Trends in Energy-Efficient Supercomputing

BoF Organizers:

Wu Feng, Erich Strohmaier, Natalie Bates, and Tom Scogland

The 22nd GREEN 500

November 2017

VirginiaTech
Invent the Future

SyNeRG
synergy.cs.vt.edu

The Ultimate Goal of "The Green500 List"

- Raise awareness (and encourage reporting) of the energy efficiency of supercomputers
  - Drive energy efficiency as a first-order design constraint (on par with performance).

Encourage fair use of the list rankings to promote energy efficiency in high-performance computing systems.

The GREEN 500

Agenda

- The Green500 and its Evolution: Past, Present, and Future *(Wu Feng)*
- Status of L2/L3 Measurements *(Erich Strohmaier)*
- Green500 L2/L3 Measurements
  - DGX SaturnV Volta *(Louis Capps)*
  - Oakforest-PACS (JCAHPC) and Reedbush-H (ITC, UTokyo) *(Masaaki Kondo)*
- The 22nd Green500 List *(Wu Feng)*
  - Trends and Evolution
- Shoubu System B, #1 on the Green500 *(Motoaki Saito)*

The Green500 and its Evolution: Past, Present, and Future

Wu Feng

The GREEN 500

VirginiaTech
Invent the Future

SyNeRG
synergy.cs.vt.edu

Brief History:

From Green Destiny to The *Green500* List

2/2002: Green Destiny (http://sss.lanl.gov/ → http://sss.cs.vt.edu/)

- "Honey, I Shrunk the Beowulf!" 31st Int'l Conf. on Parallel Processing, August 2002.

4/2005: Workshop on High-Performance, Power-Aware Computing

- Keynote address generates initial discussion for Green500 List

4/2006 and 9/2006: Making a Case for a Green500 List

- Workshop on High-Performance, Power-Aware Computing
- Jack Dongarra's CCGSC Workshop "The Final Push" (Dan Fay)

9/2006: Founding of Green500: Web Site and RFC (Chung-Hsing Hsu)

- http://www.green500.org/ Generates feedback from hundreds

11/2007: Launch of the First Green500 List (Kirk Cameron)

- http://www.green500.org/lists/green200711

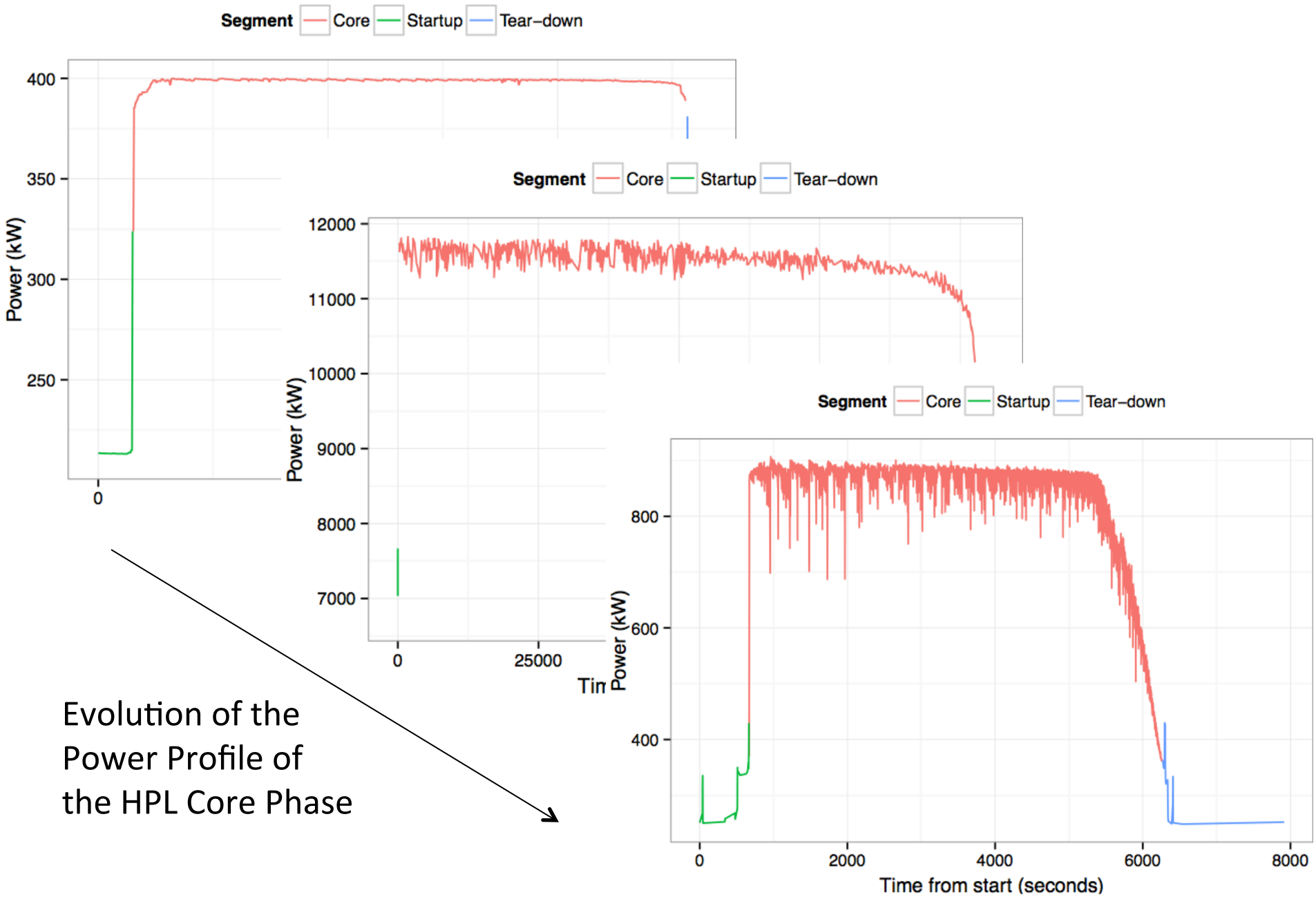
Evolution of The GREEN 500

- 11/2009: Experimental Lists Created
  - *Little Green500:* More focus on LINPACK energy efficiency than on LINPACK performance in order to foster innovation
  - ~~*HPCC Green500:* Alternative workload to evaluate energy efficiency~~
  - ~~*Open Green500:* Enabling alternative innovative approaches for LINPACK to improve performance and energy efficiency, e.g., mixed precision~~
- 11/2010: Updated Green500 Official Run Rules Released
- 06/2011: Collaborations Begin on Methodologies for Measuring the Energy Efficiency of Supercomputers
- 06/2013: Adoption of New Power Measurement Methodology, version 1.0 (EE HPC WG, The Green Grid, Green500, TOP500)
- 01/2016: Adoption of New Power Measurement Methodology, version 2.0 (EE HPC WG, The Green Grid, Green500, TOP500)

Evolution of The GREEN 500

- 05/2016: Green500 Merges with TOP500
  - Unified run rules, data collection, and posting of power measurements via the TOP500 (http://www.green500.org → http://www.top500.org/green500)
  - Enable submissions of both performance-optimized (TOP500) and power-optimized (Green500) numbers

The GREEN 500 → TOP 500 The List.



Evolution of the Power Profile of the HPL Core Phase

Legacy Assumptions
(circa 2007)

Originally designed to *encourage* reporting of power with accuracy

- Measuring a small part of a system and scaling it up does *not* introduce too much of an error
- The power draw of the interconnect fabric is *not* significant when compared to the compute system
- The workload phase of HPL will look similar on *all* HPC systems

These assumptions were re-visited by *EE HPC WG*, *The Green Grid*, *Top500*, and *Green500* (2011-2015)

D. Rohr et al., "Refining Power Measurement Methodology for Supercomputer-System Benchmarking," *International Supercomputing Conference,* July 2015.

Agenda

- The Green500 and its Evolution: Past, Present, and Future *(Wu Feng)*
- Status of L2/L3 Measurements *(Erich Strohmaier)*
- Green500 L2/L3 Measurements
  - DGX SaturnV Volta *(Louis Capps)*
  - Oakforest-PACS (JCAHPC) and Reedbush-H (ITC, UTokyo) *(Masaaki Kondo)*
- The 22nd Green500 List *(Wu Feng)*
  - Trends and Evolution
- Shoubu System B, #1 on the Green500 *(Motoaki Saito)*

Status of L2/L3 Submissions

T. Wilde, T. Scogland, W. Feng,
E. Strohmaier, and N. Bates

ISC 2017 Green500 BoF

Measuring Power:
Level 1 (L1), Level 2 (L2), Level 3 (L3)

- What is the difference between the three levels?
- Why make a L3 submission?
- Who has made a L2/L3 submission?
- What needs improving in the L2/L3 methodology and submission process?

Difference Between the Three Levels?

	Level 1	Level 2	Level 3
Subcomponents to measure	Compute nodes and network	Measure compute nodes, estimate rest (network, etc.)	Measure everything
Faction of system to measure	At least 1 / 10th of system, 15 nodes or 10 kW	At least 1 / 8 of system and 10 kW	Full system
Time period	Entire core phase	Entire core phase	Entire core phase
Measurement method	Power-sample averaged measurement	Power-sample averaged measurement	Integrated energy measurement

Why Make a L3 Submission?

ETH zürich

We do level 3 measurements because

1. It is always good to have reliable information about your data centre

2. Doing a reliable level 3 measurement is not harder than level 2 or level 1

CSCS

Green500 BoF at SC16, Salt Lake City, Wednesday, November 16, 2016 T. Schulthess 6

Why Make a L3 Submission?

