

# Power10 Memory Inception

**Building Upon OMI to Construct Tightly  
Coupled, Shared Memory Clusters**

FMS 2022

August 2-4

*SARC-102-2: Open Memory Interface (OMI)*

Baba Arimilli

Bill Starke



# Origins: Open Memory Interface (OMI) and Memory Inception (MI)

➤ CXL is now an industry standard and supports:

- Memory home agent (.mem)
- Caching capability for cache-coherent compute accelerators (.cache)
- I/O (.io)



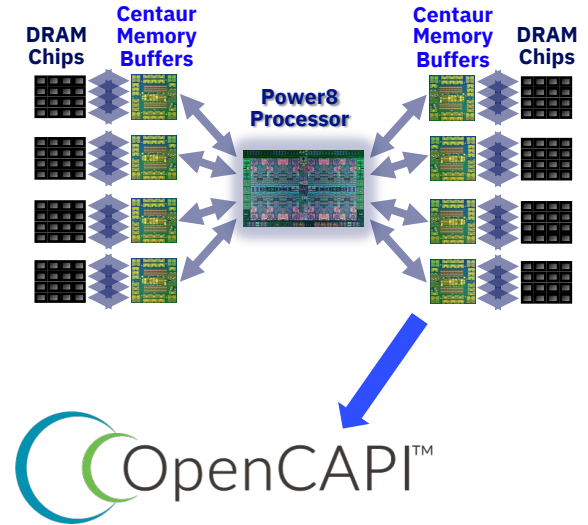
# Origins: Open Memory Interface (OMI) and Memory Inception (MI)

- **CXL is now an industry standard and supports:**
  - **Memory home agent (.mem)**
  - **Caching capability for cache-coherent compute accelerators (.cache)**
  - **I/O (.io)**
- **OpenCAPI preceded (2016) CXL and provided the similar capabilities**



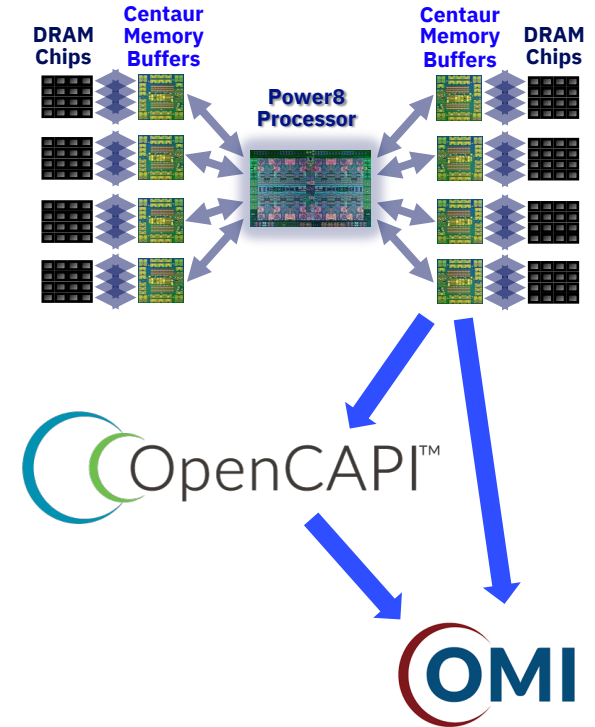
# Origins: Open Memory Interface (OMI) and Memory Inception (MI)

- CXL is now an industry standard and supports:
  - Memory home agent (.mem)
  - Caching capability for cache-coherent compute accelerators (.cache)
  - I/O (.io)
- OpenCAPI preceded (2016) CXL and provided similar capabilities
- Centaur Memory Buffer in IBM's Power8 (2014) was a precursor to the OpenCAPI memory home agent function
  - Technology agnostic i/f to processor chip
    - DDR scheduling intelligence, etc. moved to Memory Buffer chip
  - Latency optimizations over multiple generations (e.g., Explorer in IBM's Power10)



