Exceptional service in the national interest

Sandia National Laboratories

Early Experiences with Trinity - The First Advanced Technology Platform for the ASC Program

C.T. Vaughan, D.C. Dinge, P.T. Lin, S.D. Hammond, J. Cook, C. R. Trott, A.M. Agelastos, D.M. Pase, R.E. Benner, M. Rajan, and R.J. Hoekstra

Sandia National Laboratories

CUG 2016, London, UK

CCR Center for Computing Research

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Trinity

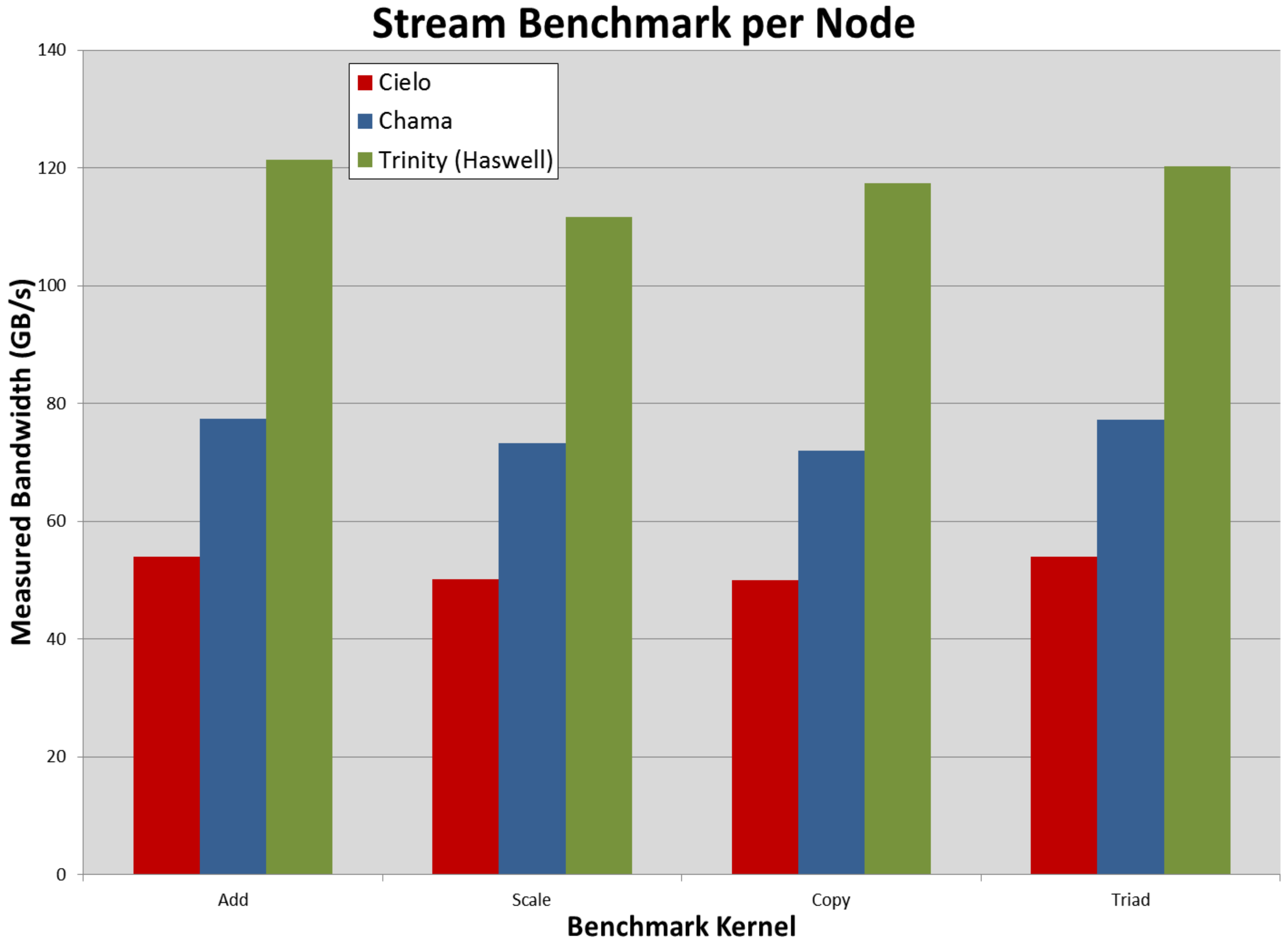
- Cray XC40
- Total of about 19000 nodes
  - About half are Intel Haswell with 2 processors per node and 16 cores per processor running at 2.3 GHz and 128 GB memory per node
  - About half are 60+ core Intel Knights Landing processors - will be delivered later this year
  - Cray Aries Dragonfly interconnect
- Expect greater than 30 PetaFlops peak

