SAND2007-0564P

# Status of 7X Applications: ASCI Red vs. Red Storm

Robert A. Ballance, Karen Haskell,
John P. Noe, Joel O. Stevenson
Sandia National Laboratories

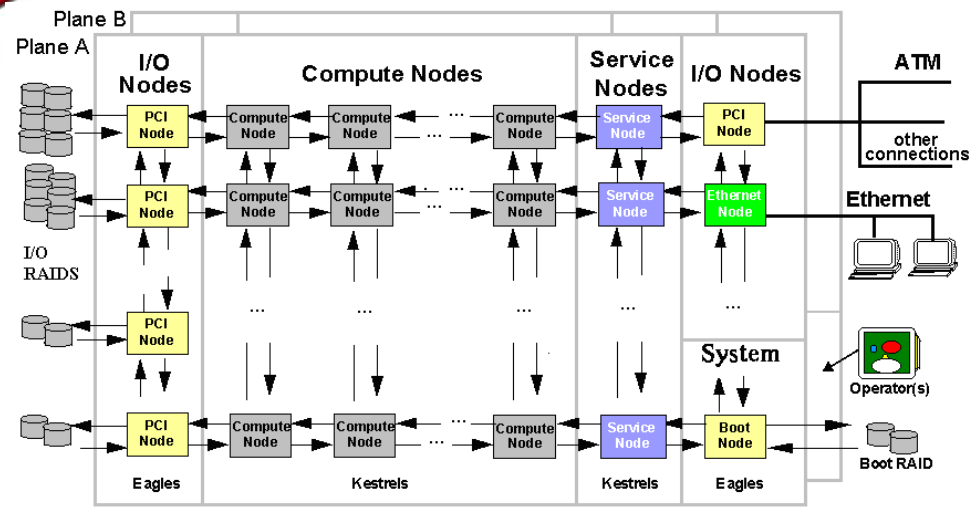
Red Storm Quarterly, Feb. 8, 2007

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

# Outline of Today's Discussion

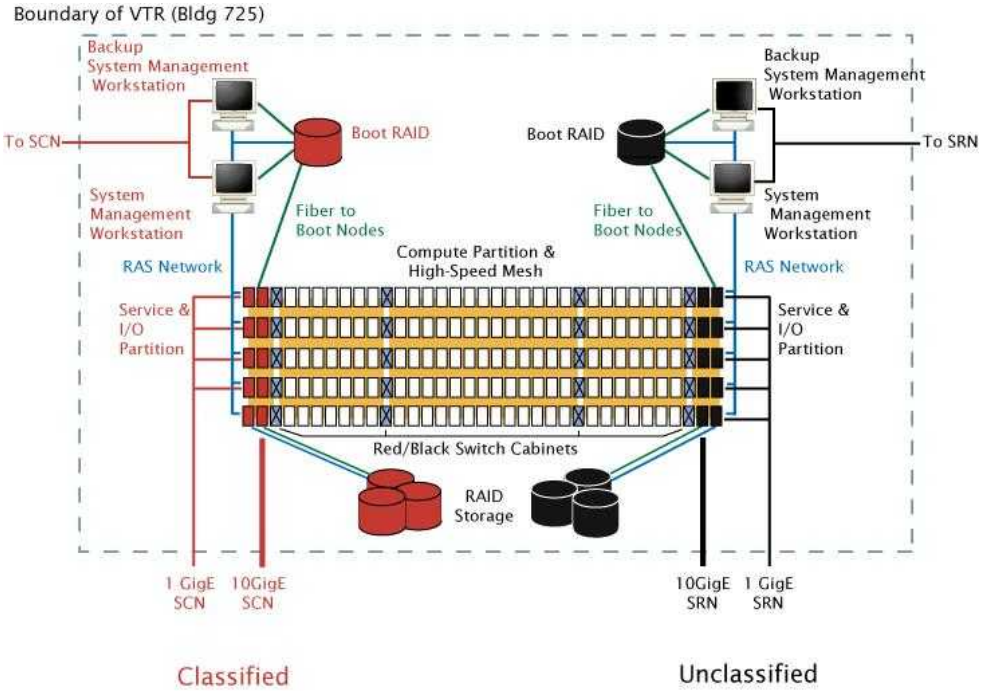
- The 7X performance suite consists of ten applications and benchmarks used in Red Storm performance testing.
- Identify one or more problems for each application, run those problems at two or three processor sizes, and compare the results between ASCI Red and Red Storm - 25 cases under study.
- Goal of 7X performance testing is to assure Sandia, Cray, and DOE that Red Storm will achieve its performance requirements.

