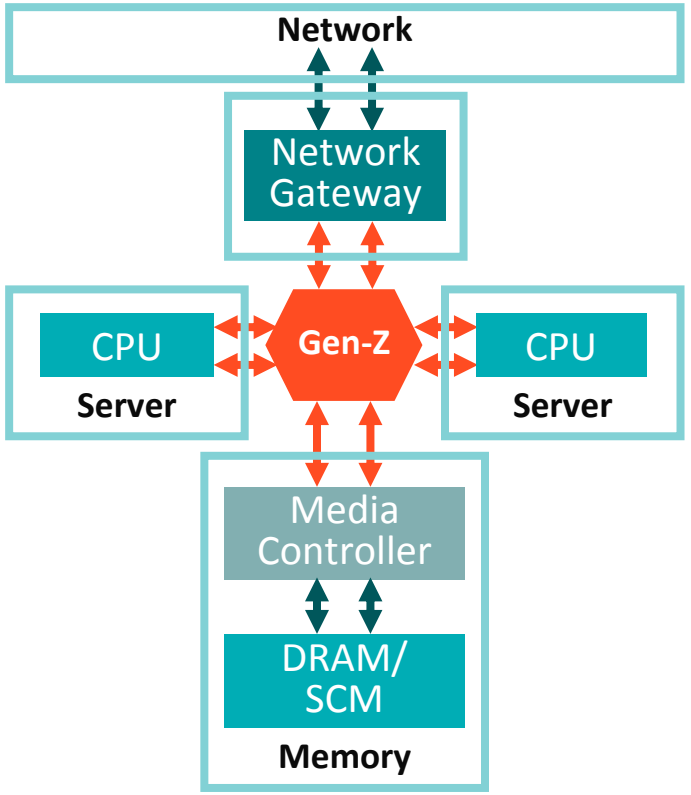




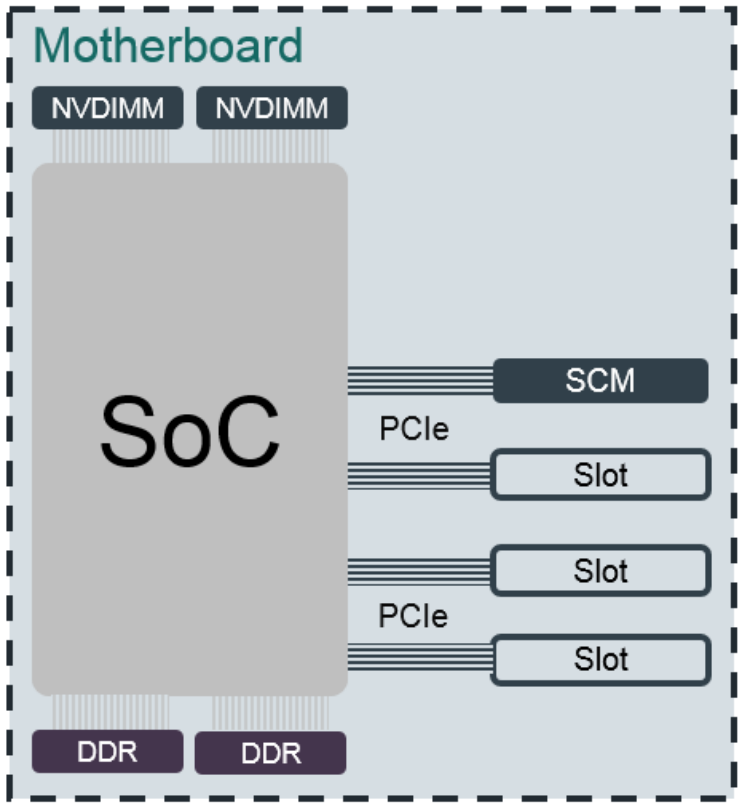


Compatibility: Ethernet Gateway

- In Gen-Z solutions, communication to Ethernet is required
  - Vast majority of clients (mobile, PC, etc.) are on Ethernet
  - Traditional data center infrastructure is on Ethernet
  - Communication between clusters of Gen-Z infrastructure
- Gen-Z Ethernet Gateways provide this connectivity
  - Translates emulated/tunneled Ethernet to native Ethernet
  - Implemented as SoCs + NICs, routers, switches, or NICs
    - Dependent on scale, bandwidth requirements of the deployment
    - Single gateway can support 10s, 100s, or 1000s of compute
  - This “North-South” traffic can scale by added gateways
- Ethernet gateway management services
  - Provide management, address resolution for gateways

