Hewlett Packard Enterprise

HPE Cray EX Architecture

Comprehensive General LUMI Course

April 23–26, 2024

Agenda

Architecture components

- Packaging
- Board-level
- Processor
- Network
- Cooling
- Lustre Storage

Recipe for a Supercomputer

- Select best microprocessor
  - Function of time
- Surround it with a bandwidth-rich environment
  - Local memory
  - Interconnection network
- Scale the system
  - Provide scalable programming and performance tools
  - Provide scalable I/O
  - Eliminate operating system interference (OS jitter)
  - Design in reliability and resiliency
  - Provide scalable system management
  - System service life

"Anyone can build a fast CPU.
The trick is to build a fast system."
— *Seymour Cray*

Nodes: Basic System Building Block

The HPE Cray EX is a massively parallel processor (MPP) supercomputer design, it is comprised of thousands of nodes.

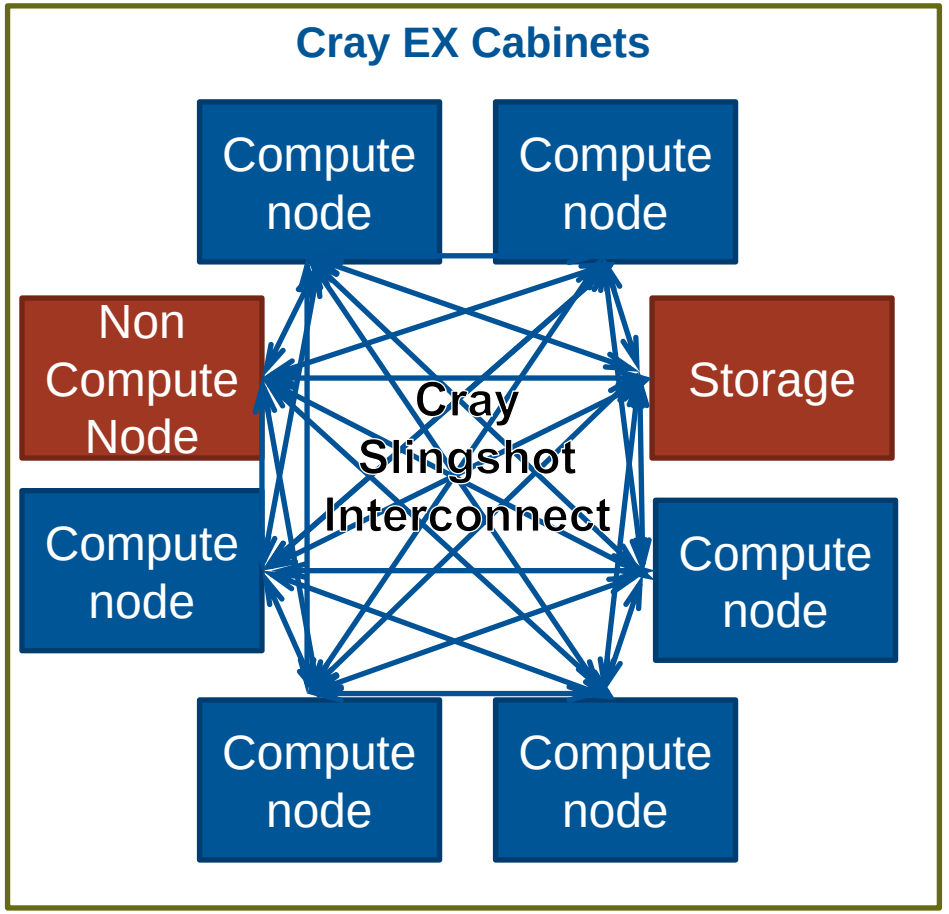
There are two basic node types:

- Compute Nodes
  - These are only used for user computation
- Non-compute Nodes
  - These perform other tasks needed for the system to function and are outside of the 'mainframe' racks that house the compute nodes. Specific roles of non-compute nodes include:
    - User access nodes (UANs)
    - Data Visualization (DVS) nodes
    - Nodes supporting various services such as WLM daemons, name services, utility file system provision etc.

Connecting Nodes Together: Slingshot

- To function as a single efficient supercomputer, individual nodes must be efficiently networked

- Nodes are interconnected by the high-speed low-latency Slingshot network



Differentiating Nodes

User Access Nodes

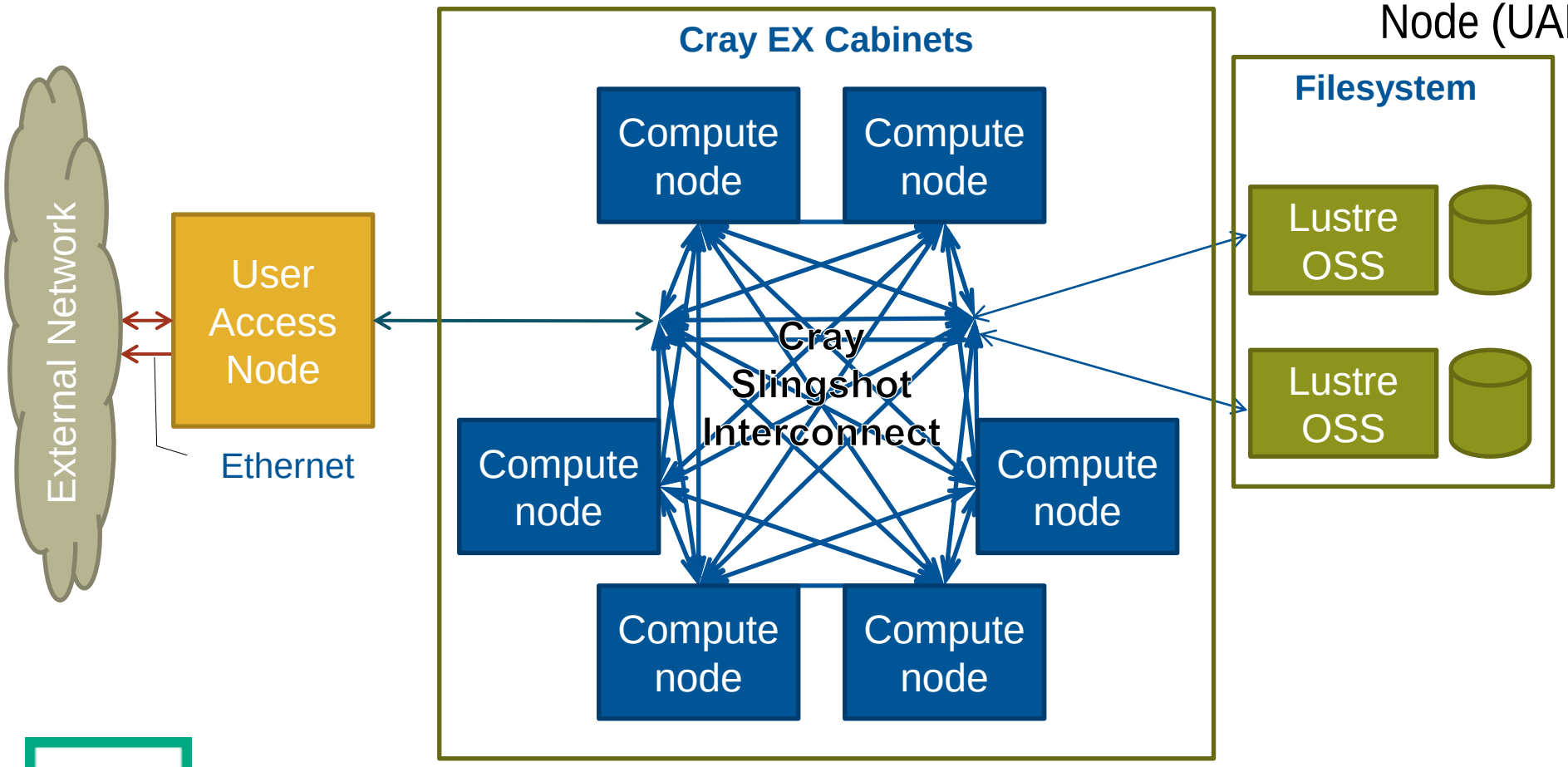
- These provide command-line access to the system
- Run full-featured version of the SLES operating system
- Used for editing files, compiling, submitting jobs and other interactive tasks
- Shared by many users concurrently

Compute Nodes

- Nodes where production jobs are allocated
- Runs the HPE Cray Operating System (COS) which is optimized for use on a compute node
- Only accessible via reservation within a resource management system like SLURM. For example a batch job.
- Most likely setup to only allow access by one user at a time (exclusive access)
  - Some systems such as LUMI allow shared access

Interacting with the System

Users interact with the system by logging into a User Access Service (UAS) or User Access Node (UAN).



System Management Services

- Many system services run on non-compute nodes
- The system management of the Cray EX supercomputer uses specialized containers run by Kubernetes as opposed to having dedicated servers.

- Containerised functions/services include:
  - Configuration Framework services
  - Boot support
  - Image management
  - Network management services
  - Slurm daemons
  - Content Projection Service (projection of filesystems to nodes, including PE)
  - System monitoring
  - Directory services (keycloak)

LUMI System

Image © CSC, Finland

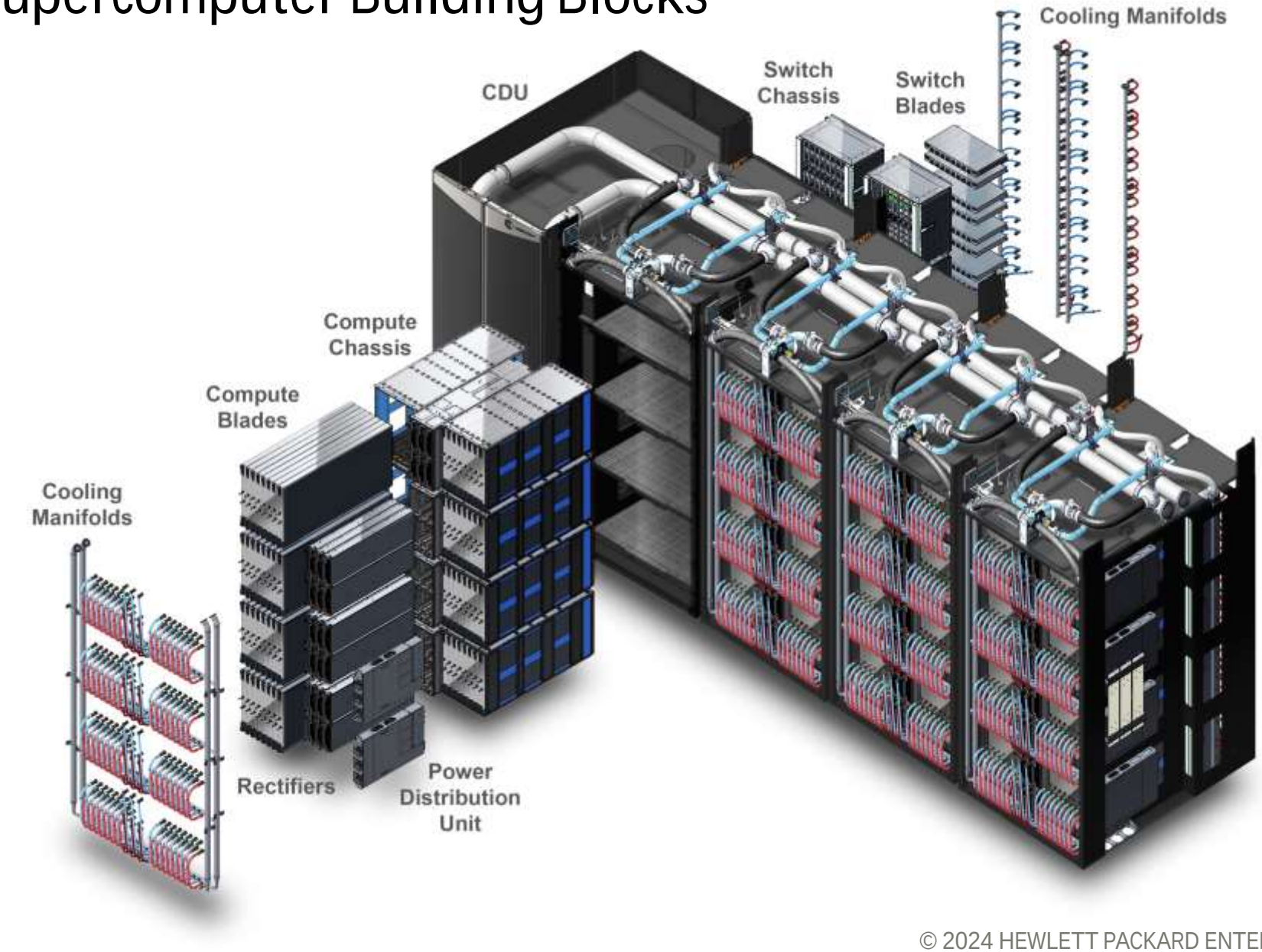
LUMI-C

- 2048 nodes
  - 1888 256GB nodes
  - 128 512GB nodes
  - 32 1TB nodes
- 2 × AMD EPYC 7763 2.45GHz base (3.5GHz boost), 64c processor
- 128 (2x64) cores per node
- 8 NUMA regions per node
- HPE Slingshot interconnect

