13th ANNUAL WORKSHOP 2017

GEN-Z AN OVERVIEW AND USE CASES

Greg Casey, Senior Architect and Strategist Server CTO Team

DellEMC

March, 2017

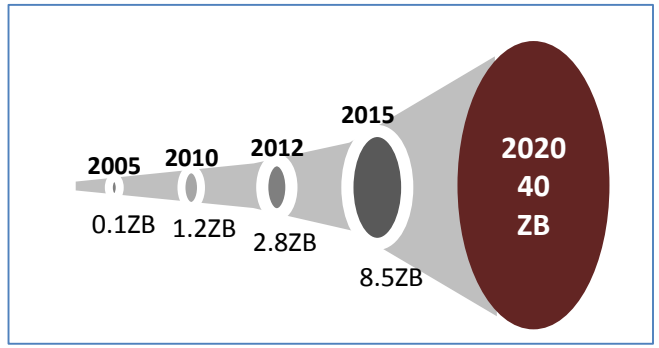
WHY PROPOSE A NEW BUS?

- **System memory is flat or shrinking**
  - Memory bandwidth per core continues to decrease
  - Memory capacity per core is generally flat
  - Memory is changing on a different cadence compared to the CPU

- **Data is growing**
  - Data that requires real-time analysis is growing exponentially
  - The value of the analysis decreases if it takes too long to provide insights

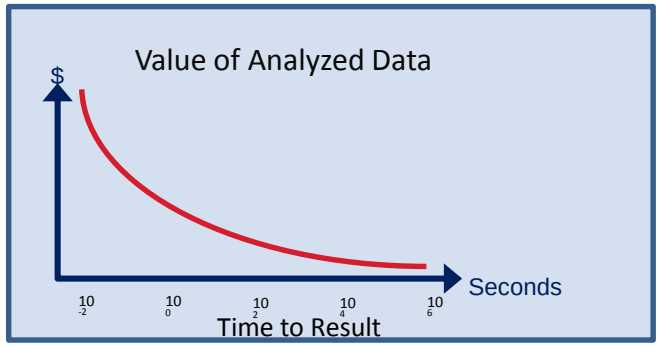
- **The industry needs an open architecture to solve the problems**
  - Memory tiers will become increasingly important
  - Rack-scale composability requires a high bandwidth, low latency fabric
    - Composability is the ability to utilize resources in an efficient manner
  - Must seamlessly plug into existing ecosystems without requiring OS changes

Explosive Growth of Data



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

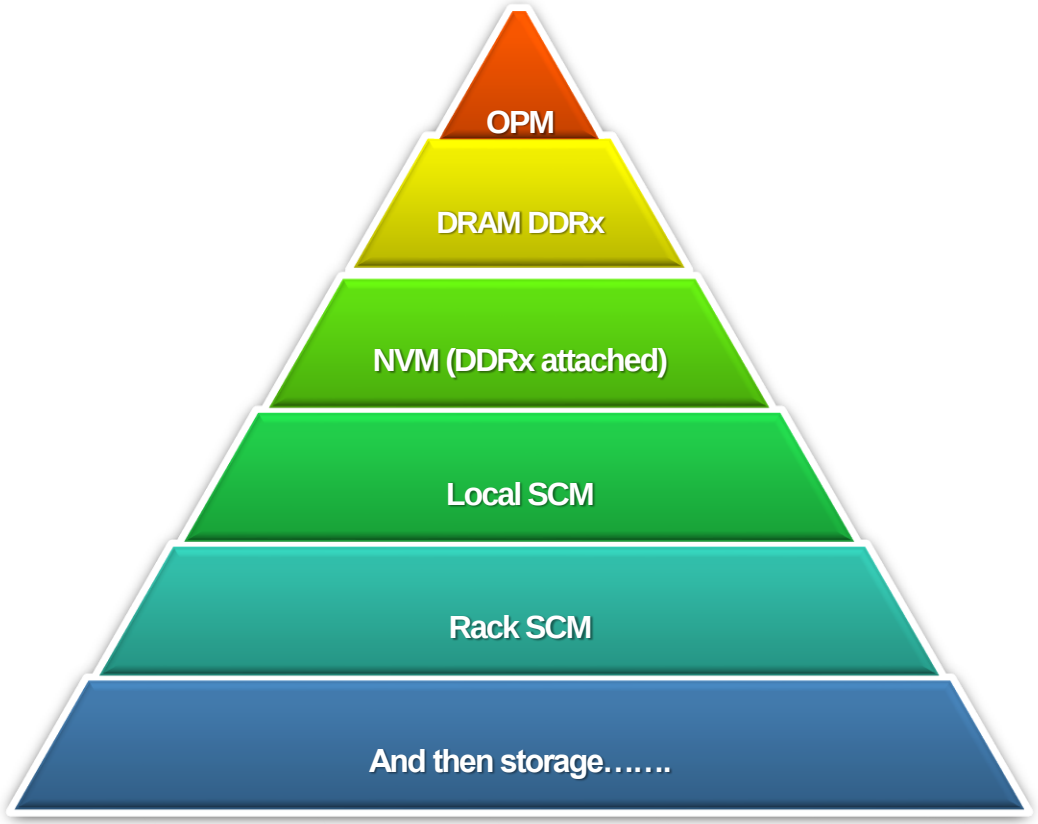
Need Answers … FAST!



