Cray XC Scalability and the Aries Network

Tony Ford

June 29, 2017

Exascale Scalability

- **Which scalability metrics are important for Exascale?**
  - Performance (obviously!)
  - What are the contributing factors?

- **How can we demonstrate these principles today?**
  - Our architectural vision needs qualification

Second Thing First… Qualification

- **How can we demonstrate and qualify scalability metrics for supercomputing?**

- **NNSA ASC Advanced Technology Platform**
  - LANL / SNL Trinity Supercomputer

We already build BIG…

LANL / SNL Trinity System

Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	National Supercomputing Center in Wuxi China	Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway NRCPC	10,649,600	93,014.6	125,435.9	15,371
2	National Super Computer Center in Guangzhou China	Tianhe-2 (MilkyWay-2) - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P NUDT	3,120,000	33,862.7	54,902.4	17,808
3	Swiss National Supercomputing Centre (CSCS) Switzerland	Piz Daint - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 Cray Inc.	361,760	19,590.0	25,326.3	2,272
10	**DOE/NNSA/LANL/SNL United States**	**Trinity - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Aries interconnect Cray Inc.**	**301,056**	**8,100.9**	**11,078.9**	**4,233**

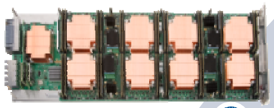
Why Trinity?

- **First instantiation of NNSA ASC Advanced Technology Platform**
  - Establishes foundation for Exascale
    - Meet future needs of current applications
    - Enable adaptation to new methodologies

Trinity Architecture Overview

Trinity			
Node architecture	KNL & Haswell	Haswell Partition	KNL Partition
Memory capacity	2.11 PB	>1 PB	>1 PB
Memory BW	>7 PB/s	>1 PB/s	>1 PB/s
Peak Flops	42.2 PF	11.5 PF	30.7 PF
Number of nodes	**19,000+**	**>9,500**	**9,500**
Number of cores	**>760,000**	**>190,000**	**>570,000**
PFS capacity	>80 PB		
Burst Buffer capacity	3.7 PB		
Network interconnect	**Cray Aries Network**		
Number of cabinets	112		