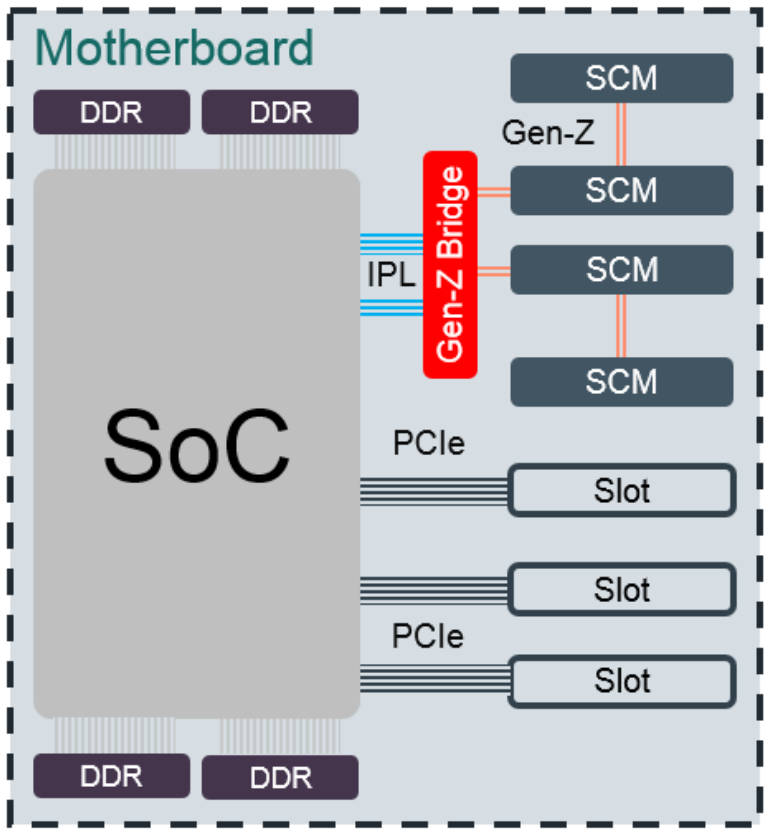


Gen-Z Evolution: Augment



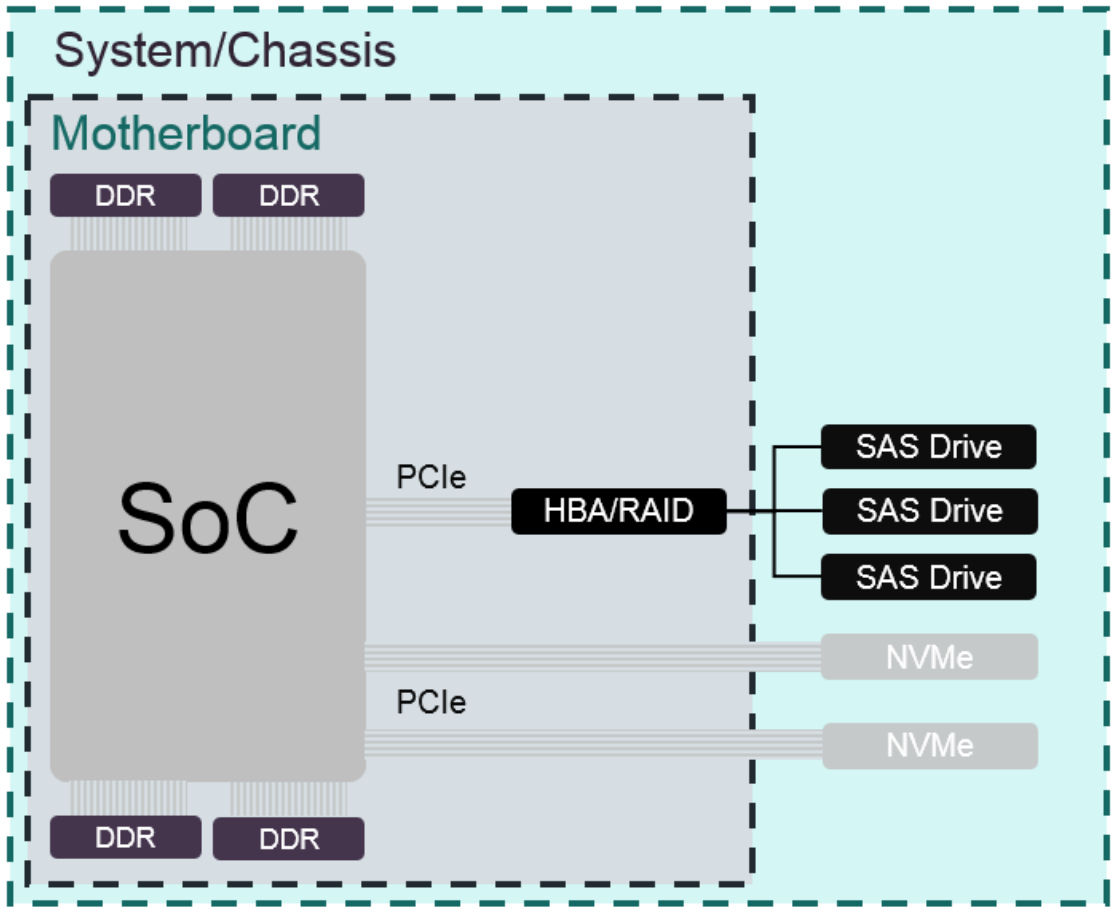
- DIMM-based Storage Class Memory (SCM)
  - Good performance, limited capacity & placement
- PCIe-based SCM
  - Good capacity, limited performance & placement

Gen-Z Evolution: Augment



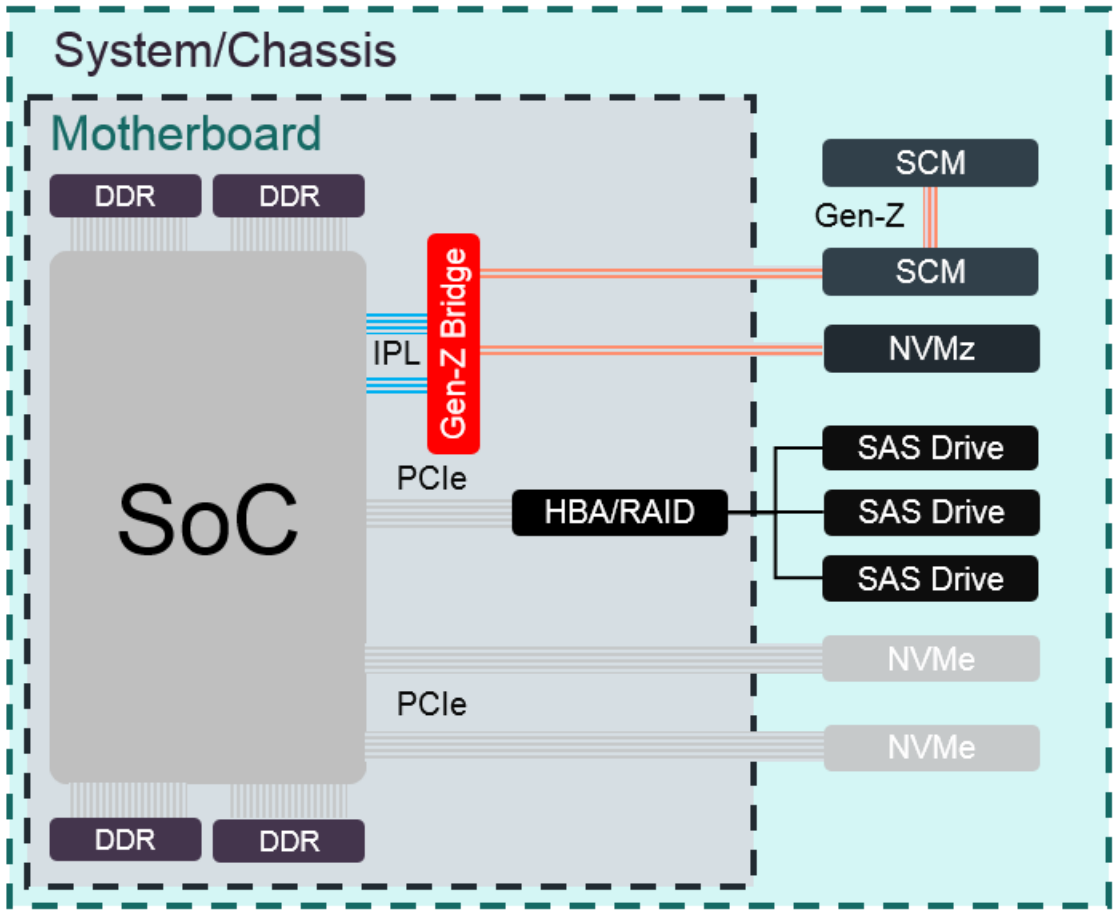
- DIMM-based Storage Class Memory (SCM)
  - Good performance, limited capacity & placement
- PCIe-based SCM
  - Good capacity, limited performance & placement
- Gen-Z based SCM
  - Bridge from Interprocessor Links (IPL) to Gen-Z
    - e.g. KTI, UPI, xGMI, HT, etc.
  - Great performance, capacity, & architectural flexibility
- What's "architectural flexibility"?
  - Let's take a look…