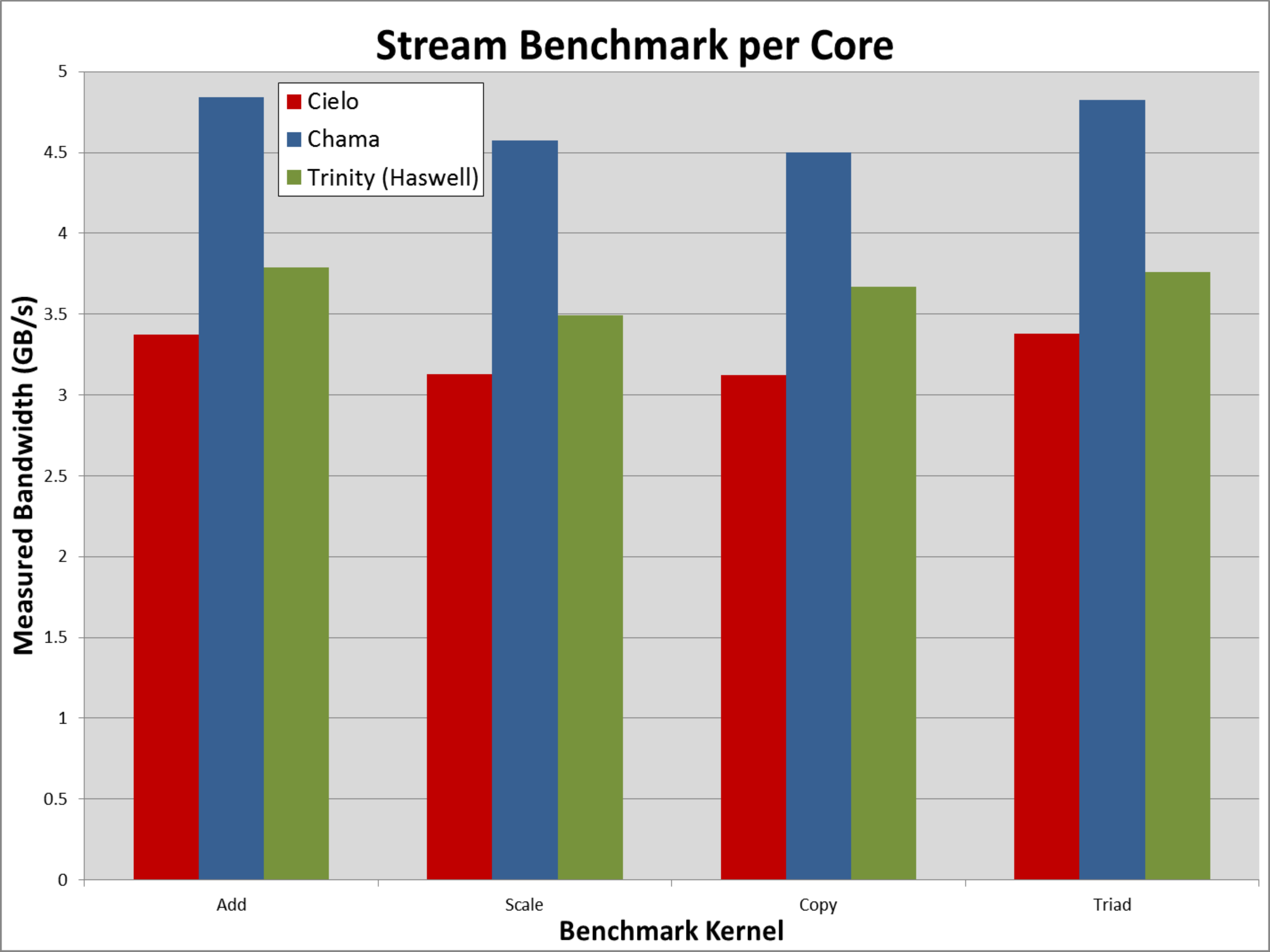
Cielo and Chama

- Cielo is a Cray XE6 and is our current generation capability machine
  - 8894 nodes with 2 oct-core AMD Magny-Cours processors running at 2.4 GHz
  - Cray Gemini 3D torus interconnect
- Chama is current generation capacity machine
  - Tri-Lab Capacity Cluster
  - 1232 nodes with 2 Intel Sandy Bridge 8 core processors running at 2.6 GHz
  - Qlogic QDR-InfiniBand Fat-Tree interconnect

Comparison of Machines

System	Cielo	Chama	Trinity (Phase-I)
Total Nodes	8,894	1,232	9,408
Total Cores	142,304	19,712	301,056
Processor	AMD Magney-Cours	Intel Sandy Bridge	Intel Haswell
Processor ISA	SSE4a	AVX	AVX-2
Clock Speed (GHz)	2.40	2.60	2.30
Cores/Socket	8 (2x4)	8	16
Cores/Node	16	16	32
Peak Node (GFLOPs)	153.6	332.8	1,177.6
Memory	DR3-1333	DDR3-1600	DDR4-2133
Channels/Socket	4	4	4
Interconnect	Cray Gemini	Qlogic QDR-InfiniBand	Cray Aries
Topology	3D-Torus	Fat-Tree	DragonFly


Stream Benchmark per Node
Cielo
Chama
Trinity (Haswell)
Measured Bandwidth (GB/s)
0
20
40
60
80
100
120
140
Add
Scale
Copy
Triad
Benchmark Kernel