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

OBLIGATORY MEMORY PYRAMID SLIDE

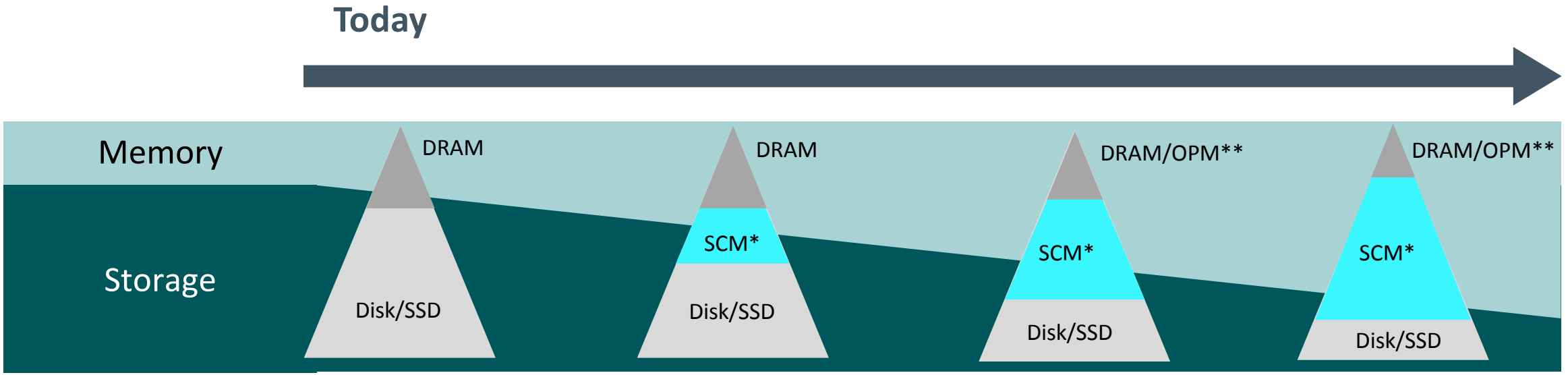
Notes: SCM -> Storage Class Memory
OPM -> On Package Memory
NVM -> Non-Volatile Memory



Memory Semantic Domain

Block/File Domain

MEMORY/STORAGE CONVERGENCE: THE MEDIA REVOLUTION



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

*SCM = Storage Class Memory

**OPM = On-Package Memory

New and Emerging Memory Technologies

- HMC
- HBM
- RRAM
- 3DXPoint™ Memory
- MRAM
- PCM
- Low Latency NAND
- Managed DRAM

HARDWARE SERVER REALITIES

- Pictured - Current Intel Xeon - 4 processors – 48 DIMMs

- Memory Bandwidth requirements are driving up the number of memory channels – Growing !
- Number of DIMMs per Channel – Shrinking !
- Power in DIMM – Growing !
- Power in CPU – Growing !
- Size of CPU – Growing !
- Size of DIMM – Growing !
- Speed Of Memory Channel – Growing !
- Number of Cores in each Socket – Growing !
- Customer Software Memory Requirements – GROWING !!!!

- Not to mention IO Busses, Integrated Graphics, GPU and FPGA Acceleration support

Gen-Z offers Architectural Opportunities

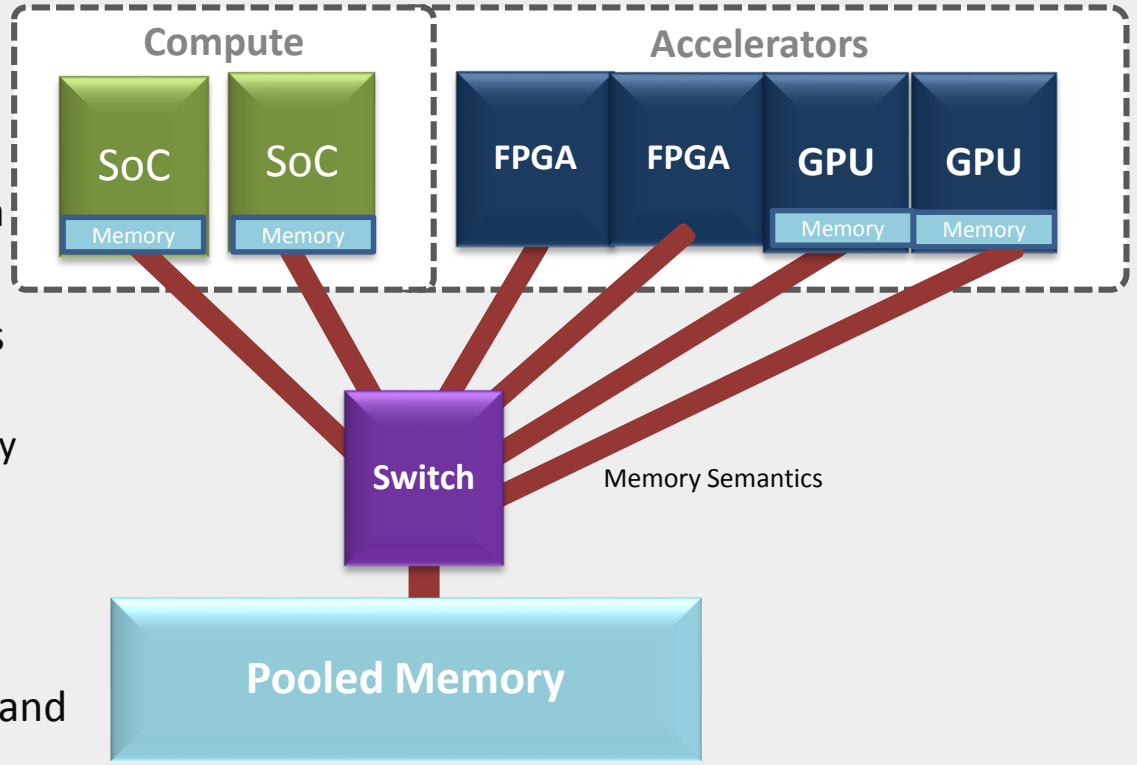
MEMORY SEMANTIC FABRIC
COMMUNICATION AT THE SPEED OF MEMORY

What is a Memory Semantic Fabric?