Aries Network Infrastructure



Compute Blade

4 Compute Nodes

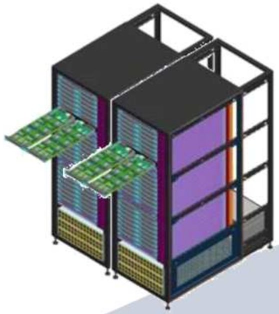
Chassis

Rank 1 Network

16 Compute Blades

No Cables

64 Compute Nodes



Group

Rank 2 Network

Passive Electrical Network

2 Cabinets

6 Chassis

384 Compute Nodes

System

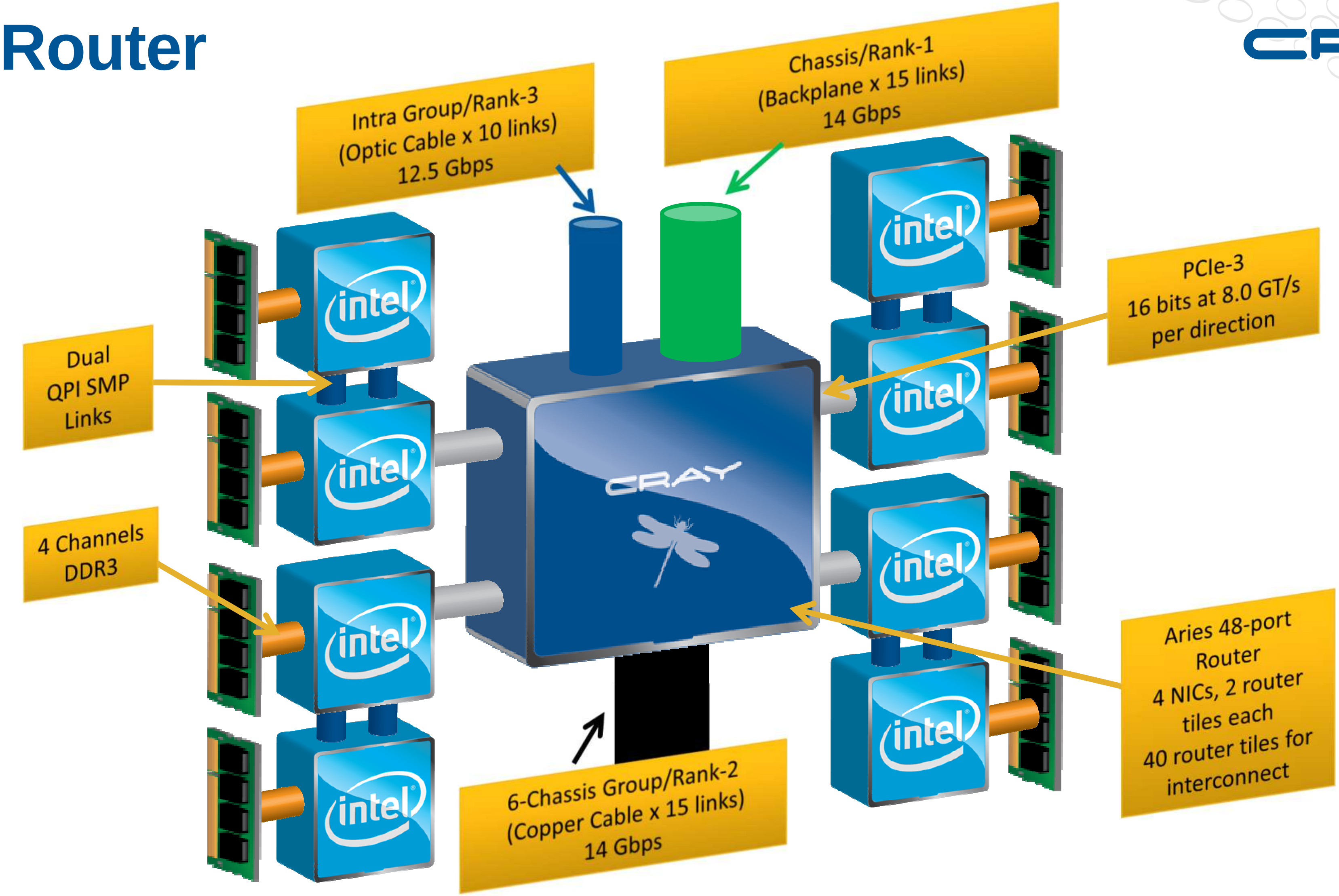
Rank 3 Network

Active Optical Network

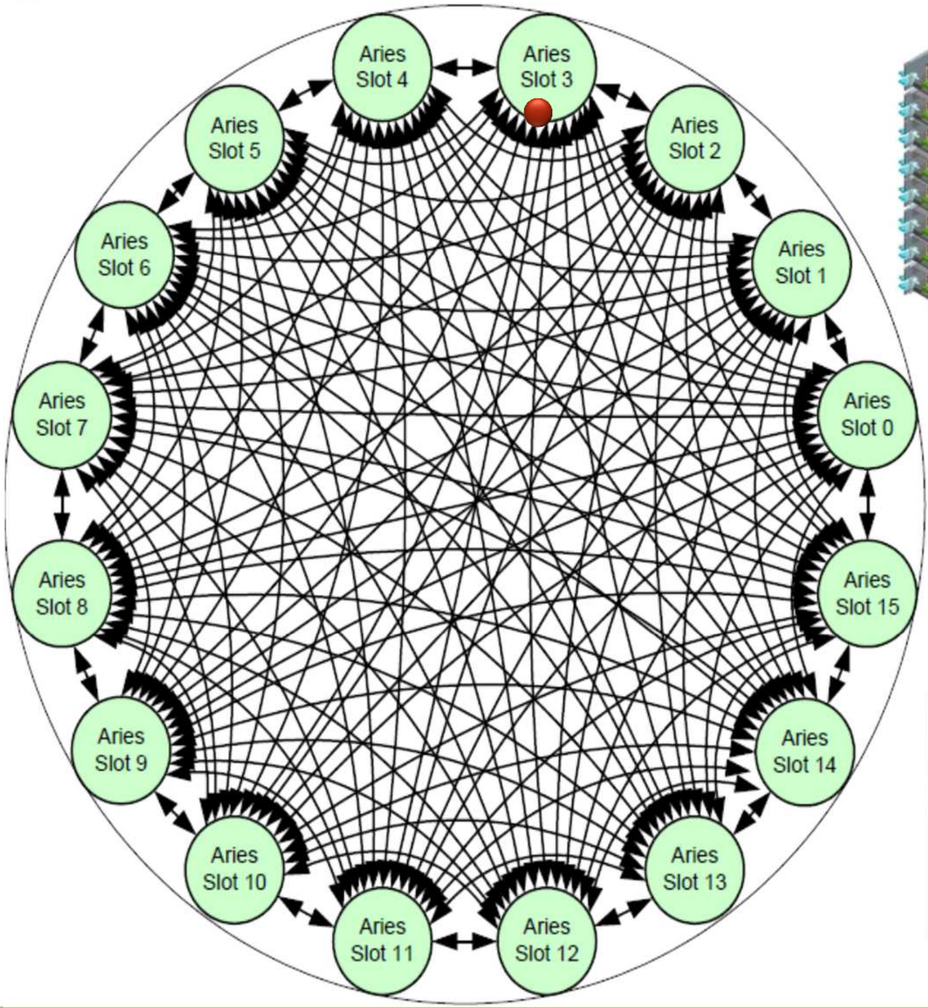
Hundreds of Cabinets

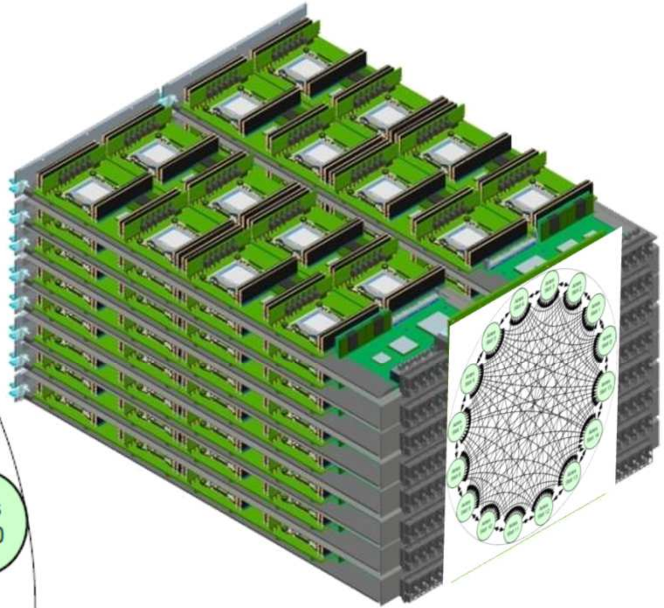
Up to 10s of thousands of nodes

Aries Router



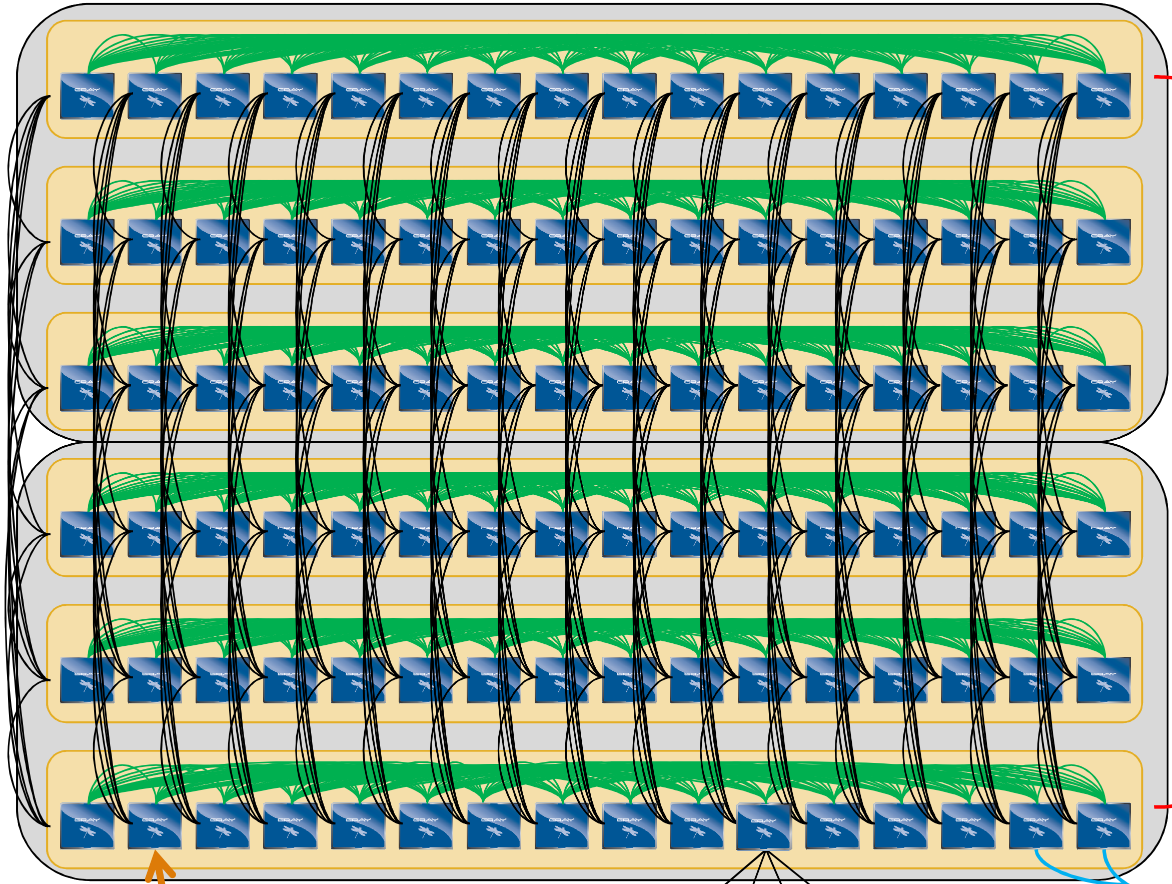
Cray XC Rank-1 Network





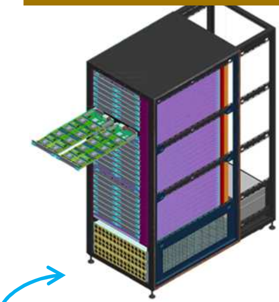