Gen-Z evolution: Augment at the system level



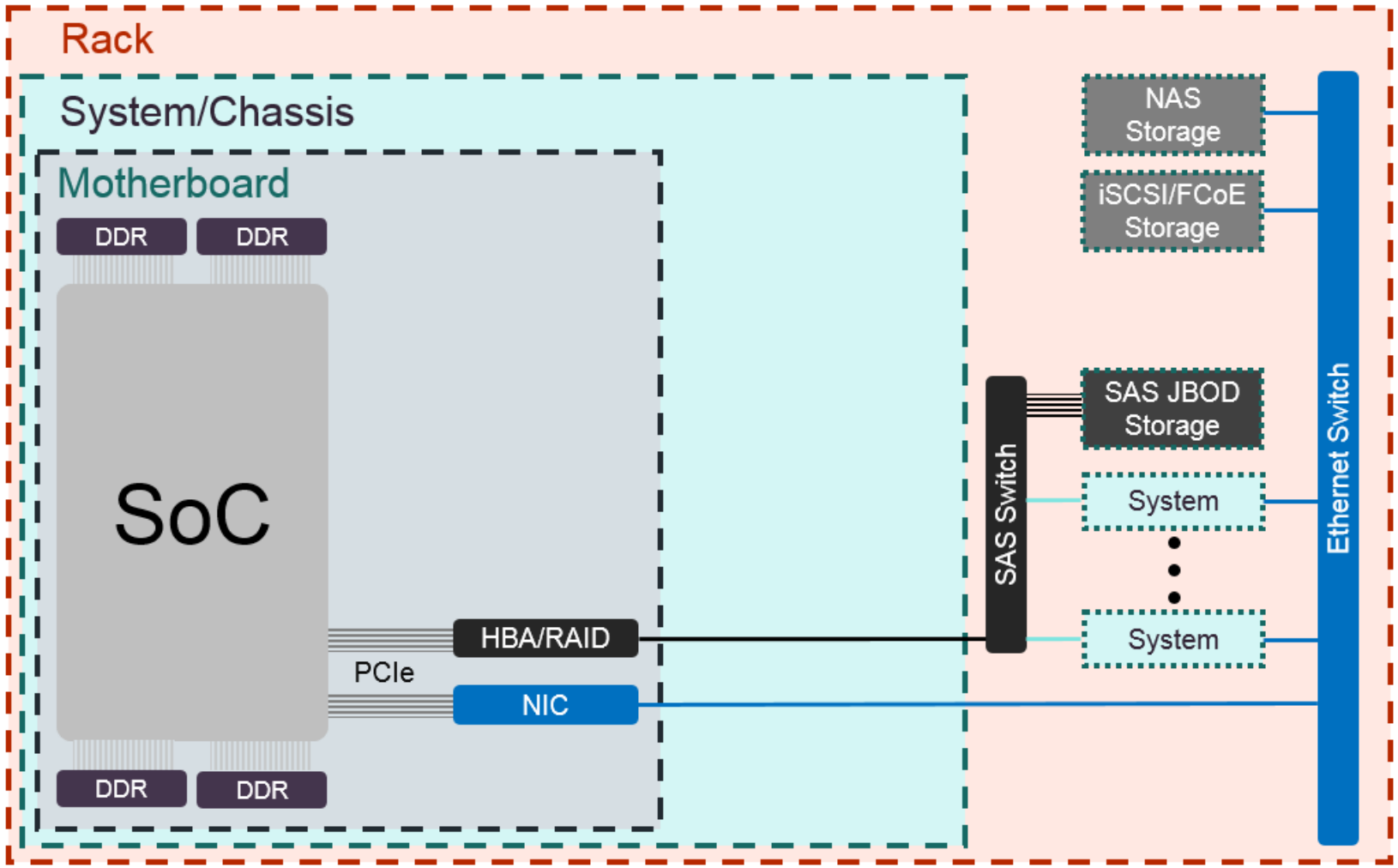
- Traditional system level memory/storage
  - DRAM/SCM typically not placed off motherboard
  - SSDs (SAS/NVMe) are primary system level options
    - Block level devices installed in drive bays

Gen-Z evolution: Augment at the system level



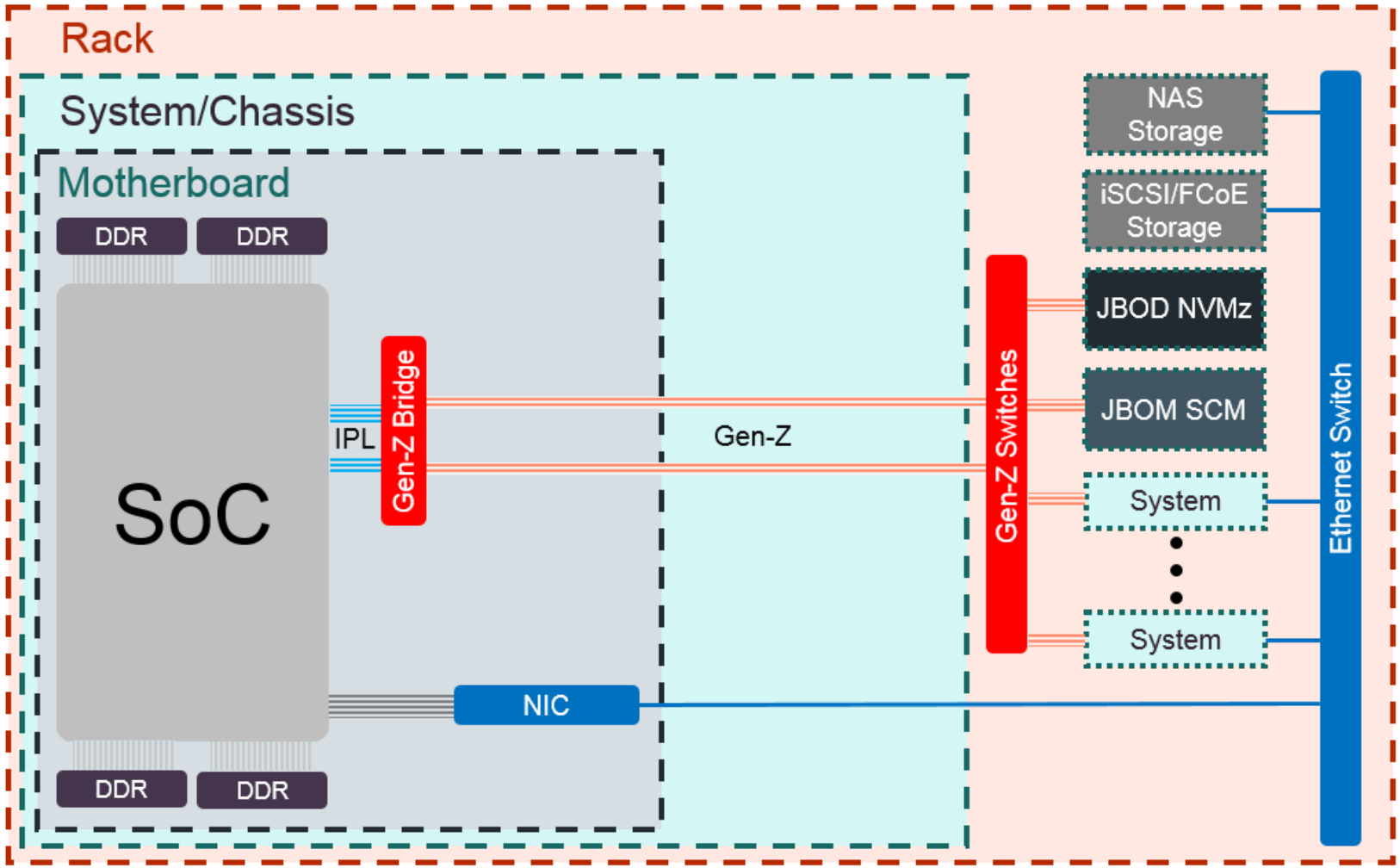
- Traditional system level memory/storage
  - DRAM/SCM typically not placed off motherboard
  - SSDs (SAS/NVMe) are primary system level options
    - Block level devices installed in drive bays
- Gen-Z SSD drives
  - Better performance, scalability, resiliency (vs NVMe)
- *New Gen-Z system level SCM modules*
  - Expandable, serviceable, high capacity memory