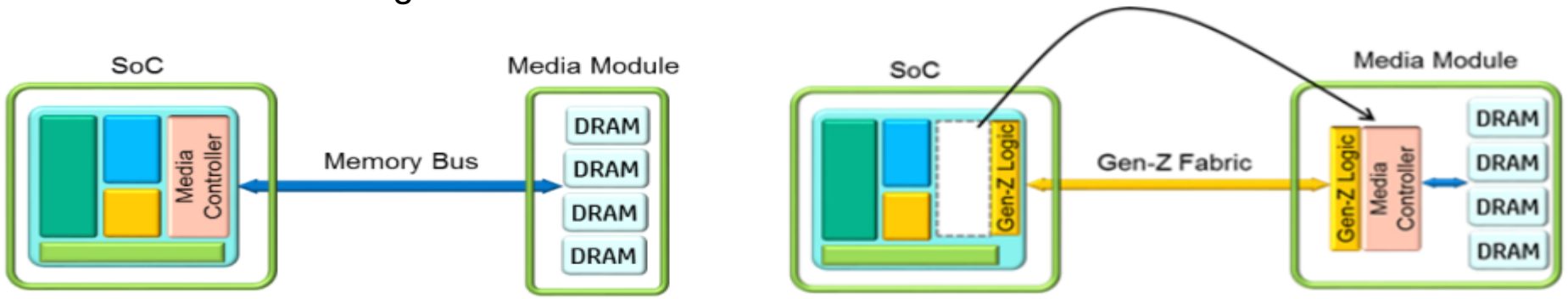
- Handles all communication as memory operations such as load/store, put/get and atomic operations typically used by a processor
- Memory semantics are optimal at sub-microsecond latencies from CPU load command to register store
  - Unlike, storage accesses which are block based and managed by complex, code intensive, software stacks

Why Now?

- The emergence of low latency, Storage Class Memory (SCM) and the demand for large capacity, rack scale resource pools, and multi node architectures



GEN-Z ATTRIBUTES

- Feature-scalable packetized transport
- Scalable and power-proportional link, physical layers, and underlying memory media access.
- Split memory controller and media controller paradigm that hides microarchitecture details and idiosyncrasies.
  - Split-controller model breaks the processor-memory interlock providing numerous technical and economic benefits while unlocking innovation
  - 

  etc.

Memory media independent.

- Solutions can transparently incorporate and evolve the optimal media for a given application while ensuring interoperability.

GEN-Z
A NEW DATA ACCESS TECHNOLOGY

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

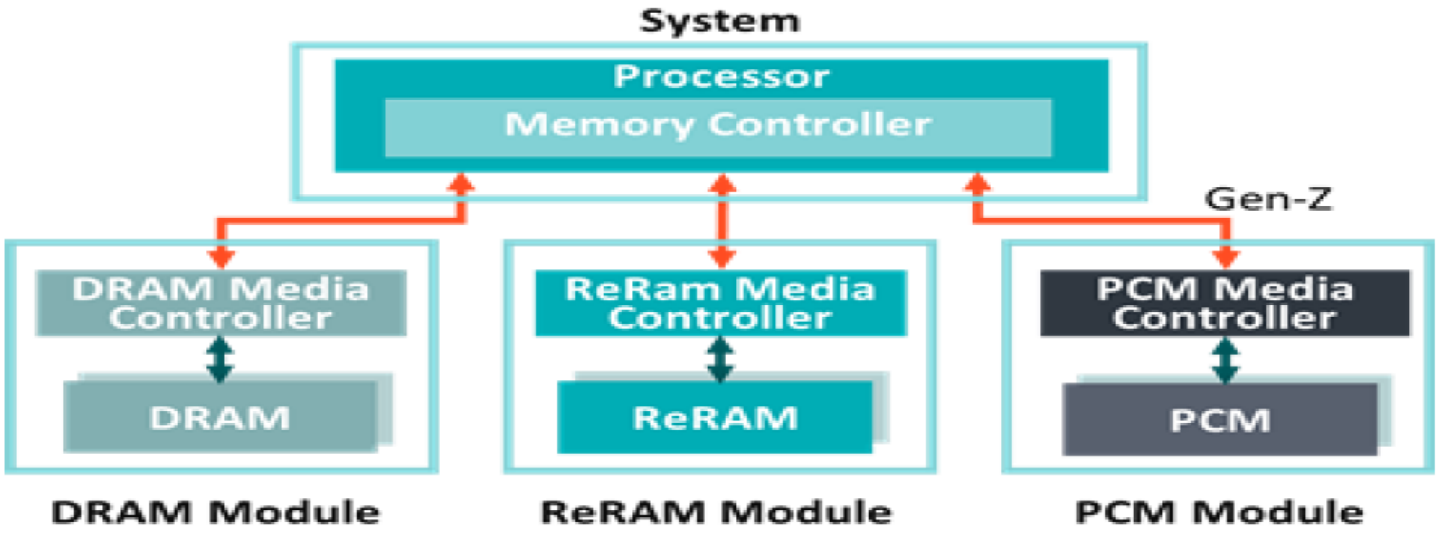
Secure Compatible Economical

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

Gen-Z *P2P, daisy or switched Topology*

GEN-Z ATTRIBUTES (CONTINUED)

- Supports processor-centric and memory-centric architectures
  - Processor-centric to ease Gen-Z transition
  - Memory-centric option to optimize memory access / movement
- Abstract physical layer interface supporting multiple physical layers and media
  - Easily tailored to market-specific needs.
  - Rapid evolution or replacement without waiting for entire ecosystem to move in lock-step



