

Titan Introduction and Timeline



Presented by:
Bronson Messer
Scientific Computing Group
Oak Ridge Leadership Computing Facility

Disclaimer: There is no contract with the vendor. This presentation is the OLCF's current vision that is awaiting final approval. Subject to change based on contractual and budget constraints.

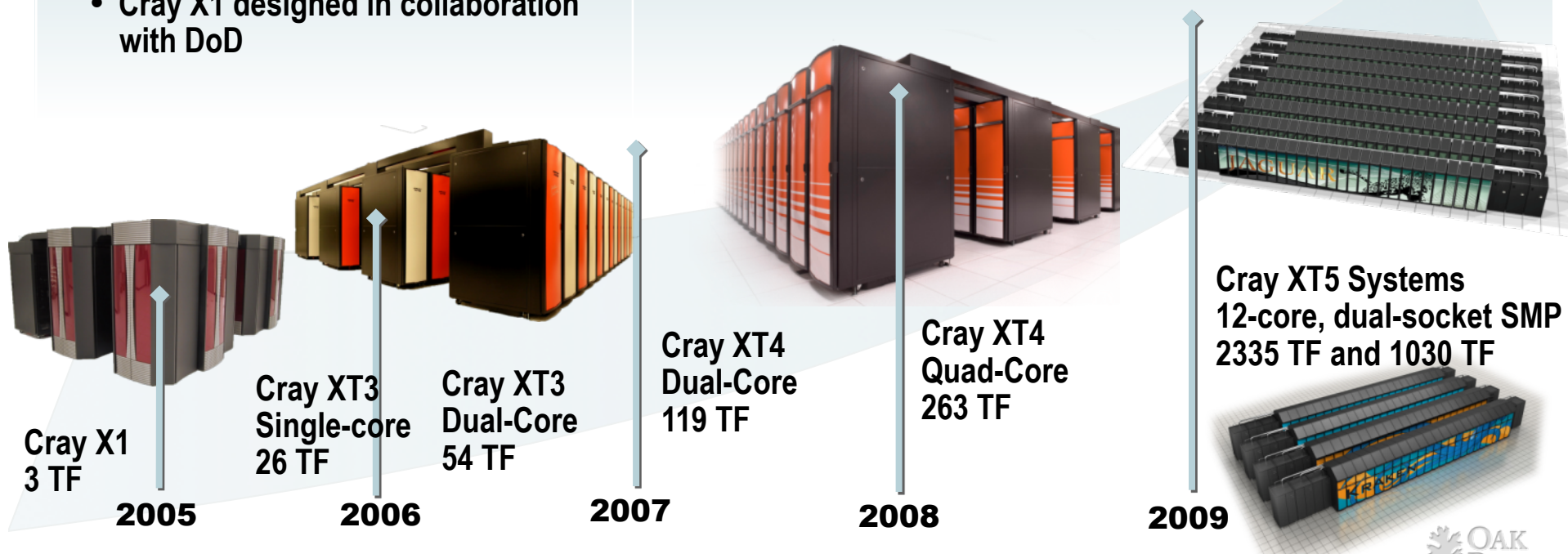
We have increased system performance by 1,000 times since 2004

Hardware scaled from single-core through dual-core to quad-core and dual-socket, 12-core SMP nodes

- NNSA and DoD have funded much of the basic system architecture research
 - Cray XT based on Sandia Red Storm
 - IBM BG designed with Livermore
 - Cray X1 designed in collaboration with DoD

Scaling applications and system software is the biggest challenge

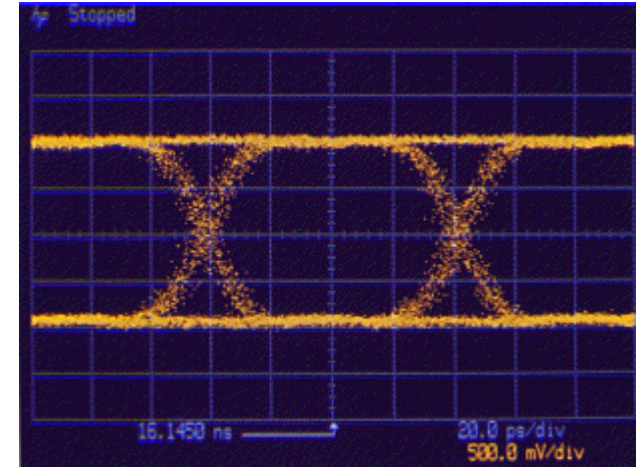
- DOE SciDAC and NSF PetaApps programs are funding scalable application work, advancing many apps
- DOE-SC and NSF have funded much of the library and applied math as well as tools
- Computational Liaisons key to using deployed systems