LANL Description of Benefits from L3 Submission

- Team Interaction
  - Our Facilities, Monitoring and Admin teams are all in different groups
  - These L3 runs forced all 3 teams to work together to collect the needed data
- These L3 Measurements Laid the Groundwork for our Future Power Monitoring Work
  - Power per job
  - Looking closer at temperature and cooling systems
- Power Monitoring is One of Our Top Issues This Year
  - Measure power usage for each Job
  - Understand power usage per cluster for new systems procurement and datacenter needs

Broader Relevance

- It is difficult to obtain a system level measurement, but it has value (beyond Green500 and Top500)
  - Do we need to go beyond lists? Of course…
- Some examples of use cases for system-level measurements include:
  - Architectural trending, system modeling (design, selection, upgrade, tuning, analysis),
  - Procurement and data-center provisioning
  - Operational improvements
  - Day-to-day workloads vs HPL
  - TCO combining HPC system and room Infrastructure
  - Validate component-level measurement (by summing them up!)

Current List: L2/L3 Submissions

Facebook
Information Technology Center, The University of Tokyo
Joint Center for Advanced High Performance Computing
Lawrence Livermore National Laboratory
Los Alamos National Laboratory
Maui High-Performance Computing Center (MHPCC)
Michigan State University
National Supercomputing Center in Wuxi
National Institute for Environmental Studies
NVIDIA
Sandia National Laboratories
Swiss National Supercomputing Centre (CSCS)
Universität Mainz

Improvements to L2/L3 Methodology?

CSCS

- No way to provide a report and supporting files on submission website

LANL

- Useful document, but intimidating for first-time users
- Need contact information for quick questions or detailed discussions
- Need a list of known metering/measurement equipment
  - We needed to contact some vendors to make sure we met the listed requirements

Thank you!

http://eehpcwg.llnl.gov

natalie.jean.bates@gmail.com

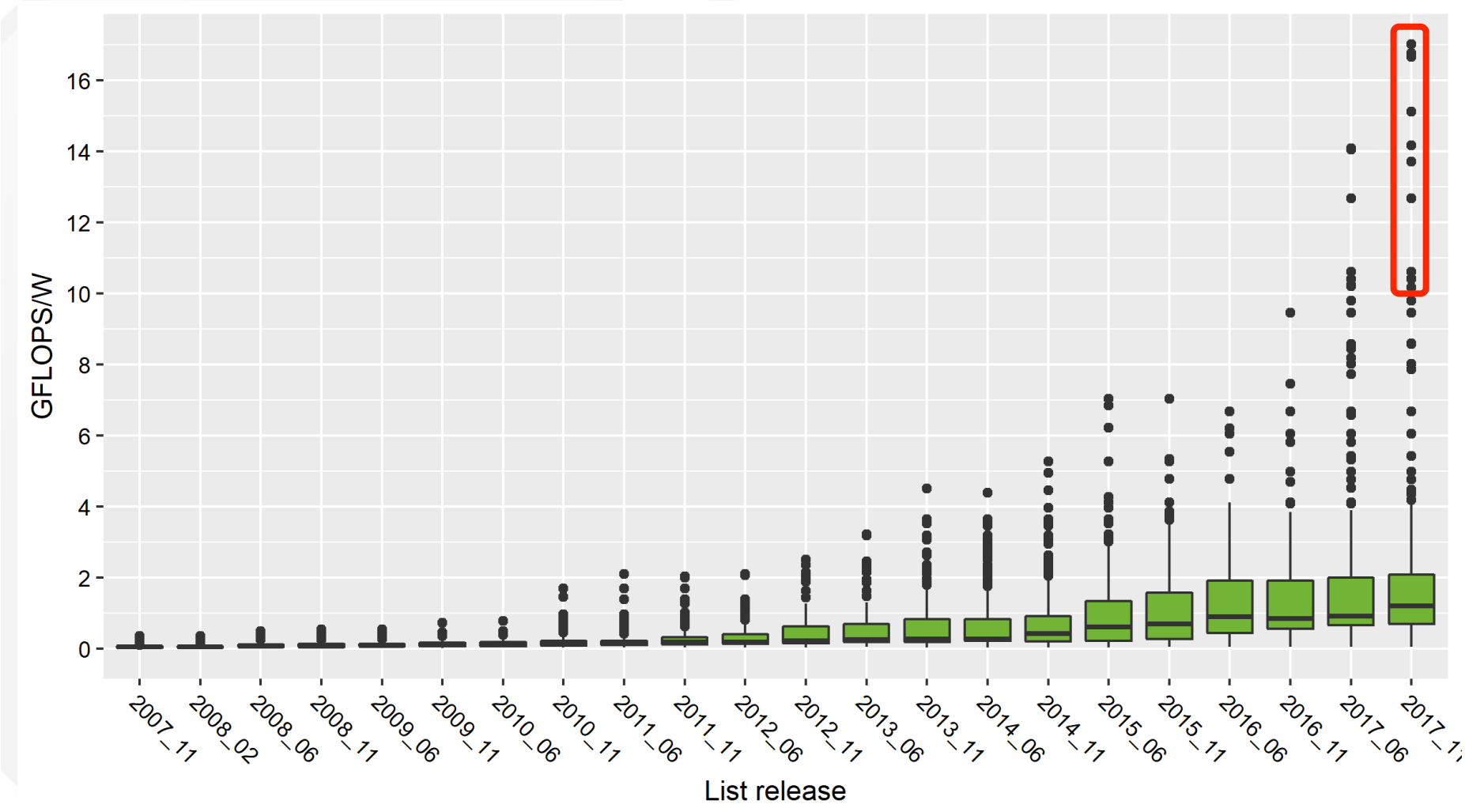
Agenda

- The Green500 and its Evolution: Past, Present, and Future *(Wu Feng)*
- Status of L2/L3 Measurements *(Erich Strohmaier)*
- Green500 L2/L3 Measurements
  - DGX SaturnV Volta *(Louis Capps)*
  - Oakforest-PACS (JCAHPC) and Reedbush-H (ITC, UTokyo) *(Masaaki Kondo)*
- The 22nd Green500 List *(Wu Feng)*
  - Trends and Evolution
- Shoubu System B, #1 on the Green500 *(Motoaki Saito)*

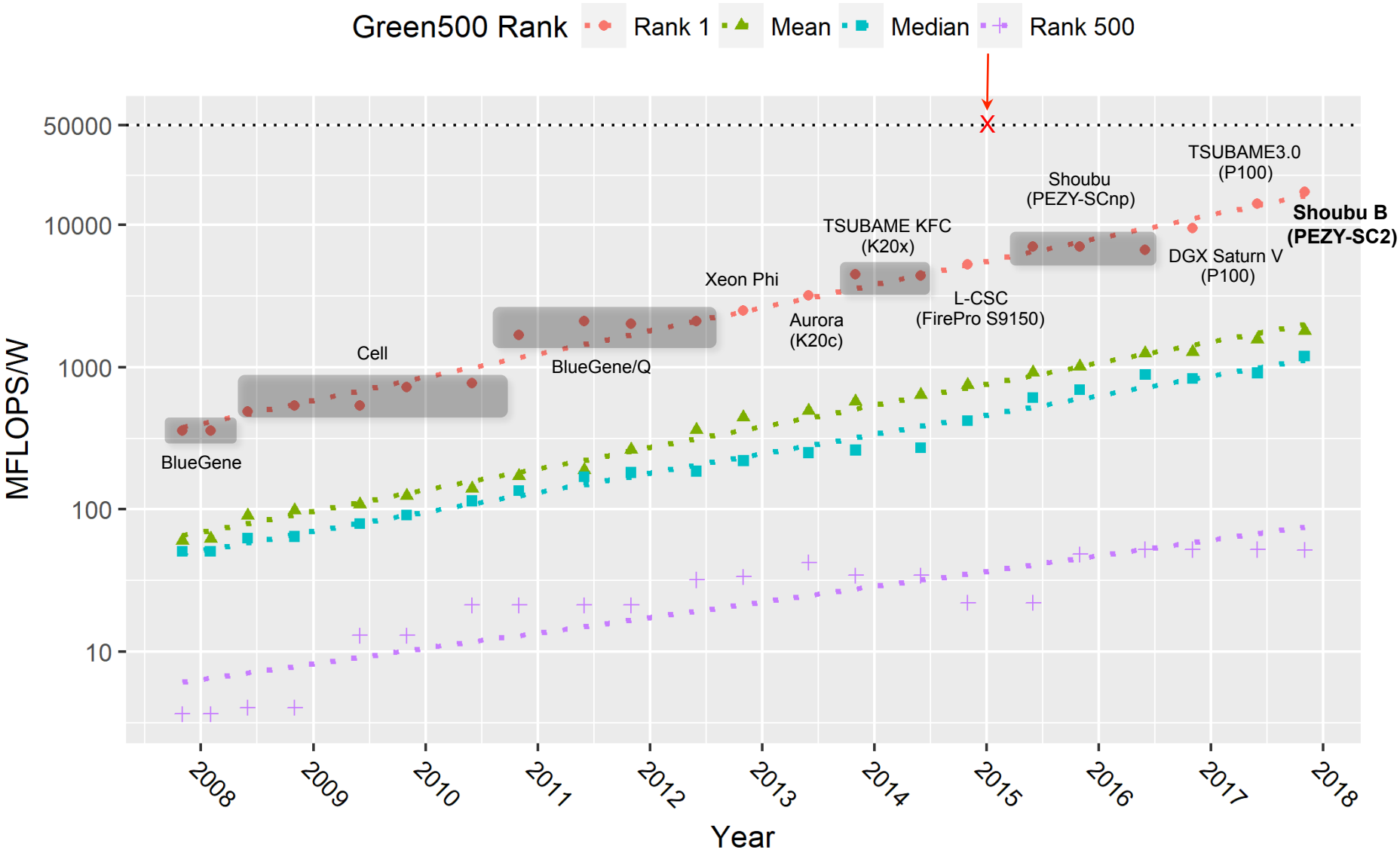
Agenda

- The Green500 and its Evolution: Past, Present, and Future *(Wu Feng)*
- Status of L2/L3 Measurements *(Erich Strohmaier)*
- Green500 L2/L3 Measurements
  - DGX SaturnV Volta *(Louis Capps)*
  - Oakforest-PACS (JCAHPC) and Reedbush-H (ITC, UTokyo) *(Masaaki Kondo)*
- The 22nd Green500 List *(Wu Feng)*
  - Trends and Evolution
- Shoubu System B, #1 on the Green500 *(Motoaki Saito)*

Trends: How Energy Efficient Are We?