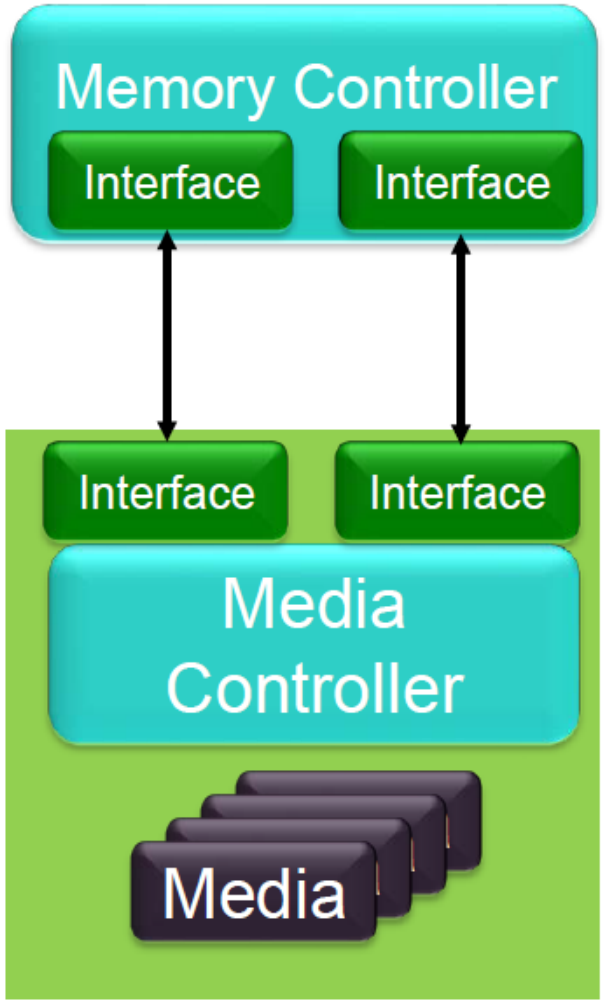
GEN-Z ATTRIBUTES (CONTINUED)

- Market-driven packaging and fabric topologies
  - Single or multi-link point-to-point topologies
  - Switched fabric topologies—component-integrated or discrete
- Common data transport with application semantic overlays to support diverse component types—processors (variety of types), memory, I/O, storage, network, FPGA, DSP, graphics, etc.
- Workload and environmentally-driven optional capabilities
  - Asymmetric interfaces and links
  - Real-time dynamic interface width and link width
  - Memory persistency
  - Hardware-based differentiated communication services.
  - Advanced and vendor-defined operations.
  - Messaging services for any-to-any communications between diverse component types
- Strong data integrity combined with transparent end-to-end packet error recovery.
- Operating system (OS) and processor independence.

GEN-Z ATTRIBUTES (CONTINUED)

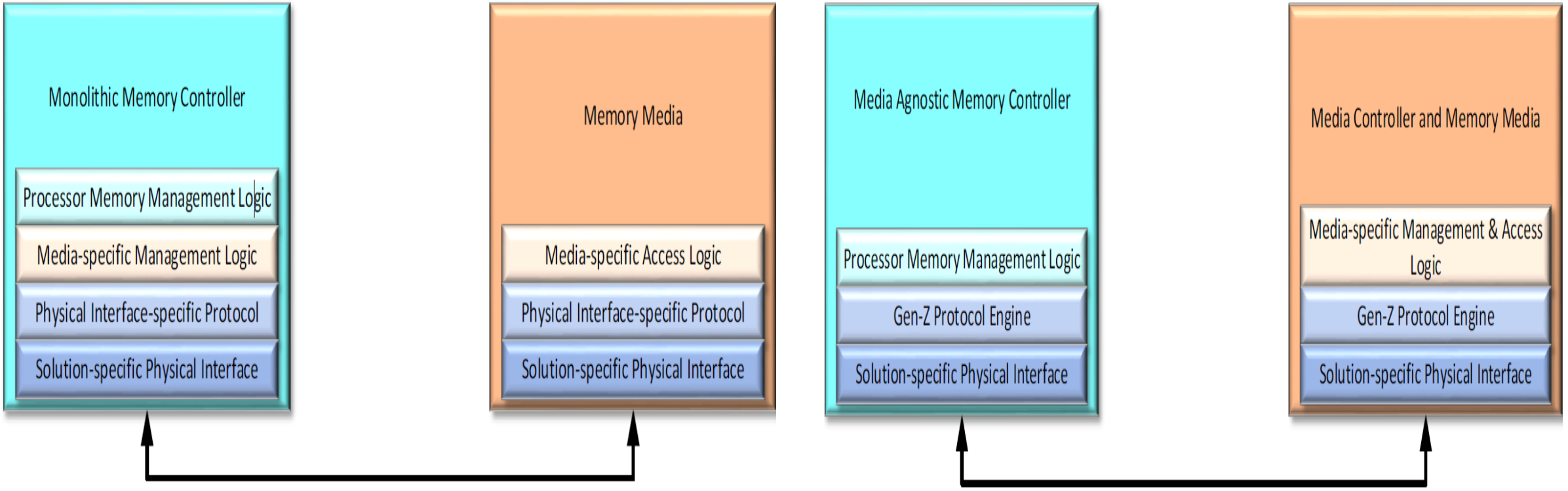
- Optional scalability:
  - Up to 2^{44}or 2^{64}per component memory addressing (zero and non-zero based)
  - Support from 2 to 2048 components per subnet.
    - Trivial subnet—point-to-point / daisy-chain / linear switch
      - Hybrid and tiered topologies supported
    - Robust subnet—many processor, memory and optionally diverse components
    - Multiple subnets per component
    - Multiple subnets joined via transparent routers
- Architected services to enable robust security solutions