# ASCI Red




1168 nodes on the unclassified side and 1166 nodes on the classified side. The middle section contains 2176 nodes. Total number of compute nodes at 4510. Each compute node contained 2 processors, bringing total processor count to 9020.

# Red Storm




5th row added bringing node count to 3360 on the unclassified side and 3360 on the classified side. The middle section contains 6240 nodes. Total number of compute nodes at 12960. Each compute node upgraded to dual core technology, bringing total processor count to 25920.

# Yesterday and Today

| | ASCI Red | Red Storm (initial op) | Red Storm (post-upgrade) |
|---|---|---|---|
| Architecture | Dist. Memory MIMD | Dist. Memory MIMD | Dist. Memory MIMD |
| Theoretical Peak Performance | 3.15 TF | 41.47 TF | 124.42 TF |
| MP-Linpack Performance | 2.38 TF | 36.19 TF | 101.4 TF |
| Total Memory | 1.21 TB | 30.95 TB | 75.9 TB |
| System Memory B/W | 2.5 TB/s | 55 TB/s | 82.9 TB/s |
| Disk Storage (Total / per Color) | 12.5 TB / 6.25 TB | 320 TB/160TB | 320 TB/160TB (Increasing) |
| Parallel File System B/W (Total / per Color) | 2.0 GB/s / 1.0 GB/s | 100 GB/s / 50 GB/s sustained disk transfer rate | 100 GB/s / 50 GB/s sustained disk transfer rate |
| External Network B/W (Total / per Color) | 0.4 GB/s / 0.2 GB/s | 50 GB/s / 25 GB/s | 50 GB/s / 25 GB/s |

# Yesterday and Today (cont.)

| | ASCI Red | Red Storm (initial op) | Red Storm (post-upgrade) |
|---|---|---|---|
| Compute Nodes (Red/Center/Black) | 4510 (1166/2176/1168) | 10368 (2688/4992/2688) | 12960 (3360/6240/3360) |
| Compute Processors (Red/Center/Black) | 9020 (2332/4352/2336) PII Xeon 333Mhz | 10368 (2688/4992/2688) Opteron 2.0Ghz | 25920 (6720/12480/6720) Opteron Dual Core 2.4Ghz |
| Service Nodes (Red/Black)<br>Disk I/O Nodes (Red/Black) | 52 (26/26)<br>73 (37/36) | 512 (256/256) Service and I/O partition (login, service, I/O, administrative nodes) | 640 (320/320) Service and I/O partition (login, service, I/O, administrative nodes) |
| System Nodes (Red/Black) | 2 (1/1) | RAS and System Management partition | RAS and System Management partition |
| Network Nodes (Red/Black) | 12 (6/6) Ethernet ATM | 100 (50/50) 10GigE to RoSE<br>20 (10/10) 1GigE to lnodes | 100 (50/50) 10GigE to RoSE<br>20 (10/10) 1GigE to lnodes |
| Number of Cabinets (Computer/Switch /Disk) | 104 (76/8/20) | 140 (108 compute/16 switch/16 service and I/O) | 175 (135 compute/20 switch/20 service and I/O) |
| Interconnect Topology | 3-D Mesh (x,y,z)<br>38 x 32 x 2 | 3-D Mesh (x,y,z)<br>27 x 16 x 24 | 3-D Mesh (x,y,z)<br>27 x 20 x 24 |

# Yesterday and Today (cont.)

| | ASCI Red | Red Storm (post-upgrade) |
|---|---|---|
| Interconnect Bandwidth | | |
| MPI Latency | 15 $\mu$s 1 hop, 20 $\mu$s max | ~4 $\mu$s 1 hop, ~7 $\mu$s max |
| Bi-Directional Link B/W | 800 MB/s | 7.68 GB/s |
| Minimum Bi-Section B/W | 51.2 GB/s | 3.6 TB/s |
| Full System RAS | | |
| RAS Network | 10 Mb Ethernet | 100 Mb and 1 Gb Ethernet |
| RAS Processors | 1 for each 32 CPUs | 1 for each 4 CPUs |
| Operating System | | |
| Compute Nodes | Cougar | Catamount |
| Service and I/O Nodes | TOS (OSFI) | Linux |
| RAS Nodes | VX-Works | Linux |
| Red Black Switching | | |
| Switches | 2/row | 4/row |