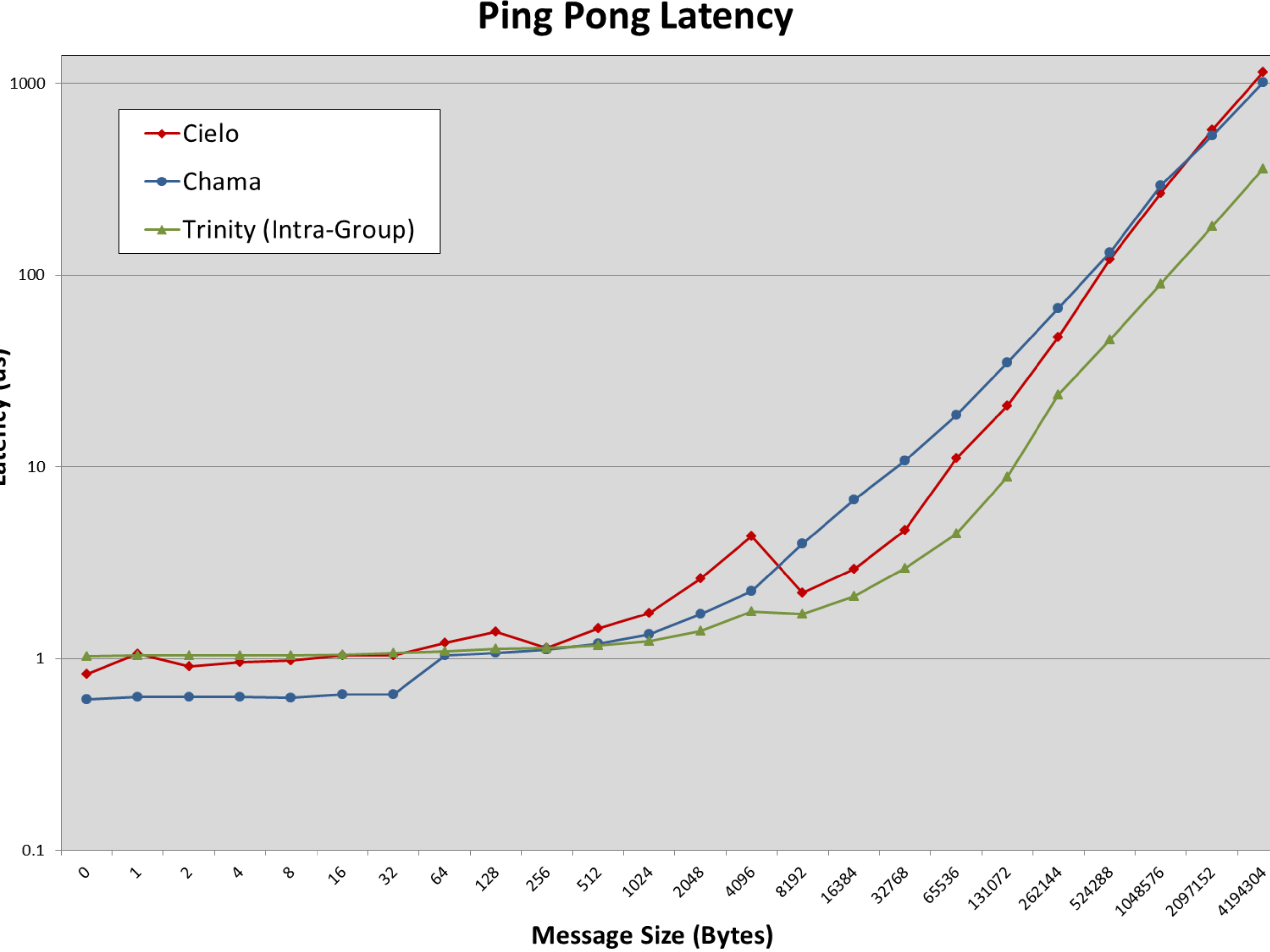

Stream Benchmark per Core
Cielo
Chama
Trinity (Haswell)
Measured Bandwidth (GB/s)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Add
Scale
Copy
Triad
Benchmark Kernel

Comments on STREAMS

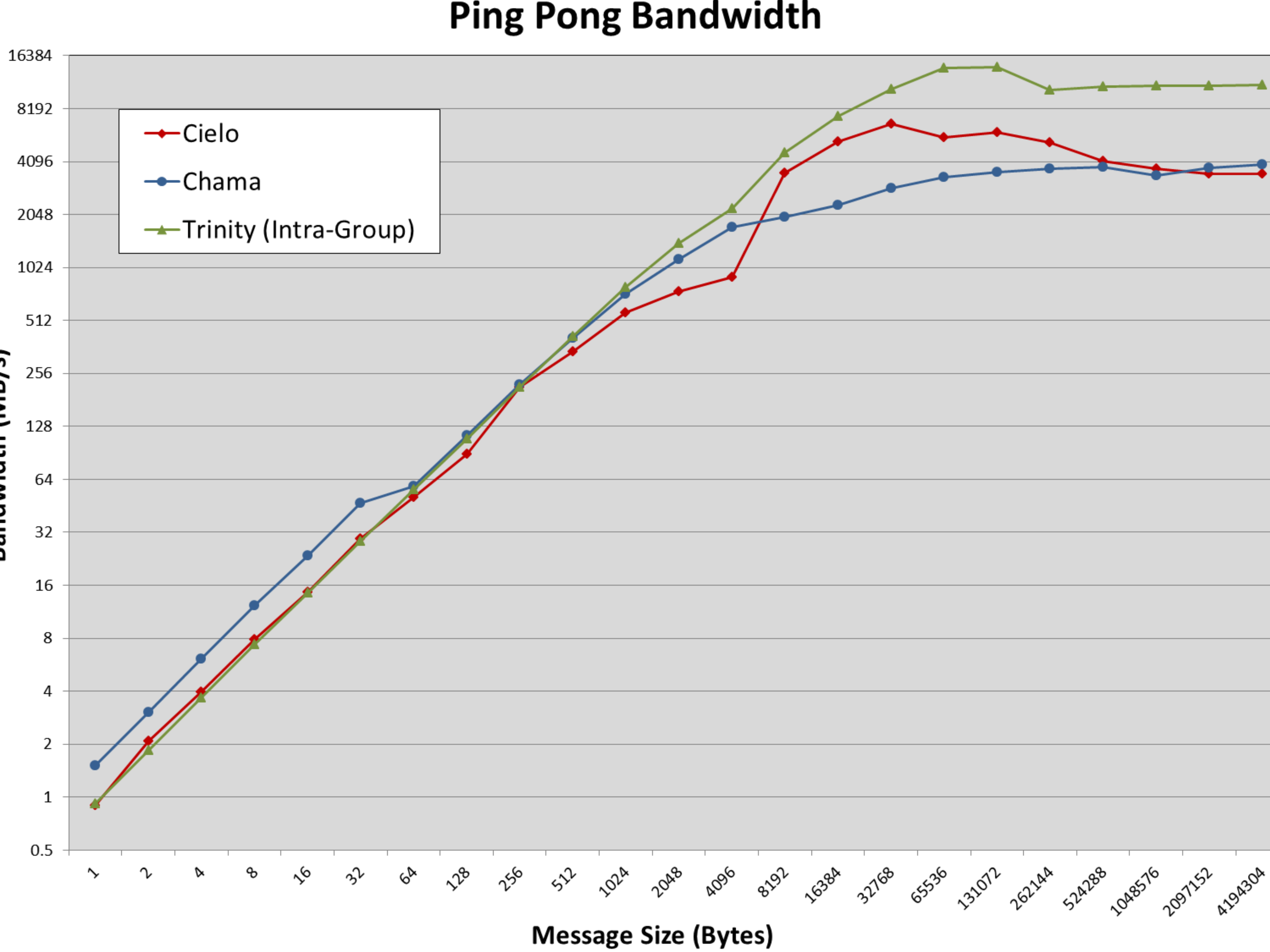
- Peak compute speed for a Trinity node is 7.7 times faster than a node of Cielo and 3.54 times faster than a node of Chama