BREAKS PROCESSOR-MEMORY INTERLOCK



- Split controller model
  - Memory controller
    - Initiates high-level requests—Read, Write, Atomic, Put / Get, etc.
    - Enforces ordering, reliability, path selection, etc.
- Media controller
  - Abstracts memory media
    - Supports volatile / non-volatile / mixed-media
  - Performs media-specific operations
  - Executes requests and returns responses
  - Enables data-centric computing (accelerator, compute, etc.)
  - Enables third-party data movement
  - Eliminates need for tight timing budgets
  - Transparent caching and acceleration to improve performance

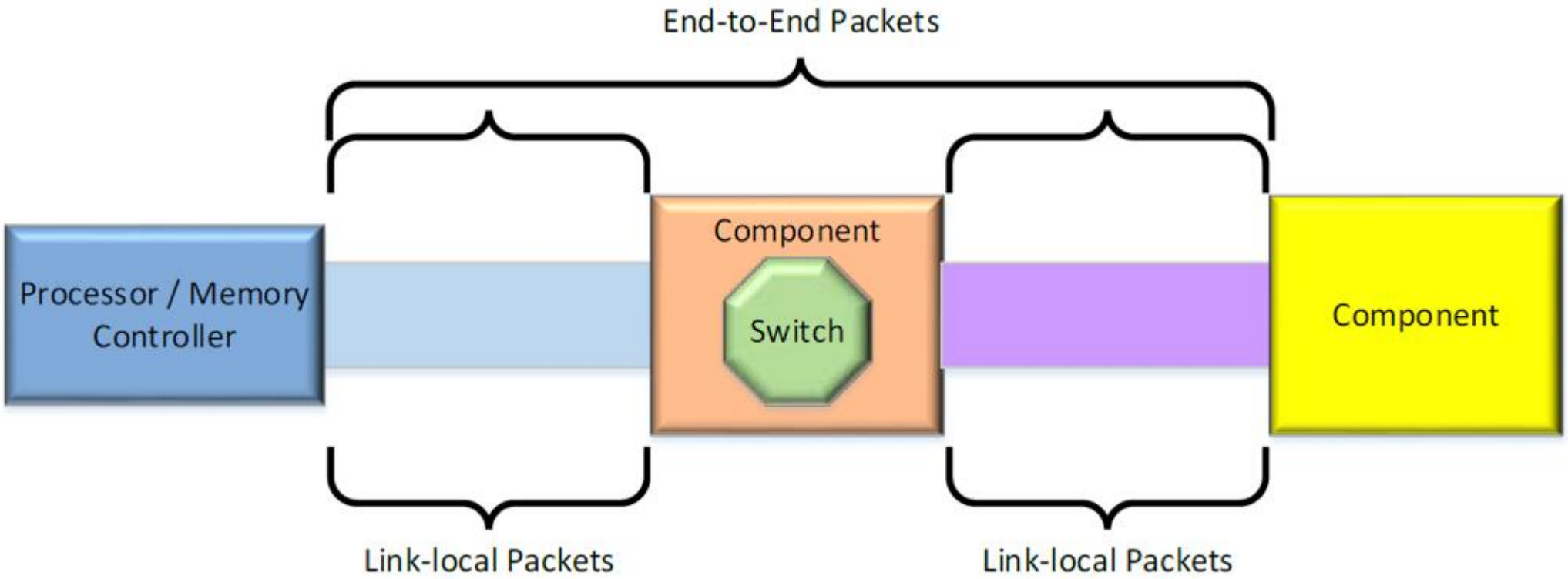
GEN-Z BREAKS THE PROCESSOR-MEMORY INTERLOCK



Today's Processor-Memory

Gen-Z Processor-Memory

GEN-Z HAS TWO PACKET TYPES



GEN-Z OPCLASS & OPCODE

- OpClass - To minimize resource requirements and provide flexibility and extensibility, requests and responses are organized into operation classes (OpClass).
  - P2P-Core OpClass - point-to-point topologies and therefore supports a simpler, non-switchable protocol.
  - Core OpClass - is implicit in that it does not contain an OpClass Label field (OCL). Packets associated with the Core OpClass may be exchanged on any interface not enabled for the P2P-Core OpClass or an implicit Vendor-defined OpClass.
  - Core 64 OpClass - extends many Core operations to support up to an effective 64-bit address.
  - Control OpClass - Control OpClass is used to access configuration resources located in Control Space.
  - Atomic1 OpClass - exchange Atomic Requests and Responses.
  - Large Data Movements - large Read Requests and Buffer requests
  - Advanced OpClass –
  - Context ID - supports operations that use a context identifier in place of an address to identify a target resource.
  - Multicast OpClass - multicast operations between components participating in a multicast group
  - Strong Order Domain OpClass - support ordered packet communications
  - Vendor Defined OpClass - enable customized operations to be exchanged between cooperating components.