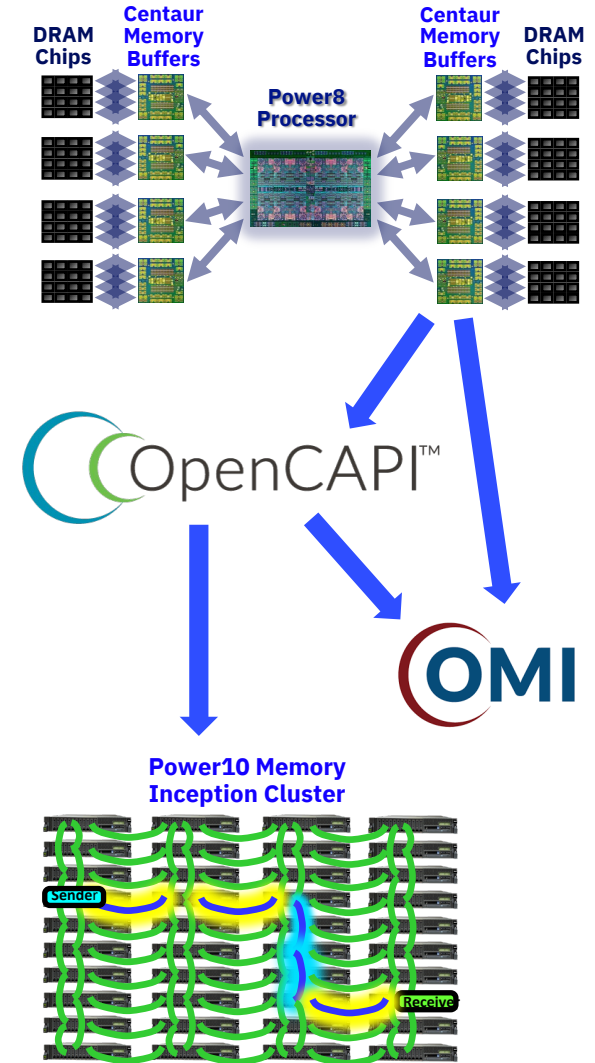
# Origins: Open Memory Interface (OMI) and Memory Inception (MI)

- CXL is now an industry standard and supports:
  - Memory home agent (.mem)
  - Caching capability for cache-coherent compute accelerators (.cache)
  - I/O (.io)
- OpenCAPI preceded (2016) CXL and provided similar capabilities
- Centaur Memory Buffer in IBM's Power8 (2014) was a precursor to the OpenCAPI memory home agent function
  - Technology agnostic i/f to processor chip
    - DDR scheduling intelligence, etc. moved to Memory Buffer chip
  - Latency optimizations over multiple generations (e.g., Explorer in IBM's Power10)
- OMI is a an optimized OpenCAPI memory home agent subset targeting main tier memory
  - Technology agnostic i/f
  - Extreme low-latency optimizations leveraging Centaur's experience enables main tier DDR memory in addition to other tiers of memory that CXL.mem and OpenCAPI support