LUMI-G

- 2978 nodes with 1 AMD CPU and 4 AMD MI-250X GPU

Cray EX Supercomputer Building Blocks



Cray EX System Building Blocks (LUMI-C)

Compute Blade

4 Compute Nodes

Chassis

8 Compute Blades

32 Compute Nodes

Cabinet

1x Rank 0 Network Group

Passive Electrical Network

8 Chassis

256 Compute Nodes

System

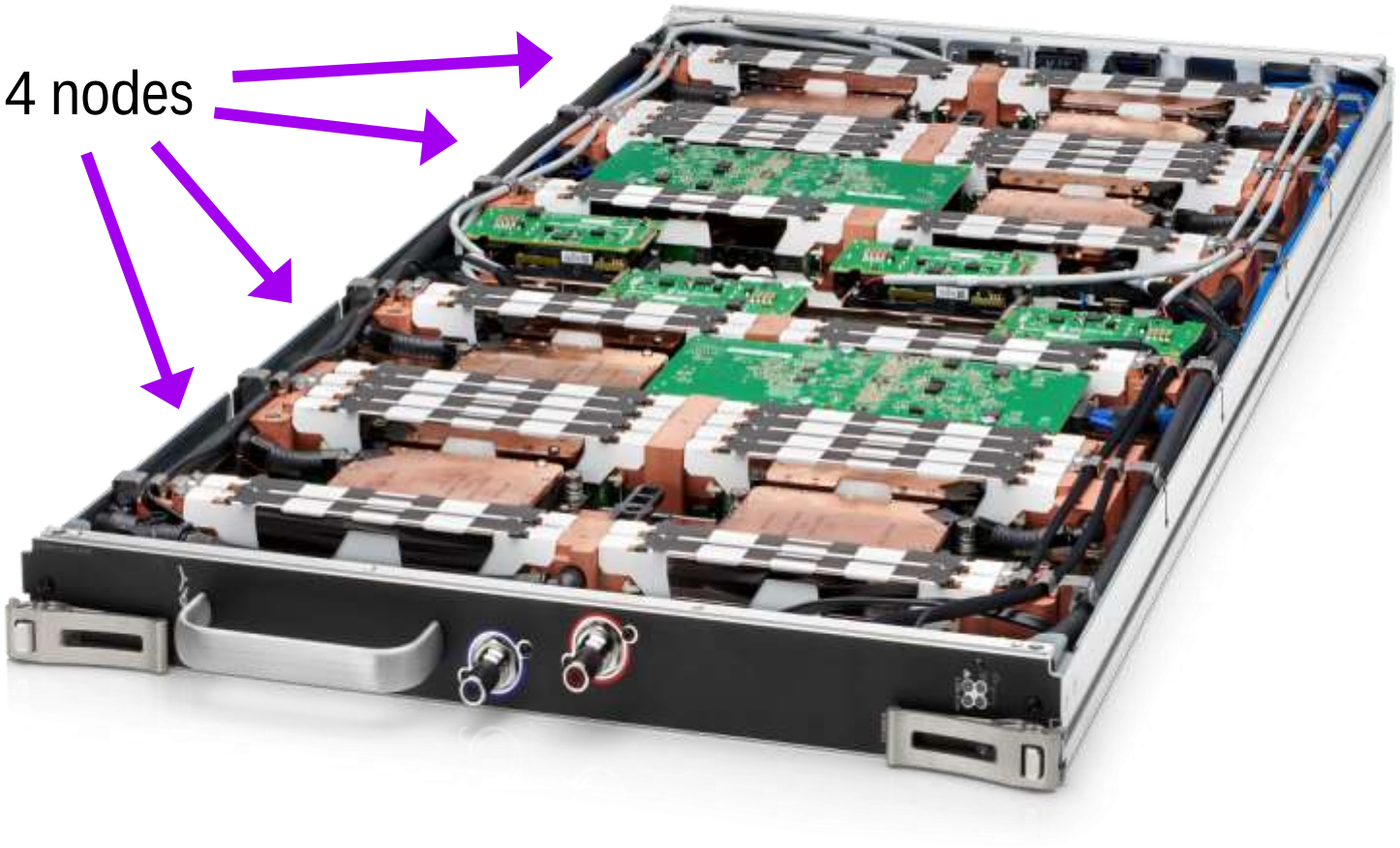
Rank 1 Network

Active Optical Network

Hundreds of Cabinets

Up to 10s of thousands of nodes

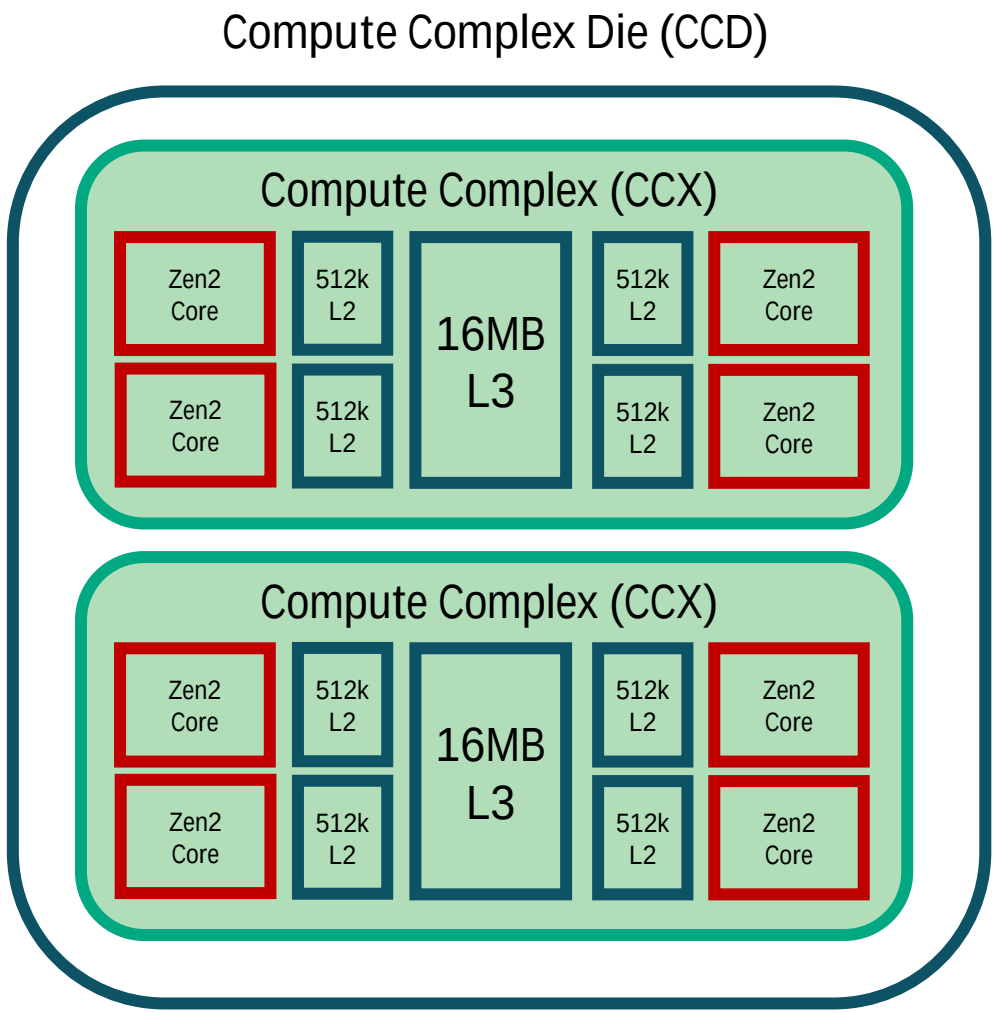
Compute Blade Architecture (LUMI-C)



Each node:

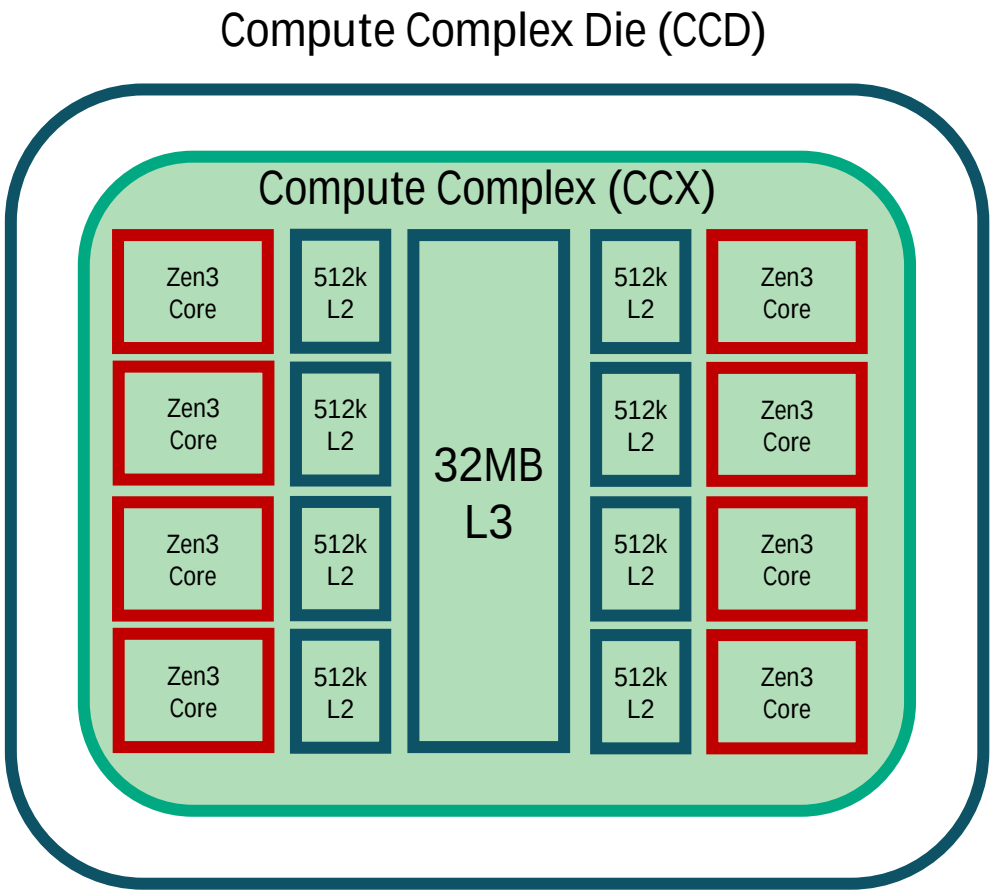
- 2 x AMD EPYC 7763 64 core 2.45GHz
- 16 x 16/32/64 GB DDR4 3200
- 256/512/1024 GB per node 2-8GB per core
- 1 x 200Gb/s injection ports per node

AMD EPYC Zen2 Rome Architecture



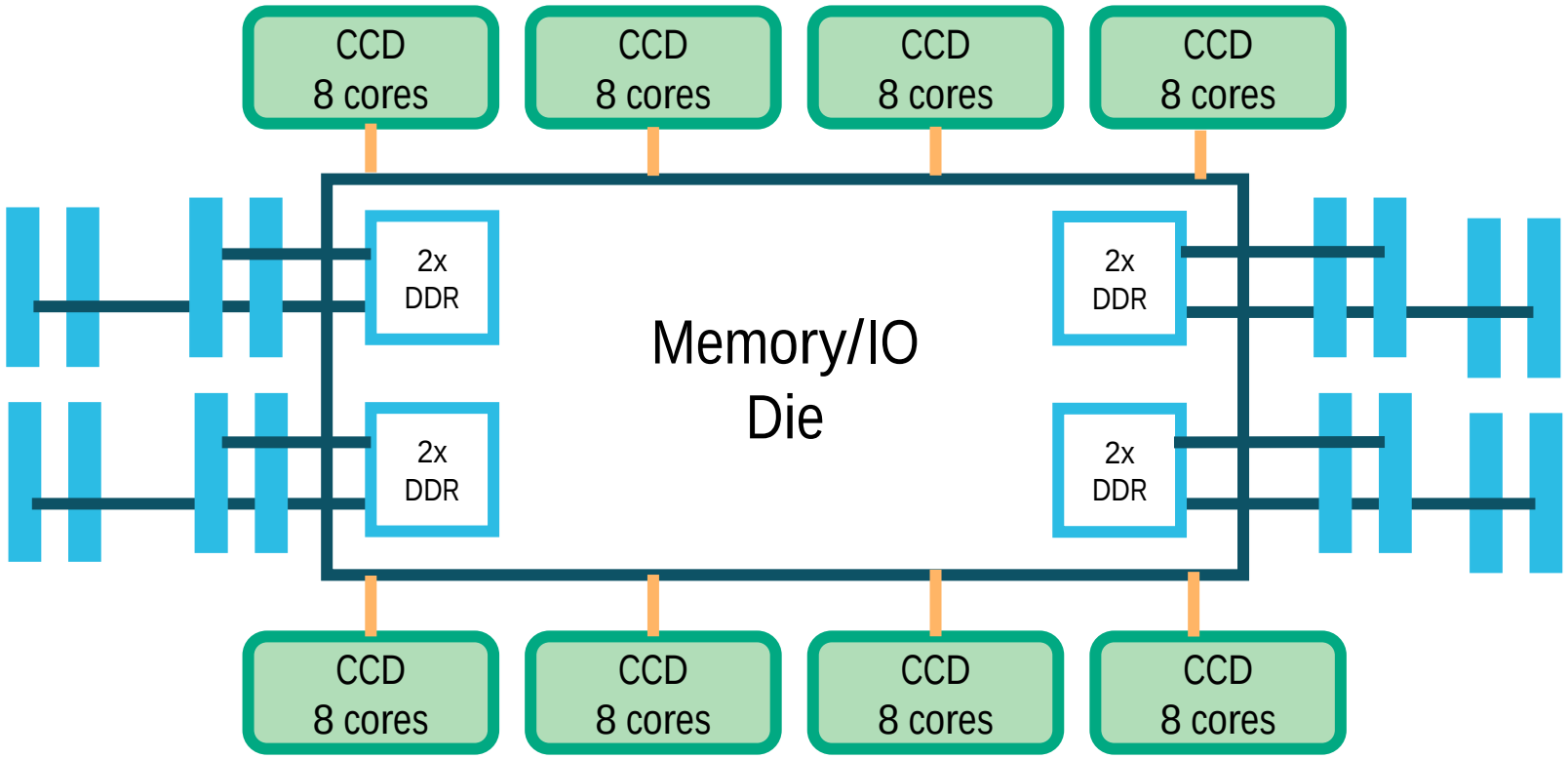
- UANs / login nodes
- Chiplet SOC design
- Infinity Fabric
- 7nm process
- Compute Complex Dies (CCDs) host cores and L2/L3 cache