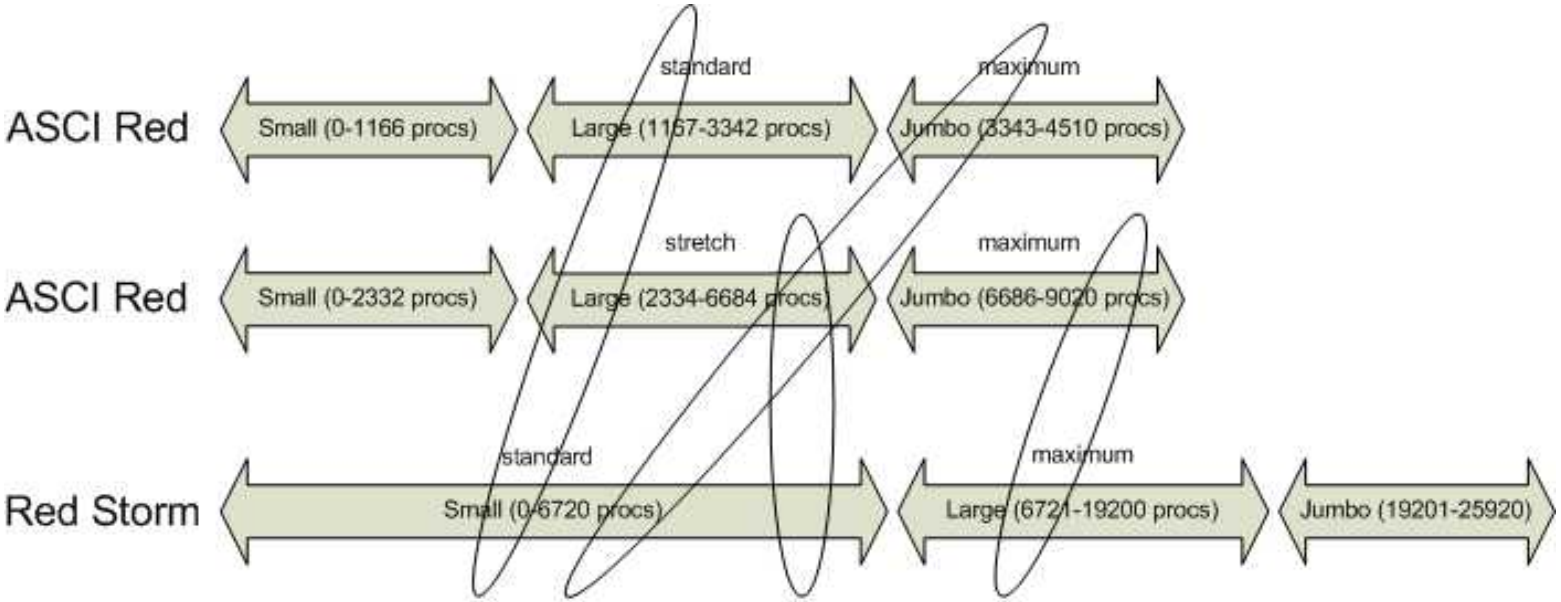
# Application Selection Criteria

- Problem sets shall be “real”.  The 7X testing effort represents production job behavior with actual input files and algorithms.
- The same calculations shall be run on ASCI Red and Red Storm. The primary metric is wall-clock time as measured by the elapsed time to execute the entire job script, including any pre and post processing.
- Calculations on ASCI Red and Red Storm should give equivalent answers.
- Problems should be chosen to use as many ASCI Red resources (processor, memory) as possible in order to place reasonable stress on Red Storm.
- Jobs run on ASCI Red should range from 3.5-8 hours.