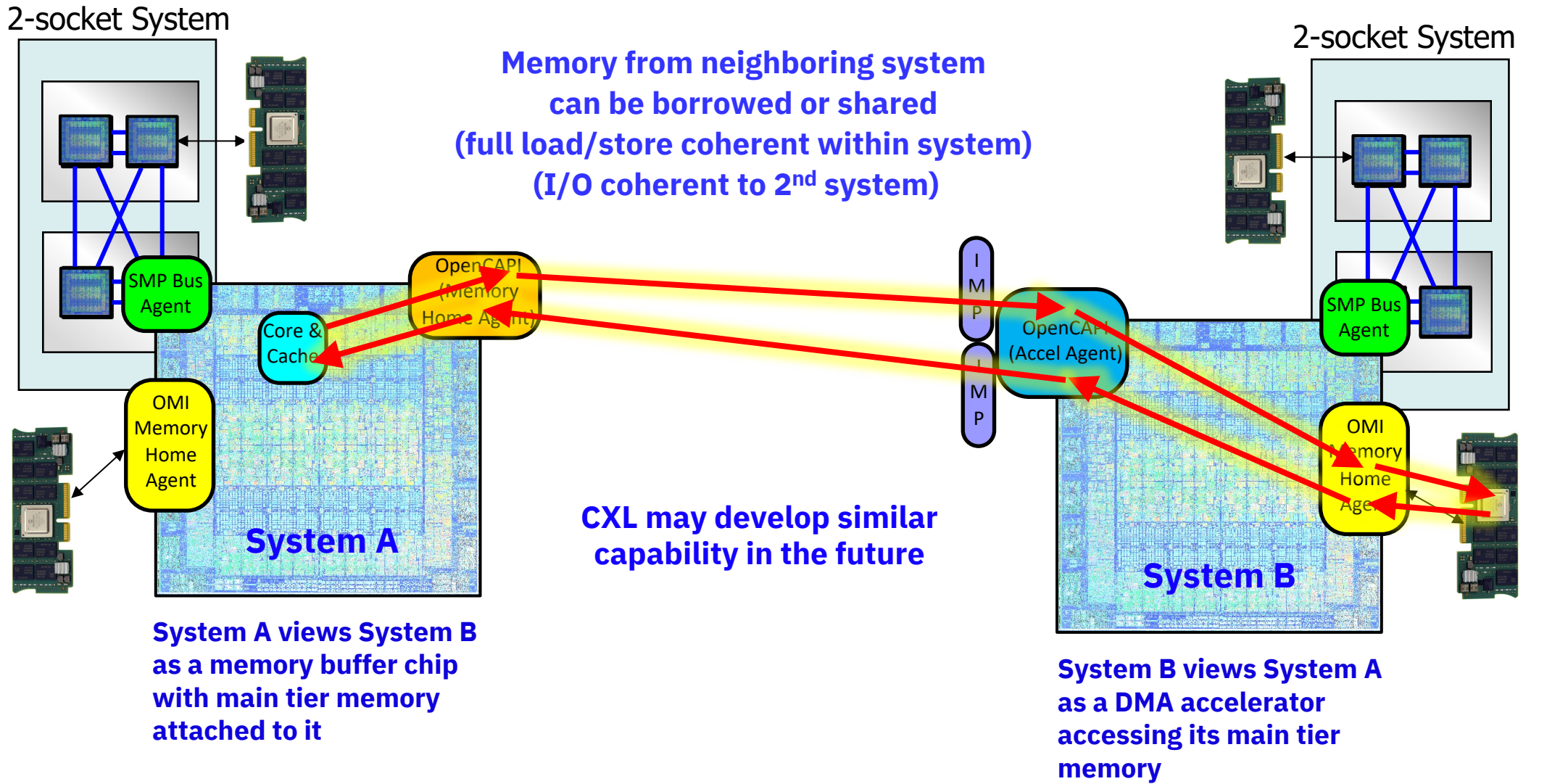


# Origins: Open Memory Interface (OMI) and Memory Inception (MI)

- CXL is now an industry standard and supports:
  - Memory home agent (.mem)
  - Caching capability for cache-coherent compute accelerators (.cache)
  - I/O (.io)
- OpenCAPI preceded (2016) CXL and provided similar capabilities
- Centaur Memory Buffer in IBM's Power8 (2014) was a precursor to the OpenCAPI memory home agent function
  - Technology agnostic i/f to processor chip
    - DDR scheduling intelligence, etc. moved to Memory Buffer chip
  - Latency optimizations over multiple generations (e.g., Explorer in IBM's Power10)
- OMI is an optimized OpenCAPI memory home agent subset targeting main tier memory
  - Technology agnostic i/f
  - Extreme low-latency optimizations leveraging Centaur's experience enables main tier DDR memory in addition to other tiers of memory that CXL.mem and OpenCAPI support
- Power10 Memory Inception (MI) built on top of OpenCAPI and includes:
  - OpenCAPI memory home agent
  - DMA accelerator
- OMI and MI are both derived from OpenCAPI – Can connect and interact together



