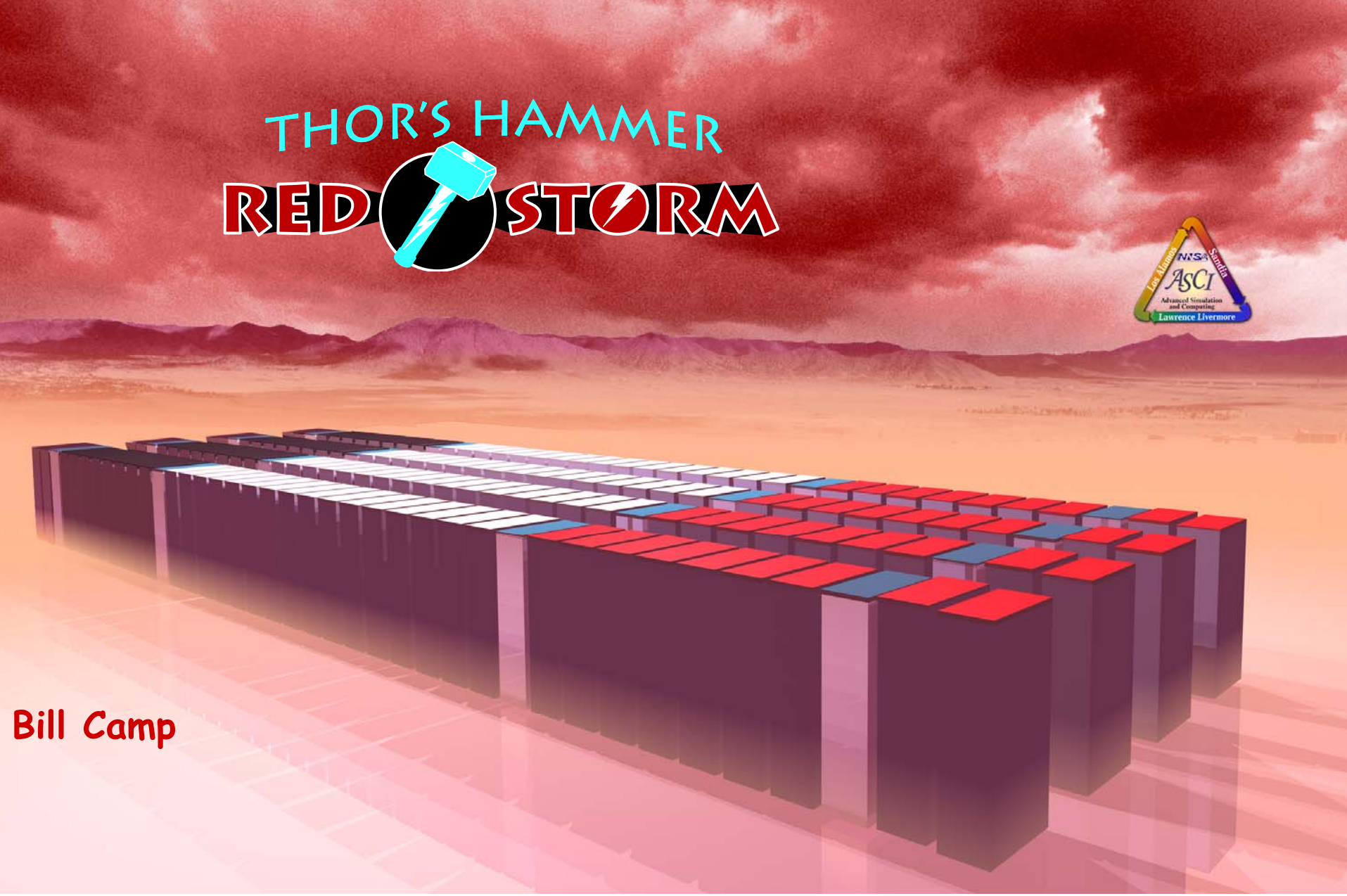
THOR'S HAMMER
RED STORM



Bill Camp

NNSA National Nuclear Security Administration

Sandia National Laboratories

THOR'S HAMMER
RED STORM

Challenges Facing the HPC Community

CUG May 16, 2005

William J. Camp
CCIM, Sandia National Laboratories
Albuquerque, NM
bill@sandia.gov

NNSA National Nuclear Security Administration

Sandia National Laboratories

Challenges to Custom System Providers

- The great base of applications perform adequately on workstations
- Most ISV's have not invested in truly scalable implementations of their products
- Customers speak with many voices
- Each of them considers themselves to be "bellweathers"
- The high end market has not really grown in 10 years
- Most designers have never carried out a real application task
- The cost of innovation is becoming unreasonable for all but one or two "giants."
- Clusters really do work well for low-end and even some mid-range applications mixes.
- Clusters are cheap to buy and to own.