Why has clock rate scaling ended?

$$\text{Power} = \text{Capacitance} * \text{Frequency} * \text{Voltage}^2 + \text{Leakage}$$

- Traditionally, as Frequency increased, Voltage decreased, keeping the total power in a reasonable range
- But we have run into a wall on voltage
 - As the voltage gets smaller, the difference between a “one” and “zero” gets smaller. Lower voltages mean more errors.
 - While we like to think of electronics as digital devices, inside we use analog voltages to represent digital states.
- Capacitance increases with the complexity of the chip
- Total power dissipation is limited by cooling



Power to move data

$$\text{Energy_to_move_data} = \text{bitrate} * \text{length}^2 / \text{cross_section_area_of_wire}$$

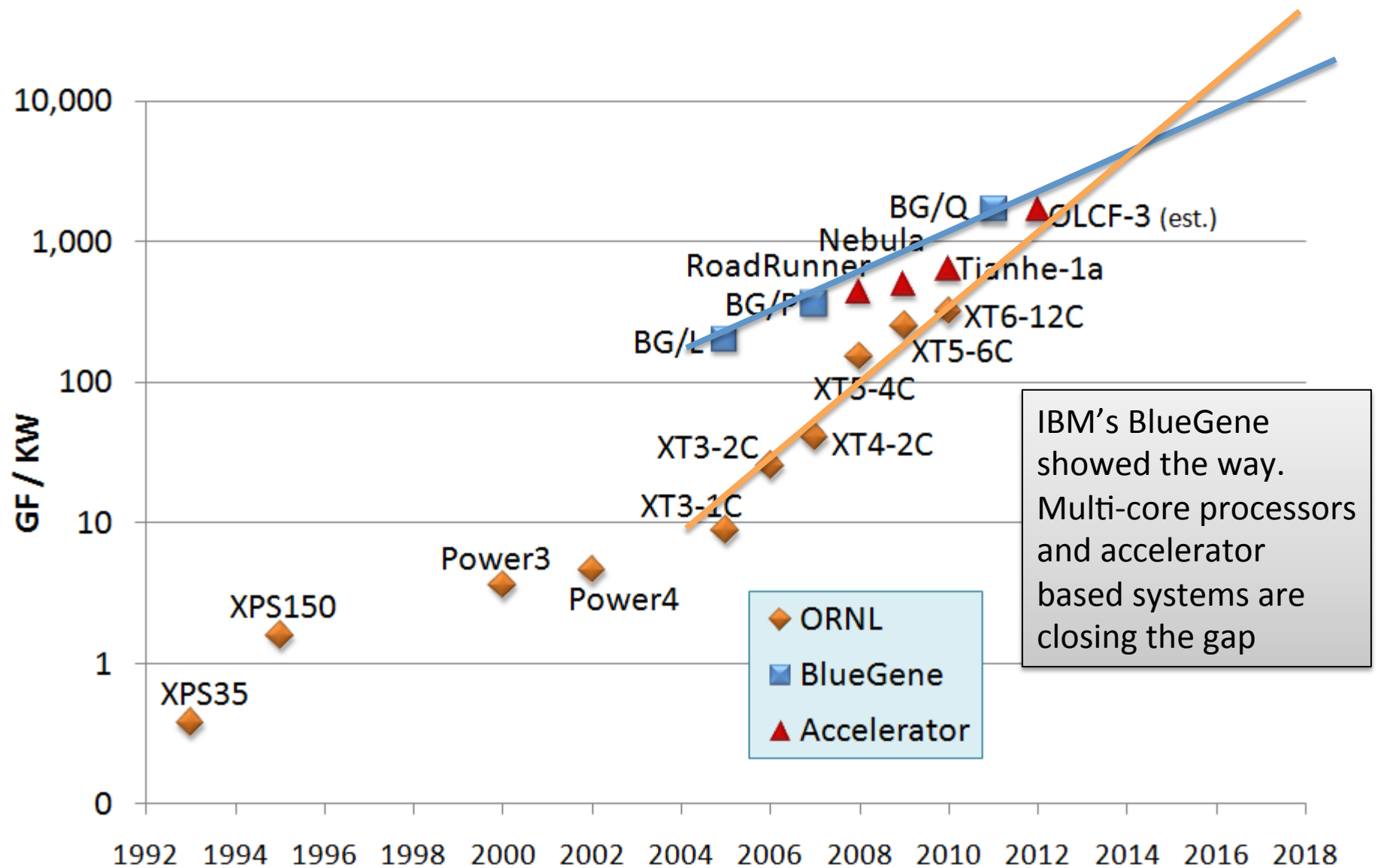
- The energy consumed increases proportionally to the bit-rate, so as we move to ultra-high-bandwidth links, the power requirements will become an increasing concern.
- The energy consumption is highly distance-dependent (the square of the length term), so bandwidth is likely to become increasingly localized as power becomes a more difficult problem.
- Improvements in chip lithography (making smaller wires) will not improve the energy efficiency or data carrying capacity of electrical wires.