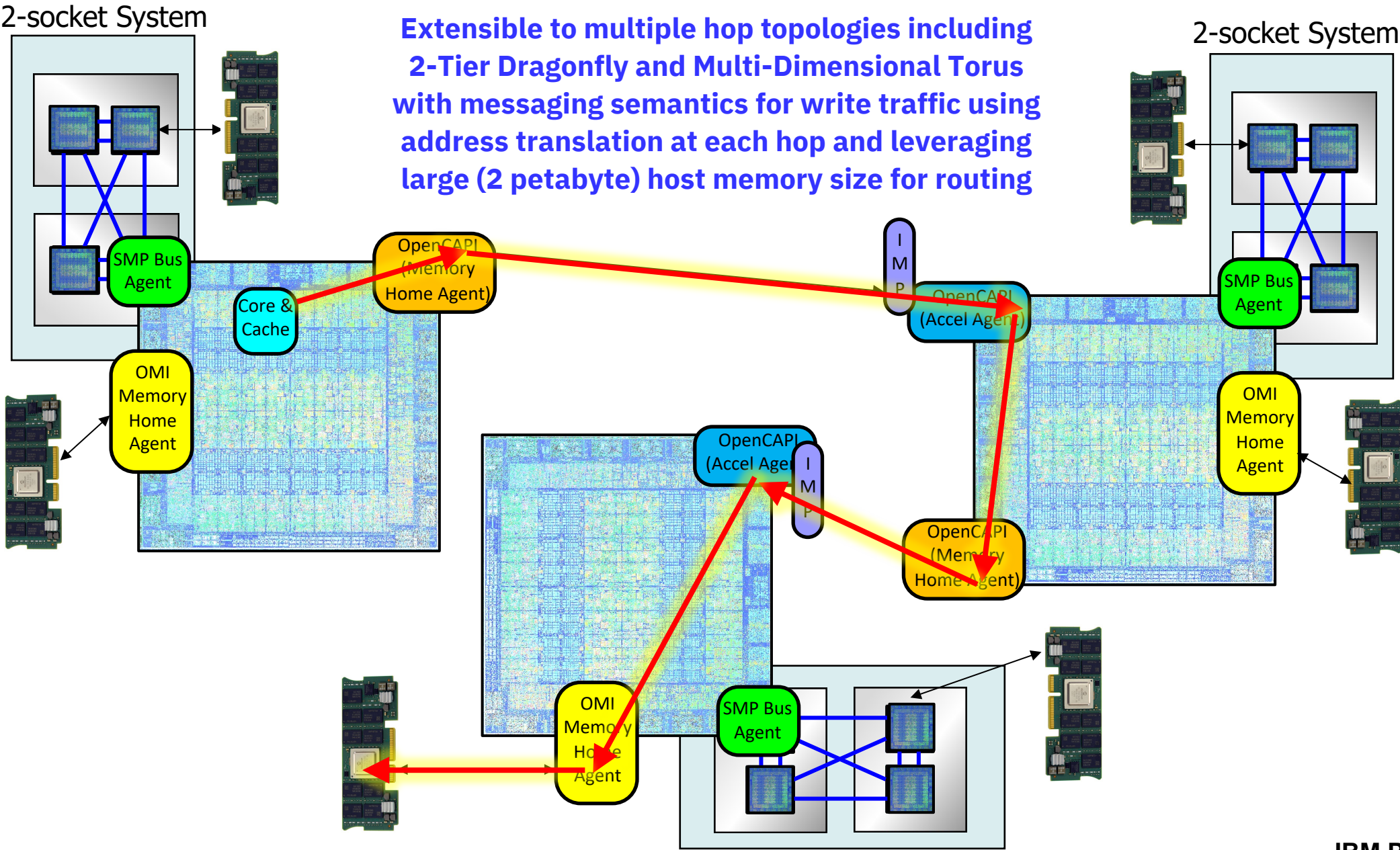
# Overview: Power10 Memory Inception



Enabling full-duplex allows all computers that are directly connected to each other via MI to dynamically access each other's memory



# Memory Clustering: Built on Memory Inception Architecture





# Memory Inception: Applications

## ➤ Infrastructure Level Memory Borrowing

- Use case: One system can borrow memory from another system in the same pod
  - Memory configured as low latency local memory
  - Or configured as NUMA latency remote memory for borrower