Gen-Z Evolution: Augment at the rack-scale



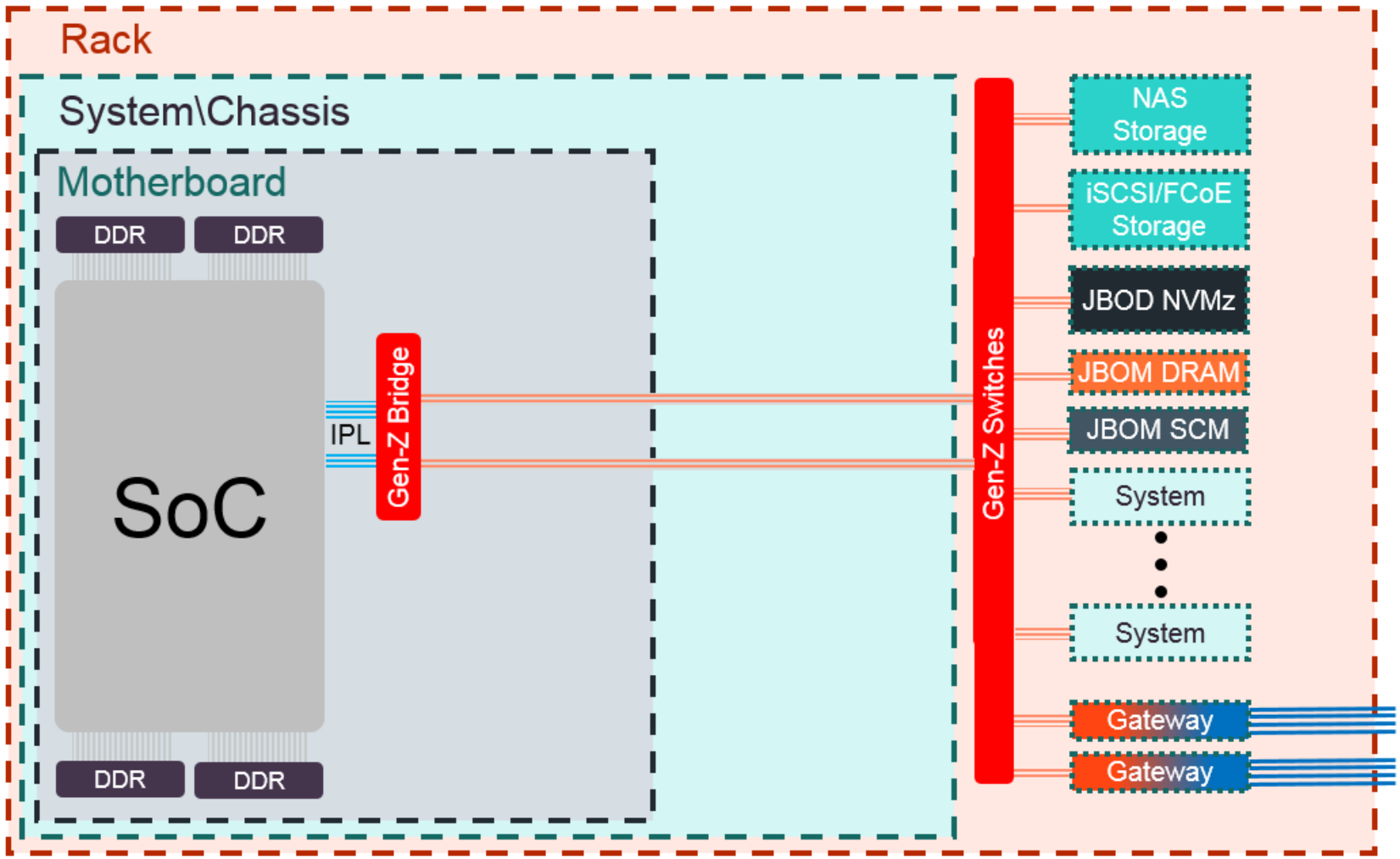
- Traditional rack-scale memory/storage
  - No DRAM/SCM supported
  - NAS & block storage systems
    - SSDs (SAS/NVMe) & disks (SAS)
  - Ethernet/FC/SAS fabric connectivity

Gen-Z Evolution: Augment at the rack-scale



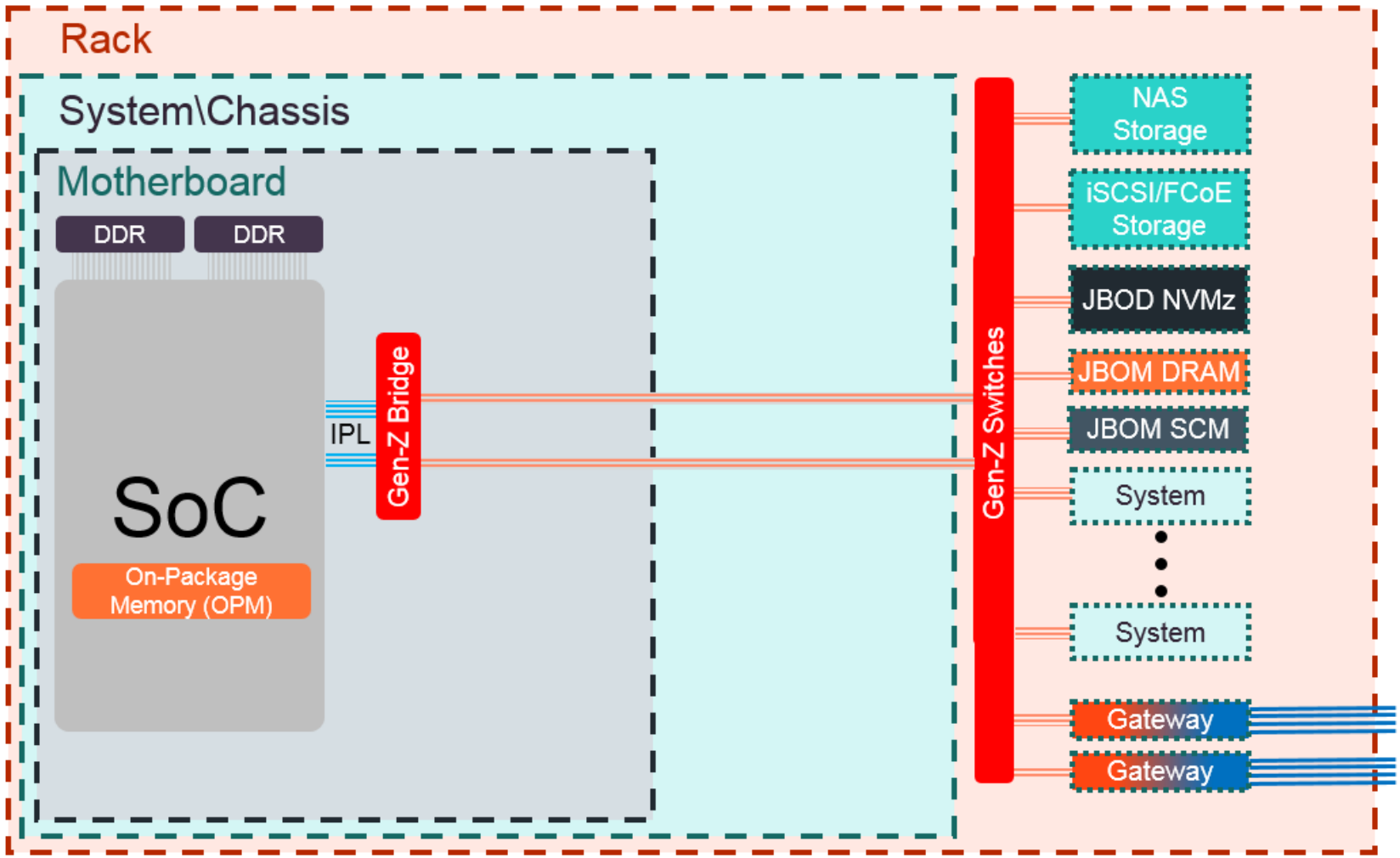
- Traditional rack-scale memory/storage
  - No DRAM/SCM supported
  - NAS & block storage systems
    - SSDs (SAS/NVMe) & disks (SAS)
  - Ethernet/FC/SAS fabric connectivity
- Gen-Z rack scale block storage
  - Gen-Z SSD array or storage array
  - Better performance, scalability, reliability
- *Gen-Z rack scale SCM modules*
  - Just a bunch of memory (JBOM)
  - Expandable, serviceable, high capacity