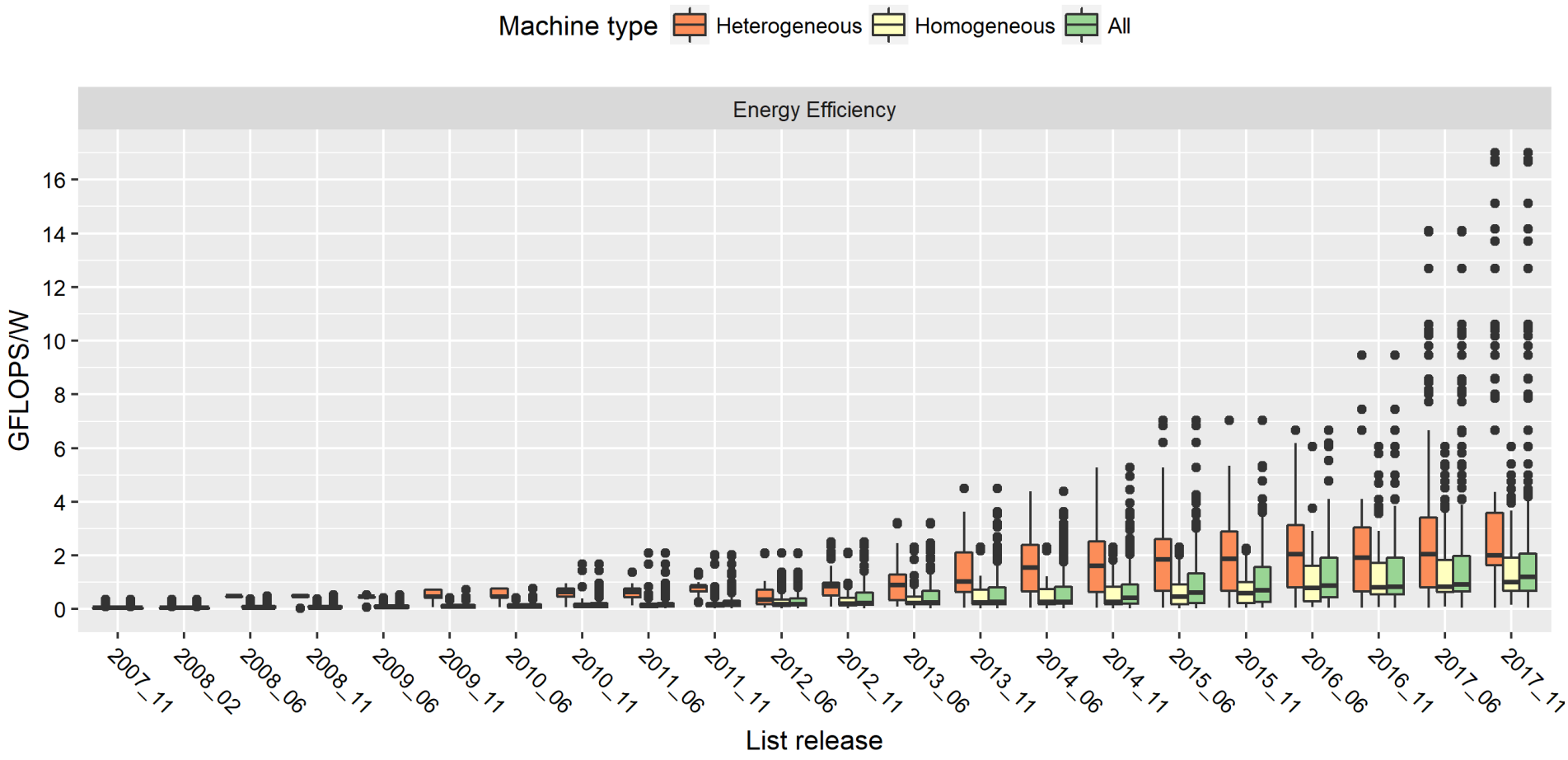


Trends: How Energy Efficient Are We?