D. A. B. Miller and H. M. Ozaktas, "Limit to the Bit-Rate Capacity of Electrical Interconnects from the Aspect Ratio of the System Architecture," Journal of Parallel and Distributed Computing, vol. 41, pp. 42-52 (1997) article number PC961285.

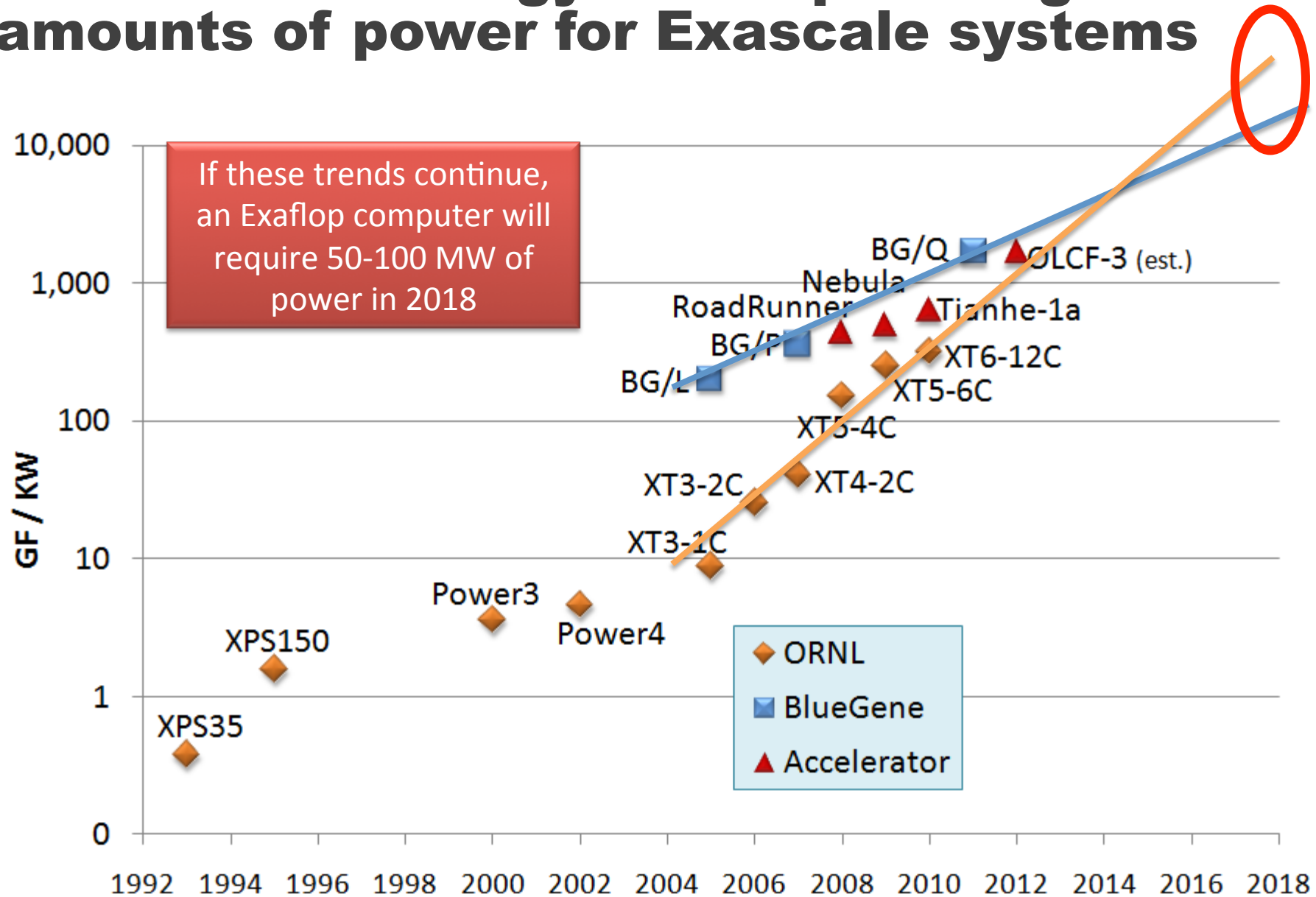
Implications for future systems

- Clock rates will stay largely the same as today, increasing the parallelism of systems to improve performance
- Energy cost of moving data is very large. We will have to explicitly manage **data locality** to limit power consumption