The great base of applications perform adequately on workstations

- What was huge problem even 10 years ago is a desktop application today.
  - E.g., for the most part, materials scientists want to run lots of relatively small calculations
- Product optimization is typically a minimax search procedure that is run as many relatively small jobs
- Most scientists & engineers are comfortable writing serial codes to solve their problems;
  - they are not computational scientists and don't want to parallelize unless they have to
- The commercial codes available to them are either serial or support low levels of parallelism.

Most ISV's have not needed to or been able to afford to invest in truly scalable implementations of their products

- Their business model is often based on "per-seat" fees, not per processor
- They have a huge investment in their serial codes
  - Robustness and Ease of Use are more important to their customers than performance
- Often their codes were designed by applications-area experts, not HPC experts
  - For the most part, they would need major restructuring for scalability
  - Unfortunately the incremental ROI is not sufficient to justify the investment
- Large customers are not willing to pay "scalable" fees-- e.g. Platform
- The high end market has not really grown in 10 years
- They worry about betting on the wrong programming model

THOR'S HAMMER
RED STORM

Customers speak with many voices

- Sandia represents the extreme high-end, "distributed-memory" end of the community
- By contrast, NASA wants large shared memory nodes
- Some customers are convinced that they have to have vectors
  - Others cite the high cost of vectors and shun them
  - SOME WANT BOTH!
- Some customers value ease of programming over performance
  - Others are committed to the high-end and comfortable with MPI
- Some want OpenMP or other thread-based shared memory-like programming tools
  - Others cite their poor scalability as a reason to avoid them
- Some like fat-tree or other multistage networks for their "flatness"
  - Others prefer meshes for their cost advantage at very large scale and higher local bandwidth per dollar
- Customers tend to favor either MPI or LOAD-STORE depending on their application mix.

Each of them considers themselves to be "bellweathers"

- Everybody wants the prestige of owning "serial # 1"
  - Serial # 2 is "okay" unless your biggest competitor bought #1.
- We all believe that our colleagues from other sites are nice bright folks…"but they really don't understand the market."
- The feature that "Mike" just has to have may make the system uninteresting to "John."
- Some customers never buy anything big but like high-tech toys and are willing to try to shape the market for you.
- Every deal sets a price precedent for the next deal.

The high end market has not really grown in 10 years

- In 1993, the market was about $2.5B and Cray had about 40% of it.
- SGI was losing visualization market share to Intel and AMD desktops with Graphics cards
- Their response was to move into the server market and take market share from Cray
- By 1996 they had bought Cray and began a downward spiral.
- IBM was quietly taking market share from everyone else and doing it without a true supercomputer!
- Today's market is not any bigger than it was in 1993
  - It might even be smaller in real dollars

- The auto industry basically no longer uses custom supercomputers
- Aerospace and the Oil industry are heading in the same direction

THOR'S HAMMER
RED STORM

Innovation is becoming unaffordable for all but "giants."

- Dataflow really would be interesting but only if IBM, Intel and/or AMD adopt it.
- Companies like Cray, SGI, and even H-P cannot afford to develop new processor designs
- If you break a widely-used programming paradigm, you had better offer orders of magnitude increases in perceived cost-performance
  - And even then be prepared to be in it for the long haul.
- The tyranny of vectors held back MPPs
- The tyranny of MPI will hold back new PIM-based programming models
- This is unfortunate but economically makes good sense.
- Inevitably, government needs to invest more on focused R&D investments at the high-end. {HPCS, PathForward, IHEC, HEC-RTF, all need to move ahead and grow}
  - Without involving the big three, IBM, Intel, and AMD, it may not be enough.
- Left to themselves, processor manufacturers are not going to drive high-end appropriate innovation.

Clusters really do work well for low-end and even some mid-range applications mixes.

- The same MPI programming model used on MPP's enables clusters.
- Most applications perform adequately on clusters out to 64--256 processors in a "weak-scaling" mode.
- Clusters work great for parallel search problems as are found in
  - Optimization
  - Uncertainty quantification
  - Parameter identification
  - Mapping out materials phase diagrams
- They do well on nearly all embarrassingly parallel problems
- They are replacing MPPs and vectors in signal processing (e.g. seismic inversion)
- Their basic components are high-volume and reliable enough at modest scale.

THOR'S HAMMER
RED STORM

Clusters are cheap to buy and to own.

- Right now most cluster providers are selling AMD or Intel based clusters for 50 CENTS A MEGAFLOPS (or less).
- At the 256--1024-node level, their other costs of ownership are similar to MPPs
- You can actually quantify the advantage of clusters for a given workload (including where it goes away).