Gen-Z Evolution: Replace at the rack level



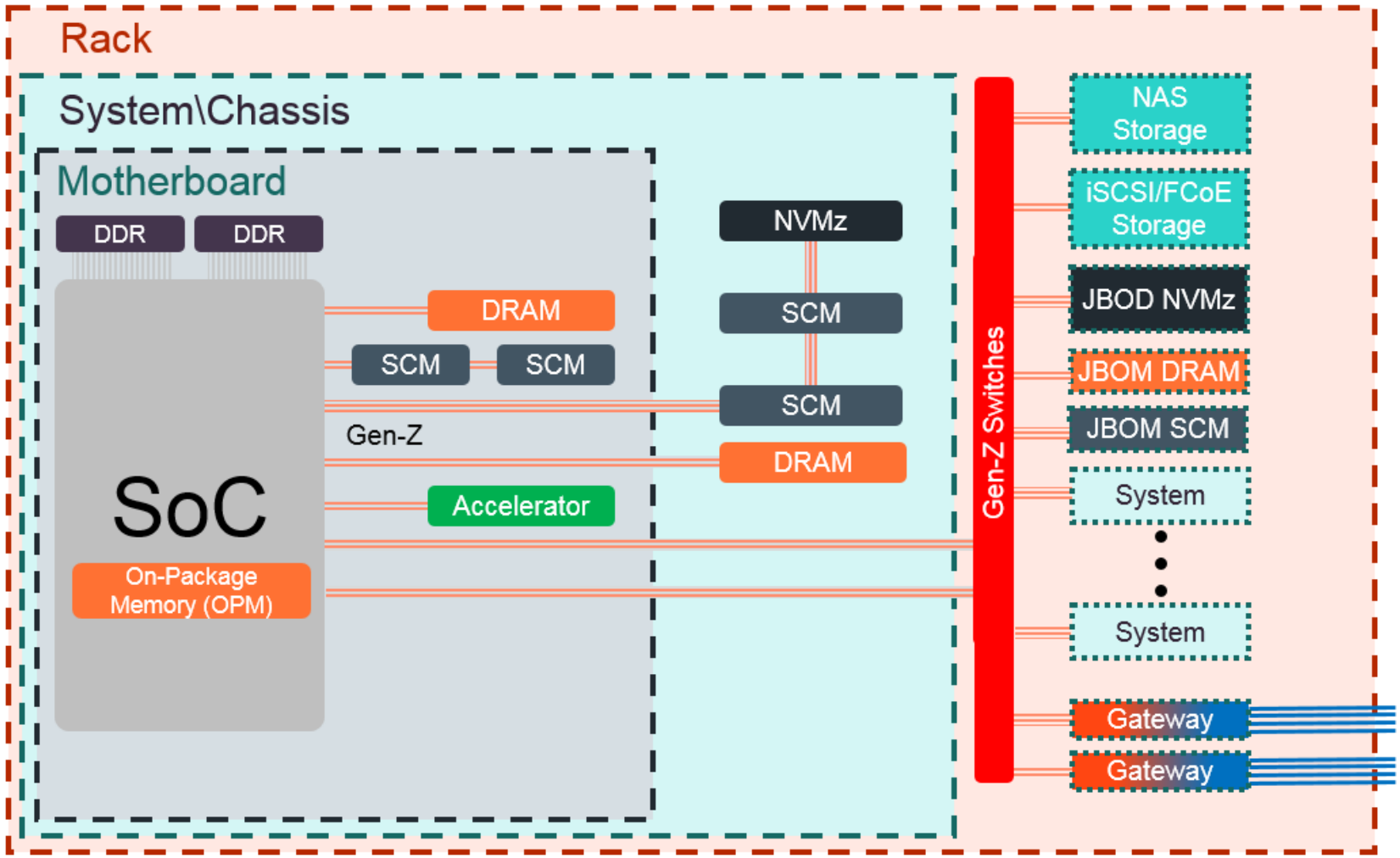
Hewlett Packard Enterprise

Gen-Z Evolution: Replace at the rack level



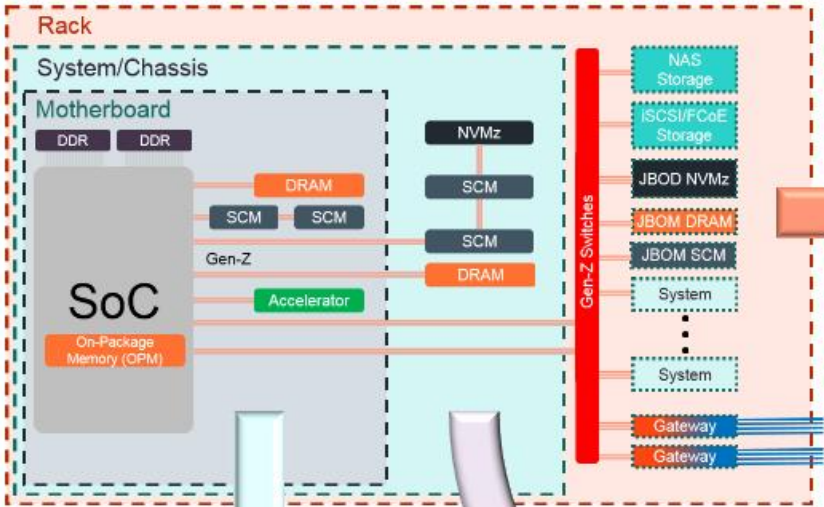
- On package memory is adopted
  - e.g. High Bandwidth Memory (HBM)
  - Freeing CPU pins (for more Gen-Z links)

Gen-Z Evolution: Replace at the rack level



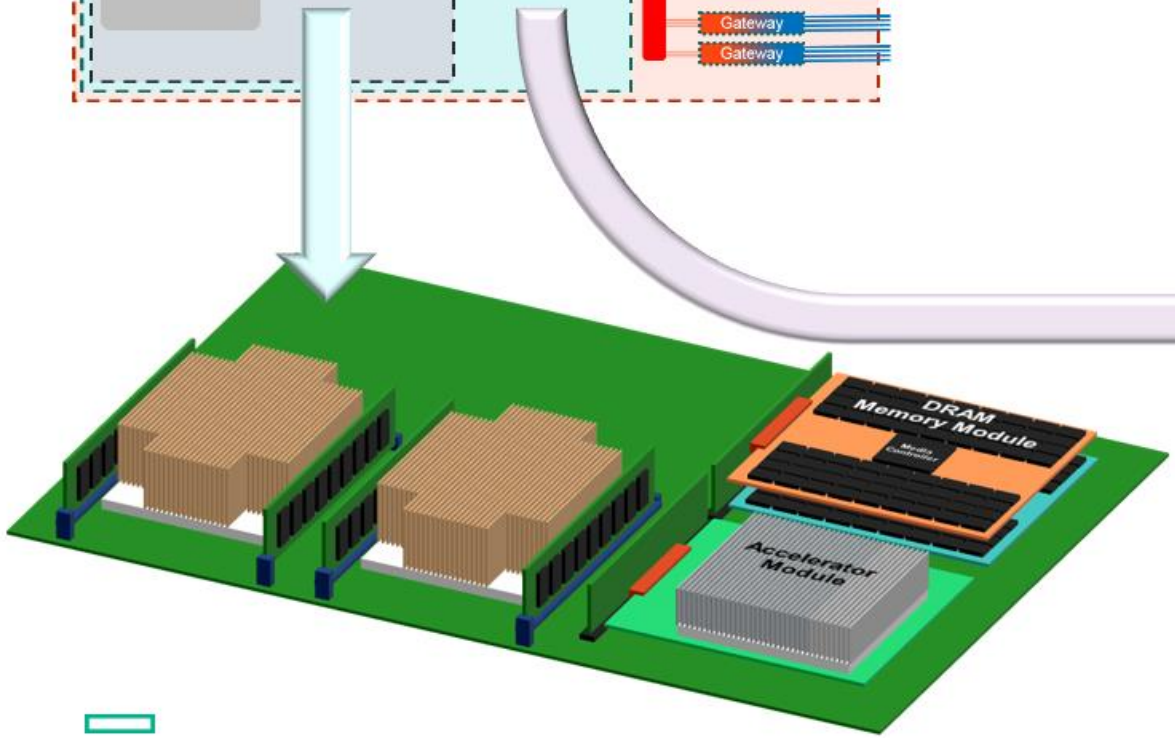
- On package memory is adopted
  - e.g. High Bandwidth Memory (HBM)
  - Freeing CPU pins (for more Gen-Z links)
- Gen-Z is integrated into SoC
  - Enables Gen-Z high performance DRAM
  - Gen-Z & PCIe pins can be shared and allocated as needed per deployment
  - ***DRAM and/or Fast SCM anywhere***
  - ***Accelerators anywhere***