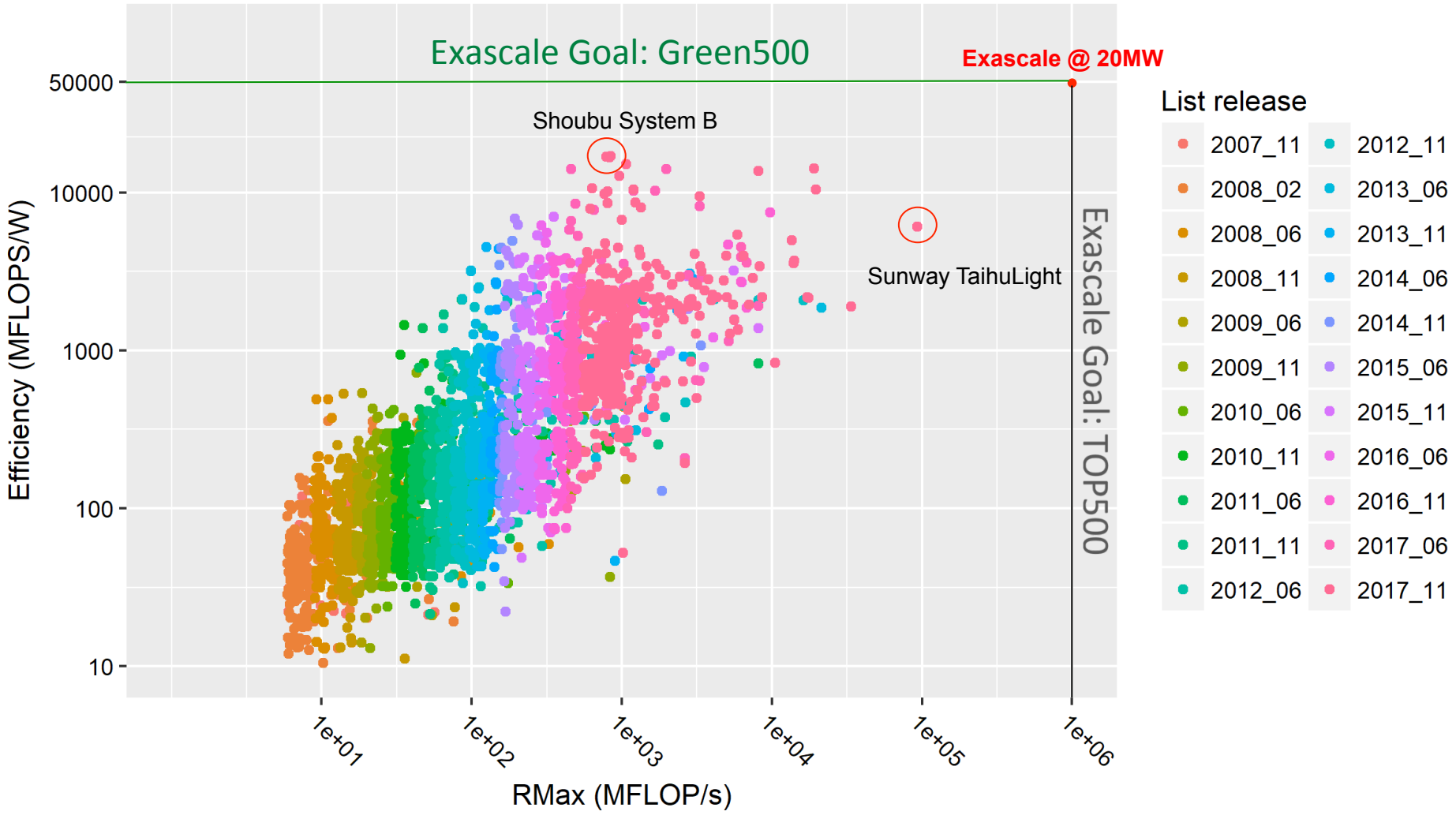


The Green500 BoF, SC|17, Nov 2017

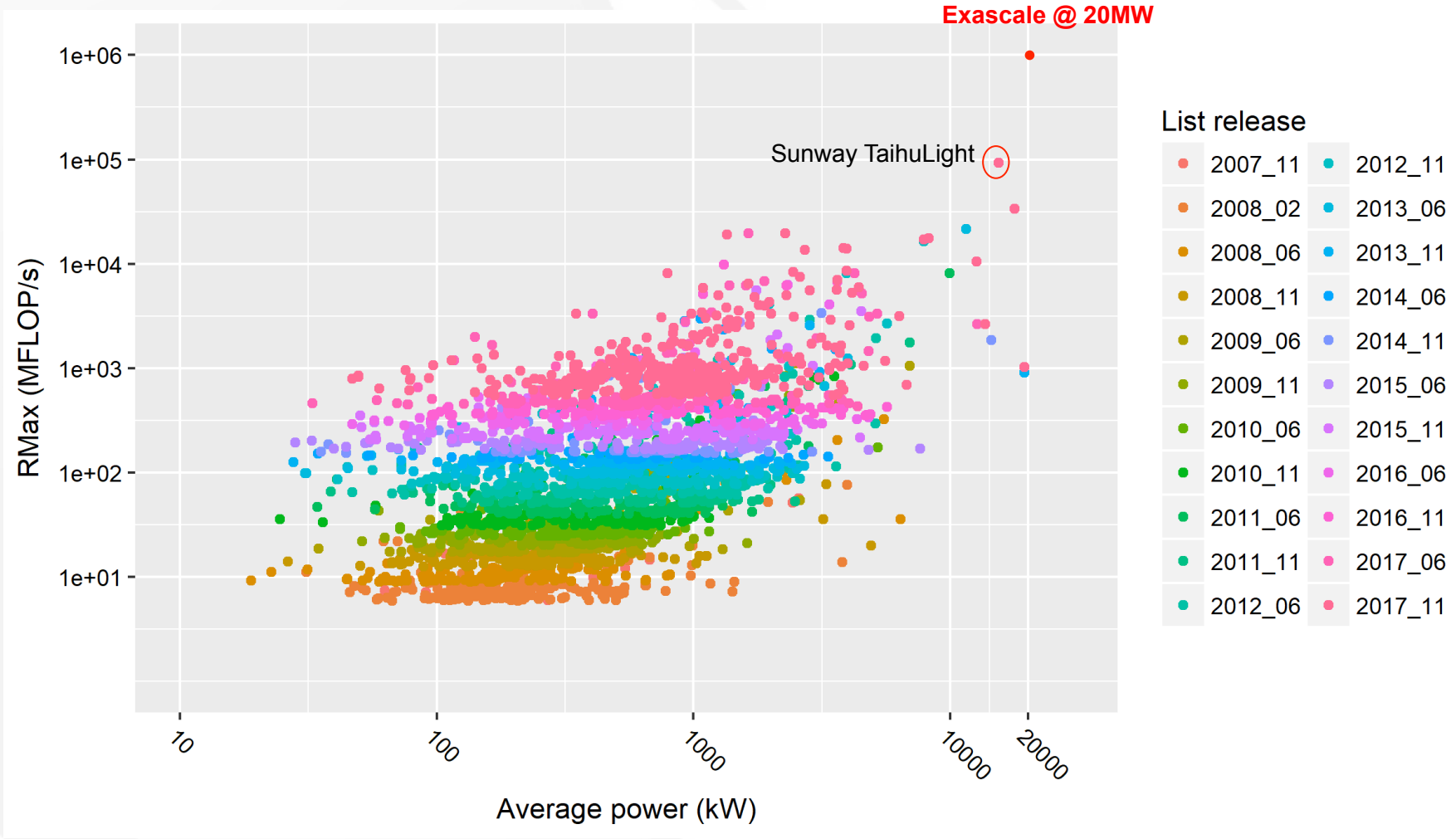
Trends in Efficiency: Homogeneous vs Heterogeneous Systems