AMD EPYC Zen3 Milan/Trento Architecture



- Chiplet SOC design
- Infinity Fabric
- 7nm process
- Compute Complex Dies (CCDs) host cores and L2/L3 cache

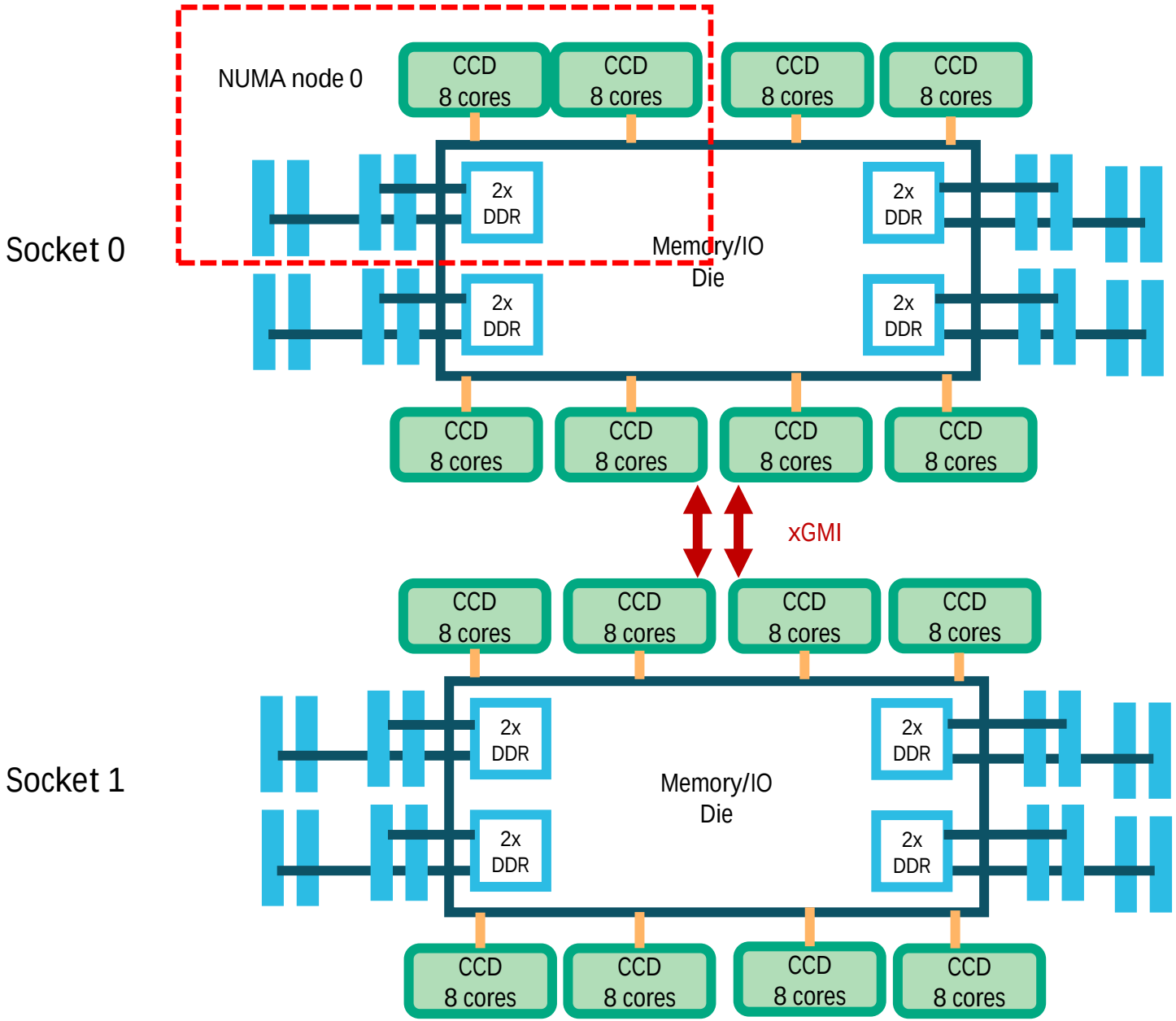
AMD EPYC Processor (Milan / LUMI-C)



AMD EPYC 7763

- Base clock 2.45GHz
- Boost clock 3.5 GHz
- 280W TDP
- 64 cores,
  128 SMT threads
- L1 cache 32kB / core
- L2 cache 512kB / core
- L3 cache 32MB / 8-cores
  256MB L3 cache in total
- 128 PCIe 4.0 lanes
- 8 channel DDR4 3200MHz,
  204.8 GB/s peak b/w
- Configured as 4
  NUMA nodes
- Vector support: AVX2

2-Socket Node



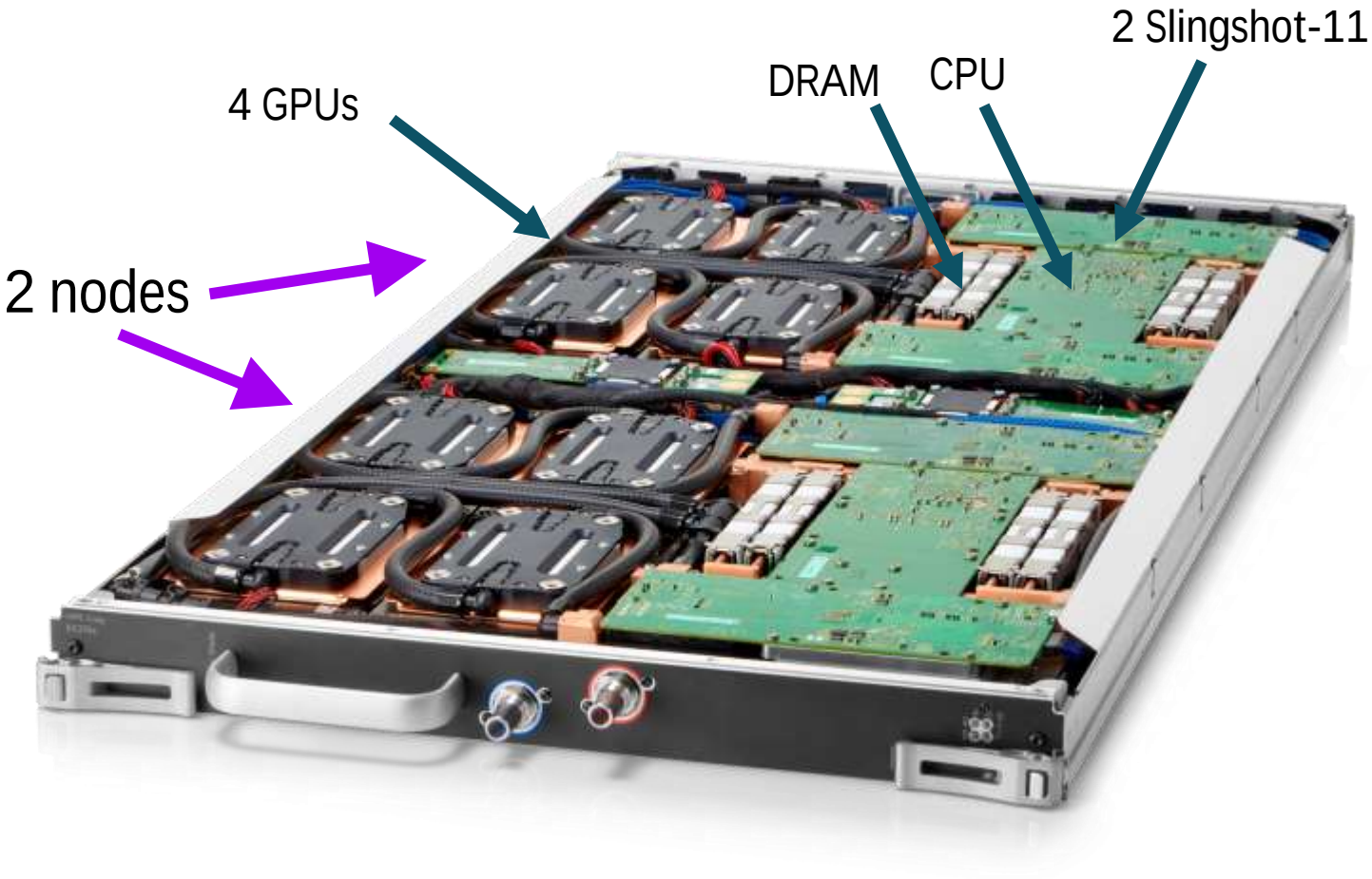
LUMI-C node NUMA Organization: 64*2 + 64*2 Linux CPUs

```
available: 8 nodes (0-7)
node 0 cpus: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143
node 0 size: 31083 MB
node 0 free: 28922 MB
node 1 cpus: 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159
node 1 size: 32249 MB
node 1 free: 26733 MB
node 2 cpus: 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175
node 2 size: 32249 MB
node 2 free: 30504 MB
node 3 cpus: 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191
node 3 size: 32237 MB
node 3 free: 31537 MB
node 4 cpus: 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207
node 4 size: 32249 MB
node 4 free: 31474 MB
node 5 cpus: 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223
node 5 size: 32249 MB
node 5 free: 31460 MB
node 6 cpus: 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238
239
node 6 size: 32249 MB
node 6 free: 30521 MB
node 7 cpus: 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 240 241 242 243 244 245 246 247 248 249 250 251 252 253
254 255
node 7 size: 32247 MB
node 7 free: 30961 MB
```

LUMI-C node NUMA Distances

```
node distances:
node   0   1   2   3   4   5   6   7
  0:  10  12  12  12  32  32  32  32
  1:  12  10  12  12  32  32  32  32
  2:  12  12  10  12  32  32  32  32
  3:  12  12  12  10  32  32  32  32
  4:  32  32  32  32  10  12  12  12
  5:  32  32  32  32  12  10  12  12
  6:  32  32  32  32  12  12  10  12
  7:  32  32  32  32  12  12  12  10
```