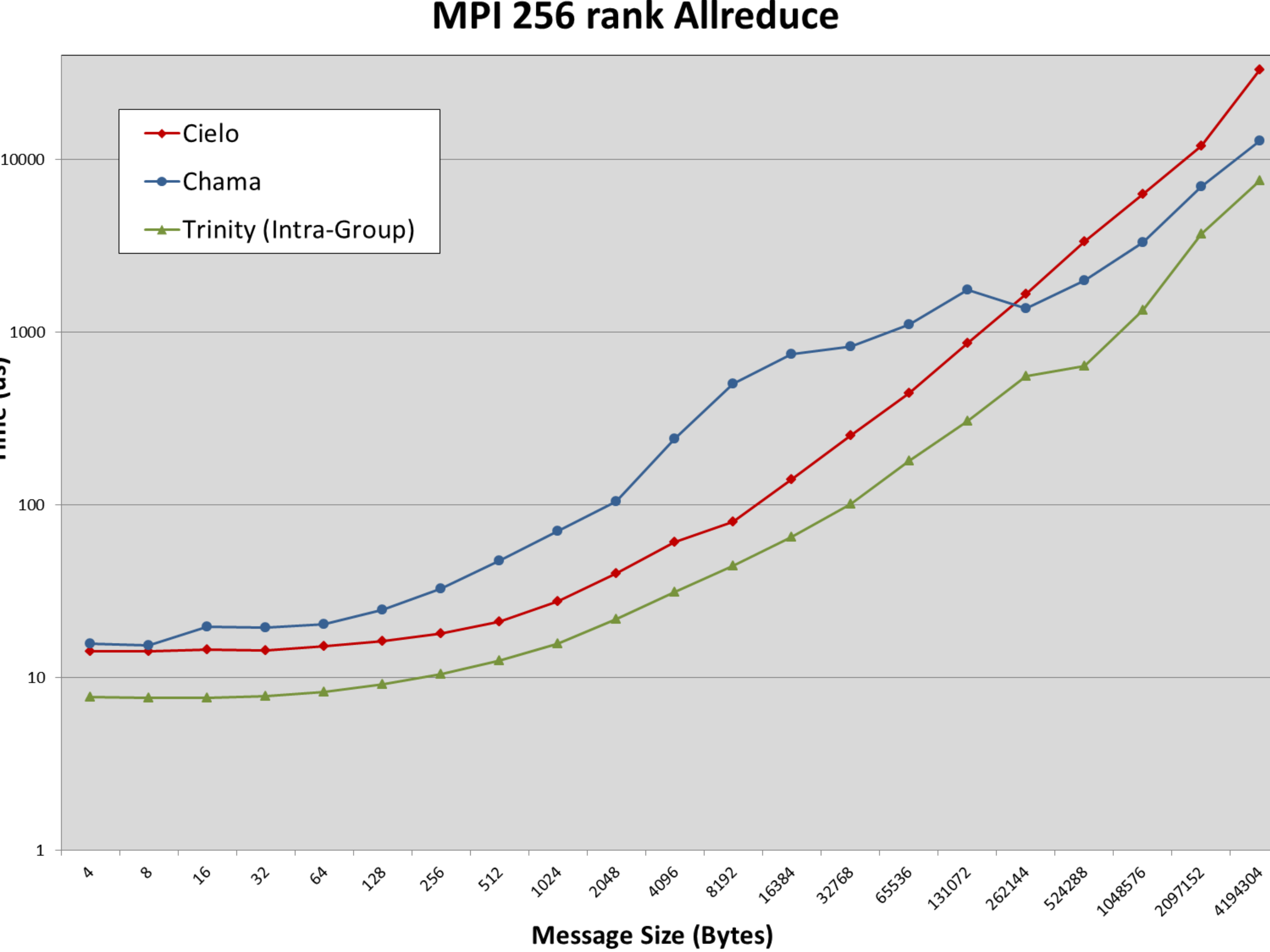
- STREAM results per node on Trinity are >2X that of Cielo and more than 1.5X that of Chama

- STREAM results per core on Trinity are about 11% higher than a core of Cielo and about 22% lower than a core of Chama



Ping Pong Latency
Cielo
Chama
Trinity (Intra-Group)
1000
100
10
1
0.1
Latency (us)
0
1
2
4
8
16
32
64
128
256
512
1024
2048
4096
8192
16384
32768
65536
131072
262144
524288
1048576
2097152
4194304
Message Size (Bytes)