THOR'S HAMMER
RED STORM

Challenges to HPC Users

- Companies keep going away
- IBM is not the (only) answer
- MPI really does provide their only portable, scalable programming model
- They may not like MPI but every day their investment in it grows
- Programming is way too fragile
- It should not be cheaper and easier to test a new part than to simulate its testing reliably
- Computers keep getting more and more unbalanced

Some solutions for Manufacturers

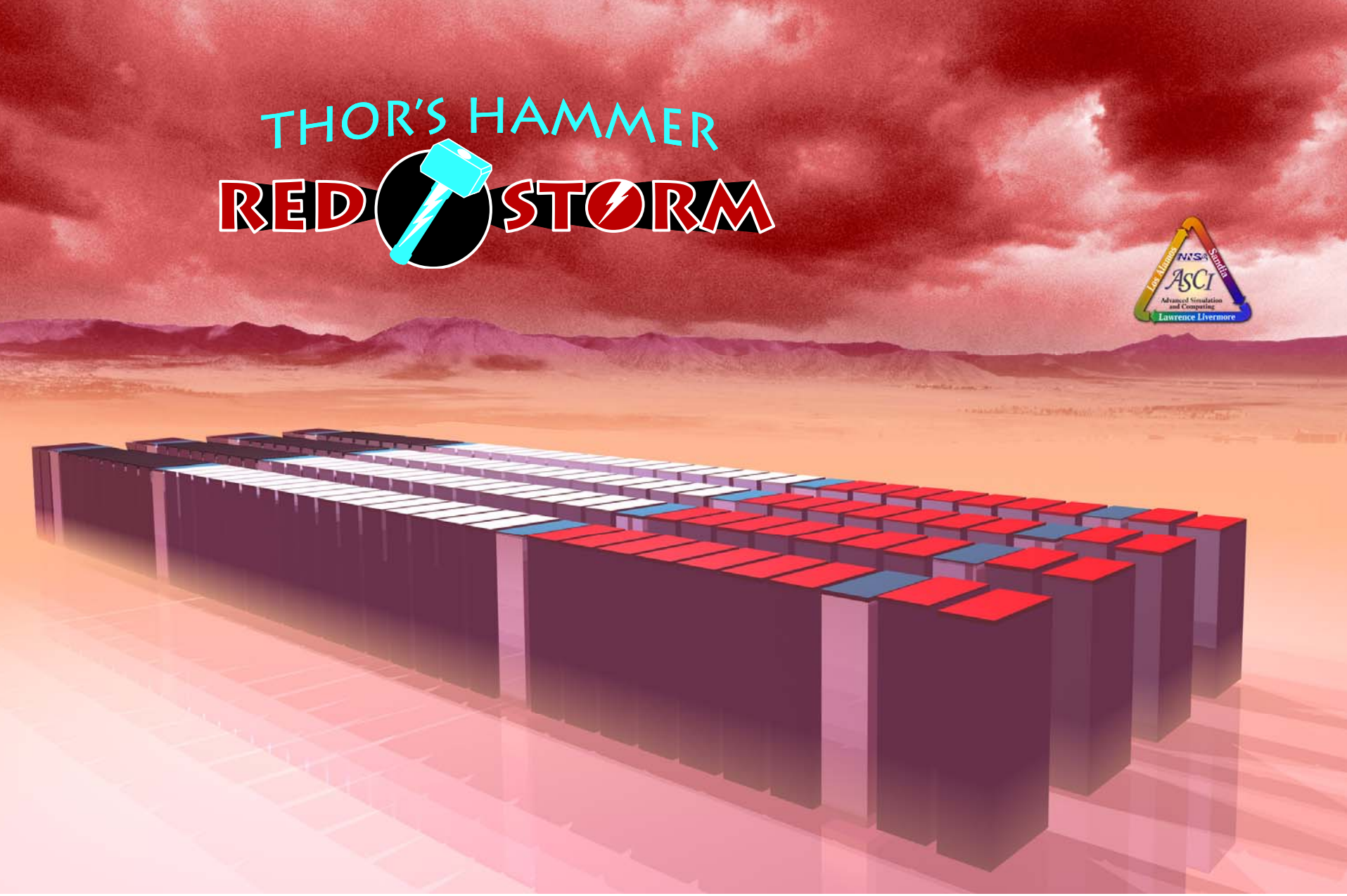
- Vendors need to define their niche and win in it before expanding beyond it.
- Companies need to be market focused rather than customer compelled!
- Vendors like Cray and SGI need to work with their customers to influence Intel and AMD
- Use high-volume, high-reliability, low-cost technology everywhere you can
- Resist the temptation to build something because you can or because it would be “interesting”
  - The two “geeks” who encourage you to do so do not represent people with money to spend
- Execution is more important than “great technology”-- witness IBM
- Use Open-source strategically but sparingly
- Work with your customers to delineate where the point is that you can beat clusters on a cost/performance basis;
  - Make sure the market understands your simple story of how you do that
- Ask the ROI question before you make any investment and have someone other than the proponent evaluate the ROI.
- Invest in bandwidth over processing speed wherever you can. (A blatant attempt at customer compulsion)

Some suggestions for HPC users

- Make intelligent decisions about cost of ownership-- you will win
- Resist tyrannizing your vendor partner over “bells and whistles”
- Think about managing the transition beyond MPI
  - “A stitch in time saves nine”
- Think about your computing environment as a system of systems
- Besides asking how much more quickly your current computer needs can be met,
  - Think about what can you do tomorrow that you cannot do today
  - Quantify the cost and the payoff relative to alternatives

Buy Red Storm!

;>)



NNSA
National Nuclear Security Administration

Sandia National Laboratories