- **Chassis with 16 compute blades**
- **128 Sockets**
- **Inter-Aries communication over backplane**
- **Per-Packet adaptive Routing**

Cray XC Rank-2 Network



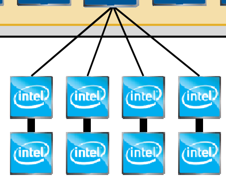
2 Cabinet Group
768 Sockets

6 backplanes connected with copper cables in a 2-cabinet group: “Black Network”



Active optical cables interconnect groups “Blue Network”

16 Aries connected by backplane “Green Network”



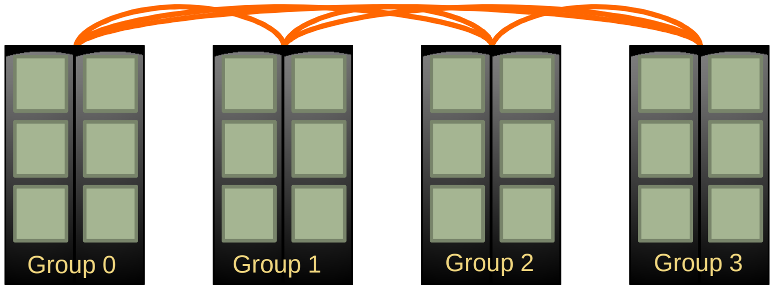
4 nodes connect to a single Aries

Cray XC40 Rank-2 Network

- **Cray XC40 two-cabinet group**
  - 768 Sockets
  - 96 Aries Chips
- **All copper and backplanes signals running at 14 Gbps**