Ping Pong Bandwidth
Cielo
Chama
Trinity (Intra-Group)
16384
8192
4096
2048
1024
512
256
128
64
32
16
8
4
2
1
0.5
1
2
4
8
16
32
64
128
256
512
1024
2048
4096
8192
16384
32768
65536
131072
262144
524288
1048576
2097152
4194304
Message Size (Bytes)


MPI 256 rank Allreduce
Cielo
Chama
Trinity (Intra-Group)
10000
1000
100
10
1
4
8
16
32
64
128
256
512
1024
2048
4096
8192
16384
32768
65536
131072
262144
524288
1048576
2097152
4194304
Message Size (Bytes)

Comments on MPI Benchmarks

- Chama shows better Ping Pong Latency than Cielo and Trinity for small messages, but for large messages, Trinity has better latency

- Likewise, Chama has better Ping Pong Bandwidth for small messages and Trinity has more than twice the Bandwidth for large messages

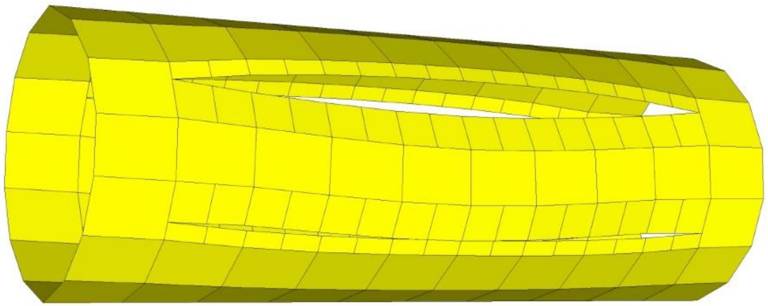
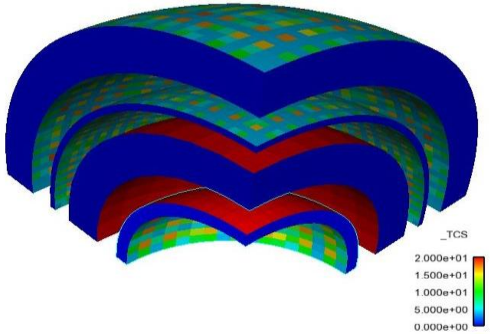
- Allreduce operations are 2 to 10 times faster on 256 ranks of Trinity

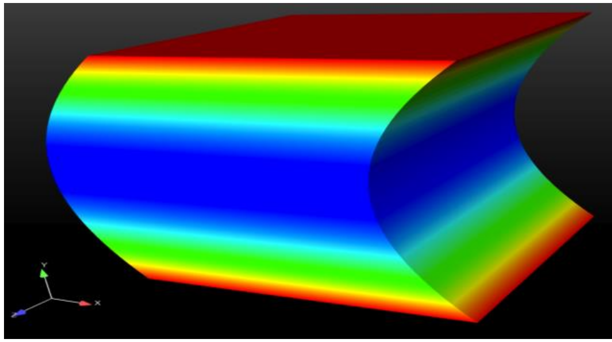
Focus Codes

- Focus on Production SIERRA applications
  - SIERRA/Solid Mechanics (SM)
  - SIERRA/Aerodynamics
  - SIERRA/Structural Dynamics (SD)

- SIERRA is a large C++ framework
  - provides framework for several codes
  - Includes several Third Party Libraries
  - Contains common C++ classes and methods
  - Common infrastructure for parallel codes

SIERRA/SM (Solid Mechanics)

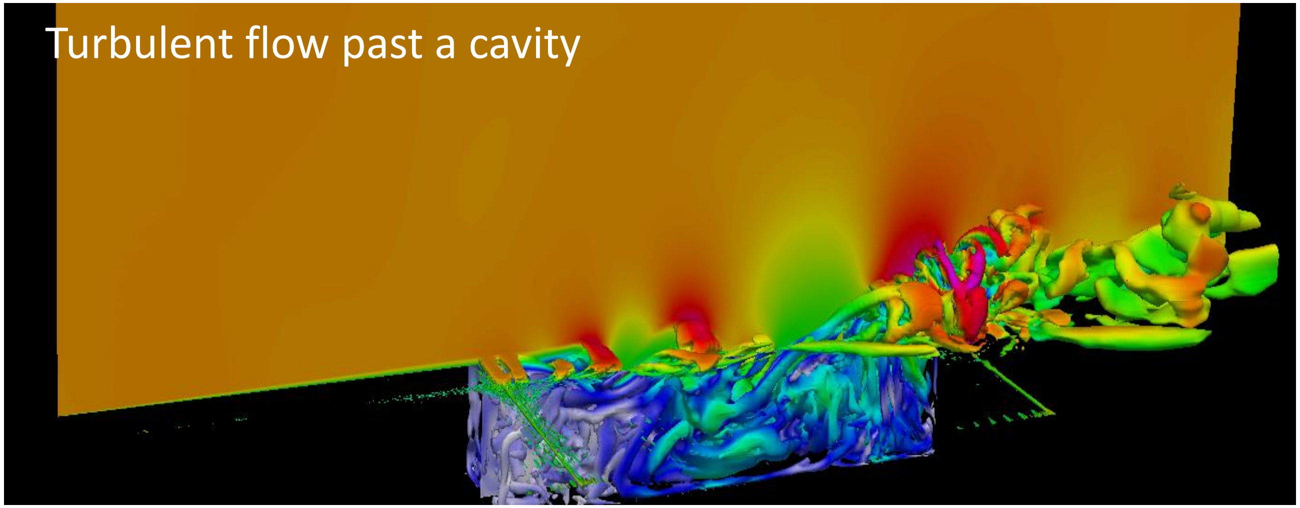





- A general purpose massively parallel nonlinear solid mechanics finite element code for explicit transient dynamics, implicit transient dynamics and quasi-statics analysis.
- Built upon extensive material, element, contact and solver libraries for analyzing challenging nonlinear mechanics problems for normal, abnormal, and hostile environments.
- Similar to LSDyna or Abaqus commercial software systems.