# Memory Inception: Applications

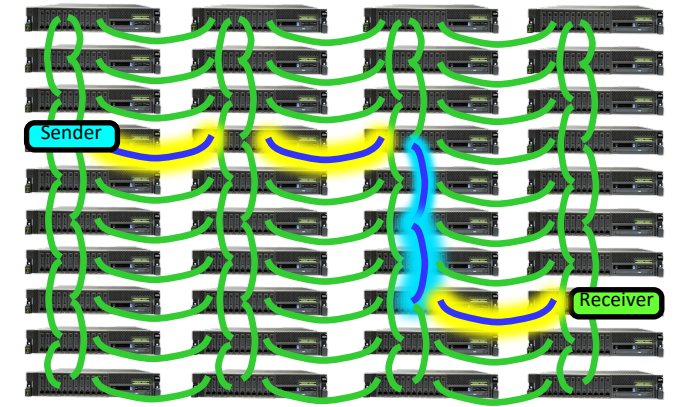
## ➤ Infrastructure Level Memory Borrowing

- Use case: One system can borrow memory from another system in the same pod
  - Memory configured as low latency local memory
  - Or configured as NUMA latency remote memory for borrower



## ➤ Cluster Communication Fabric

- Low latency messaging scaling to 1000's of nodes
- Support for torus and dragonfly topology multi-hop routing
- Optimized for short messages



# Memory Inception: Applications

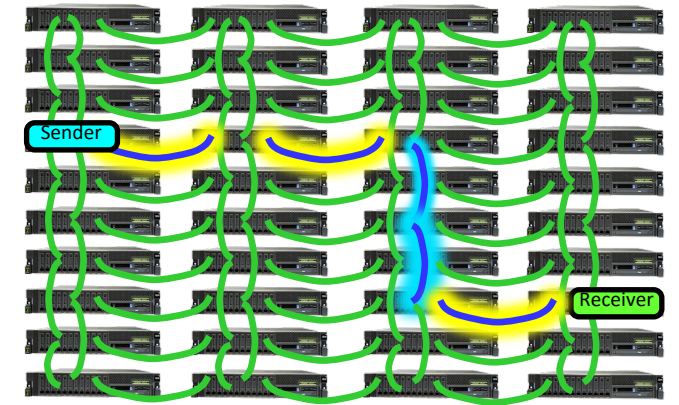
## ➤ Infrastructure Level Memory Borrowing

- Use case: One system can borrow memory from another system in the same pod
  - Memory configured as low latency local memory
  - Or configured as NUMA latency remote memory for borrower



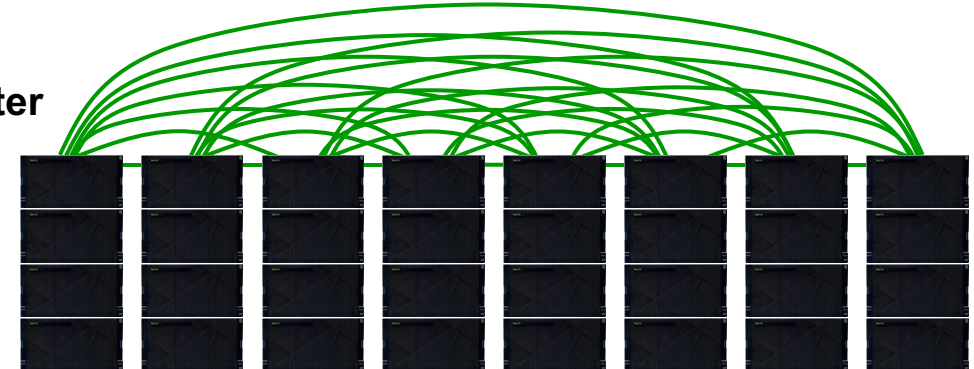
## ➤ Cluster Communication Fabric

- Low latency messaging scaling to 1000's of nodes
- Support for torus and dragonfly topology multi-hop routing
- Optimized for short messages