Cray XC40 Rank-3 Network

- An all-to-all pattern is wired between the groups using optical cables (blue network)
- Up to 240 ports are available per 2-cabinet group
- The global bandwidth can be tuned by varying the number of optical cables in the group-to-group connections



Example: A 4-group system is interconnected with 6 optical "bundles". The "bundles" can be configured between 20 and 80 cables wide

Scalability Metrics for Supercomputing

- **Standard metrics**
  - Infrastructure
  - Compute
  - Power

- **Performance of collective operations across system**
- **MPI stack memory footprint**
- **Job start-up times**

Why Dragonfly?

- **Cost**
  - Dragonfly minimizes the use of active optical components
  - Eliminates need for Director switches
- **Scalability**
  - Topology scales to very large systems
  - Flat average hop count and latency
- **Simplicity**
  - Implemented without external switches
  - No HBAs or separate NICs and Routers
- **Performance**
  - More than just a case of clever wiring, this topology leverages state-of-the-art adaptive routing that Cray developed with Stanford University

Why Dragonfly?

- **Comparison with Fat-Tree**
  - Cost
    - Fat-Tree increases cost per node with system size
    - 2x optical links for the same global bandwidth
    - Requires external ToR and Director class routers
  - Latency
    - Fat-Tree requires 2 optical hops per route vs. 1 for Dragonfly
  - Load balancing
    - Application traffic patterns (all-to-all, uniform random) self load-balance using Dragonfly
    - Large electrical group captures most of local load

Infrastructure Topology

- **Dragonfly average hop-count is low and flat**
  - Up to 2 hops within source group
  - One hop to destination group
  - Up to 2 hops within destination group
  - Hop-count stays low out to very big networks

- **Per-node bisectional bandwidth is flat**
  - Half of all-to-all connected global links
  - Grows linearly with system size out to very big networks

Power Consumption – Energy to Solution

Energy to solution = time to solution x system power x power utilization efficiency

- **Time to solution**
  - Reducing time is the most effective element
  - *Improvements in scalability directly affect time to solution*

- **System power**
  - XC system rack consumes ~80kW under load
  - 420W per dual socket node
  - 15W per node Aries Dragonfly
  - 48W per node Fat-Tree
  - *Reduces consumption by 625kWH - $3.5m running costs*

- **Power utilization efficiency (PUE)**
  - 480V distribution, 48V rack, ~1V component
  - Warm water cooling
  - *Cray XC is typically 1.1 – 1.25 PUE*
  - *Google data center – 1.21 PUE*
  - *Microsoft data center - 1.22 PUE*

PUE = Facility Energy / System Energy

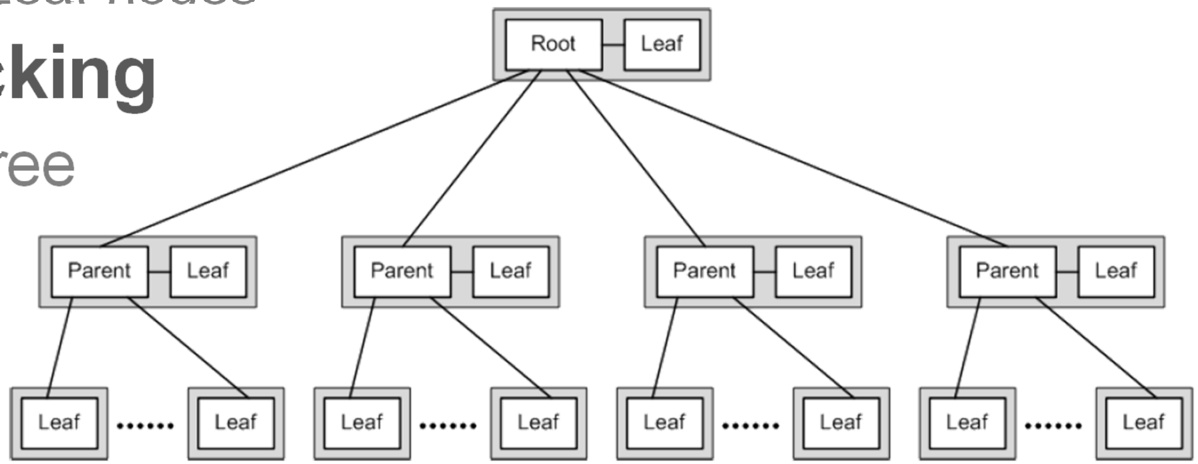
Collective Operations

- **Aries NICs includes a Collective Offload Engine**
  - No CPU involvement
  - Latency optimized
  - Scales with network size
  - Up to radix 32
  - 32 / 64 bit integer and floating point add
  - Min/max, compare & swap, bit operations
  - No topological tree dependency

- **High branching ratio = shallow trees**
  - Low latency
  - Require only 3 stage reduction tree for common operations on Trinity