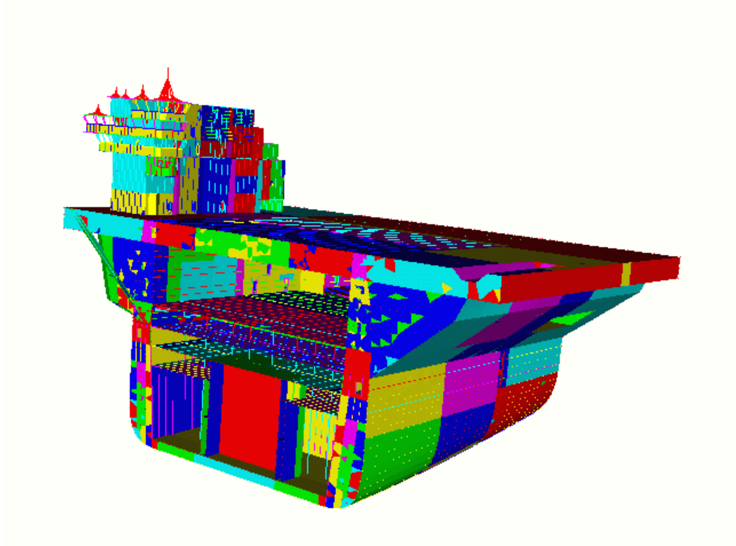
Summary of Sierra/Aero

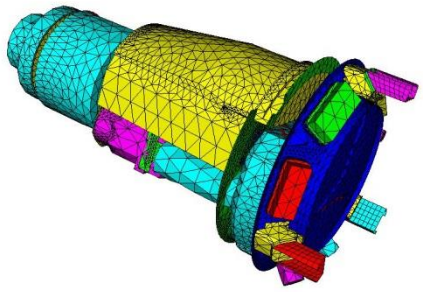
- Unstructured meshes
- One and two equation turbulence models
- LES and Hybrid RANS
- Uses either FETI or Trilinos for sparse matrix operations and solvers.
- Assembly is substantial portion of the computational cost.



SIERRA/SD Domain Areas

- General Structural Dynamics, Finite Elements
  - Vibrations, normal modes, implicitly integrated transient dynamics, frequency response analysis
  - Shells, Solids, Beams, Point Masses
  - Complicated Large Structures
  - Typically many constraint equations
- Acoustics and Structural Acoustics
  - Even larger systems
  - More constraints
  - Infinite Elements (nonsymmetric)
- Optimization, UQ and Inverse Methods
  - Adjoint methods
  - Material and Parameter inversion
  - Verification and Validation



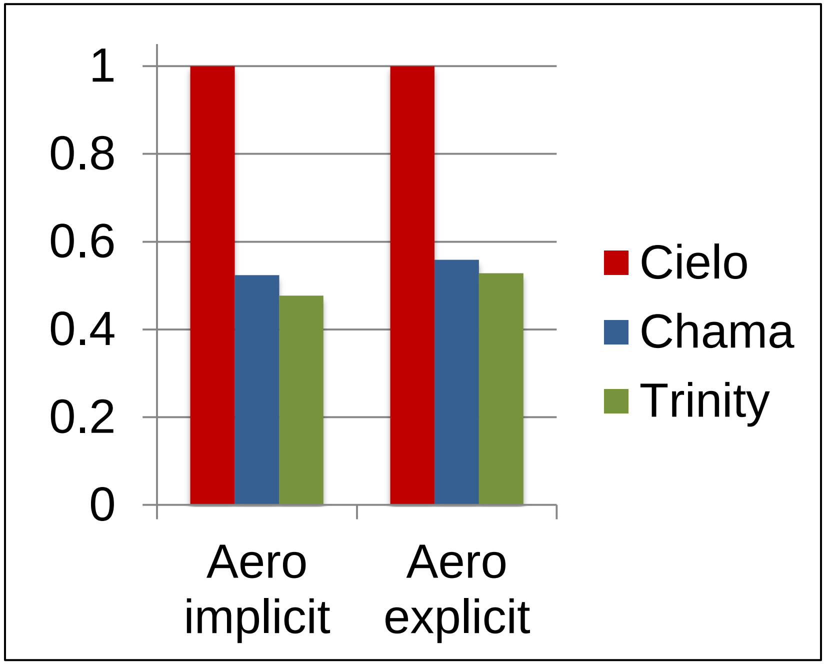


Code Characteristics

- SIERRA/SM extensively uses sparse direct solvers
  - the iterative solve requires a local solve, coarse solve, and a preconditioner
  - The preconditioning step dominates the cost (>90%).
- SIERRA/Aero uses Trilinos solvers
  - GMRES for solver with Symmetric Gauss-Seidel preconditioner
- SIERRA/SD uses Domain Decomposition solver for eigensolve
  - About 84% of non-MPI time spent in solver which makes use of BLAS routines

Aero Times (in seconds)