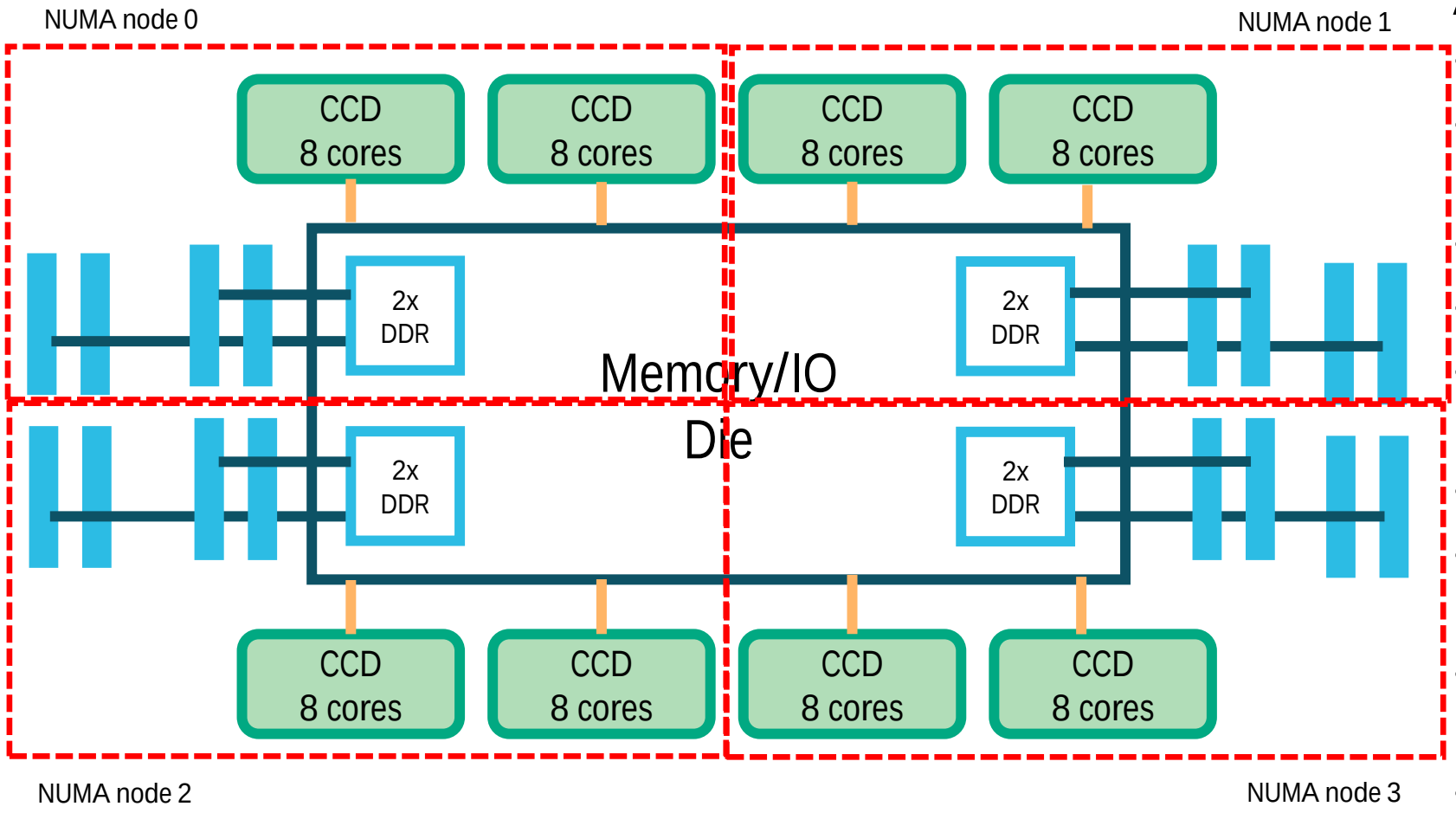
Compute Blade Architecture (LUMI-G)



Each node:

- AMD EPYC 7A53 “Optimized 3rd Gen EPYC” 64-Core Processor, 2.00 GHz
- 512 GB DDR4 memory
- 4x AMD MI-250X GPU
- Each GPU connected to a Slingshot 200Gb/s NIC

AMD EPYC processor (LUMI-G)

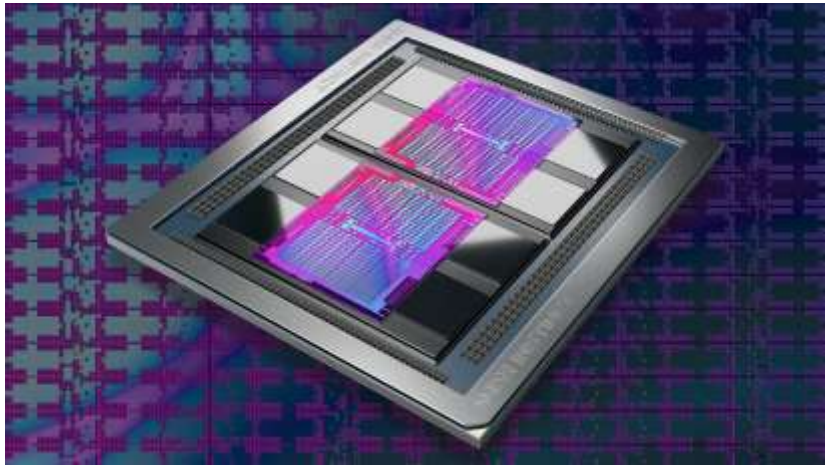


AMD EPYC 7A53

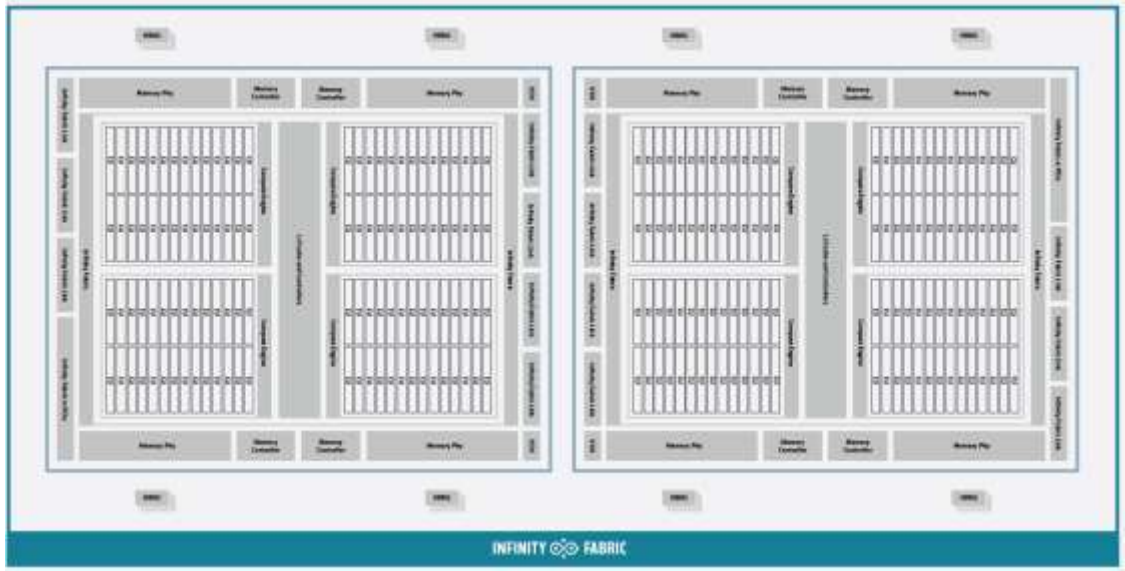
- Base clock 2.00 GHz
- 64 cores,
  128 hardware threads
- L1 cache 32kB / core
- L2 cache 512kB / core
- L3 cache 32MB / 8-cores
  256MB L3 cache in total
- 128 PCIe 4.0 lanes
- 8 channel DDR 3200MHz,
  204.8 GB/s peak b/w
- Configured as 4
  NUMA nodes
- Vector support: AVX2

Mi250X GPU Architecture

- Two compute dies (Graphics Compute Dies (GCDs))
  - Interconnected with 200 GB/s per direction
- Dedicated Memory (HBM2e) Size: 128 GB
  - High bandwidth device memory (up to 3.2 TB/s)
  - Memory Clock: 1.6 GHz
- AMD CDNA2 Architecture
  - 110 Compute Units (CU) per each die = 220 CUs
  - 64 SIMD threads per each CU = 14080 Stream Processors
  - Peak FP64/FP32 Vector: 47.9 TFLOPS
  - Peak FP64/FP32 Matrix: 95.7 TFLOPS
  - Total L2 cache: 8 MB per each die (64kB per CU)
  - Frequency: up to 1700 MHz
  - Max power: up to 560 Watts
- Memory coherency with CPU



Source: https://www.amd.com/en/products/server-accelerators/instinct-mi250

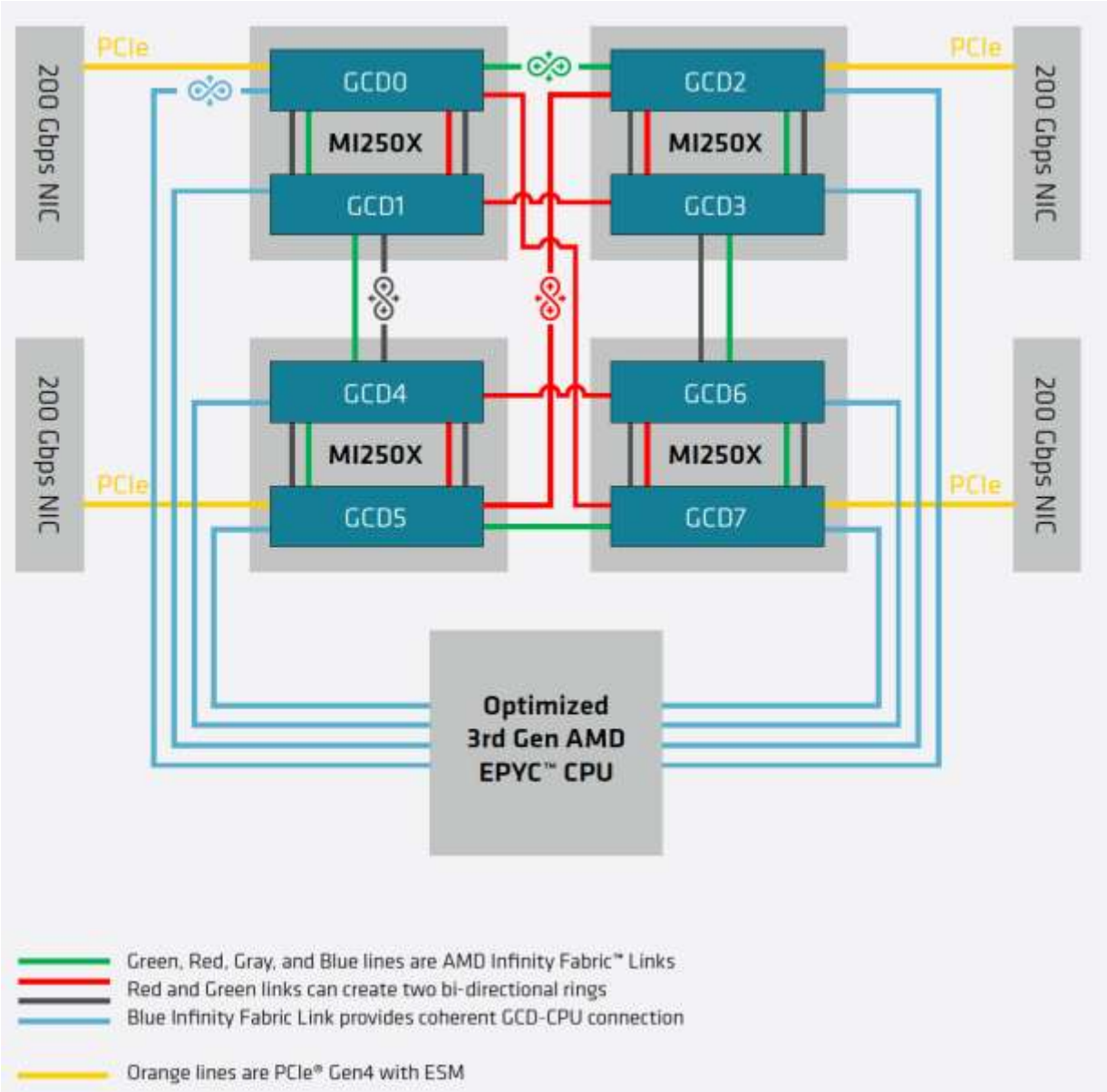


Source: (https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf)

Node Architecture (LUMI-G)

- The programmer can think of the 8 GCDs as 8 separate GPUs, each having 64 GB of high-bandwidth memory (HBM2E)
- The CPU is connected to each GCD via Infinity Fabric CPU-GPU, allowing a peak host-to-device (H2D) and device-to-host (D2H) bandwidth of 36+36 GB/s
  - Coherent memory CPU-GPU
- The 2 GCDs on the same MI250X are connected with Infinity Fabric GPU-GPU
- The GCDs on different MI250X are connected with Infinity Fabric GPU-GPU in the arrangement shown in the diagram on the right, where the peak bandwidth ranges from 50-100 GB/s based on the number of Infinity Fabric connections between individual GCDs