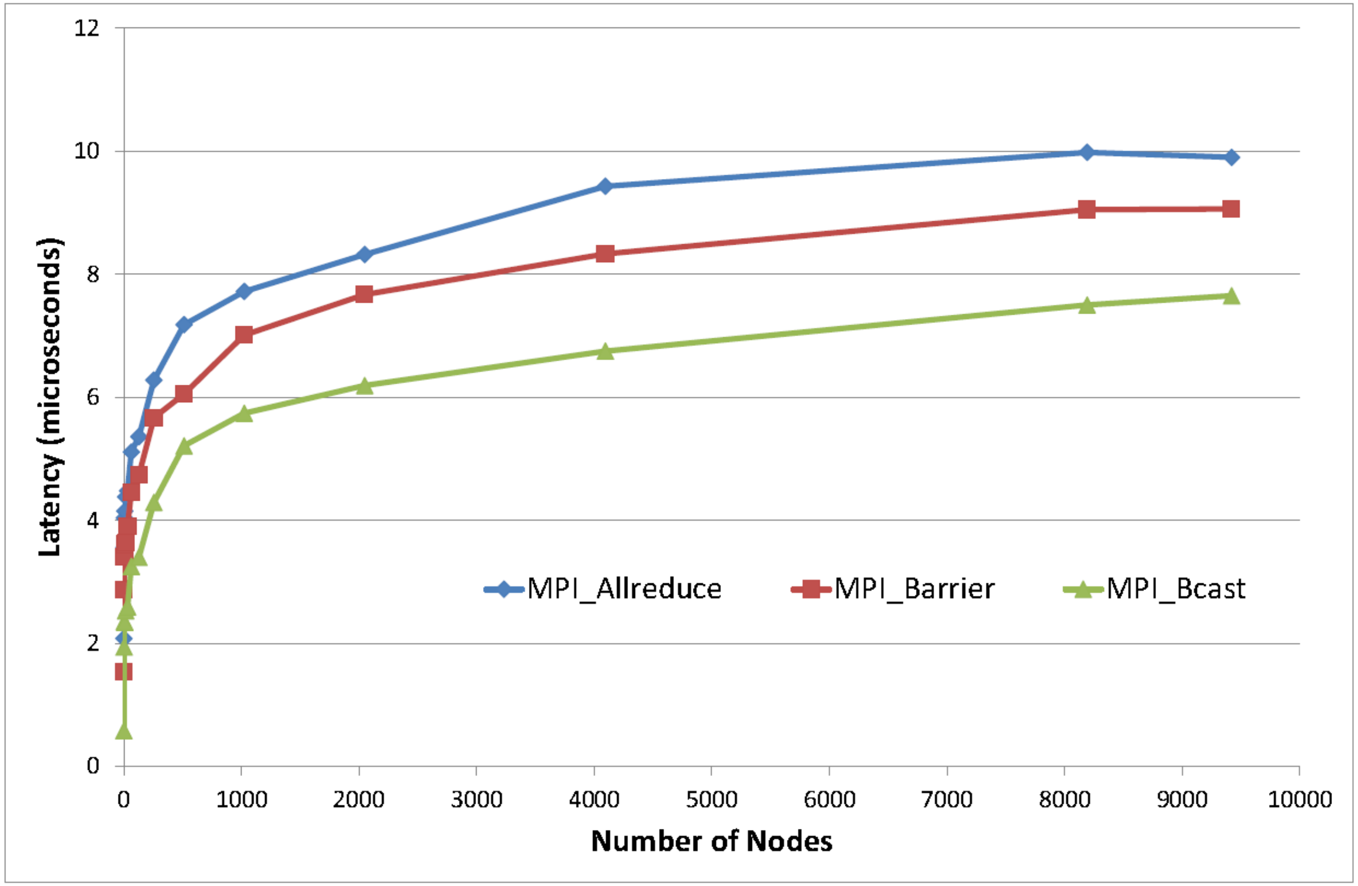
Collective Operations Offload

- **Typical 2 phase operation**
  - Ready phase
    - Leaf Nodes join the reduction
    - Reduction operator applied as data moves towards the Root Node
  - Multicast phase
    - Results pass from Root to Leaf nodes
- **Operations are non-blocking**
  - 128 per job for a radix-32 tree