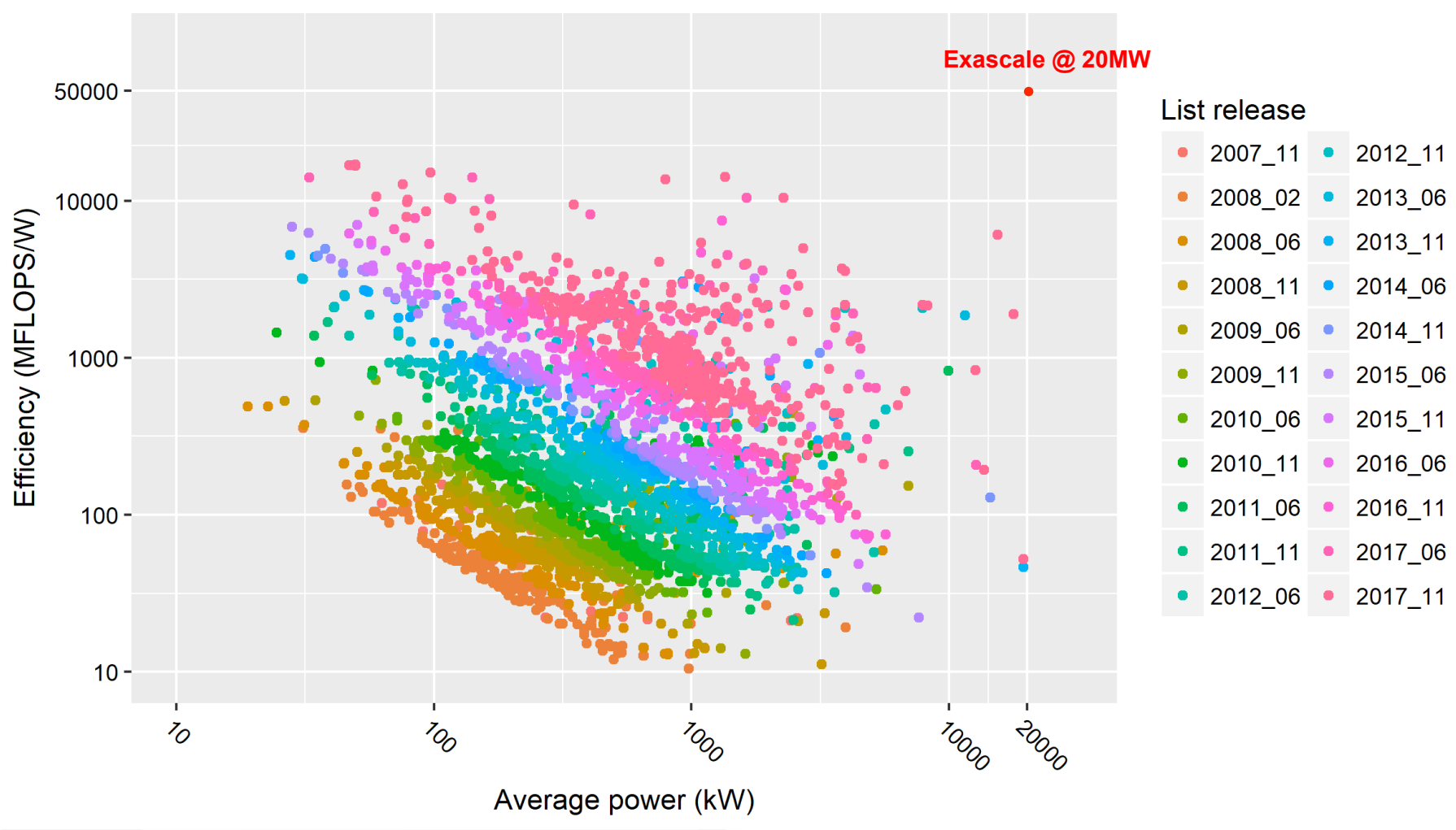
Efficiency vs. Performance



Performance vs. Power



Efficiency vs. Power

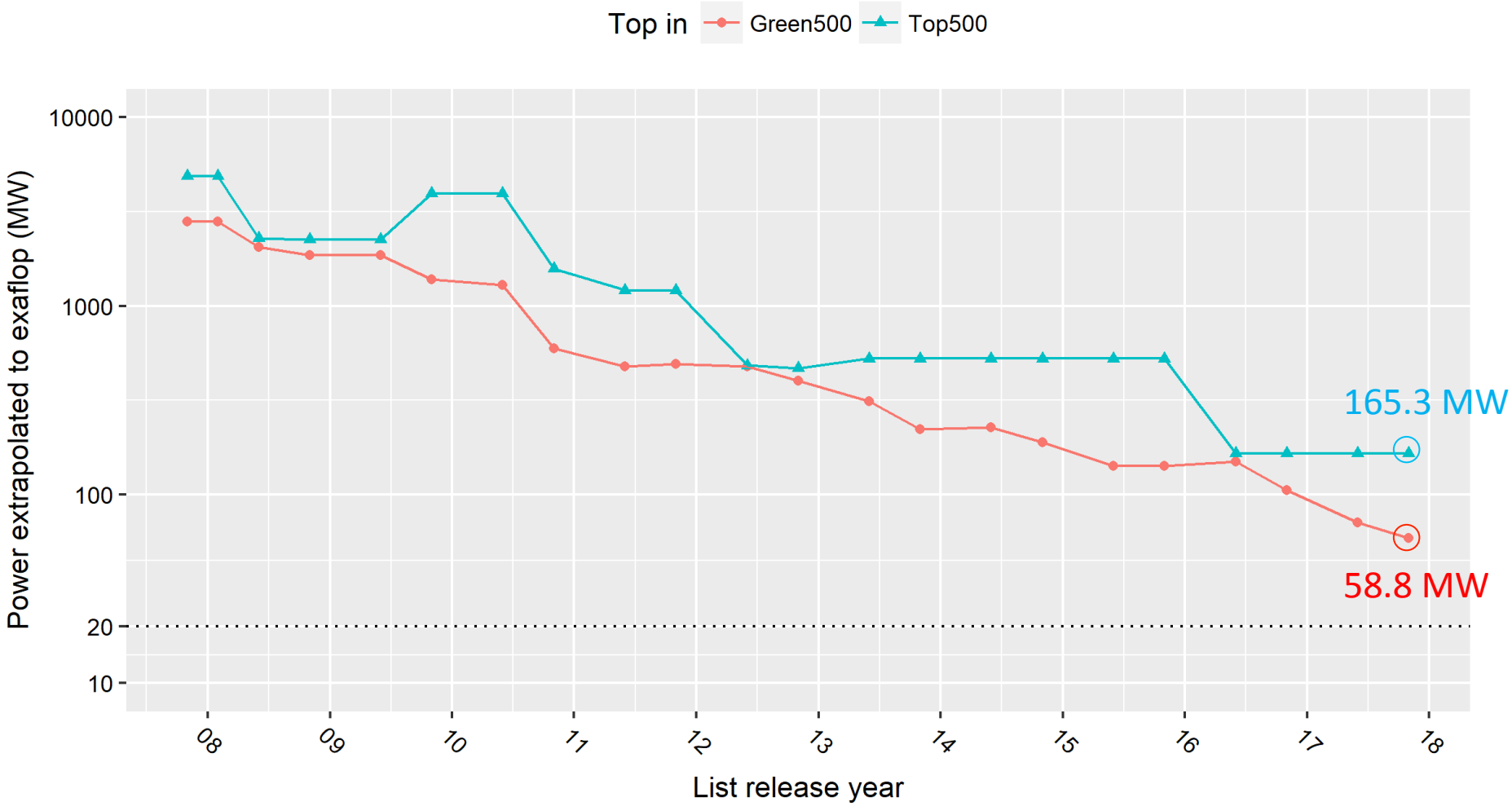


Trends Towards Exascale

Exascale Computing Study: Technology Challenges in Achieving Exascale Systems

- Goal
  - "Because of the difficulty of achieving such physical constraints, the study was permitted to assume some growth, perhaps a factor of 2X, to something with a maximum limit of 500 racks and *20 MW* for the computational part of the 2015 system."
- Realistic Projection?
  - "Assuming that Linpack performance will continue to be of at least passing significance to real Exascale applications, and that technology advances in fact proceed as they did in the last decade (both of which have been shown here to be of dubious validity), then […] an Exaflop per second system is possible at around *67 MW*."

Trends: Extrapolating to Exaflop



Green500 Rank	TFLOPS/ W	Name	Site	Computer
1	17.009	Shoubu System B	Advanced Center for Computing and Communication, RIKEN	ZettaScaler-2.2, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2
2	16.759	Suiren2	High Energy Accelerator Research Organization /KEK	ZettaScaler-2.2, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2
3	16.657	Sakura	PEZY Computing K.K.	ZettaScaler-2.2, Xeon E5-2618Lv3 8C 2.3GHz, Infiniband EDR, PEZY-SC2
4	15.113	DGX SaturnV Volta	NVIDIA Corporation	NVIDIA DGX-1 Volta36, Xeon E5-2698v4 20C 2.2GHz, Infiniband EDR, NVIDIA Tesla V100
5	14.173	Gyoukou	Japan Agency for Marine-Earth Science and Technology	ZettaScaler-2.2 HPC system, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 700Mhz
6	13.704	TSUBAME3.0	GSIC Center, Tokyo Institute of Technology	SGI ICE XA, IP139-SXM2, Xeon E5-2680v4 14C 2.4GHz, Intel Omni-Path, NVIDIA Tesla P100 SXM2
7	12.681	AIST AI Cloud	National Institute of Advanced Industrial Science and Technology	NEC 4U-8GPU Server, Xeon E5-2630Lv4 10C 1.8GHz, Infiniband EDR, NVIDIA Tesla P100 SXM2
8	10.603	RAIDEN GPU subsystem	Center for Advanced Intelligence Project, RIKEN	NVIDIA DGX-1, Xeon E5-2698v4 20C 2.2GHz, Infiniband EDR, NVIDIA Tesla P100
9	10.428	Wilkes-2	University of Cambridge	Dell C4130, Xeon E5-2650v4 12C 2.2GHz, Infiniband EDR, NVIDIA Tesla P100
10	10.398	Piz Daint	Swiss National Supercomputing Centre (CSCS)	Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100

Brief Analysis of Top 10 Machines on The GREEN 500

- All 10 machines are accelerator-based
  - 4 PEZY-SC2, 1 NVIDIA V100, 5 NVIDIA P100
- From June 2017 Green500,
  - Stayed in top 10 (6)
    - Ranks #5 to #10 in this year's list
  - Slid out of top 10 (3)
    - Now in top #20 in this year's list
  - Dropped out due to not meeting performance requirements (1)
    - Kukai -- #2 in June 2017 list
- New in Top 10 (4)
  - The top 4 positions (Shoubu System B, Suiren2, Sakura, and DGX Saturn)
  - 3 power-optimized runs (#4 DGX SaturnV Volta, #6 TSUBAME3.0, and #10 Piz Daint)

The GREEN 500 CERTIFICATE

Shoubu System B, a PEZY Computing / Exascaler ZettaScaler-2.2 System at the Advanced Center for Computing and Communication, RIKEN, Japan

is ranked

No. 1 in the Green500

among the World's TOP500 Supercomputers

with 17.0 GFlops/Watt Linpack Power-Efficiency

on the Green500 List published at the SC17 Conference, November 13, 2017

Congratulations from the Green500 Editors

Acknowledgements

- Key Contributor
  - Vignesh Adhinarayanan

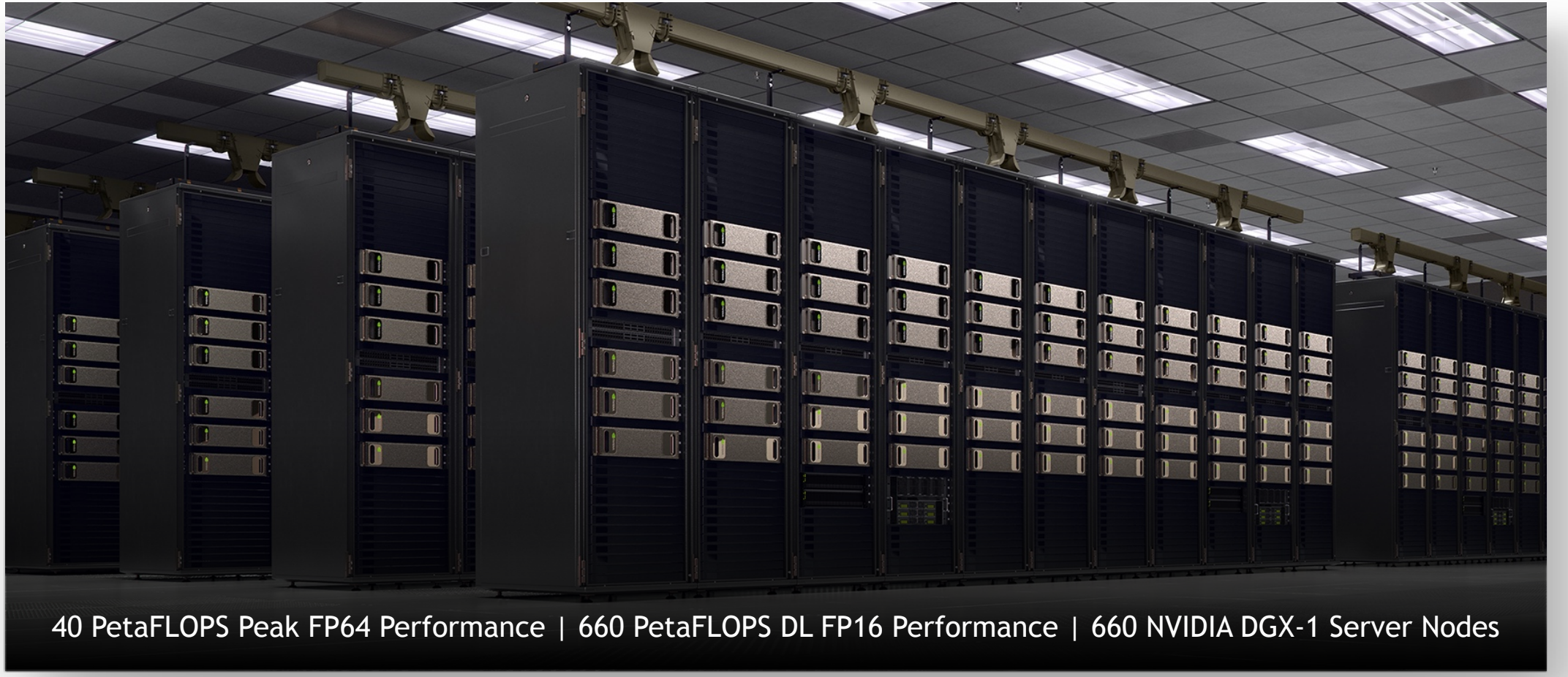
- Energy-Efficient HPC Working Group (Lead: Natalie Bates) and TOP500 (Erich Strohmaier, Jack Dongarra, Horst Simon)

- **YOU!**
  - For your contributions in raising awareness in the energy efficiency of supercomputing systems