Trends in power efficiency

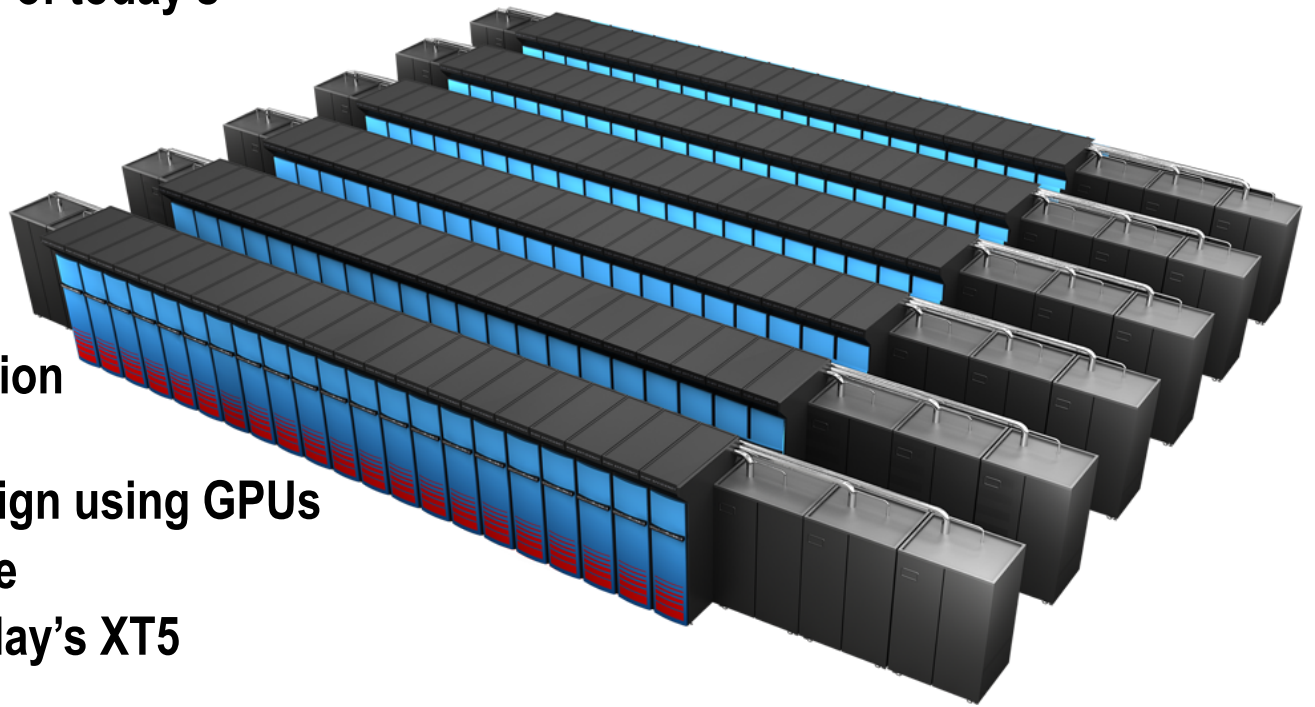


Current Technology will require huge amounts of power for Exascale systems



ORNL's "Titan" 20 PF System Goals

- Designed for science from the ground up
- Operating system upgrade of today's Linux Operating System
- Gemini interconnect
 - 3-D Torus
 - Globally addressable memory
 - Advanced synchronization features
- New accelerated node design using GPUs
- 10-20 PF peak performance
 - 9x performance of today's XT5
- Larger memory
- 3x larger and 4x faster file system