# Application Selection Criteria (cont.)

- Simplified geometries are preferred in order to simplify input file creation and to avoid meshing problems during benchmarking.
- All applications should use standard production-use capabilities including I/O, checkpoint/restart, and visualization files.
- When an application can be run using alternative algorithms, such as Alegra with and without contact, that application may have more than one benchmark problem in the suite.
- We will test applications in three modes : standard, stretch, maximum.

# Modes: Standard, Stretch, Maximum




ASCI Red: 4510 compute nodes (9020 processors). Proc 0 mode uses one processor per node and the full 256 MB of memory. Proc 3 mode uses two processors per node but only 128 MB of memory is available to each process.

Red Storm: Upgrading to 12960 compute nodes (25920 processors). Each node is dual core topology with minimum 2 GB of memory per node. Memory available to each process is halved when using two processors per node.

# Application Descriptions

- Alegra with Contact
  - Quasistatic electromechanics (QSEM) problem in which a curved impactor depoles a potted active ceramic element.
    - Standard - 2048 processors
    - Stretch - 6484 processors
- Alegra without Contact
  - QSEM problem identical to the contact problem except the boundary condition is a prescribed displacement rather than an impactor, eliminating the need for contact.
    - Standard - 2048 processors
    - Stretch - 6484 processors
- CTH
  - Shock physics (3D of a large conical shaped charge).
    - Standard - 2000 processors
    - Stretch - 6480 processors
    - Maximum - 9000 processors

# Application Descriptions (cont.)

- ITS
  - Monte Carlo solution of linear time-independent coupled electron/photon radiation transport problems, with or without the presence of macroscopic electric and magnetic fields of arbitrary spatial dependence.
    - Standard - 3200 processors
    - Maximum - 4500 processors
    - Stretch - 6500 processors
    - Maximum - 9000 processors
- PARTISN
  - Sntiming problem - flux and eigenvalue convergence as monitored by partisn (Parallel Time-dependant SN transport).
    - Maximum - 4096 processors
    - Stretch - 6480 processors
    - Maximum - 8930 processors

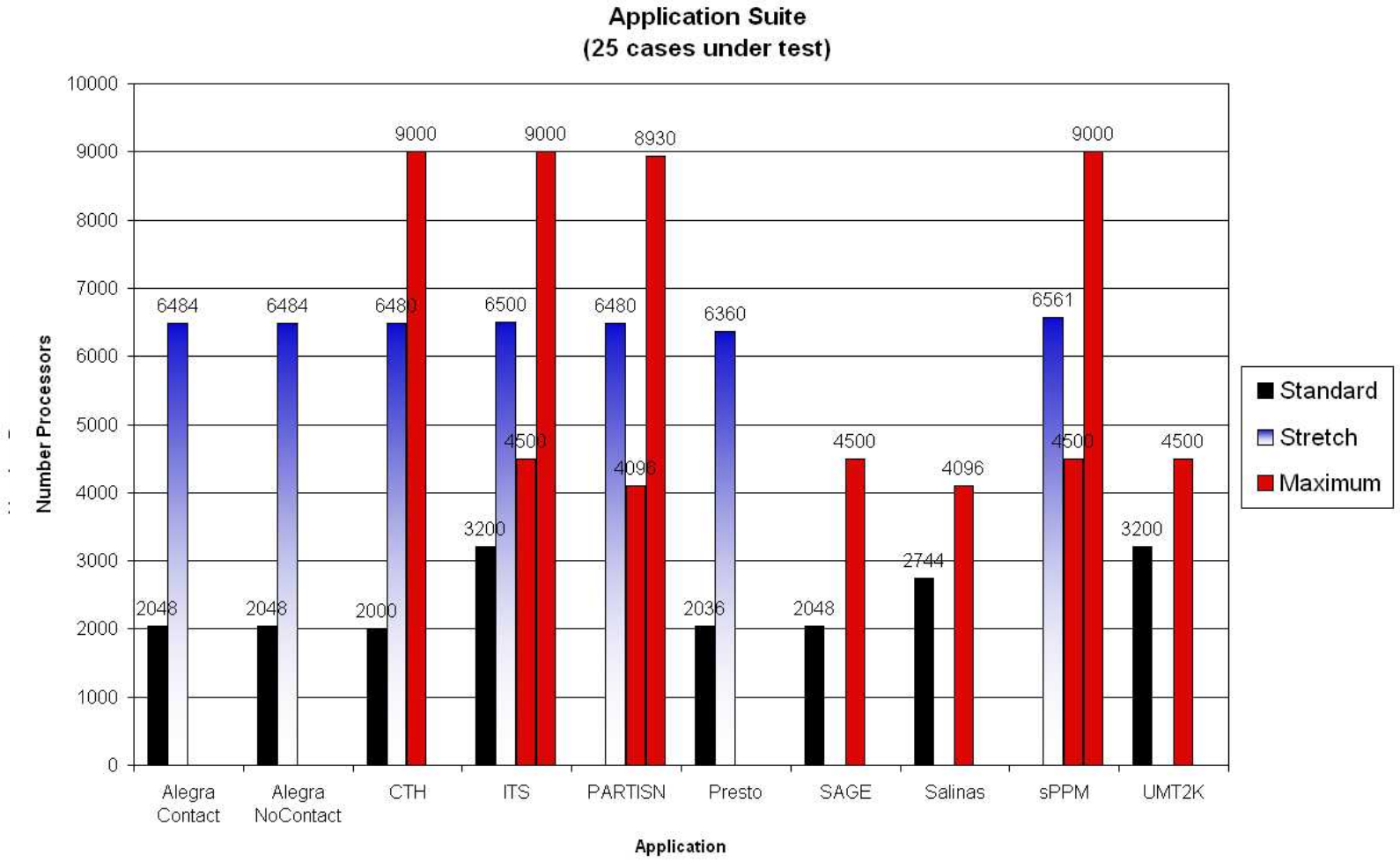
# Application Descriptions (cont.)