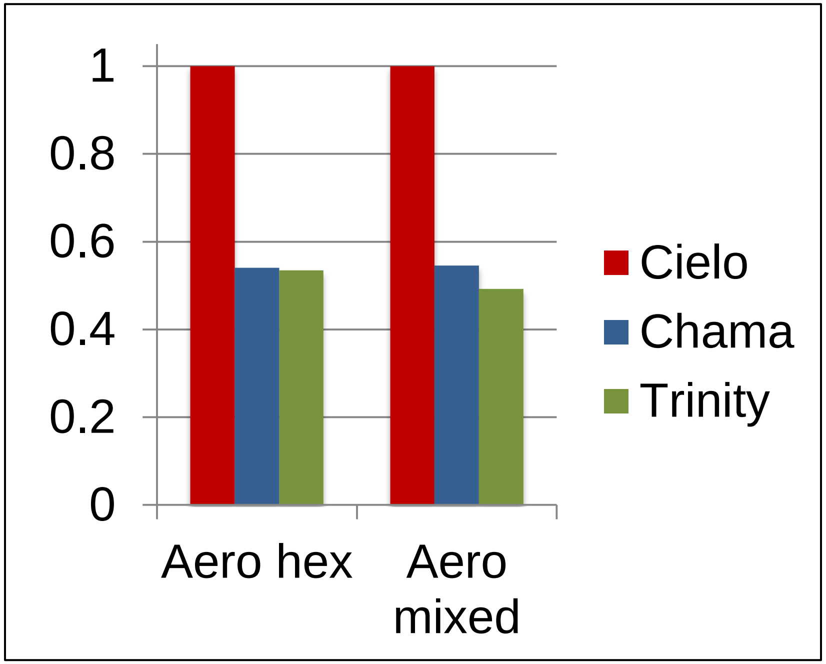
Code/problem	Cores	Cielo	Chama	Trinity
Aero implicit	128	1834.0	961.2	874.4
Aero explicit	128	527.0	294.6	278.2



- Both Chama and Trinity are about twice as fast as Cielo
- For explicit problem, compute time is similar, but MPI time is difference between Chama and Trinity
- For implicit problem, Trinity is 10% faster than Chama for compute and 14% to 20% faster for MPI time

Aero Times (in seconds)

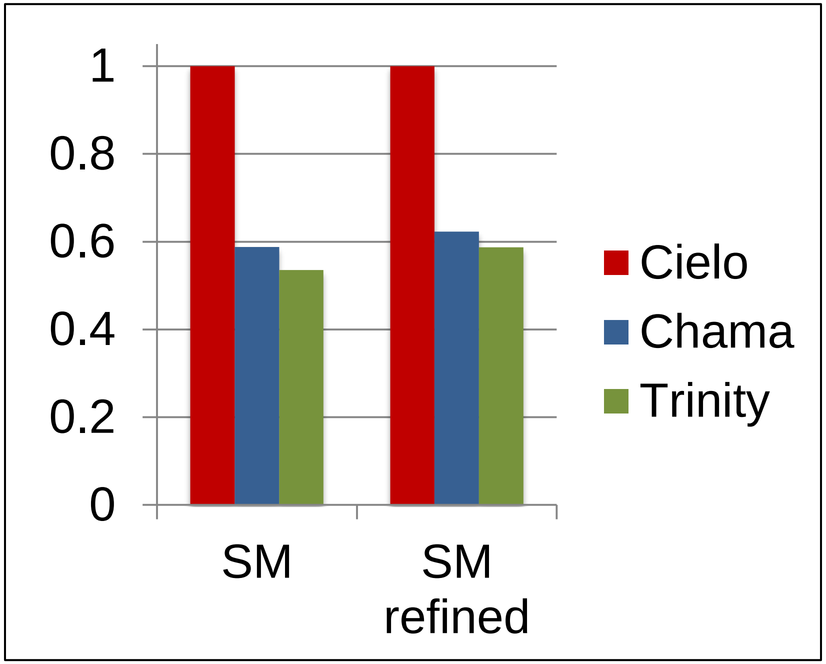
Code/problem	Cores	Cielo	Chama	Trinity
Aero hex	512	658.2	355.9	351.9
Aero mixed	512	390.7	213.3	192.4



- Both Chama and Trinity are about twice as fast as Cielo
- For both problems, compute time is similar on Chama and Trinity and MPI time is about 20% less on Trinity than on Chama
  - MPI time is about 20% for the hex problem and 55% for the mixed

SM Times (in seconds)

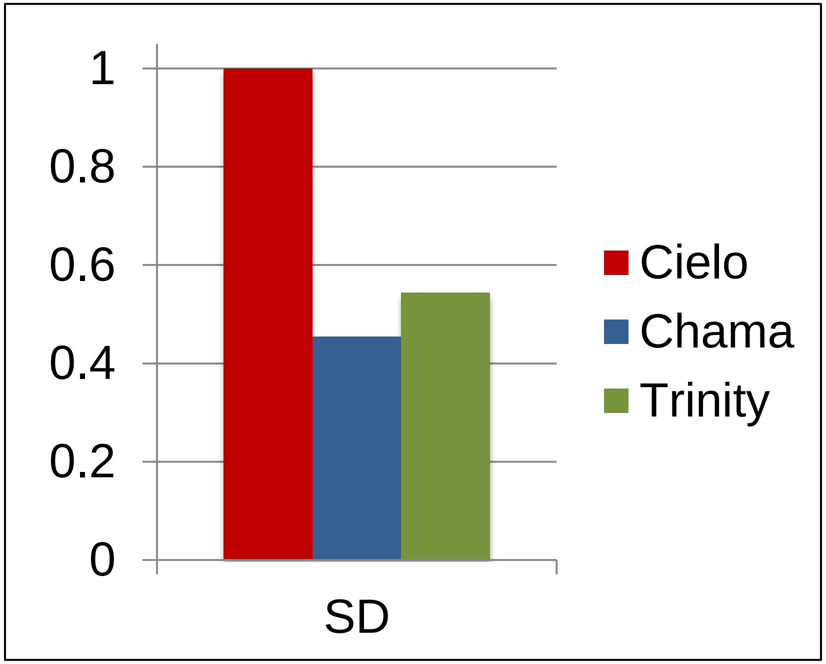
Code/problem	Cores	Cielo	Chama	Trinity
SM	16	1118.3	657.5	598.4
SM refined	128	2332.1	1452.1	1369.1