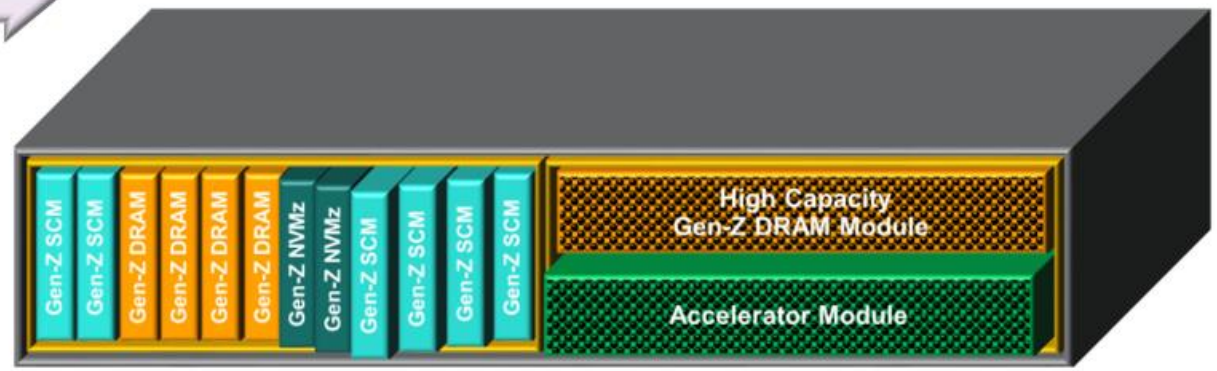
Gen-Z Evolution: Innovation



Accelerator

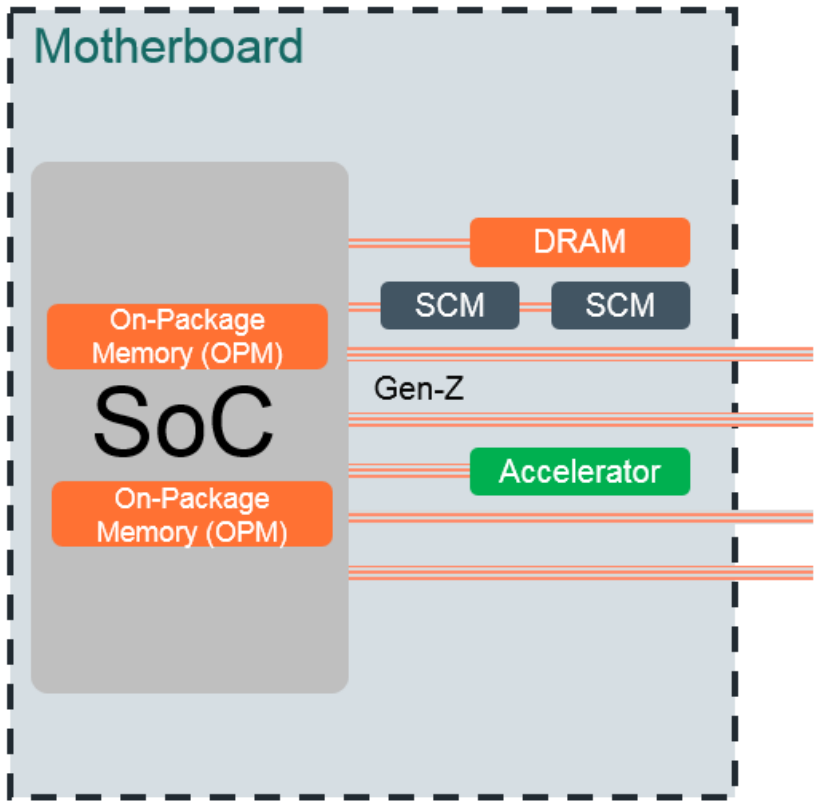
Fabric Attached DRAM





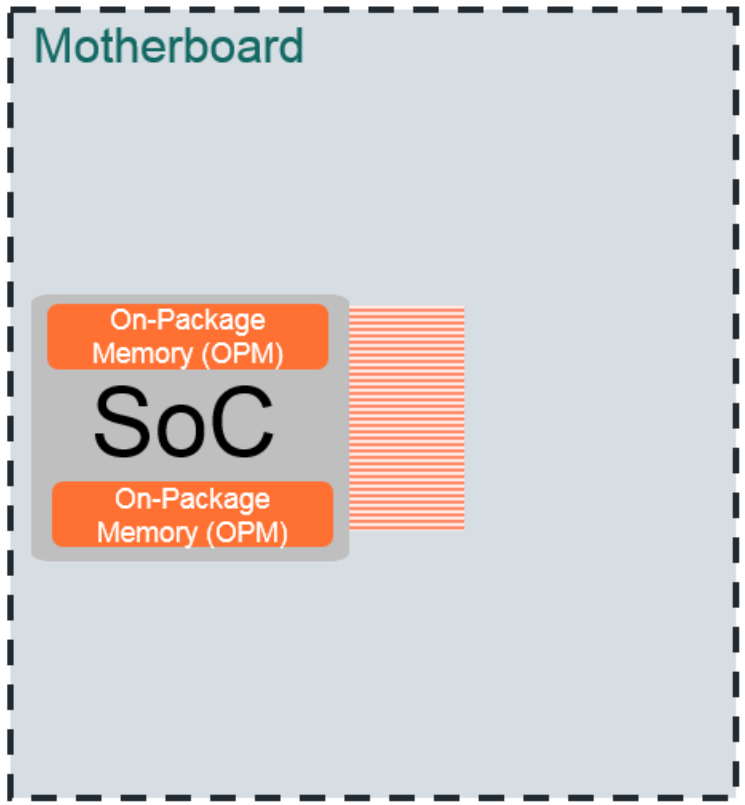
Hewlett Packard Enterprise

Gen-Z Evolution: Composability



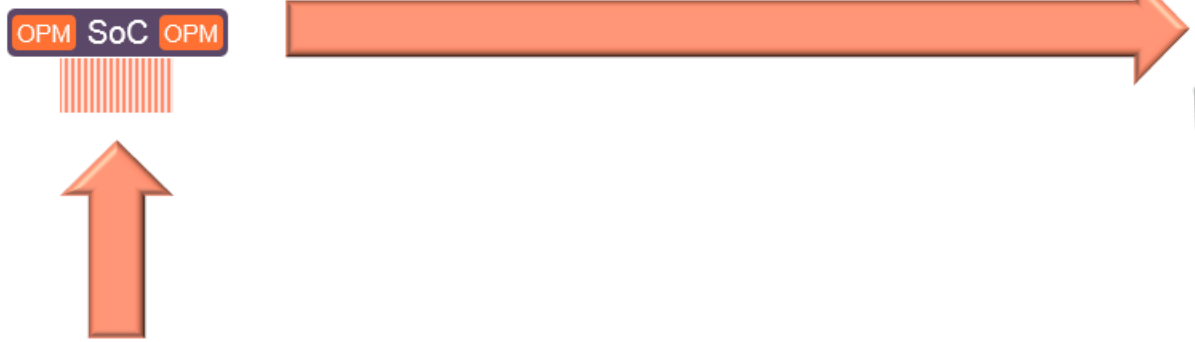
- On package memory replaces external DDR
  - Freeing more CPU pins (for more Gen-Z links)