- Presto
  - Rectangular bricks stacked in an alternating fashion in a plane to produce a wall which is three elements thick. Four of these walls are lined up in the thin direction. The walls are then given a sudden pressure loading such that they compress against each other.
    - Standard - 2036 processors
    - Stretch - 6360 processors
- SAGE
  - Asteroids simulation - 45 degree (Cx45j), 3D, granite asteroid impact into stratified medium of water, calcite, granite crust, and mantle.
    - Standard - 2048 processors
    - Maximum - 4500 processors
- Salinas
  - Transient dynamics problem – one unit cube model.
    - Standard - 2744 processors
    - Maximum - 4096 processors

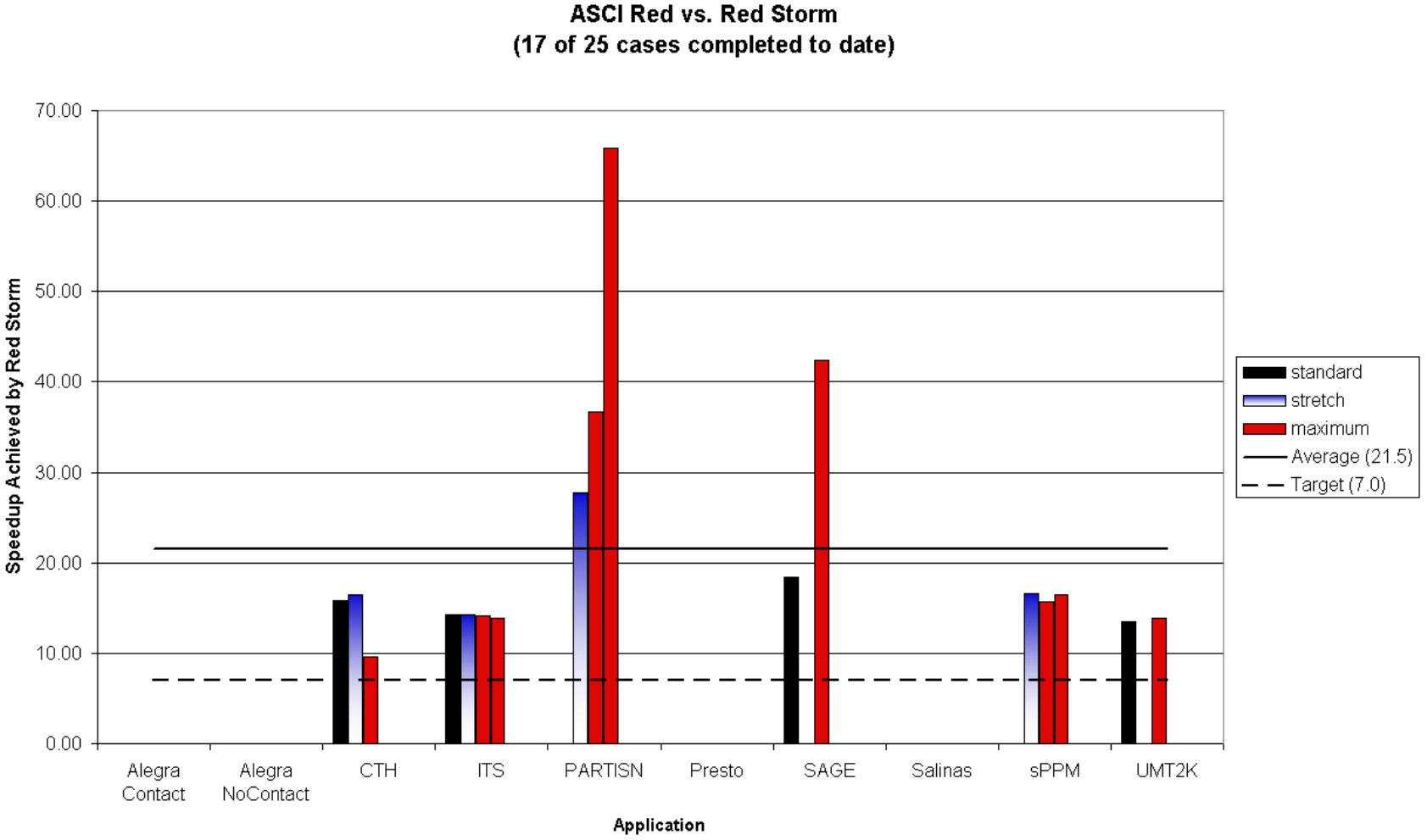
# Application Descriptions (cont.)

- sPPM
  - Shock physics - solves a 3D gas dynamics problem on a uniform Cartesian mesh, using a simplified version of 3D hydrodynamics code Piecewise Parabolic Method.
    - Maximum - 4500 processors
    - Stretch - 6561 processors
    - Maximum - 9000 processors
- UMT2000
  - 3D, deterministic, multigroup, photon transport code for unstructured meshes.
    - Standard - 3200 processors
    - Maximum - 4500 processors

# Selected Applications/Benchmarks



# How Much Faster is Red Storm?



# Summary

- The 7X test suite consisted of 10 applications and benchmarks that were used in Red Storm performance testing.