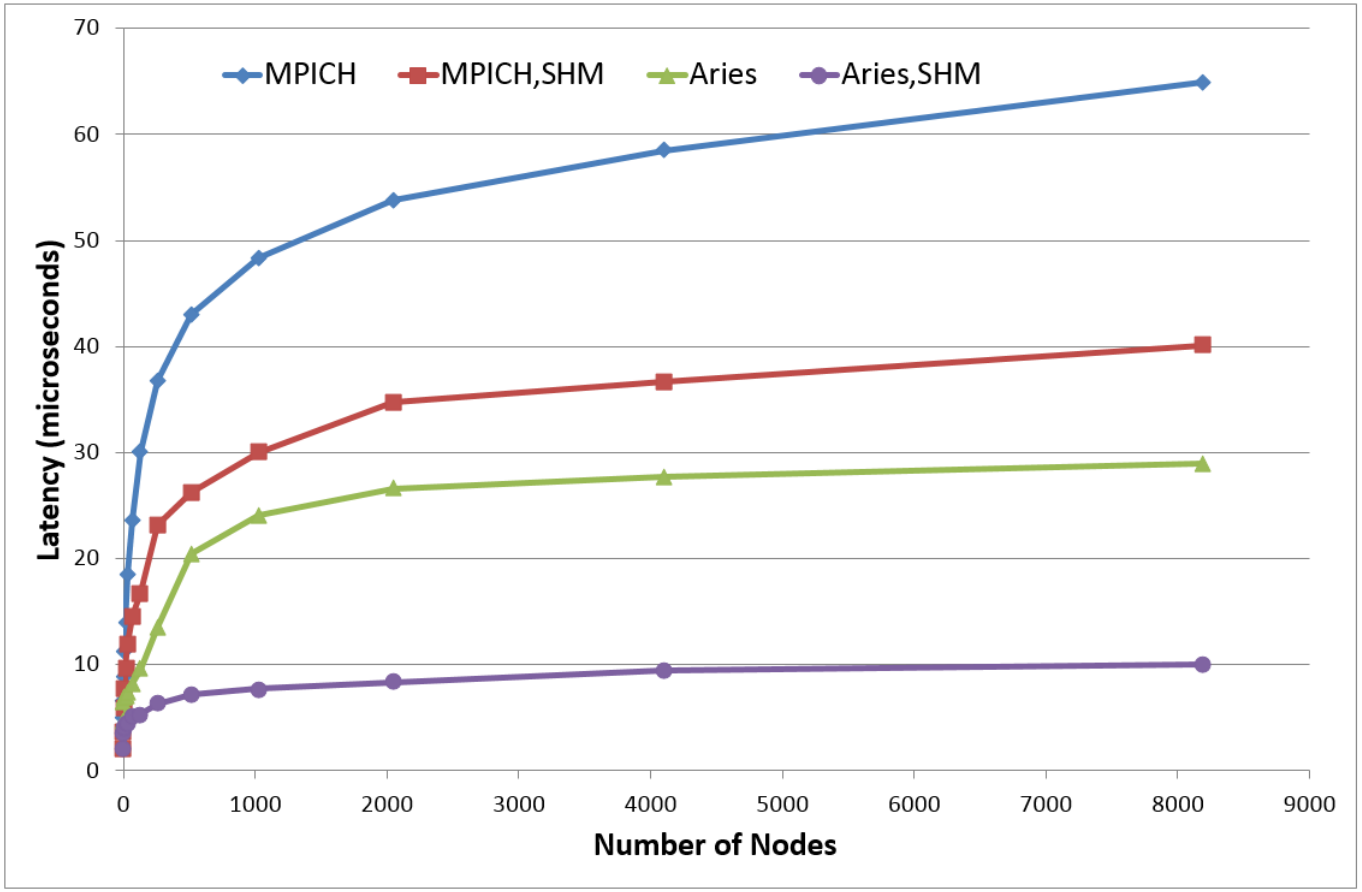
Collective Operations Offload

- **Trinity latencies**
  - High branching ratio
  - High scalability
- **All 8 byte operations under 10 usecs**