## ➤ Massive Shared (software coherent) Cluster Memory

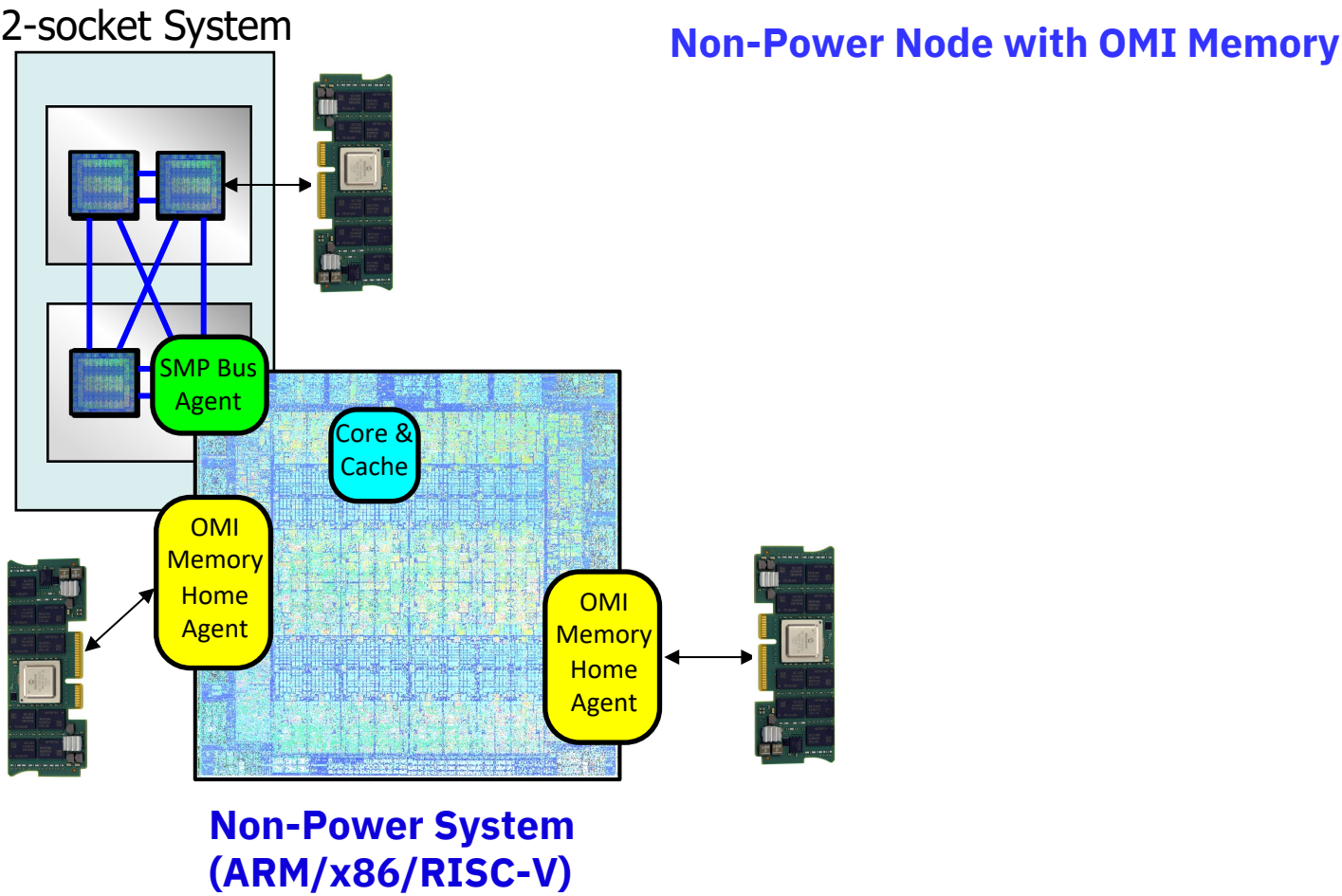
- Up to 2 Petabytes directly load/store accessible to any process in cluster
- Up to over 60,000 HW compute threads (with 32 racks)
- Enables new types of applications