- Both Chama and Trinity are about 1.7 times as fast as Cielo
- For refined problem, MPI takes about 55% of the time
- On Trinity and Chama, the compute time is similar, but the MPI time is larger on Chama than Trinity
    - Lots of small to medium sized messages

SD Times (in seconds)

Code/problem	Cores	Cielo	Chama	Trinity
SD	120	993.0	451.0	540.0

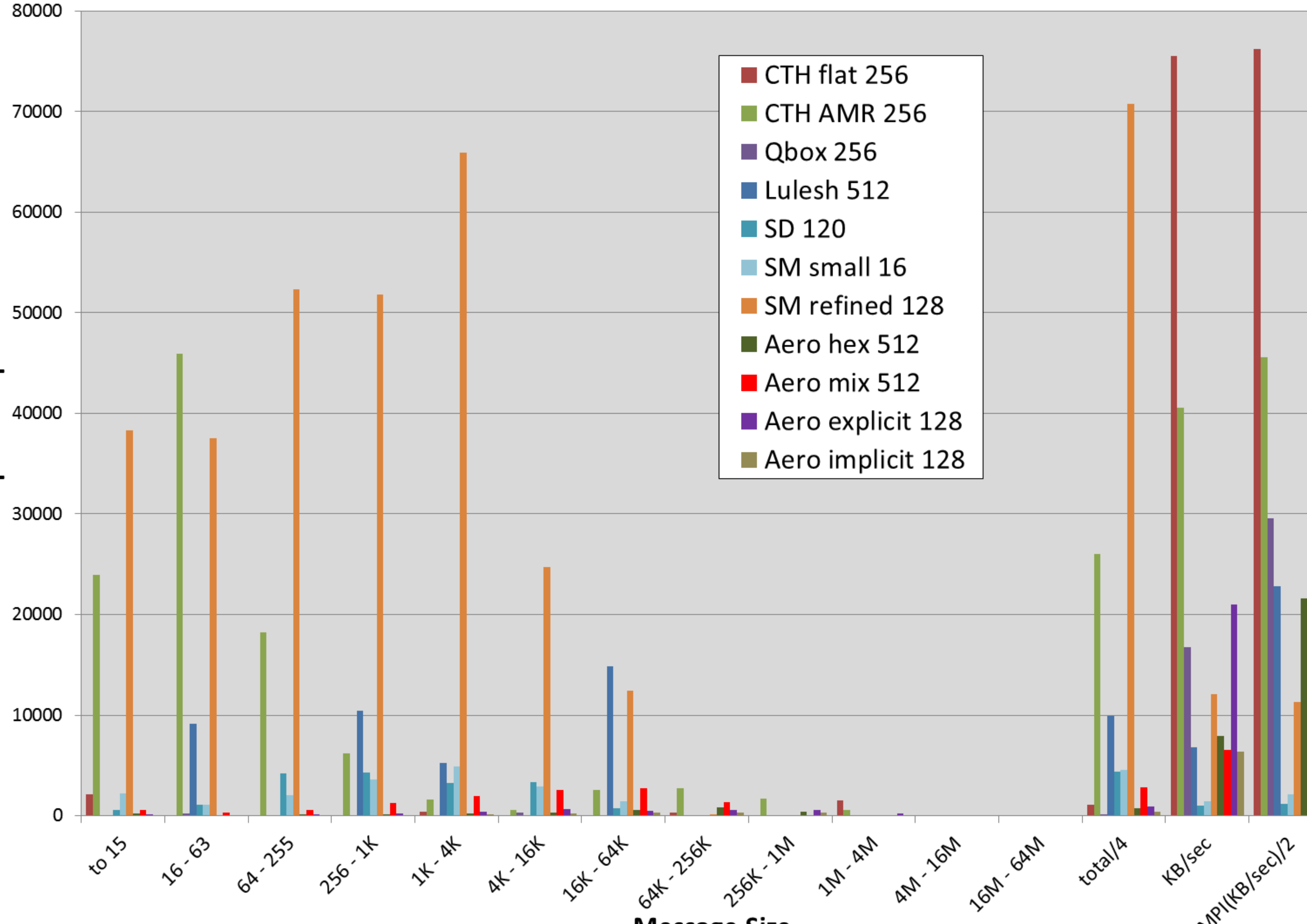


- Both Chama and Trinity are about twice as fast as Cielo
- MPI and compute times are larger on Trinity than Chama, but the MPI time includes wait times, which is an indication of load imbalance
  - Most of compute time is spent in a solver which has a large number of DGEMM calls

Comments on Performance

- All of the codes show Trinity faster than Cielo by a factor of about 2 on half as many nodes
  - Effectively a factor of four per node
- Most of the codes show Trinity 5% to 12% faster than Chama on half as many nodes
  - About a factor of two per node
- Sierra/SD about 18% slower on Trinity than on Chama
  - Slightly worse than the difference in clock speeds

Message sizes on Trinity



Summary

- We ported three production codes to our new capability machine and compared its performance to our current machines
  - We got about a factor of four performance improvement per node over our current capability machine and a factor of two over our current capacity machine

- We are investigating the performance of SIERRA/SD

- We have started working on the Knight's Landing portion of the machine

Sandia National Laboratories

Exceptional service in the national interest