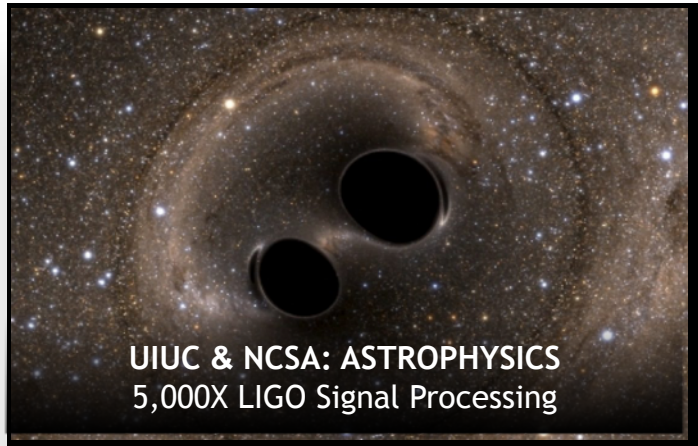
SUPERCOMPUTING 2017 – G500

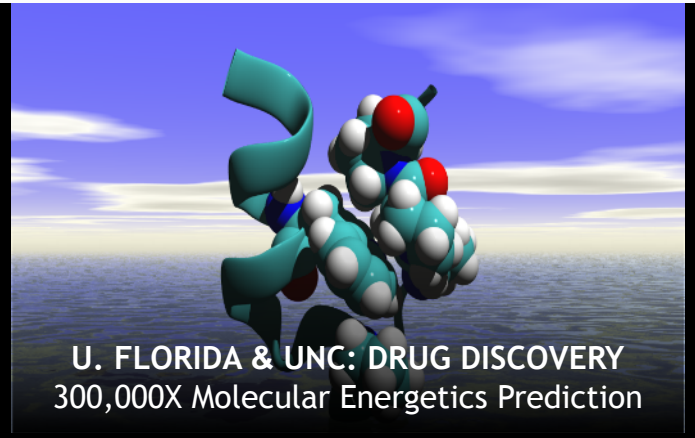
NVIIDA DGX SATURNV Volta

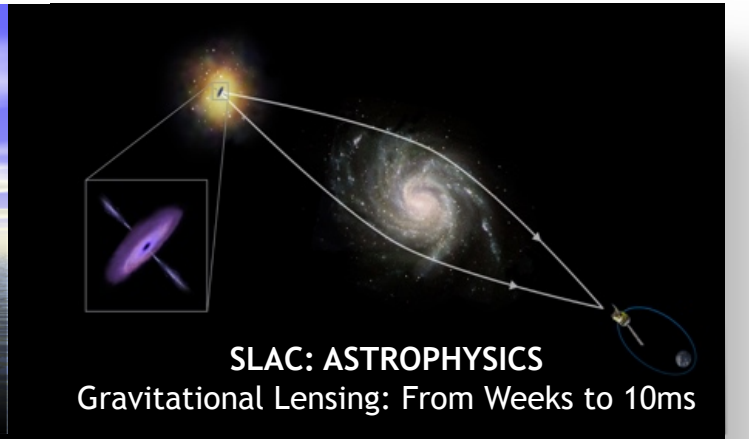
NVIDIA SATURNV WITH VOLTA
PATH TO EXASCALE

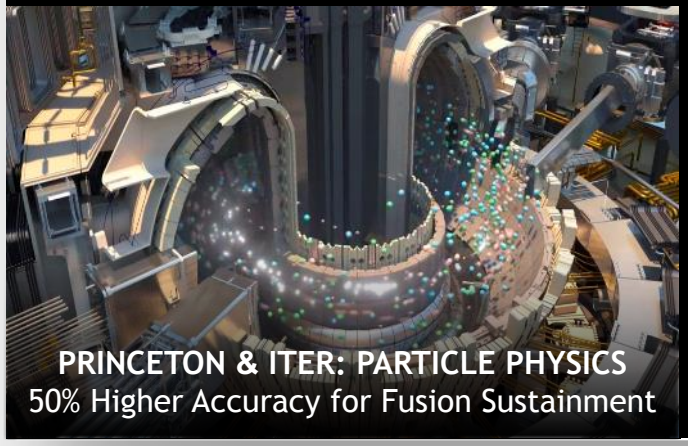


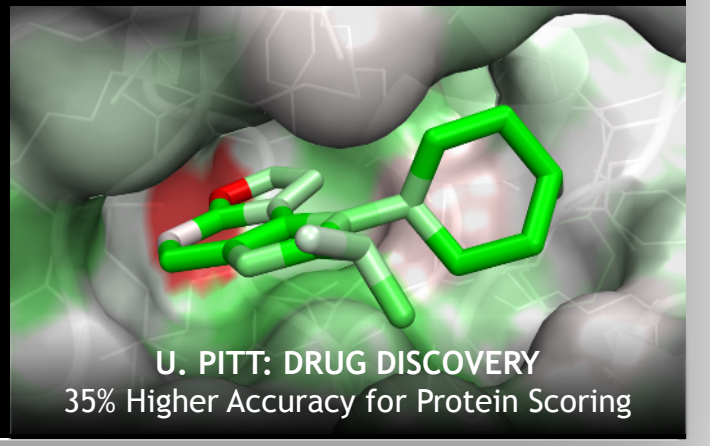
DEEP LEARNING COMES TO HPC











NVIDIA DGX-1 VOLTA

DEEP LEARNING SUPERCOMPUTER

Engineered for deep learning

- 1000/126/62/ TF TENSOR/FP32/FP64
- 8x Tesla V100 SXM
  NVLINK Cube Mesh
- 7 TB SDD RAID, 4x EDR IB

Accelerates major AI frameworks

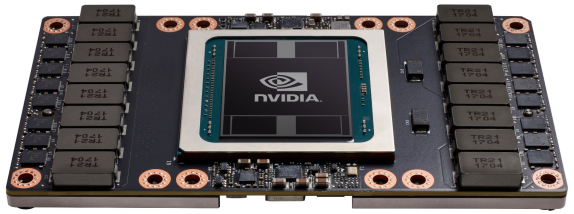
Deep Learning and HPC Containers

CUDA, OpenACC

Visualization

nvidia.com/dgx1

GPU PERFORMANCE COMPARISON



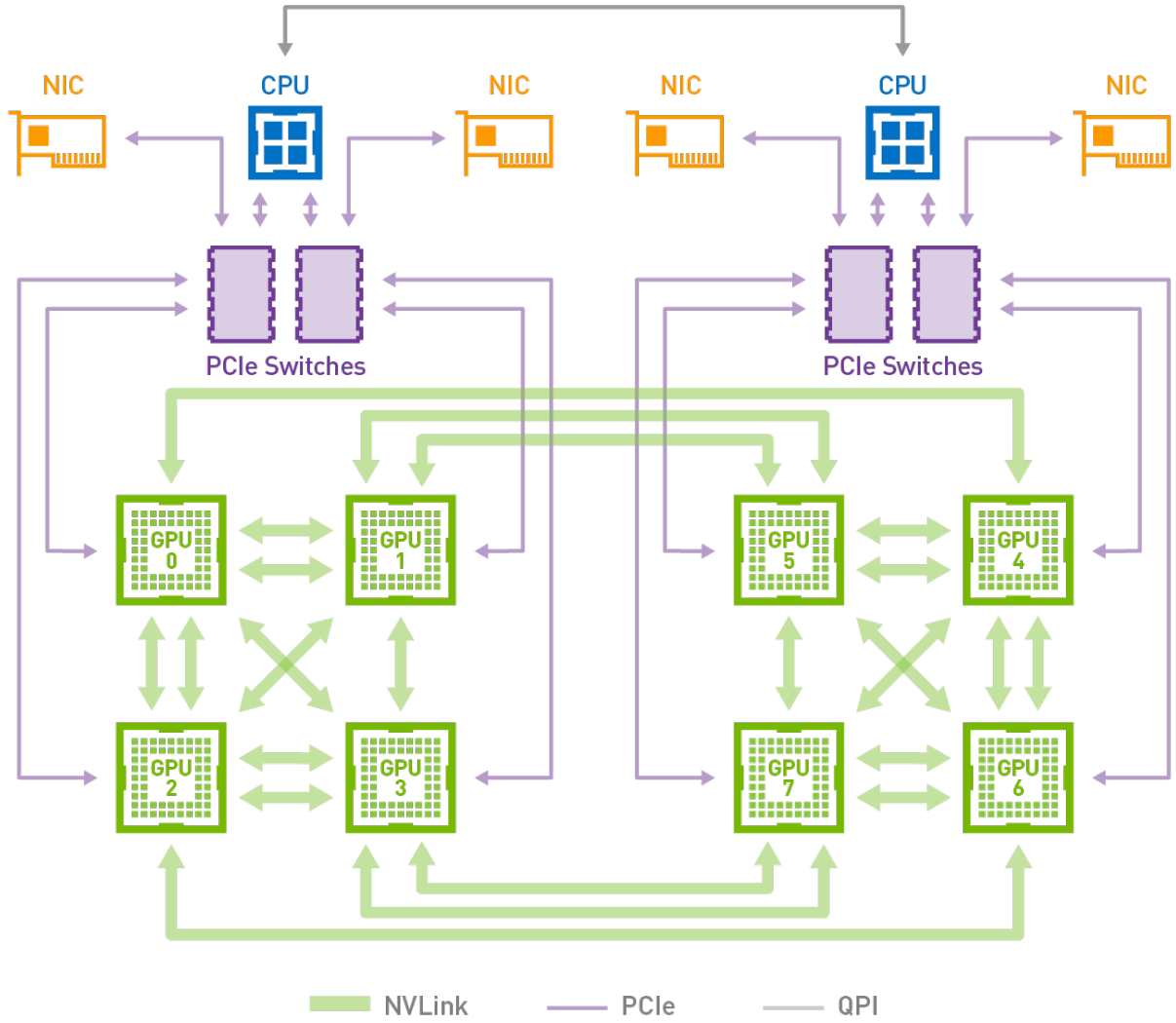
	P100	V100	Ratio
Training acceleration	10 TOPS	125 TOPS	12x
Inference acceleration	21 TFLOPS	125 TOPS	6x
FP64/FP32	5/10 TFLOPS	7.8/15.7 TFLOPS	1.5x
HBM2 Bandwidth	720 GB/s	900 GB/s	1.2x
NVLink Bandwidth	160 GB/s	300 GB/s	1.9x
L2 Cache	4 MB	6 MB	1.5x
L1 Caches	1.3 MB	10 MB	7.7x