Source: https://www.amd.com/system/files/documents/amd-cdna2-white-paper.pdf

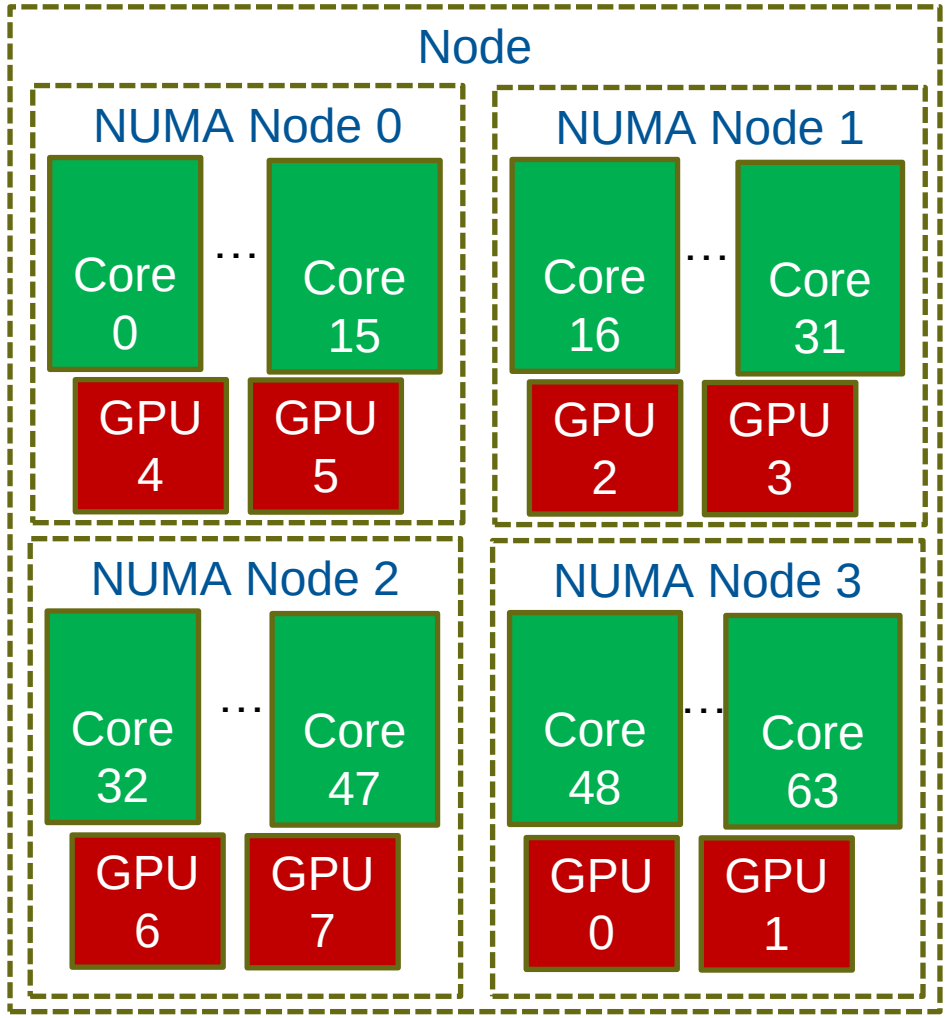


GPUS TO NUMA DOMAINS MAPPING

- GPUs are associated to NUMA nodes
- Can use rocm-smi to see the topology

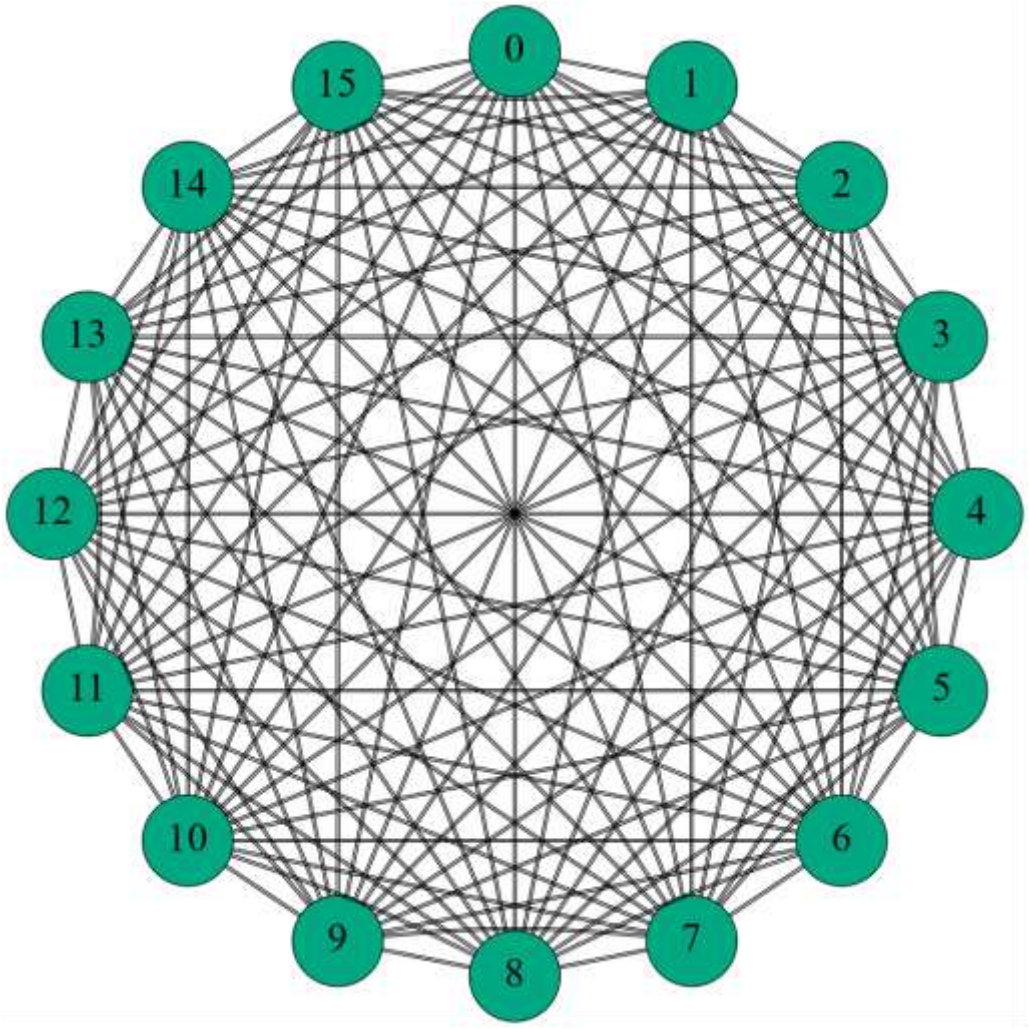
```
> srun --nodes=1 -p <partition> -A <your_project> -t "00:02:00" --ntasks=1 --gres=gpu:8 rocm-smi --showtopo
…
========================================= Numa Nodes =========================================
GPU[0] : (Topology) Numa Node: 3
GPU[0] : (Topology) Numa Affinity: 3
GPU[1] : (Topology) Numa Node: 3
GPU[1] : (Topology) Numa Affinity: 3
GPU[2] : (Topology) Numa Node: 1
GPU[2] : (Topology) Numa Affinity: 1
GPU[3] : (Topology) Numa Node: 1
GPU[3] : (Topology) Numa Affinity: 1
GPU[4] : (Topology) Numa Node: 0
GPU[4] : (Topology) Numa Affinity: 0
GPU[5] : (Topology) Numa Node: 0
GPU[5] : (Topology) Numa Affinity: 0
GPU[6] : (Topology) Numa Node: 2
GPU[6] : (Topology) Numa Affinity: 2
GPU[7] : (Topology) Numa Node: 2
GPU[7] : (Topology) Numa Affinity: 2
================================== End of ROCm SMI Log ==================================
```

GPUS TO NUMA DOMAINS MAPPING



NUMA ID	0	0	1	1	2	2	3	3
Optimal GPU ID	4	5	2	3	6	7	0	1

High Speed Network: Slingshot



Cray EX uses the Slingshot network

- Dragonfly topology
- 64-port switches
- Compatible with Ethernet (Supports RoCE)
- 200Gbps signaling
- Nodes connected with 1 x 200Gbps NICs (LUMI-C), 4 x 200Gbps NICs (LUMI-G)
- Quality of Service features to target latency at full load
  - Adaptive Routing
  - Congestion Control

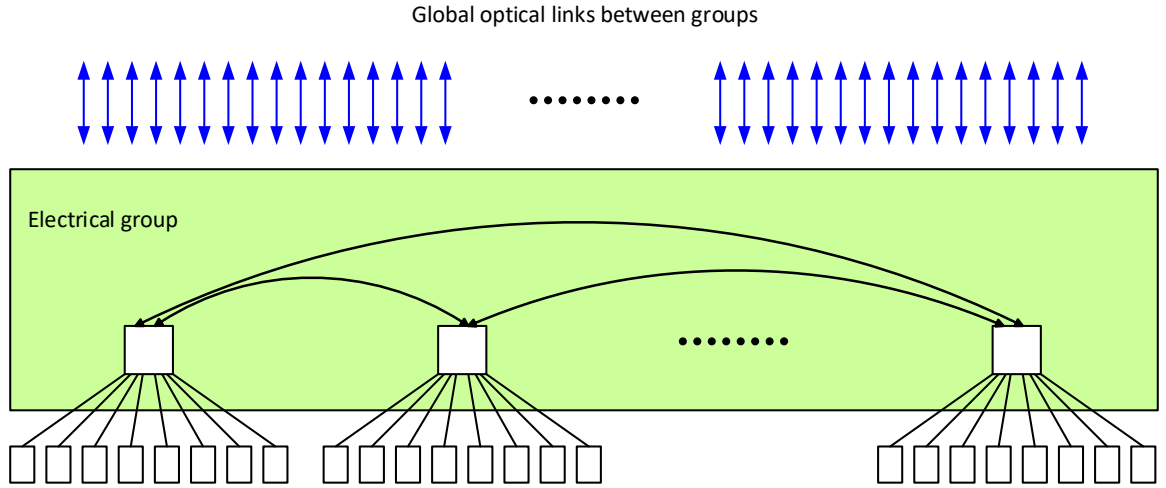
What is a Dragonfly Network?

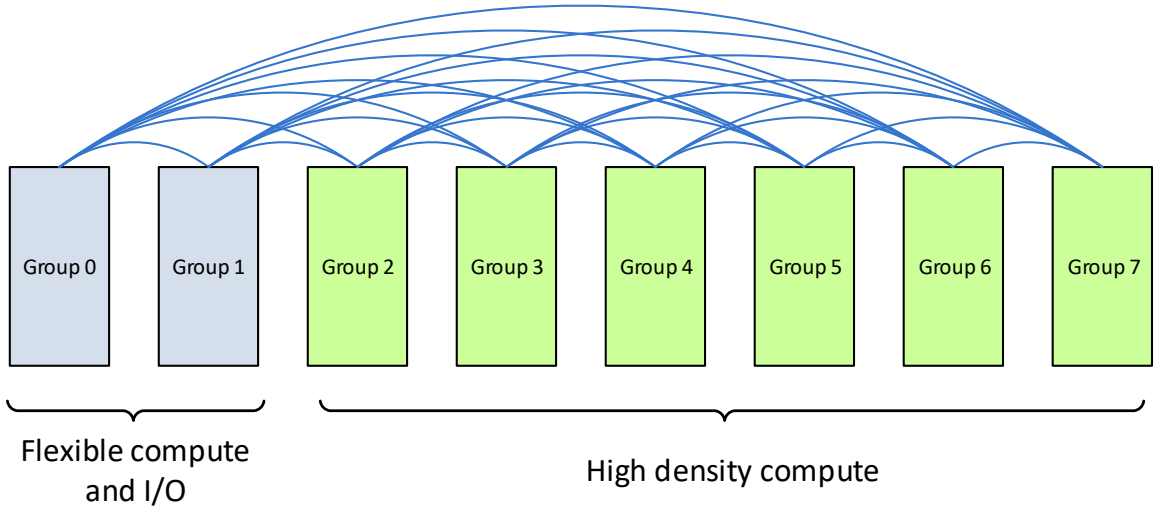
- Nodes are organized into groups
  - Usually racks or cabinets
  - Electrical links from NIC to switch

- All-to-all amongst switches in a group
  - Electrical links between switches

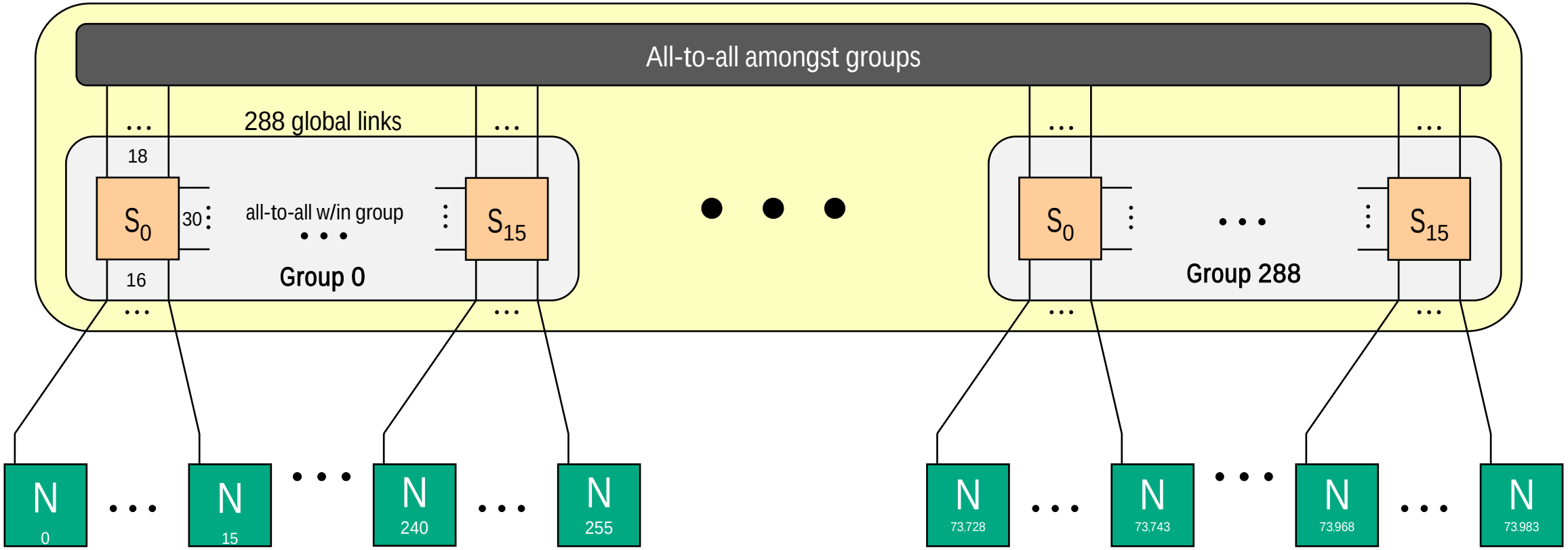
- All-to-all between groups
  - Optical links