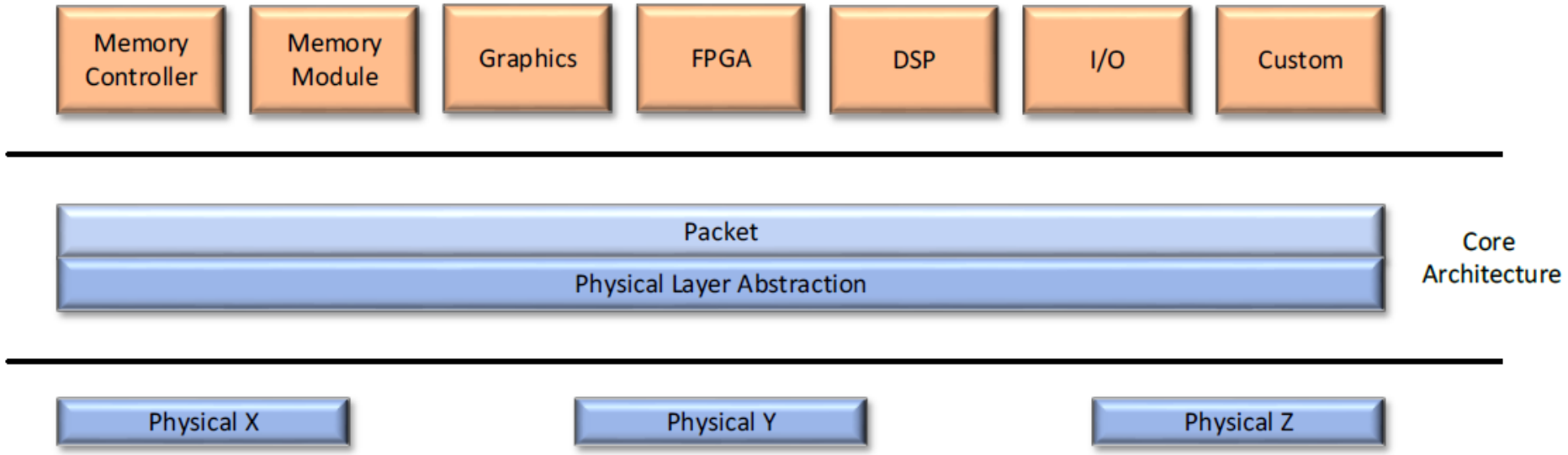
CACHE COHERENCY PROTOCOL

- Gen-Z supports a set of operations to allow coherent communications
  - Invalidate and Writeback, Write with Target Cache, Read Modified, Read Exclusive, etc.

- Cache coherency protocols are customized to a given processor ISA
  - Source of innovation and differentiation
  - Standard coherency protocol would require a per ISA translation bridge chip adding solution cost / latency / complexity

- Off-chip coherency protocols are difficult to efficiently scale
  - Requires complex coherency schemes, e.g., directories
  - Requires complex error and resiliency schemes to avoid cascade failures

LAYERED ARCHITECTURE



- Core architecture defines operations, protocol, and physical layer abstraction10s-100s GB/s to TB/s (future) per link bandwidth
- Multiple physical layers and signaling rates specified per market
  - Leverage existing standards and map to Gen-Z specificsCurrent signaling proposal is (16 / 25 / 28 (NRZ) / 56 GT/s (PAM 4) / 112 GT/s (PAM 4)
  - Supports electrical and optical medias (VCSEL / SiP) with multiple lambda
  - Unidirectional links (separate Tx and Rx lanes in symmetric or asymmetric configurations)

DETAILED SPECIFICATION

UniCast

Buffer Requests

Atomic Requests

Out-of-Band Discovery

Flow Control

LLR

A-Keys

Tags

Global Component IDs

Precision Time

R-Keys

Vendor Defined Packets

Virtual Channels

Interrupts

Unsolicited Event Packet

In-of-Band Discovery

Atomics

Link-Local Flow Control

Wake Threads

Component IDs

MultiCast

Encapsulation

Pattern Requests

Lightweight Notification (LN)