- One or more problems were identified for each application, and those problems were run at two or three processor sizes, comparing the results between ASCI Red and Red Storm - 25 data points under study - 17 completed to date.
- Red Storm has achieved its performance requirements of 7X performance over Red Storm, posting an average speed-up of 21.5X.

# Acknowledgments


- The authors thank Courtenay Vaughan, Bob Benner, John Van Dyke, Sue Goudy, and Mahesh Rajan for their assistance with compiling, configuring, and troubleshooting on ASCI Red and Red Storm. Many thanks also to the ASCI Red (Frank Jaramillo, Paul Sanchez, Mike Martinez, Sean Taylor) and Red Storm system administrators and support staff for their assistance. Thanks also to Mark Hamilton (1543) for assistance in setting up the Sourceforge repository.
- The authors thank Cray Research engineers Paul Burkhardt, Doug Enright, Ron Pfaff, and Mike Davis for their assistance with compiling and optimizing the application codes for Red Storm benchmark runs.
- Sue Goudy, Sue Kelly, Mike McGlaun, Jim Tompkins, and Courtenay Vaughan have all provided help, suggestions, and guidance as the predecessor to this report was assembled (The 7X Cookbook).
- We also thank the application code developers for their assistance: Brian Franke (ITS), Garth Reese (Salinas), Galen Gizler (SAGE), John Daly (SAGE), Kevin Brown (Presto), Arne Gullerud (Presto), Bruce Bainbridge (Calore), Tolulope Okusanya (Calore), Allen Robinson (Alegra), Rich Drake (Alegra), and Josh Robbins (Alegra).