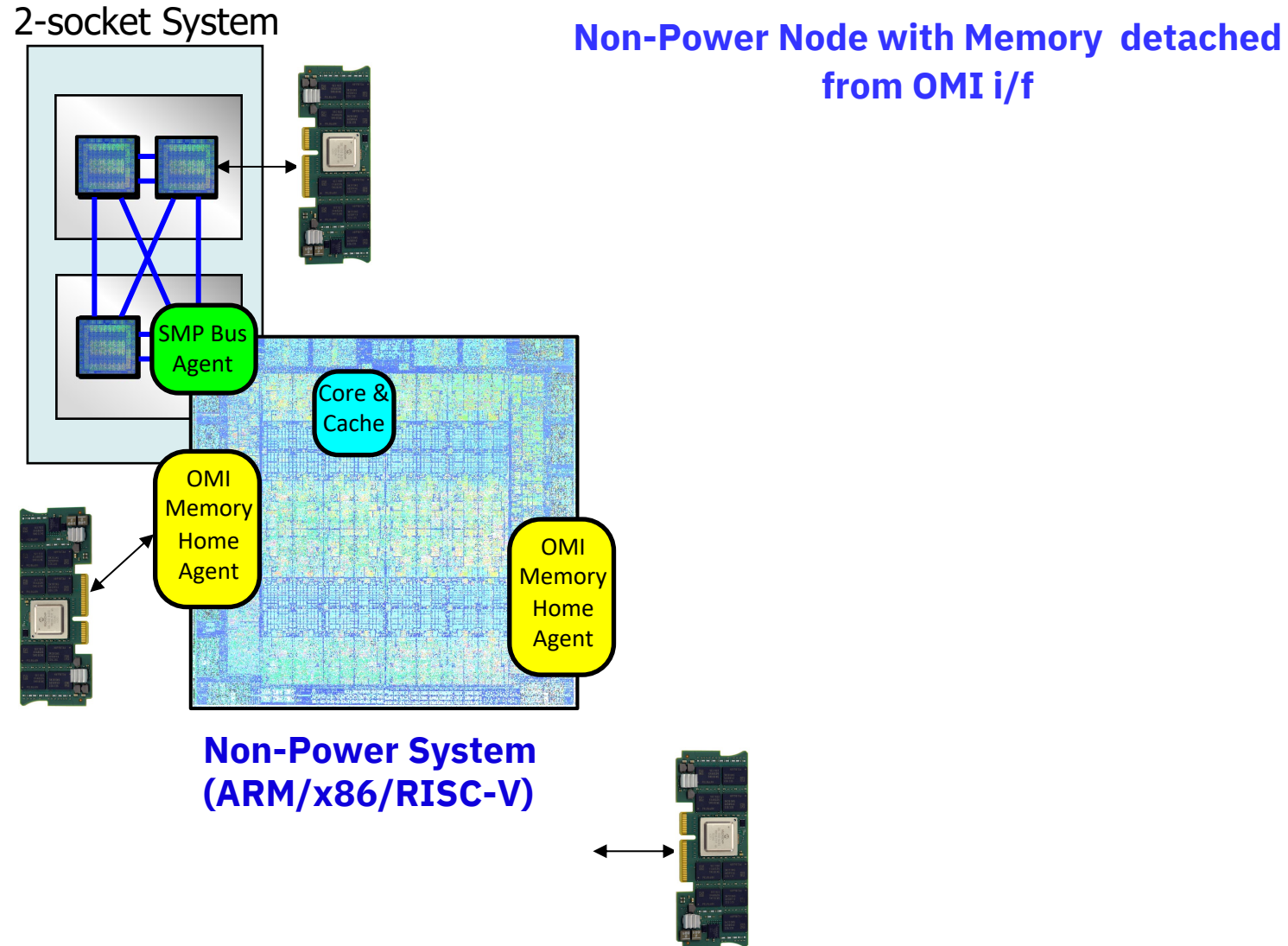
(Memory cluster configurations show processor capability only, and do not imply system product offerings)