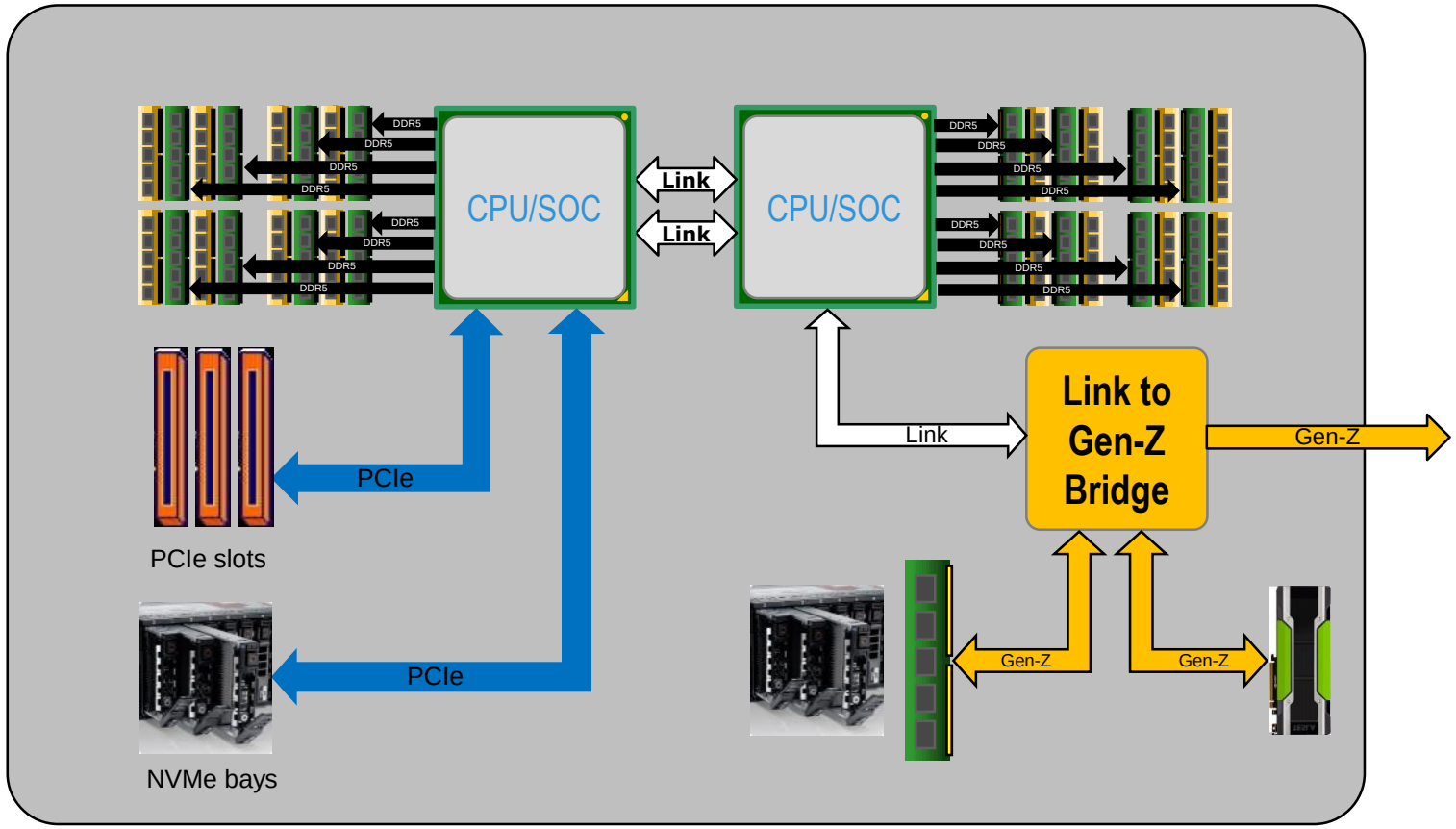
C-State Power Control

SECURITY

- **Architecture supports multiple hardware-enforced isolation mechanisms**
  - Isolation mitigates probability of error or failure ripple effects
    - All violations immediately visible to detecting component and peer (when applicable)
    - All violations immediately reported to management
  - Isolation does not equal security

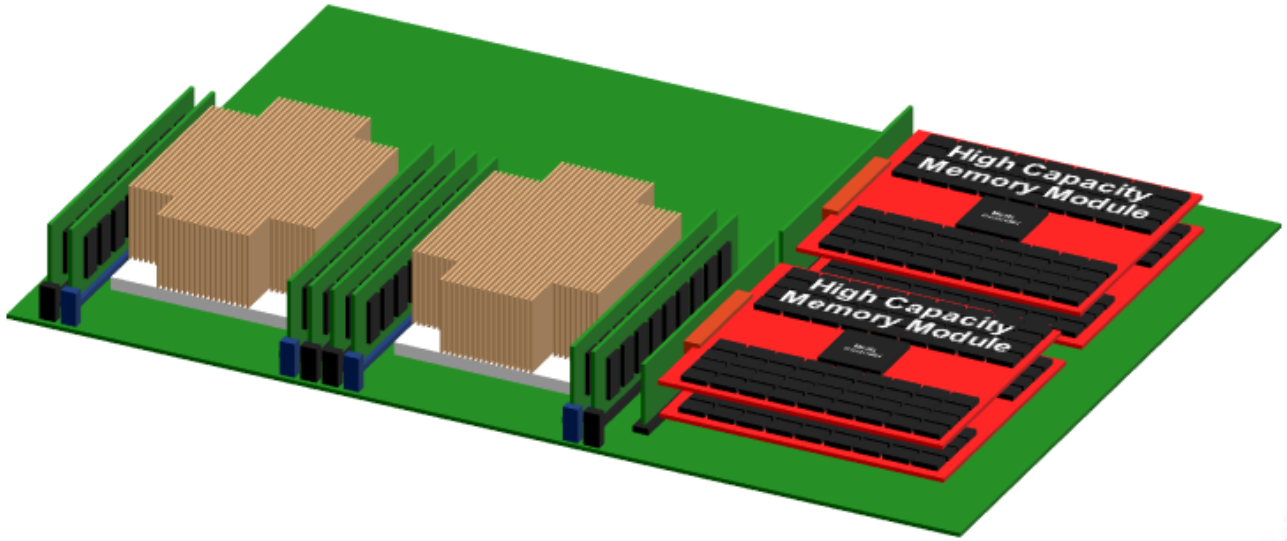
- **Architecture supports authenticated communications**
  - Packets may contain a HMAC (Hash-based Message Authentication Code) and Anti-replay Tag
    - Keys protected by AES-256
    - Multiple secured hash techniques supported
  - Communicating components validate the security fields
    - Authorized peer component, untampered packet, non-replayed packet
    - All violations immediately visible to detecting component and peer (when applicable)
    - All violations immediately reported to management
  - Endpoints are responsible for privacy, e.g., encryption
    - Gen-Z is responsible for ensuring packets are not tampered with or replayed.