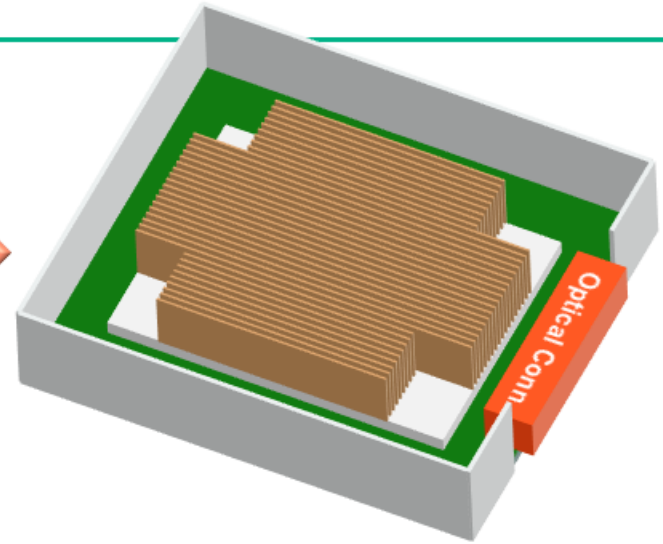
Gen-Z Evolution: Composability

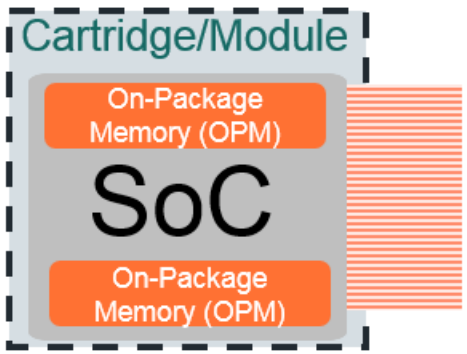


- On package memory replaces external DDR
  - Freeing more CPU pins (for more Gen-Z links)
- Integrated Gen-Z enables high speed access anywhere
  - No need to place peripherals locally on motherboard
- Results in smaller CPU package

Gen-Z Evolution: Composability



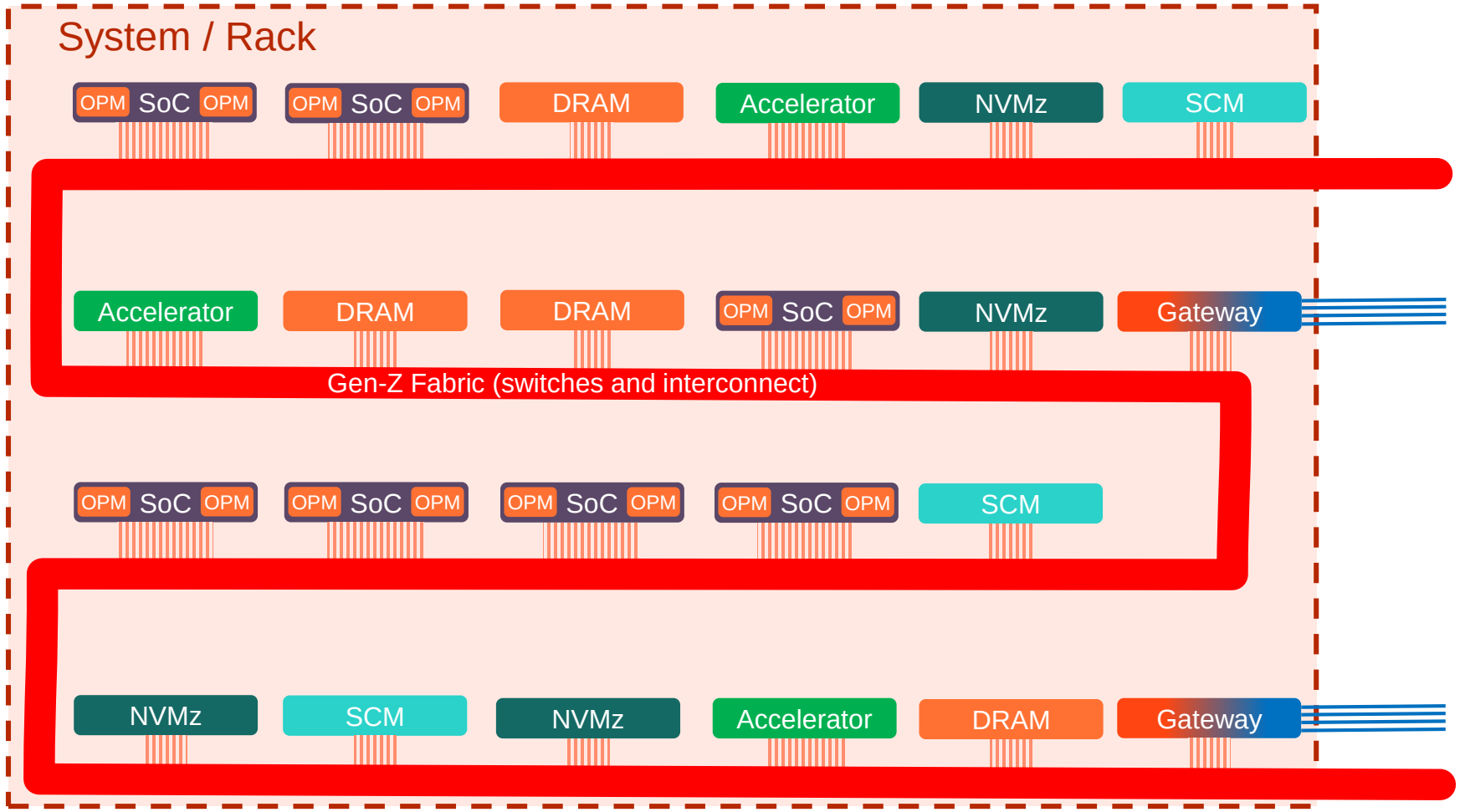


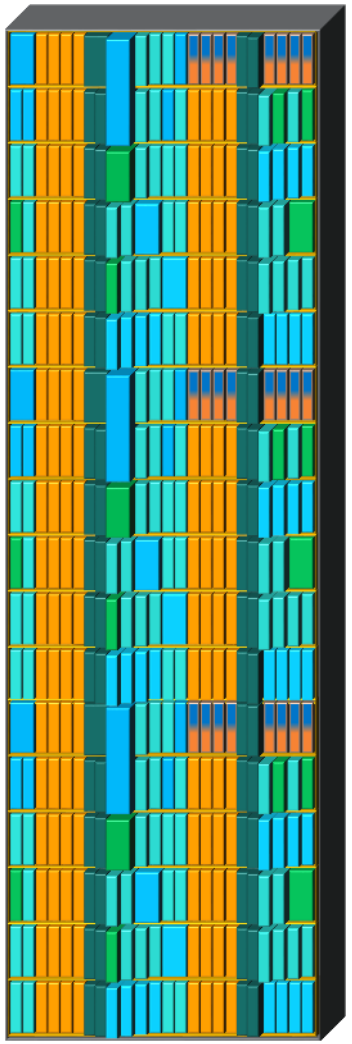


- On package memory replaces external DDR
  - Freeing more CPU pins (for more Gen-Z links)
- Integrated Gen-Z enables high speed access anywhere
  - No need to place peripherals locally on motherboard
- Results in smaller CPU package
- Server motherboard becomes a cartridge or module!!
  - ***Can be placed anywhere just like any other peripheral !!***

Hewlett Packard Enterprise

Gen-Z and Composable Infrastructure





Compose physical infrastructure as easily as composing virtual infrastructure

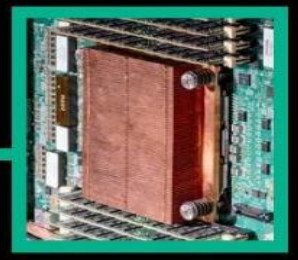
Hewlett Packard Enterprise

Q: How do we get to Memory-Driven Computing?

