

Steve Scott
Cray CTO





Slingshot is Designed for HPC





Used in all three announced US Exascale systems

© 2019 Cray Inc.

But Also for New Data-Centric Workloads



Standards-driven Ethernet networks for interoperability



State-of-the-art HPC networks for performance

Groundbreaking congestion control that provides strong performance isolation between applications (major advance over traditional ECN mechanism)











Slingshot: Bringing HPC to Ethernet



HPC Networks Slingshot **Ethernet** Standards based Standards based Proprietary (single / interoperable / interoperable vendor) Commodity Non-commodity Commodity technology technology technology Converged Converged HPC interconnect network network only Limited HPC Full set of HPC Full set of HPC features features features High latency Low latency Low latency Efficient for large Efficient for small Efficient for small to payloads only to large payloads large payloads Limited scalability Very scalable for Very scalable for for HPC **HPC & Big Data** HPC & Big Data

© 2019 Cray Inc.

Slingshot Overview





Slingshot is Cray's 8th generation scalable interconnect

Earlier, Cray pioneered:

- Adaptive routing
- High-radix switch design
- Dragonfly topology

64 ports x 200 Gbps

Over 250K endpoints with a diameter of just three hops

Ethernet Compliant

Easy connectivity to datacenters and third-party storage;

"HPC inside"

World class Adaptive Routing and QoS

High utilization at scale; flawless support for hybrid workloads

Efficient Congestion Control

Performance isolation between workloads

Low, Uniform Latency

Focus on tail latency, because real apps synchronize

© 2019 Cray Inc.

Slingshot's Rosetta Switch

- TSMC 16nm FF
- 64 ports x 200 Gbps/dir
 - PAM4 56G
- ~250W
- Tiled architecture:
 - 32 peripheral function blocks
 - Network SerDes, MAC/PCS/LLR
 - Ethernet Lookup functions
 - 32 tile blocks
 - All other port functionality
 - Management Block (MB)
- First silicon Sept '18, Production Q4'19



Intra-Switch Routing

- 4 rows of 8 tiles
- Two switch ports per tile
- Distributed crossbars based on row busses, column channels, and per-tile crossbars
- No global arbitration

Port A

© 2019 Cray Inc.

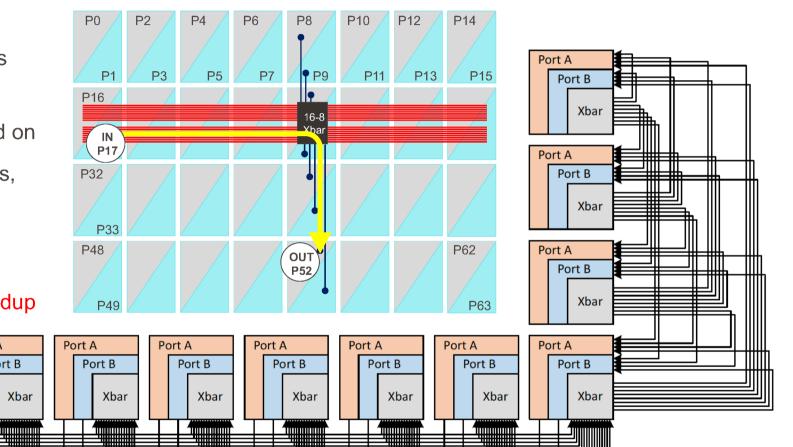
Port B

Xbar

4x internal speedup

Port A

Port B



Aurora: Architecting Argonne's First Exascale Supercomputer for Accelerated Scientific Discovery

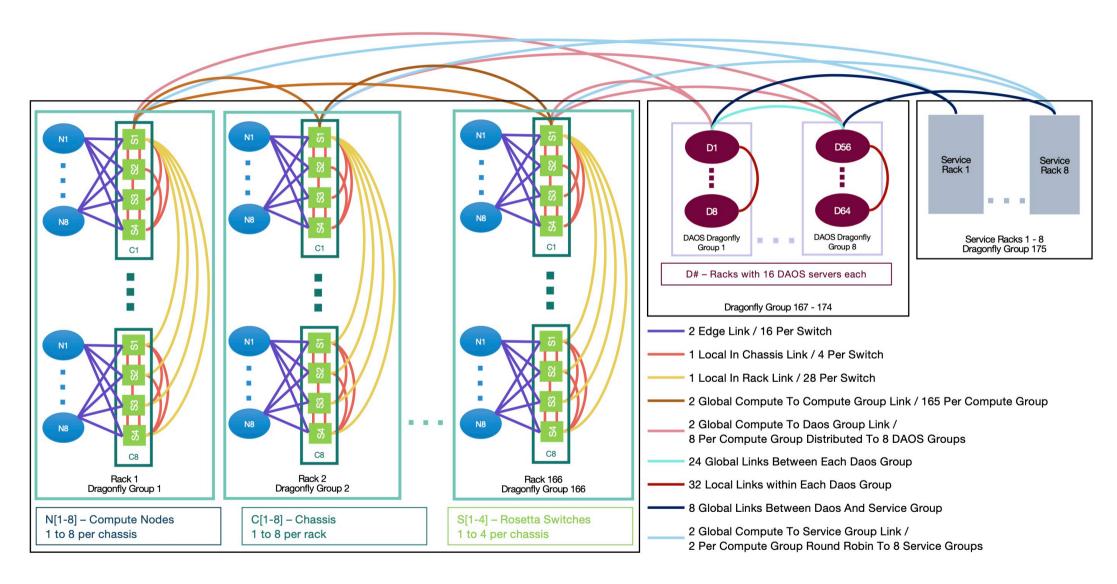


Figure 6: Aurora 1-D Dragonfly Topology.