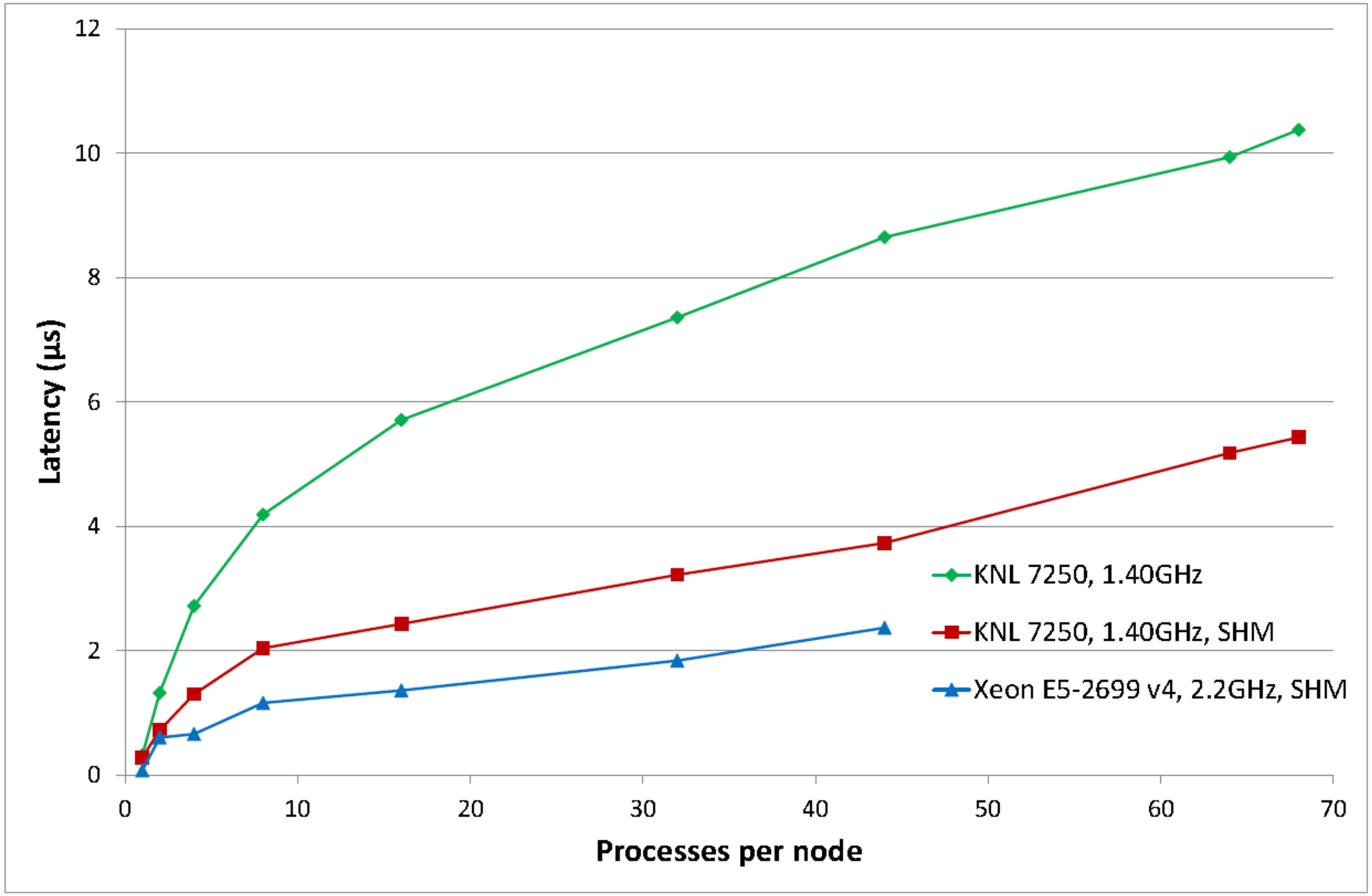


Collective Operations – Local shared-memory

- **MPI_Allreduce**
- **Offload changes the balance**
  - Now intra-node reduction becomes significant
  - Especially on many-core nodes



Collective Operations – Local shared-memory

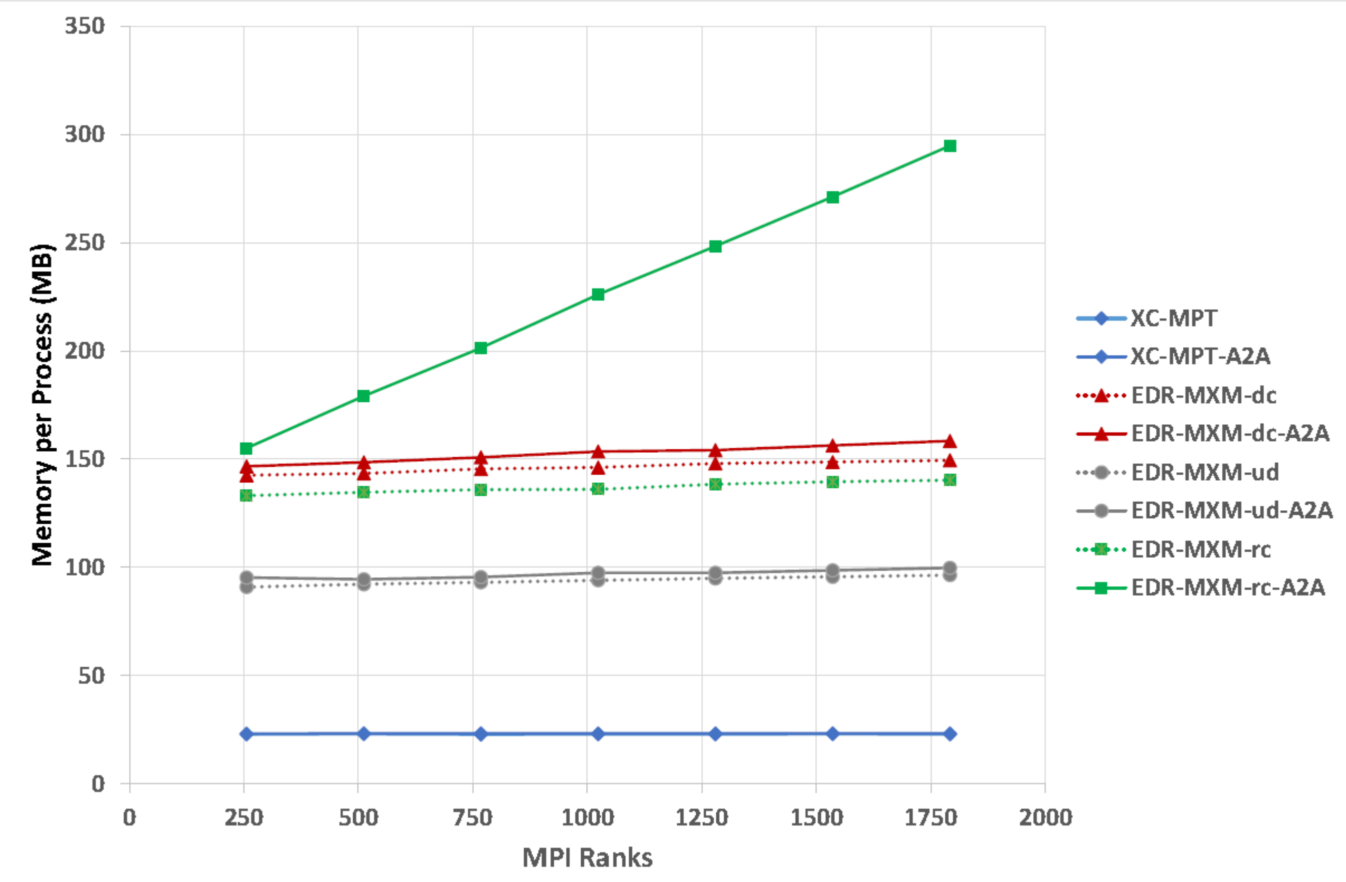


MPI Stack Memory Footprint

- **Memory used by MPI stack presents scalability barrier**
  - Each process maintains static state for every other process
  - Total memory used for inter-processor comms includes:
    - MPI Virtual Channel structures – peer-to-peer state
    - Process Management Interface
    - Transport Layer memory
    - Per-node shared state