VOLTA NVLINK

300GB/sec

50% more links

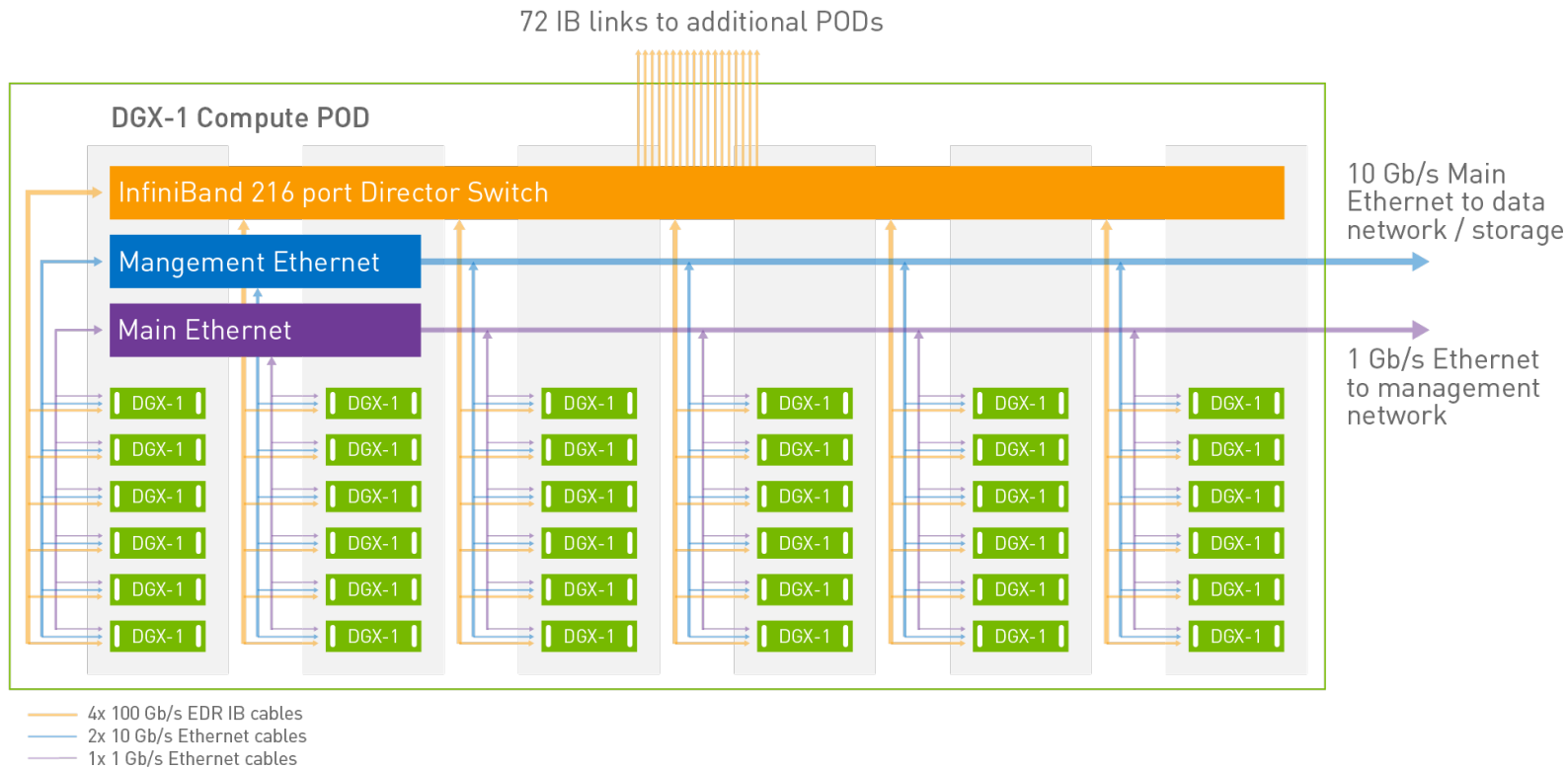
28% faster signaling



DGX POD

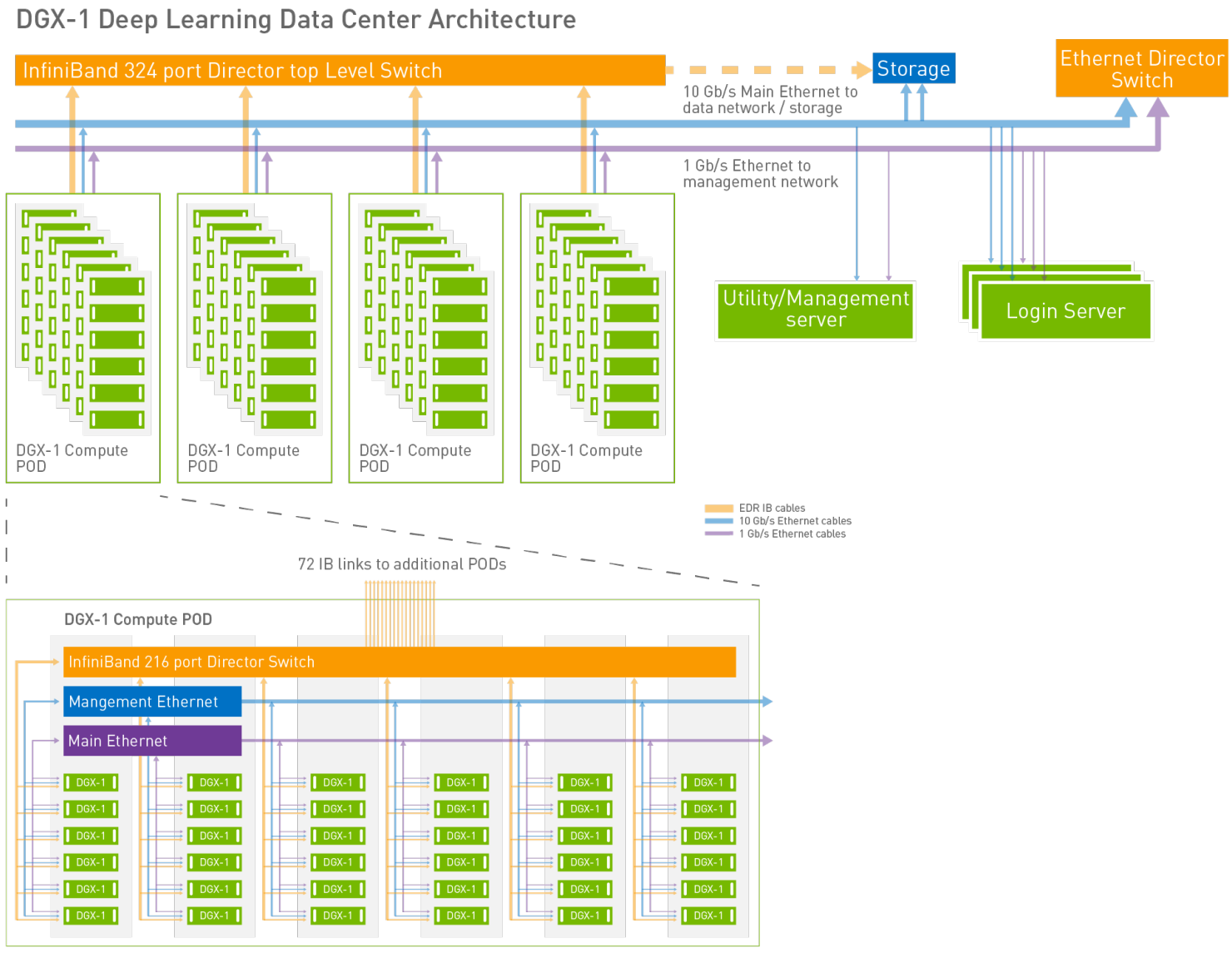
36 DGX-1 nodes

- Defines a DGX-1 “POD”
- Can be replicated for greater scale, ex: large cluster configuration
- 6 racks, 6 nodes per rack
- Larger IB director switch (216 ports) with capacity for more pods via unused ports



DGX-1 DATA CENTER

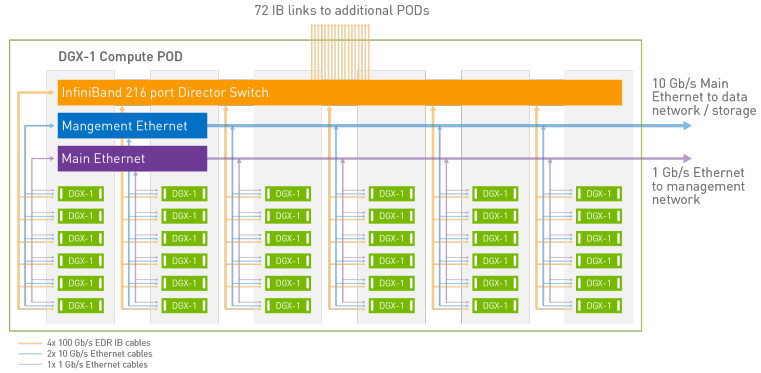
Reference Architecture



NVIDIA DGX SATURNV VOLTA

33 node cluster

- 33 NVIDIA DGX-1 Nodes – 264 V100 GPUs
  - 8x NVIDIA Tesla V100 SXM GPUs – NVLINK CubeMesh
  - 2x Intel Xeon 20 core GPUs
  - 512TB DDR4 System Memory
  - SSD – 7 TB scratch + 0.5 TB OS



http://nvidia.com/dgx1

- Mellanox 216 port EDR L1 and L2 switches
  - 4 ports per system
  - Full Fat tree topology