Cray XK6 Compute Node

XK6 Compute Node Characteristics

AMD Opteron 6200 Interlagos
16 core processor

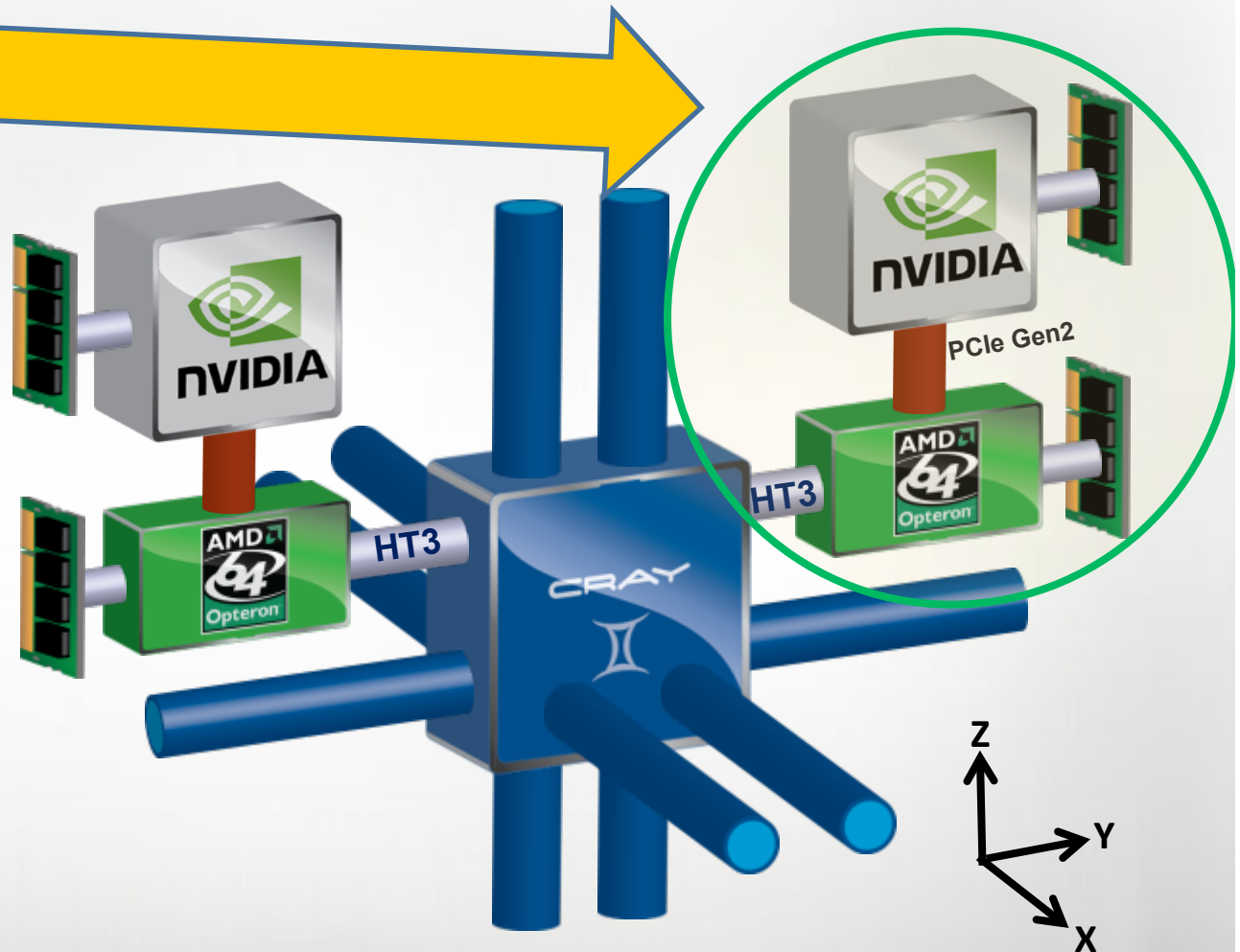
Tesla X2090 @ 665 GF

Host memory
16 or 32GB
1600 MHz DDR3

Tesla X090 memory
6GB GDDR5 capacity

Gemini high speed Interconnect

Upgradeable to NVIDIA's
Kepler many-core processor



File System

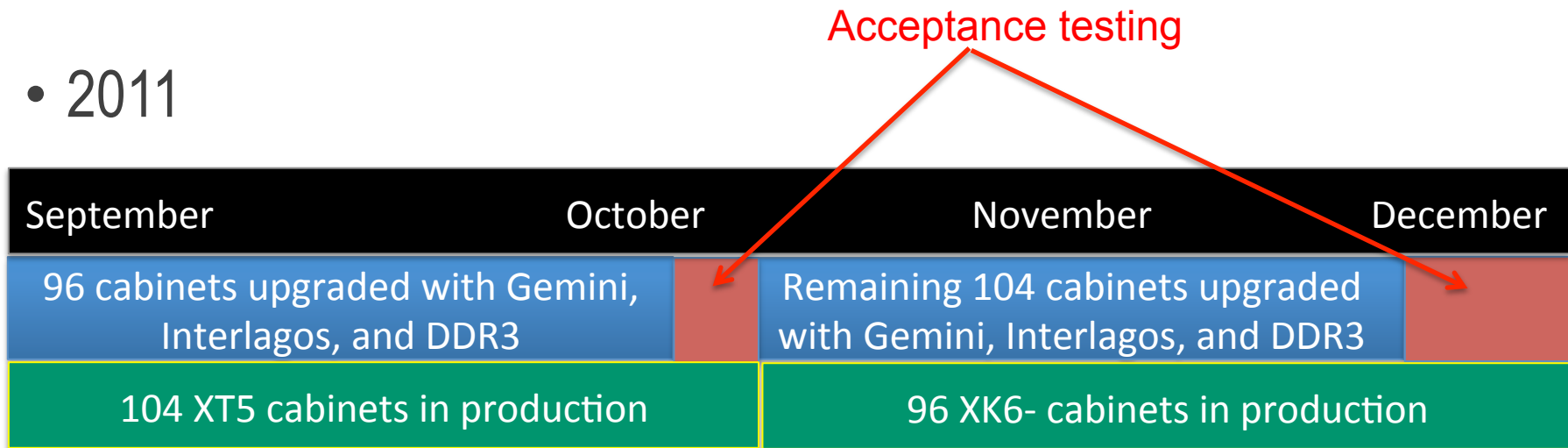
- We will continue to use Lustre for our file system for Titan
- Plan to use Lustre version 2.x
 - Much more scalable metadata
 - Etc. etc. etc.
- Competitive procurement for the storage
 - Expect to get between 400 and 700 Gigabytes per second of bandwidth
 - Expect to add between 10 and 30 Petabytes of storage

Titan is a staged upgrade of jaguar

1 st Upgrade (Sept. – Dec. 2011)	Final System
• Jaguar will be upgraded in 2 stages • 1st stage • Roughly 3rd week of September through the end of October • 96 of jaguar's 200 cabinets will be removed from service and upgraded to XK6, minus the GPUs • 2nd stage • November – December • 96 XK6- cabinets returned to service • Remaining 104 jaguar XT5 cabinets upgraded to XK6- • 2 “halves” joined and acceptance test run in late December – no user access during this time • ~10 cabinets will contain Tesla X2090 GPU	• 2nd half of 2012 • 2nd socket in each XK6 board populated with Kepler GPU • Precise timeline unknown at this time