- Group can have different characteristics
- Global network can be tapered to reduce cost



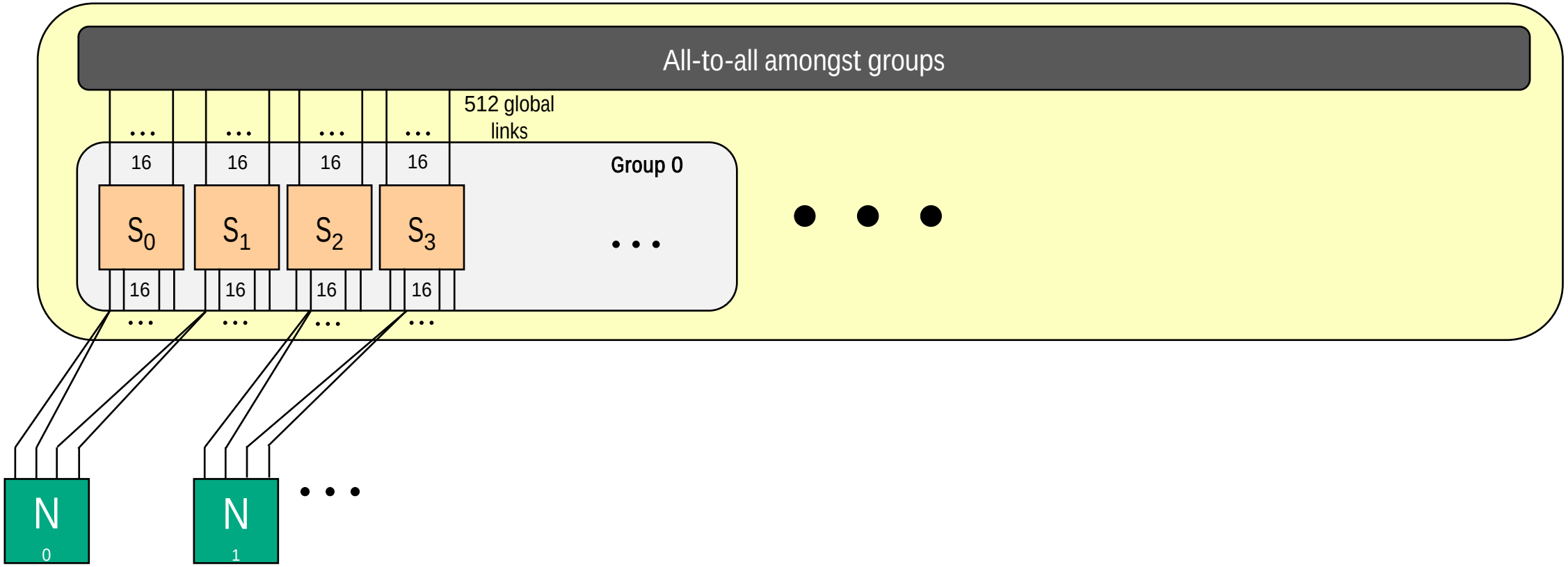


Slingshot Network – 16 Switch Group 1 NIC per Node (LUMI-C)



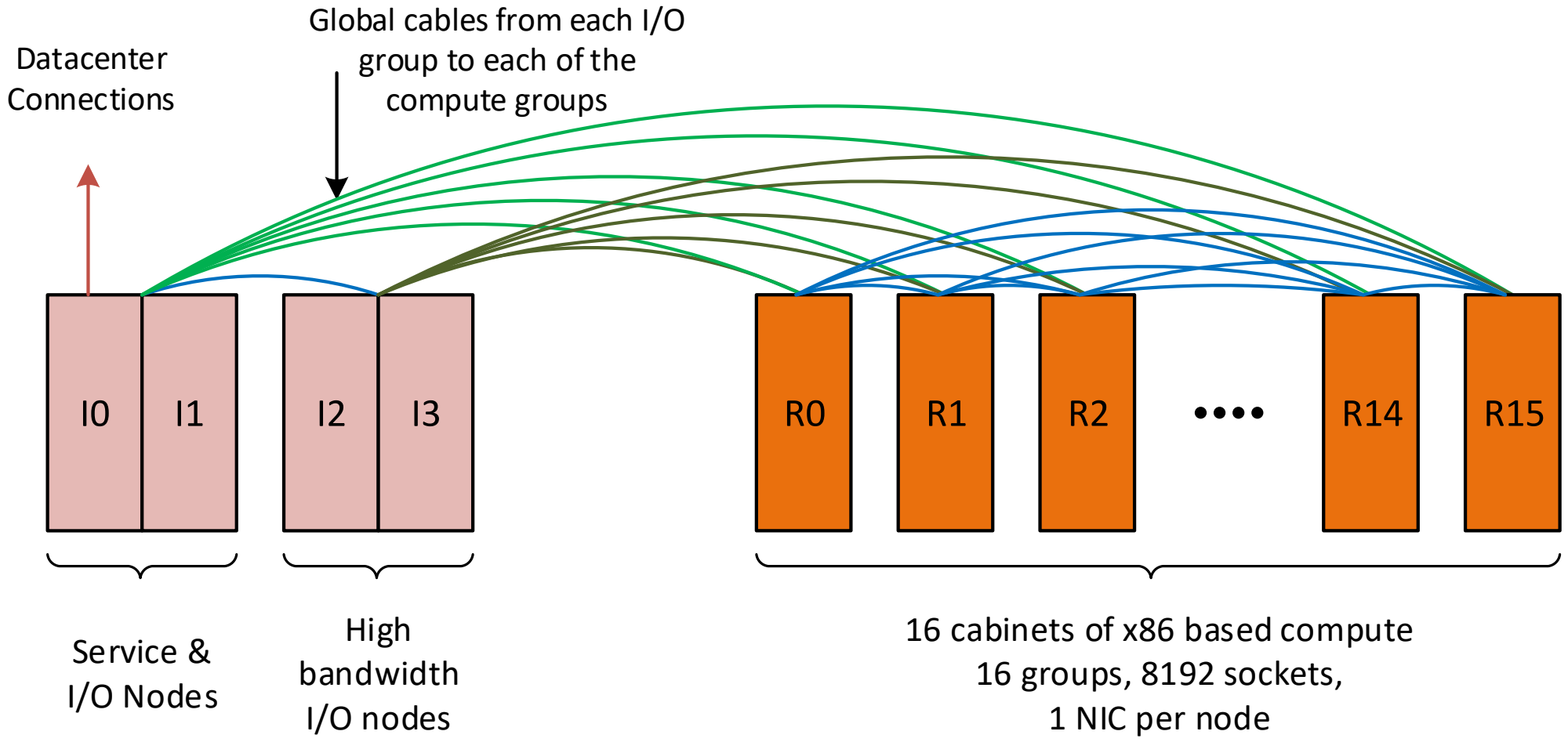
16 switch group, 1 group per cabinet, 1 NIC for each of 256 nodes, up to 73,984 nodes (289×256)

Slingshot Network – 32 Switch Group 4 NICs per Node (LUMI-G)

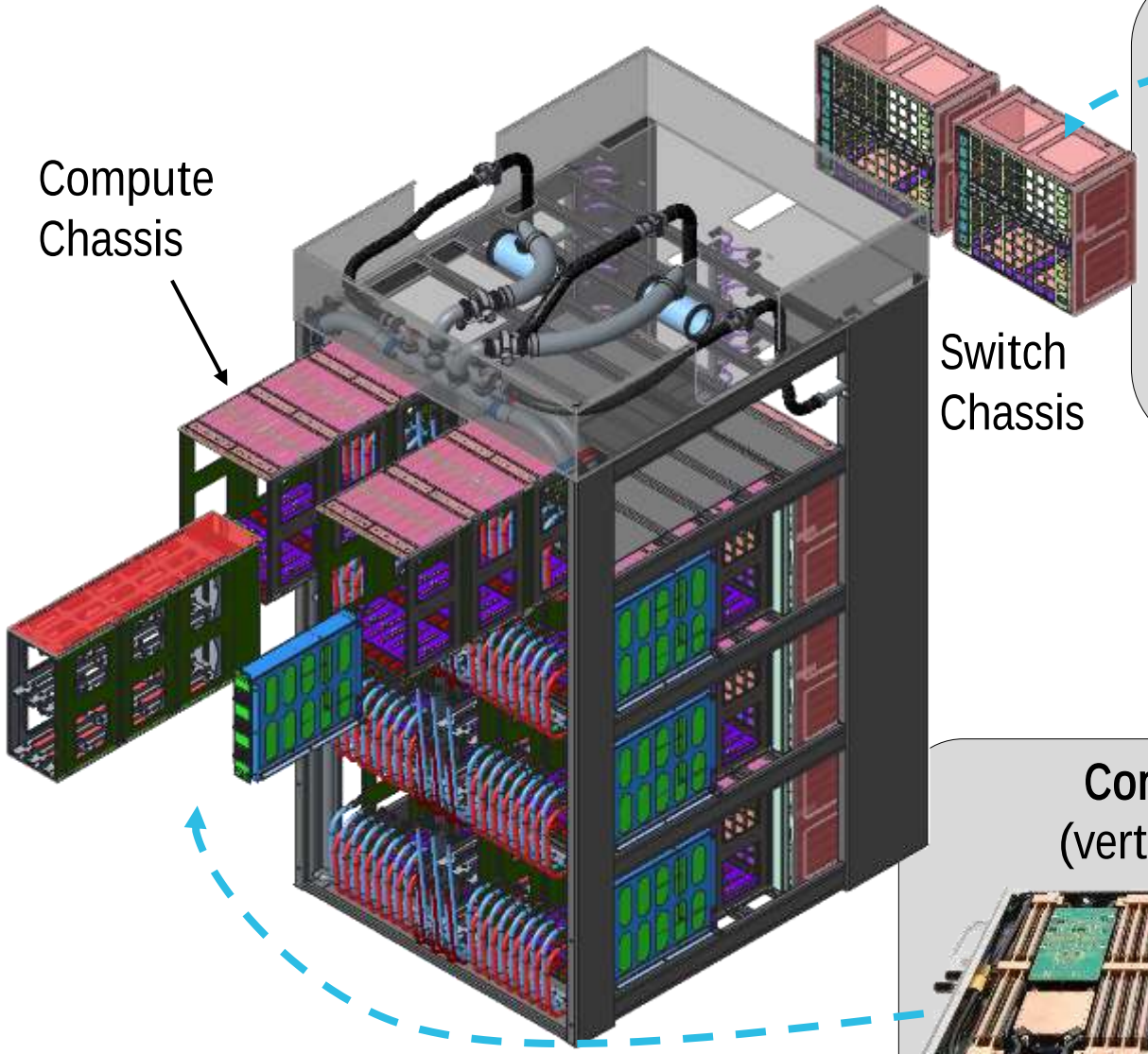


32 switch group, 1 groups per cabinet, 4 NICs for each of 128 nodes

Cray Slingshot Network in a Shasta System



Slingshot in HPE Cray EX Supercomputer



"Colorado" Switch (horizontal in chassis)

Rear - To Fabric

Group/Global links

Internal – mate to blade

Switch and NICs mate at 90° orientation

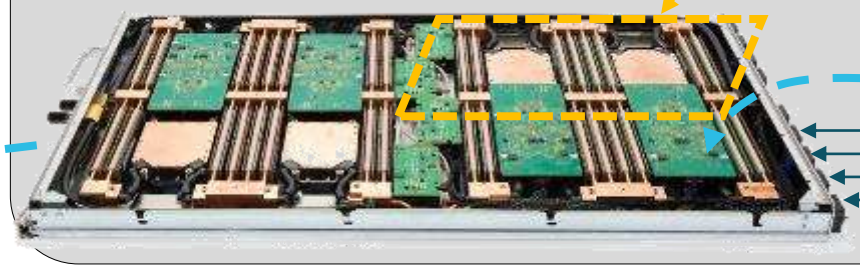
Group cables

Global cables

QSFP-DD Cables

Compute blade
(vertical in chassis)