Hewlett Packard Enterprise

GEN-Z and the Machine

MEMORY FABRIC SWITCH
Enables processors to access Fabric-Attached Memory across any node on the system.

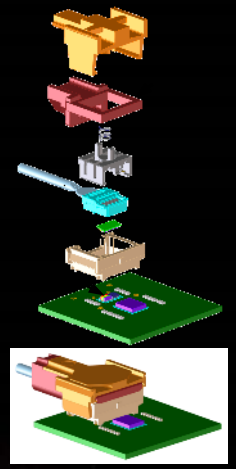


TASK-SPECIFIC PROCESSING
Flexible Memory-Driven Computing architecture can match compute tasks to different types of processor to optimize performance and efficiency.

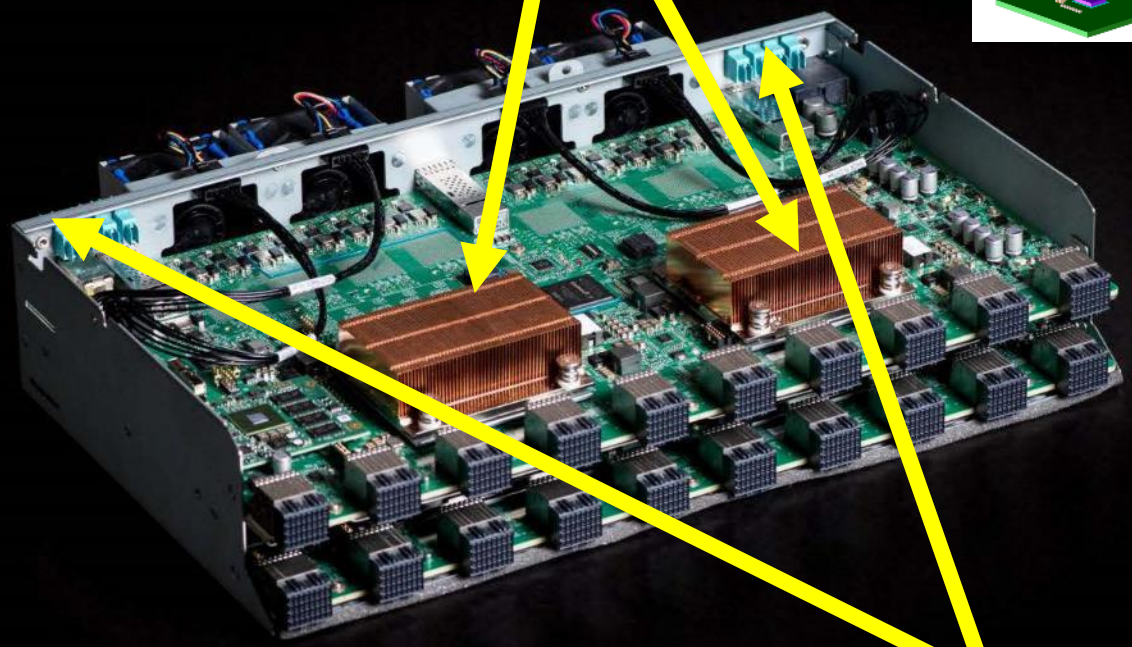
PHOTONICS INTERCONNECTS
Rapidly transfers data between enclosures with light instead of electricity to access shared memory.

MEMORY AT THE CENTER
Combines memory and storage into a vast pool of Fabric-Attached Memory to radically increase computing efficiency and speed by enabling multiple processors to share memory.

Memory Fabric
Switches



High-speed
High-density
Small form factor
Low power
Low cost
VCSEL
optics



HPE introduces the world's largest single-memory computer – The prototype contains 160 terabytes of memory

•160 TB of shared memory spread across 40 physical nodes, interconnected using a high-performance fabric protocol.

•An optimized Linux-based operating system running on ThunderX2, Cavium's flagship second generation dual socket capable ARMv8-A workload optimized System on a Chip.

•Photonics/Optical communication links, including the new X1 photonics module, are online and operational.

•Software programming tools designed to take advantage of abundant of persistent memory.

Hewlett Packard Enterprise

How big is 160 TB?

160 Terabyte prototype can simultaneously work with over

160 million

books worth of content

=

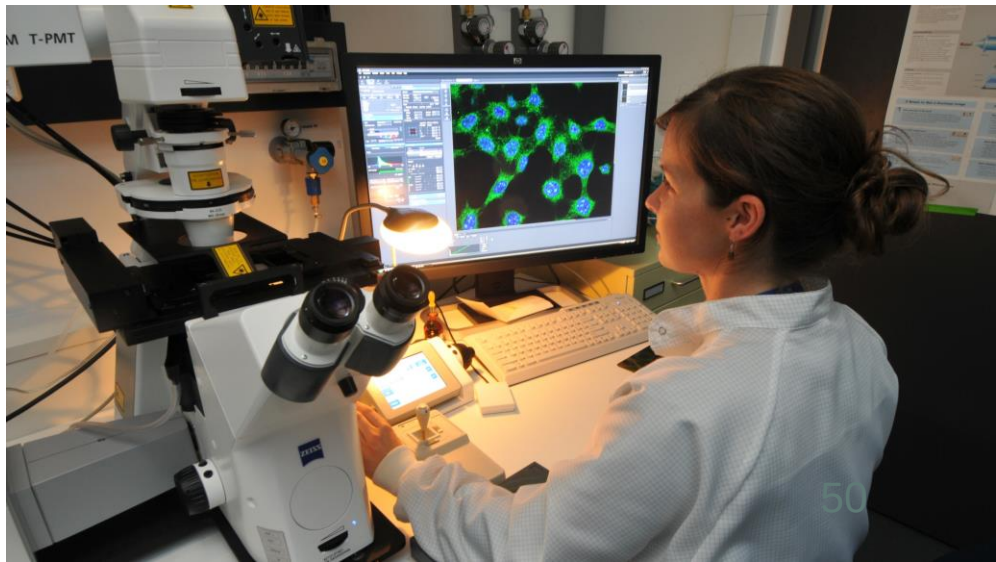
- Library of Congress
- Library of Congress
- Library of Congress
- Library of Congress
- Library of Congress

Hewlett Packard Enterprise

First collaboration: DZNE

- Collaborating with German Center for Neurodegenerative Diseases (DZNE) needed a new kind of computer to manipulate their massive data sets to accelerate finding a cure for Alzheimer's, a disease that impacts one in 10 people age 65 or older in the world. The initial findings from our collaboration are powerful and promising:
  - We've only started to scratch the surface with one component of their overall data analytics pipeline and are already **getting 9X speed improvements.**
  - We're getting **results that used to take more than 22 minutes in less than 3.**
  - We believe these **gains could increase up to 100X** when we expand our learnings to the other components of their pipeline. Saved time translates to save lives, these efficiencies could change the game.
  - DZNE has never been able to work with so much data at one time, which means **different correlations and better answers than ever before** – ultimately resulting in new discoveries to help cure Alzheimer's.

DZNE
Deutsches Zentrum für Neurodegenerative Erkrankungen in der Helmholtz-Gemeinschaft



Hewlett Packard Enterprise

U.S Department of Energy works with HPE to design a Memory-Driven SuperComputer

- Develop a reference design for an exascale supercomputer that will enable a broad set of modeling and simulation applications unachievable today

- Accelerating breakthroughs in science, medicine, technology, engineering and many other fields.

- Scientific applications would impact nearly every corner of research, from the physics of star explosions to precision medicine for cancer.

"We see this DOE grant as a vote of confidence in the ability of HPE and Hewlett Packard Labs to help overcome daunting technology challenges that are impeding everyone's progress toward exascale computing," - Steve Conway, IDC research vice president of high performance computing

Hewlett Packard Enterprise

Thank You

Questions?