Jaguar upgrade timeline

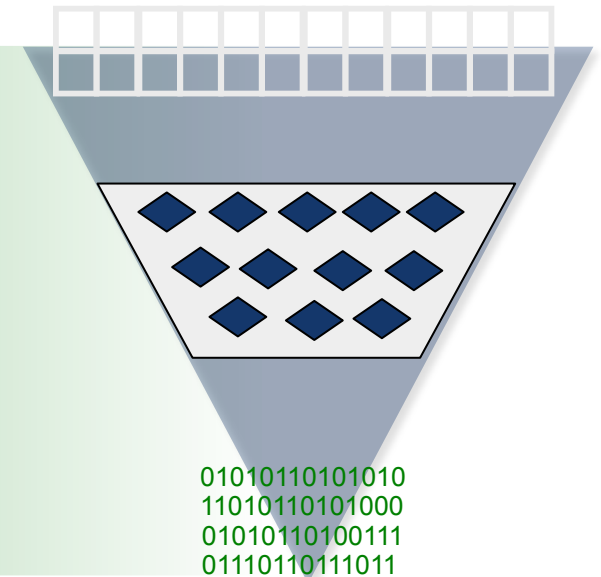
- 2011



- Jan 2012 – Full, upgraded, 200-cabinet jaguar system returned to service
- 2012 final upgrade to Titan
 - Roughly mirrors the 2011 initial upgrade path

Hierarchical Parallelism

- MPI parallelism between nodes (or PGAS)
- On-node, SMP-like parallelism via threads (or subcommunicators, or...)
- Vector parallelism
 - SSE/AVX on CPUs
 - GPU threaded parallelism



- **Exposure of unrealized parallelism is essential to exploit all near-future architectures.**
- **Uncovering unrealized parallelism and improving data locality improves the performance of even CPU-only code.**

Questions?

Bronson Messer

Email: bronson@ornl.gov

The research and activities described in this presentation were performed using the resources of the Oak Ridge Leadership Computing Facility at Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC0500OR22725.



Titan Training and Education



Robert Whitten

Disclaimer: There is no contract with the vendor. This presentation is the OLCF's current vision that is awaiting final approval. Subject to change based on contractual and budget constraints.



U.S. DEPARTMENT OF
ENERGY



OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE FOR THE DEPARTMENT OF ENERGY

Training Strategy

- Expose parallelism
 - Case studies
 - Lessons learned
- Compiler / Directive-based
 - Cuda Fortran
 - HMPP
- Debugging / Performance Analysis
- Optimization
- Accelerator programming
 - Cuda
 - OpenCL

Expose
Parallelism

Use Tools

Examine
performance

Optimize

Low level
programming

Planned Upcoming Events

- Titan Summit - August 15-17, 2011
- CAPS Workshop - September 19-23, 2011
- Fall Training - October 2011
- Titan Workshop - December 2011
- Spring Training - March 2011



Titan Summit

- August 15 – 17, 2011
- OLCF Staff
 - Case studies / lessons learned
- Tool vendors
- Users
 - Application roadmaps
 - Experiences
 - Expectations



The Portland Group



- More information and registration:
 - <https://www.olcf.ornl.gov/event/titan-summit/>

We need volunteers for the Summit!

- Present current state of their application codes
- Present planned future roadmap
- Talk about expectations for Titan and the OLCF
- Tell OLCF what needs you have or will have
- Tell me what training you would like to see

CAPS Workshop

- September 19 – 23, 2011
- HMPP Basics
- Advanced features
- Data distribution
- HMPP++
- One-on-one sessions
- <http://www.caps-entreprise.com/index.php>
- Registration and website coming soon!

In the works - TBD

- Cray Workshop
 - Cray compiler
 - Craypat
- Cuda Workshop
- Vampir Workshop
- Hybrid-programming Workshop
- DDT Workshop
- Totalview Workshop
- Tau Workshop
- Some will be webcast-only events

Questions?

TITAN

<http://www.olcf.ornl.gov>