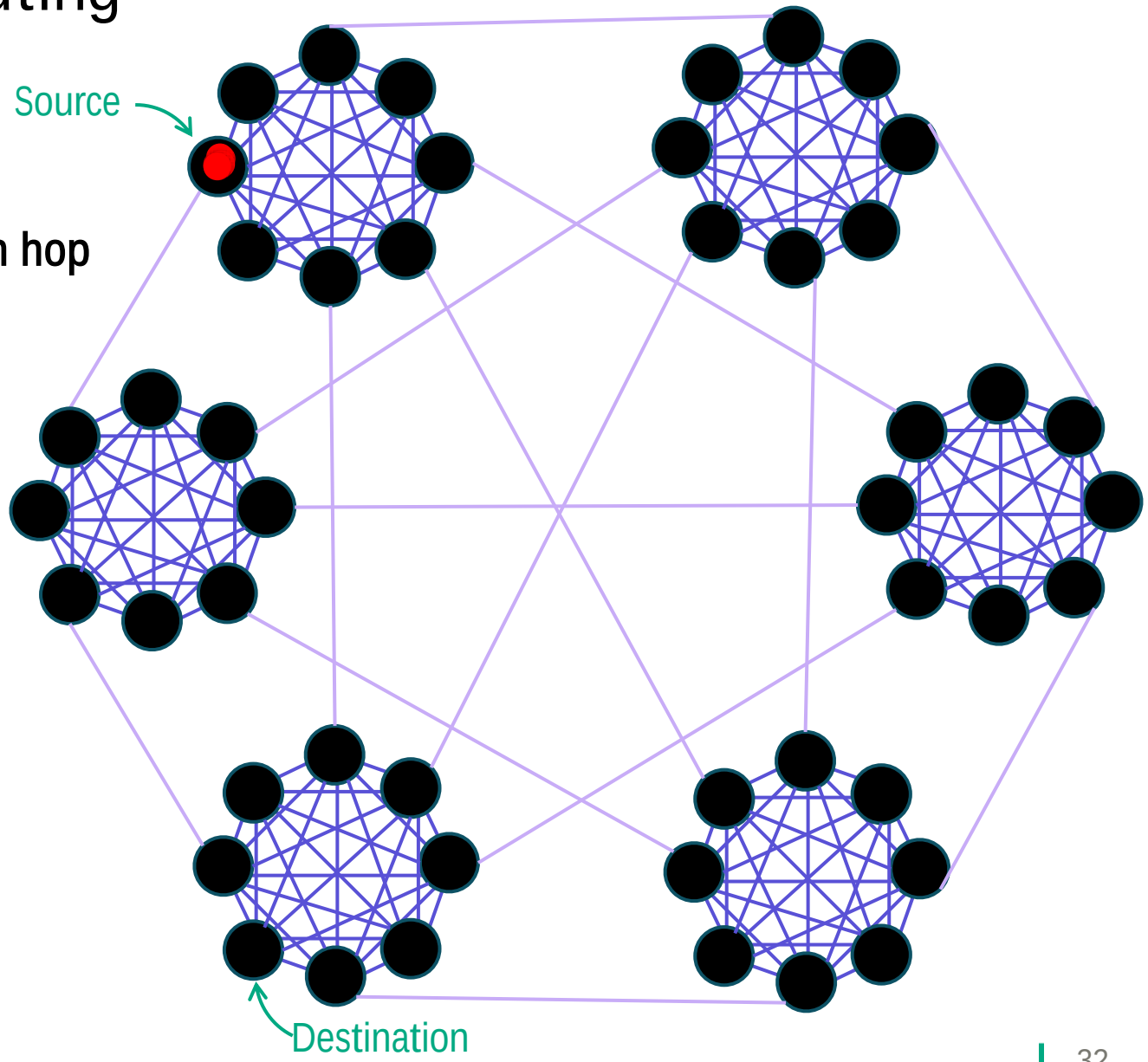
One 2-socket server node
(4 nodes per tray)



Dual CX5 NIC Mezz Card

Slingshot's Fine-Grain Adaptive Routing

Packet-by-packet routing of unordered traffic
(e.g. MPI/Lustre bulk data) optimally routed at each hop



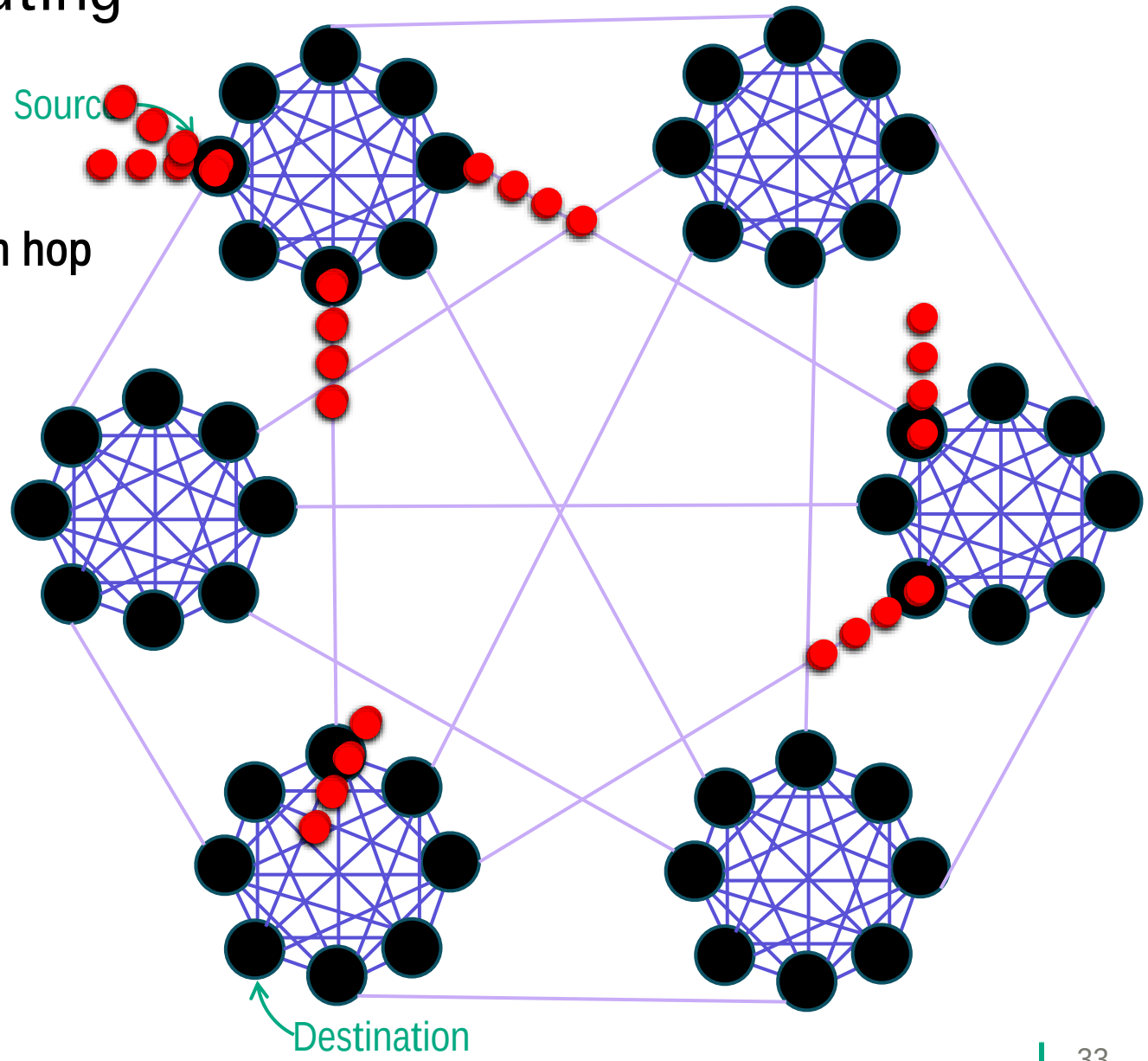
Slingshot's Fine-Grain Adaptive Routing

Packet-by-packet routing of unordered traffic
(e.g. MPI/Lustre bulk data) optimally routed at each hop

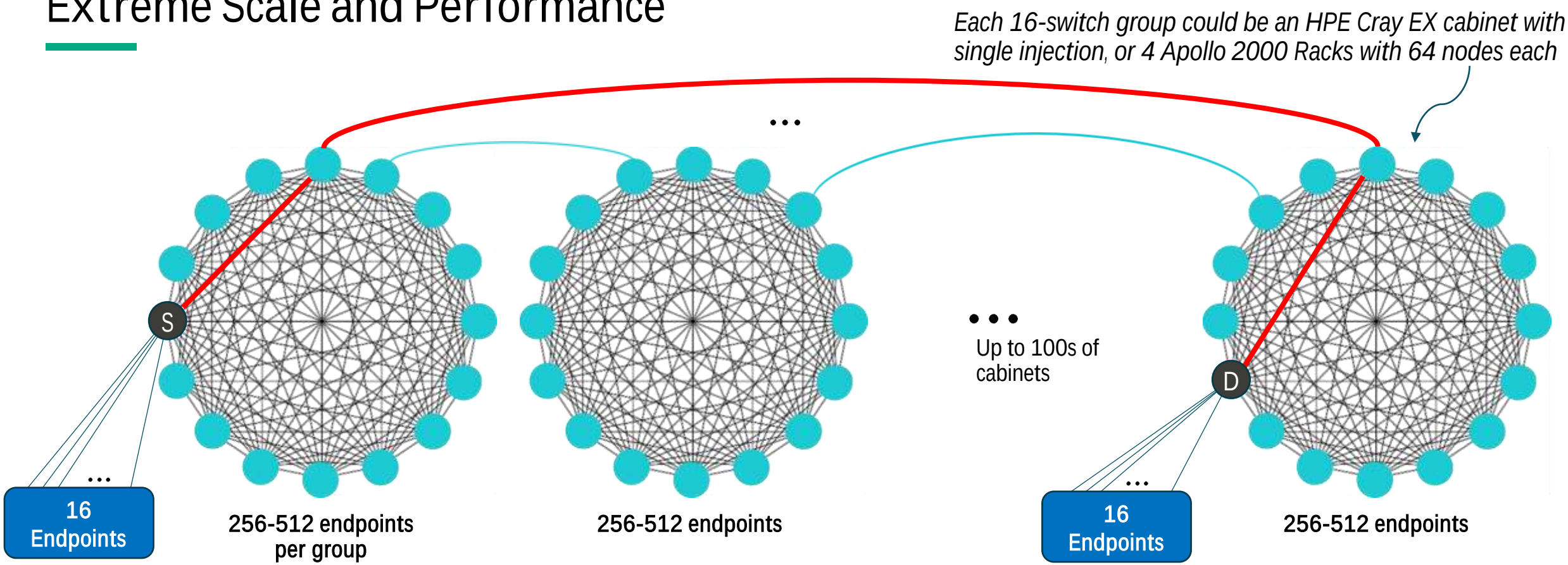
Adaptive routing of ordered traffic (e.g. Ethernet)
Each new flow can take an optimal new path

Impact:

- Performance on ordered flows with rerouting is nearly as good as that for unordered traffic
- Realize near full B/W of the topology at any scale
- Utilize cost effective topologies like dragonfly

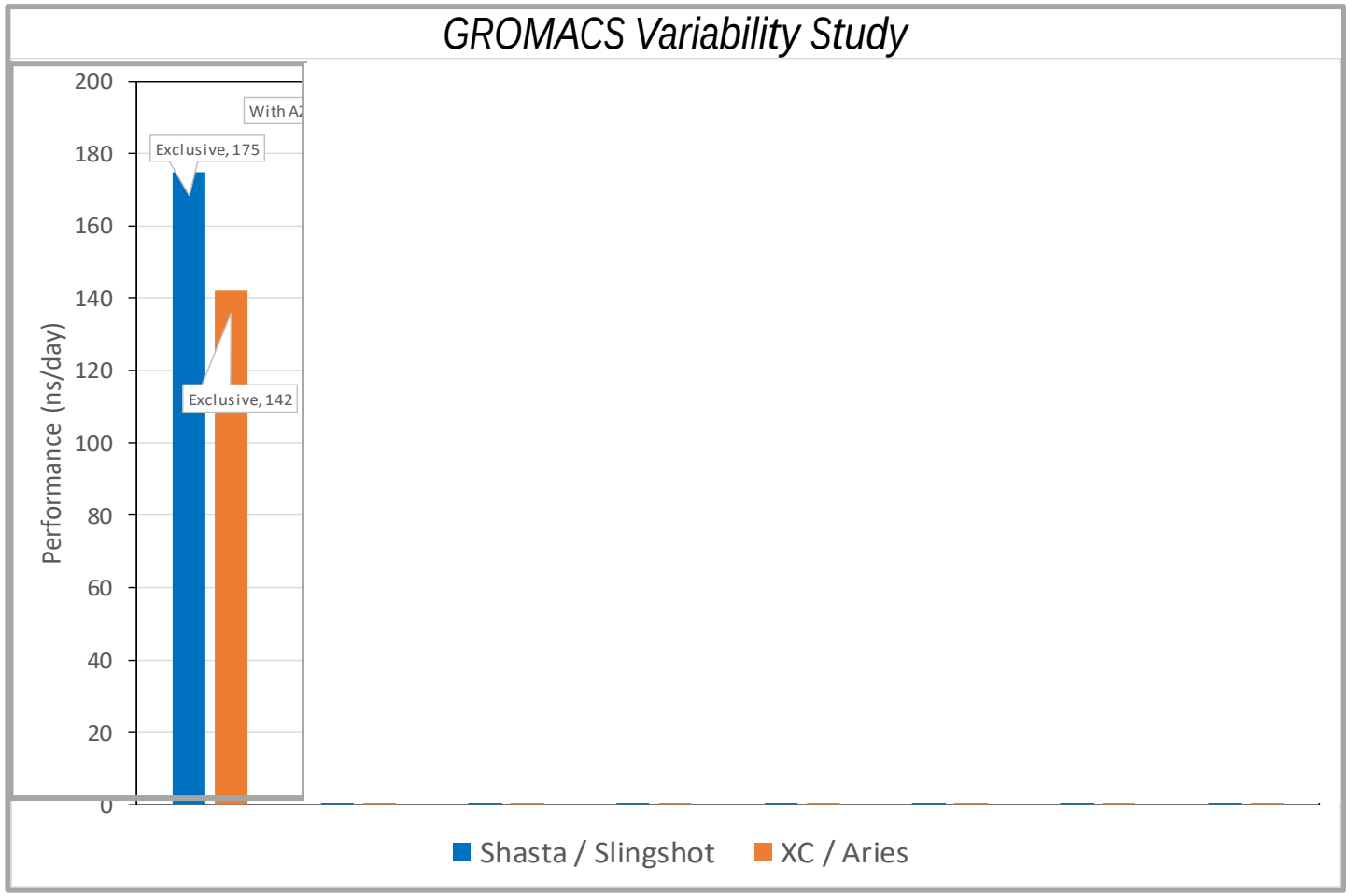


Extreme Scale and Performance



Scales to over 250K endpoints with just ***three*** switch-switch hops!