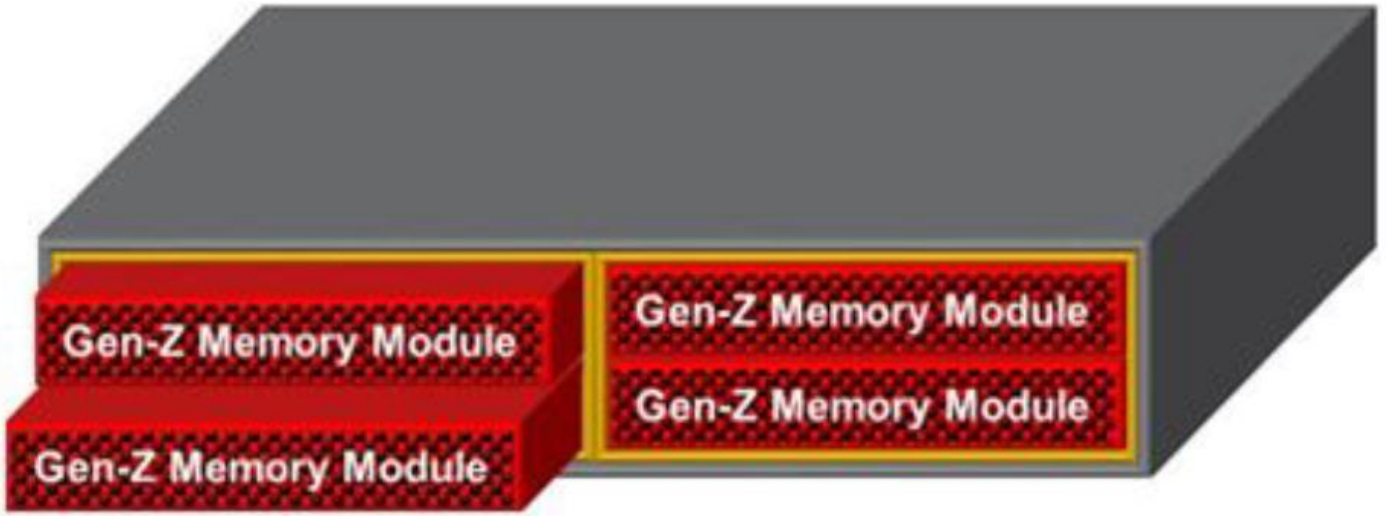
2020 SERVER VISION



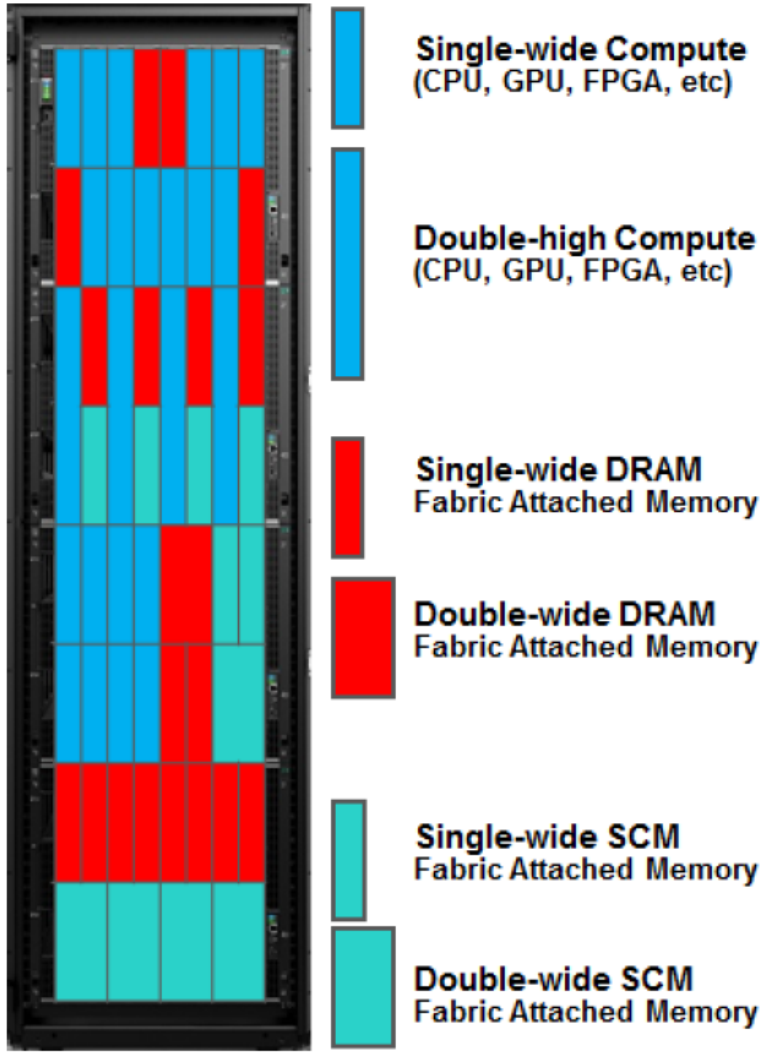
x4, x8?

GENZ MECHANICAL CONCEPTS





GENZ RACK SCALE CONCEPTS



GEN-Z INDUSTRY CONSORTIUM

Mission

- Create a next generation interconnect that will bridge existing solutions while enabling new unbounded innovation
- Develop in an open, non-proprietary standards body where adoption, differentiation and innovation is promoted as an industry standard.

- **A Transparent Organization**: Gen-Z has been formed as a not-for-profit organization, its ongoing development occurs on the basis of an open decision-making procedure available to all interested parties.
- **Wide availability**: The Gen-Z standard will be published and available free of charge.
- **End-User Choice**: There are no constraints on the re-use of the standard. Gen-Z creates a fair, competitive market for implementations of the standard.
- **Equality**: Gen-Z does not favor one implementer over another.

AMD ARM BROADCOM CRAY Micron SAMSUNG SK hynix XILINX

DELL EMC Hewlett Packard Enterprise HUAWEI IDT

Board Members

GEN-Z SUMMARY

- Scalable, universal system interconnect and protocol
- Optimized for memory-semantic communications
  - Highly-extensible and easily customized
  - Delivers transparent aggregation and resiliency services
- Breaks processor-memory interlock
  - Increases solution agility and innovation opportunities
- Enables and optimizes hybrid and data-centric computing
- Opportunity to simplify / reduce software overhead and complexity
- Unmodified OS support
- Scales across solution segments—mobility-to-client-to-server-to-enterprise-to-cloud
- Common modular connector and mechanical form factors—memory, I/O, storage, etc.
- …and much, much more

MEMBERSHIP UPDATES: 31 CURRENT MEMBERS

- Alpha Data
- AMD
- Amphenol Corporation
- ARM
- Broadcom Ltd.
- Cavium Inc.
- Cray
- Dell
- FoxxConn Interconnect Technologies
- HPE
- Huawei R&D USA
- IBM
- IDT
- Lenovo
- Lotes Ltd
- Mellanox Technologies, Ltd
- Mentor Graphics
- Micron
- Microsemi
- Nokia
- PLDA Group
- Red Hat
- Samsung
- Seagate
- SK Hynix
- Spin Transfer Technologies
- TE Connectivity Corporation
- VMWare
- Western Digital Technologies, Inc. (Sandisk)
- Xilinx

13th ANNUAL WORKSHOP 2017

THANK YOU

Greg Casey, Senior Architect and Strategist Server CTO Team

DellEMC