IBM Power10 OMI-MI

# Hybrid Memory Clustering: **Non-Power** and **Power Systems**

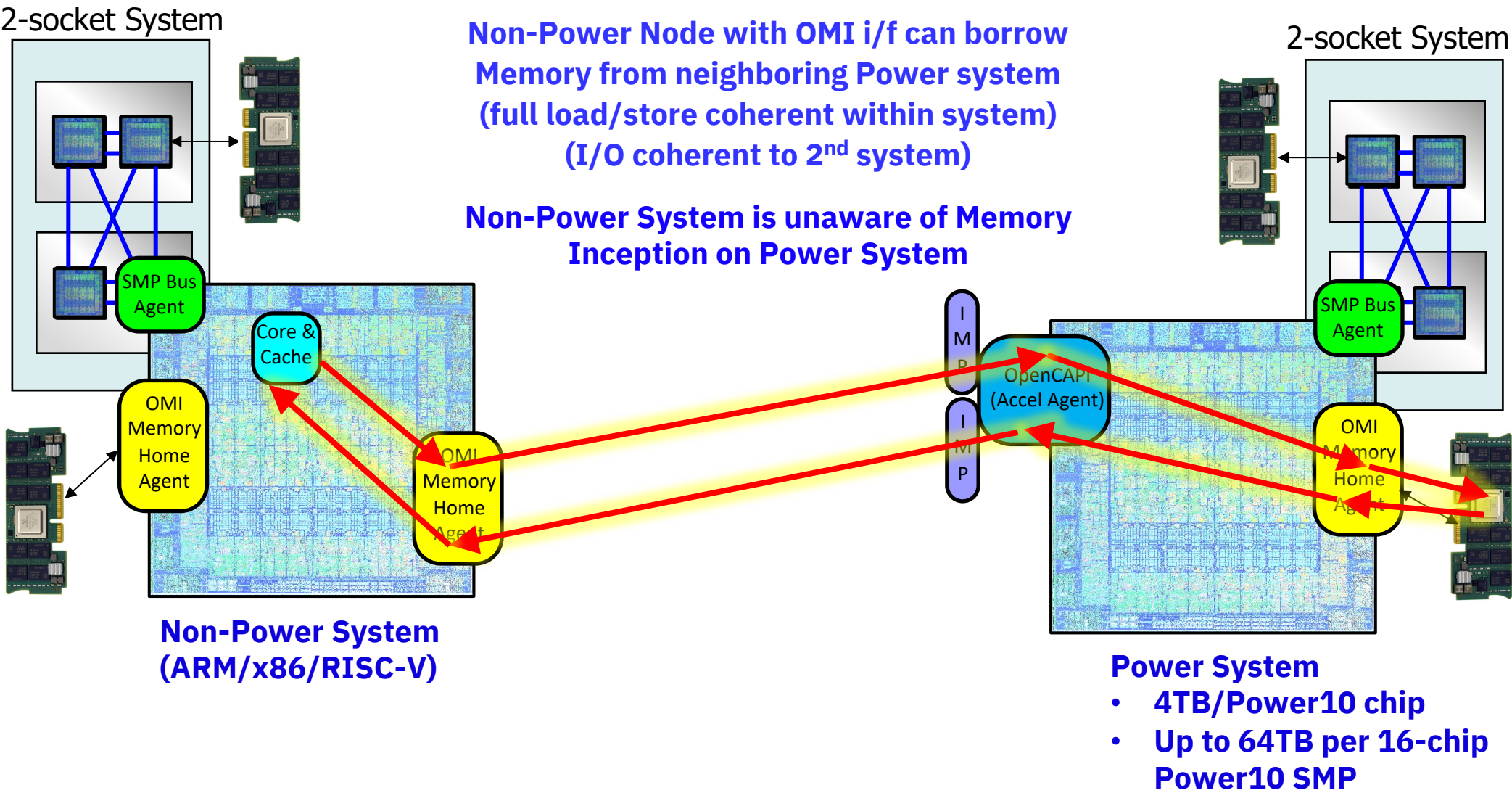


# Hybrid Memory Clustering: **Non-Power** and **Power Systems**





# Hybrid Memory Clustering: Non-Power and Power Systems





When you interact with IBM, this serves as your authorization to Flash Memory Summit or its vendor to provide your contact information to IBM in order for IBM to follow up on your interaction.

IBM's use of your contact information is governed by the IBM Privacy Policy.