Congestion Management



"At the same time that HPC centers are getting increasingly in need of congestion control is precisely the moment when Cray-now-HPE has a new switch that is doing congestion control in a new fashion…

The congestion control features in HPE Slingshot seem to be working like a charm."

Timothy Prickett Morgan – The Next Platform

Eliminate run-to-run variability and achieve better real-world performance

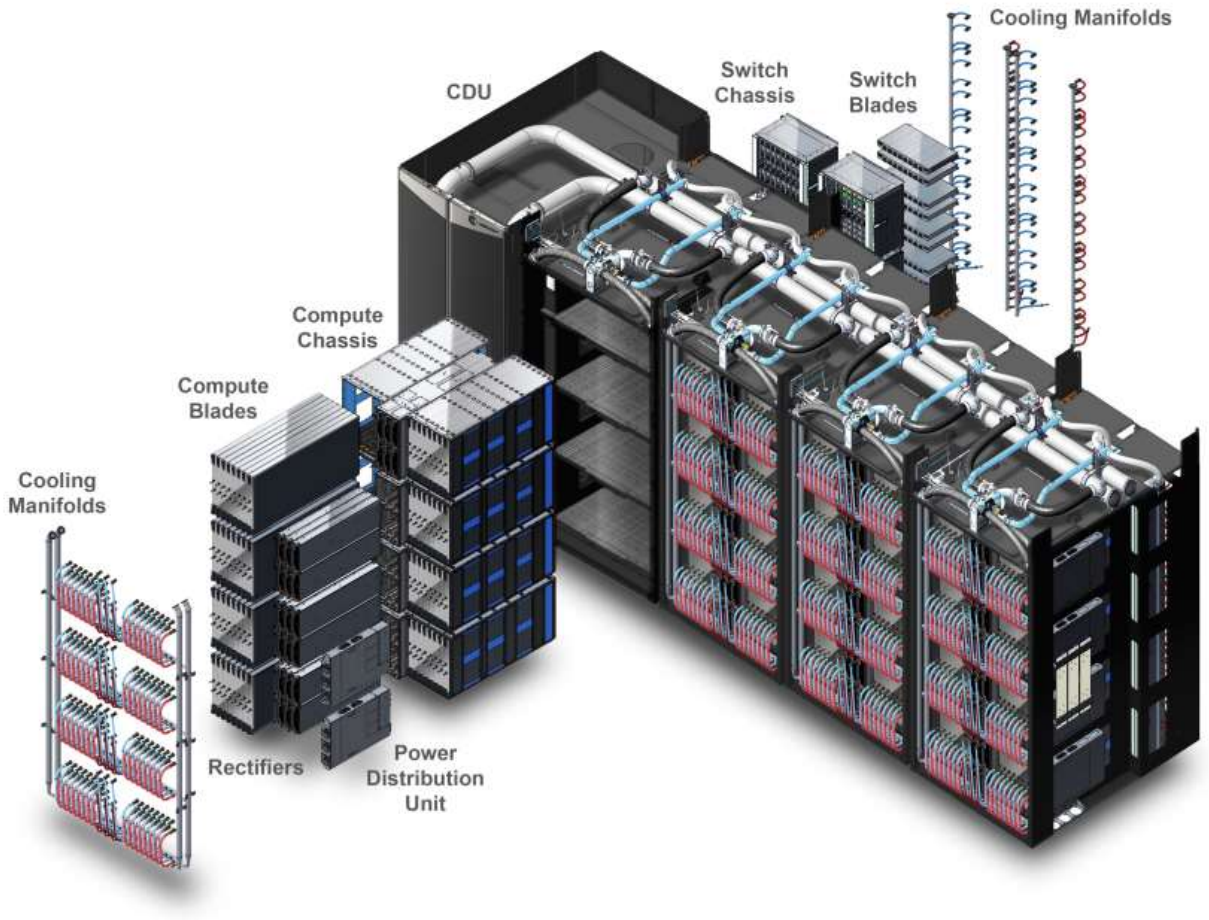
GPCNet on Frontier

NetworkLoad Tests v1.3

- Test with 72000 MPI ranks (9000 nodes)
- 1800 nodes running Network Tests
- 7200 nodes running Congestion Tests (min 1800 nodes per congestor)

Network Tests running with Congestion Tests - Key Results		
Test Name	**Congestion Impact Factor**	
	Avg	99%
RR Two-sided Lat (8 B)	1.0X	1.0X
RR Two-sided BW+Sync (131072 B)	1.0X	1.0X
Multiple Allreduce (8 B)	1.0X	1.0X

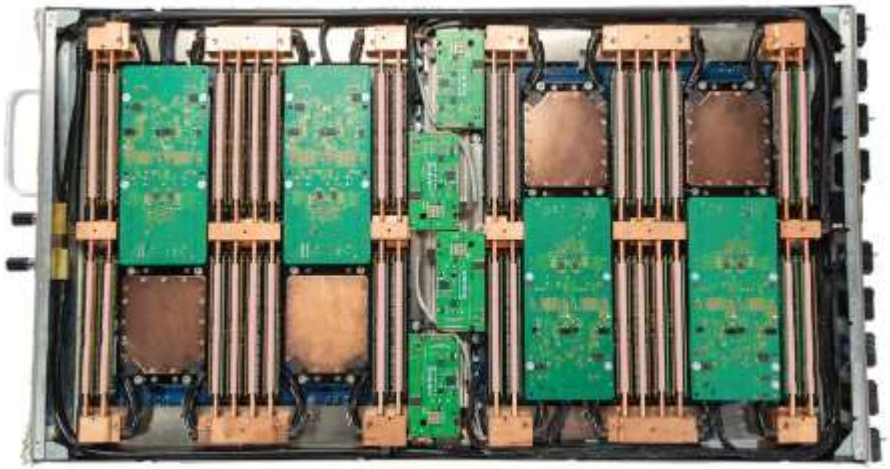
Cooling Overview



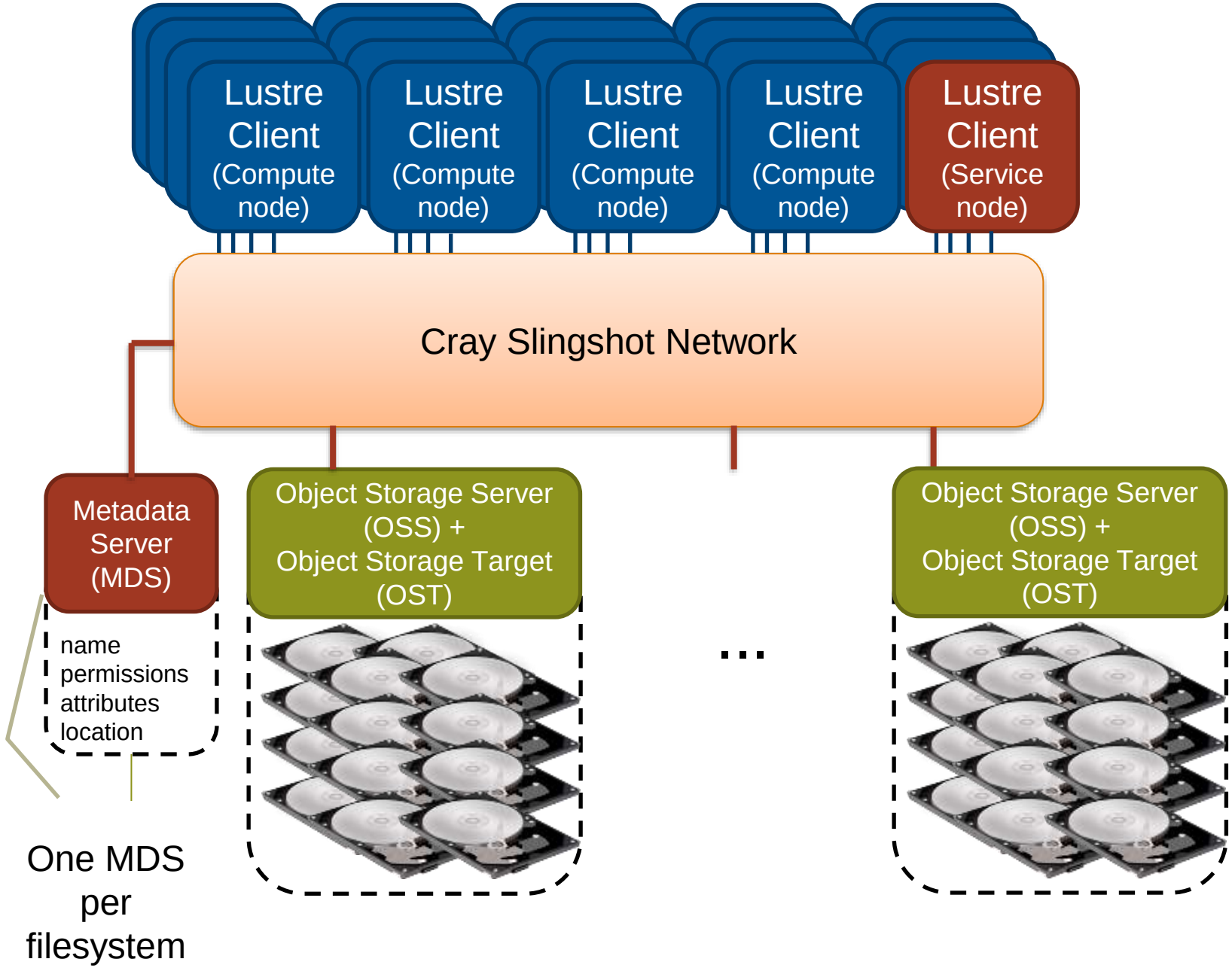
Cooling Distribution Unit removes heat from up to 4 cabinets into data centre water

Cray EX System Cooling

- Efficient direct liquid cooling is used to cool the system
- Cooling Distribution Unit (CDU) supports up to 4 cabinets
- CDU removes heat from compute loop into data centre water
- Data centre inlet temperature up to 32C
- Cooling loop traverses CDU and cabinets
- Compute blades, switch blades and rectifiers are liquid cooled
- Quick connect, dripless connectors are used to facilitate maintenance
- Both cooling plates (eg for CPU / mezzanine) and capillary tubes (memory) are used on compute blades



Storage



One MDS per filesystem

Clusterstor E1000 Lustre Storage

- Single building block

- System Management Unit (SMU)
  - Boot and management services
- Fully integrated Metadata Unit (MDU)
  - Lustre Metadata software
  - Metadata NVMe storage
  - Dual redundant management servers
  - Metadata storage target
- Fully integrated Scalable Storage Unit (SSU-D)
  - Storage controller, Lustre server
  - Disk Storage Modules NVMe SSDs and SAS HDDs for main storage
- Fully integrated in storage rack

Clusterstor E1000F

- Single building block

- System Management Unit (SMU)
  - Boot and management services
- Fully integrated Metadata Unit (MDU)
  - Lustre Metadata software
  - Metadata NVMe storage
  - Dual redundant management servers
  - Metadata storage target RAID10
- Fully integrated Scalable Storage Unit (SSU-F)
  - Storage controller, Lustre server
  - NVMe gen 4 based storage
  - Optimized for throughput
  - GridRAID
- Fully integrated in storage rack