- **Cray MPI implements dynamic MPI Virtual Channels**
  - Significantly reduces MPI stack memory footprint
  - Memory allocated only when ranks contact each other
  - Utilizes connectionless RMAs – no VC usage

MPI Stack Memory Footprint



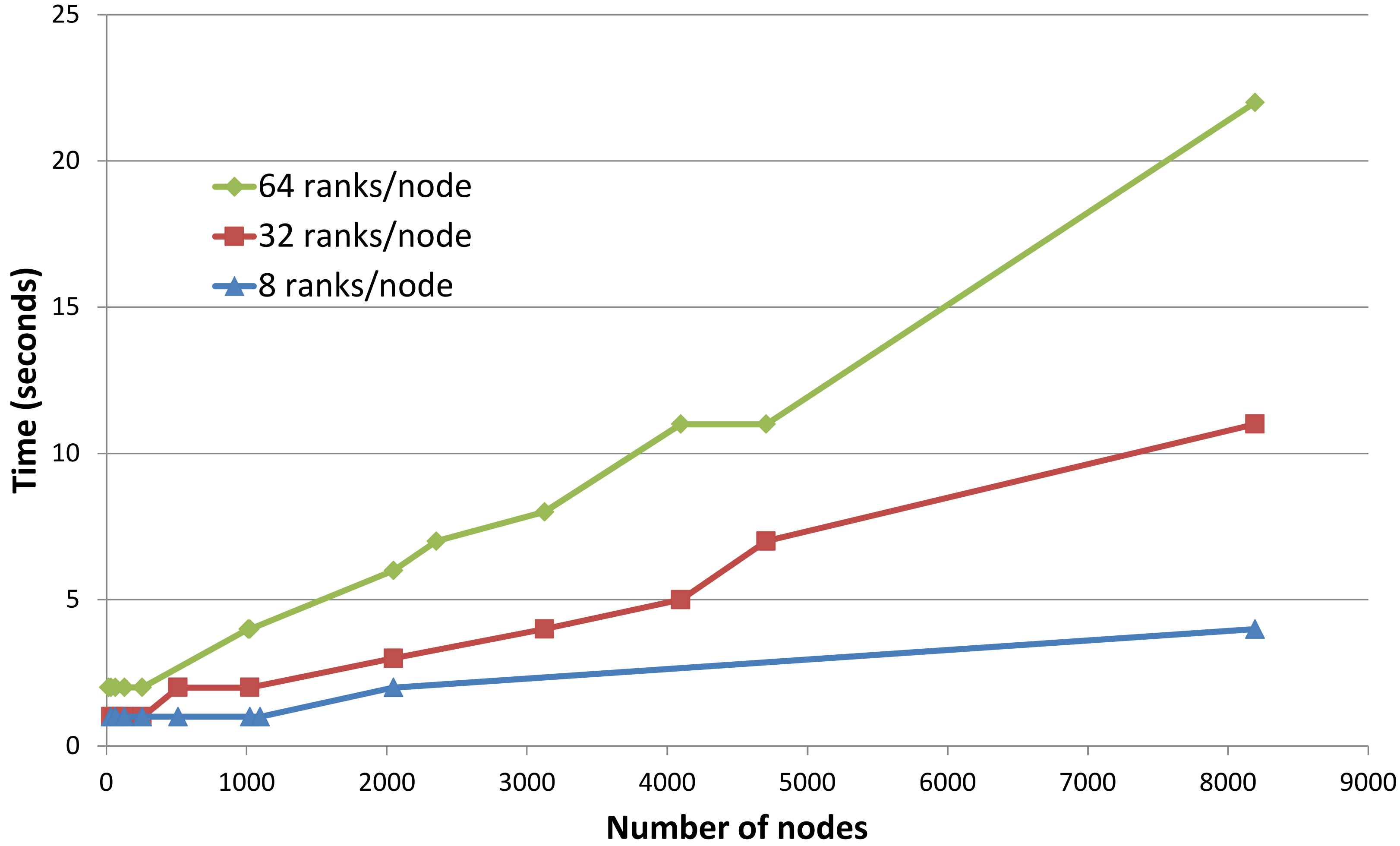
Job Startup Times

- **Startup times rise with job size**
  - Reduced efficiency for large jobs
  - Decreased system utilization

- **Cray XC startup times illustrated**
  - 301,248 process job, 32 ranks/node - 12 secs
  - 64 ranks/node – 24 secs

- **Dynamic libraries must provision I/O resources to ensure fast loading**

Job Startup Times on Trinity



Conclusions

- **Key aspects of scalability identified**
  - Infrastructure, MPI stack, collectives & cost
- **Investigated effects at scale using Trinity**
  - Running >300,000 MPI ranks
- **Excellent performance demonstrated in key metrics**
  - Benchmarks, acceptance tests & early application use
- **Interconnect & its software are critical factors**
  - Dragonfly network & software stack are key elements of Trinity
  - Results shown could not have been achieved otherwise

Thank You

tford@cray.com