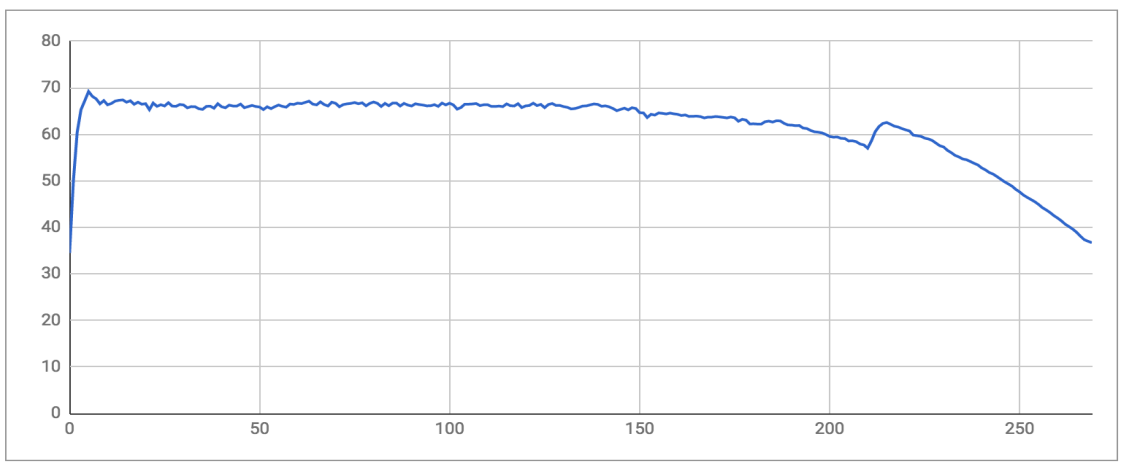
- Ubuntu 16.04, CUDA 9, OpenMPI 1.10.7
  - NVIDIA GPU BLAS + Intel MKL (NVIDIA GPU HPL)

- The HPL Team
  - Massimilano Fatica
  - Everett Phillips
  - Louis Capps
  - Craig Tierney
  - Allison Seidel
  - Doug Holt

NVIDIA DGX SATURNV TOP500

33 node Green 500 run and Level 2

- **HPL Setup**
  - Problem contained mainly in GPU memory (~16GB / GPU)
  - 33 nodes * 8 GPU/node * 16 GB mem/GPU = 15,872 GB mem --- N = 723456
- **Measurement**
  - Level 2 (partial level 3)
    - 2 PDUs per rack (compute, storage, switches)
    - PDU input power time-stamped during full run @ 1s intervals – ipmi reading from all PDUs
      - Log scripts kicked off at beginning and stoped at end of run
      - Coorelated time stamp of start/stop of core compute with logs
    - Measured all cluster hardware – nodes, switches, storage
- **Performance**
  - HPL Rpeak – 1,820 TF
  - HPL Rmax - 937 TF
  - Pwr avg – 62 KW



- **15.1 GF / Watt**

NVIDIA DGX SATURNV TOP500

Rank	TOP500 Rank	System	Cores	Rmax (TFlop/s)	Power (kW)	Power Efficiency (GFlops/watts)
1	259	**Shoubu system B** - ZettaScaler-2.2, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 , PEZY Computing / Exascaler Inc. Advanced Center for Computing and Communication, RIKEN Japan	794,400	842.0	50	17.009
2	307	**Suiren2** - ZettaScaler-2.2, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 , PEZY Computing / Exascaler Inc. High Energy Accelerator Research Organization /KEK Japan	762,624	788.2	47	16.759
3	276	**Sakura** - ZettaScaler-2.2, Xeon E5-2618Lv3 8C 2.3GHz, Infiniband EDR, PEZY-SC2 , PEZY Computing / Exascaler Inc. PEZY Computing K.K. Japan	794,400	824.7	50	16.657
4	149	**DGX SaturnV Volta** - NVIDIA DGX-1 Volta36, Xeon E5-2698v4 20C 2.2GHz, Infiniband EDR, NVIDIA Tesla V100 , Nvidia NVIDIA Corporation United States	22,440	1,070.0	97	15.113

SC17 SCC

Student Cluster Competition

- **Just in...Student Measurement**
- **Performance**
  - HPL Rmax - 51.8TF TF
  - Pwr avg – 3 KW

- **17.26 GF / Watt**

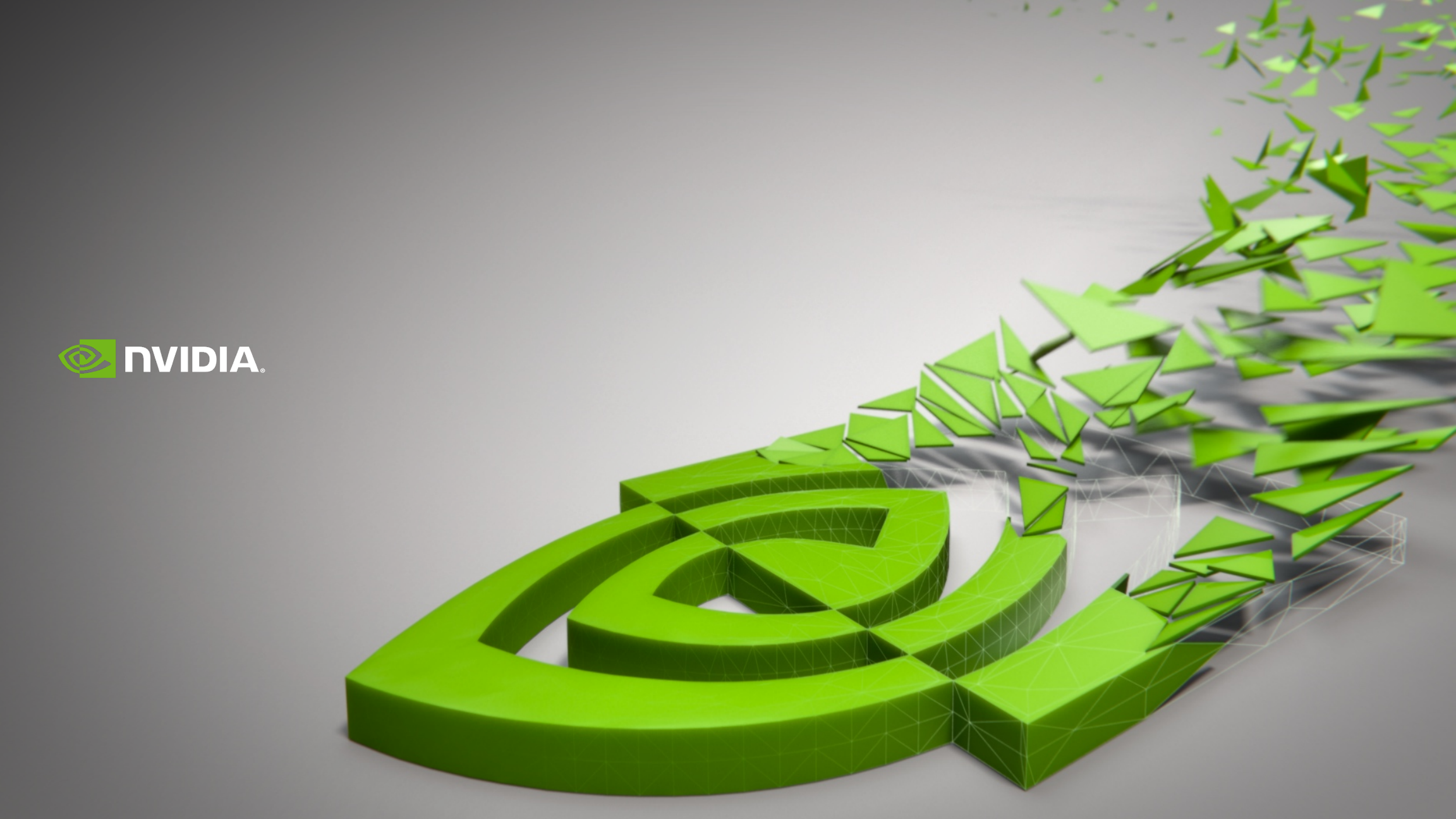
NVIDIA DGX SATURNV G500 RUN

33 node Green 500 run – Measurement Challenges

- **Green500 Level 2**
  - Mainly ensuring the power was partitioned correctly to measure
    - Includes all switches, storage, nodes
    - In some case, extra power was seen during run found to be nodes that were not part of the run (had to ensure these nodes are powered off)
- **Measurement**
  - Can be difficult to work with ipmi to read power devices
  - Starting and stopping logging to line up with runs can be difficult
    - Instrument in scripts (power monitor and averaging)

NVIDIA DGX SATURNV G500 RUN

- **Value of Green500 for post submission studies**
  - Found systems using too much power during idle
    - Study system components
  - CPU using too much power during low activity times
    - Review code and CPU power state activity
  - Tuning of power vs performance is critical for future


NVIDIA

JCAHPC

Green500 Measurement on Oakforest-PACS (JCAHPC) and Reedbush systems (ITC, UTokyo)

Masaaki Kondo and Toshihiro Hanawa

Joint Center for Advanced High Performance Computing /

Information Technology Center

The University of Tokyo

東京大学情報基盤センター
INFORMATION TECHNOLOGY CENTER, THE UNIVERSITY OF TOKYO

筑波大学
計算科学研究センター
Center for Computational Sciences

Oakforest-PACS

- Full Operation started on December 1, 2016
- 8,208 Intel Xeon/Phi (KNL), 25 PF Peak Performance
  - Fujitsu
- Nov. 2016: TOP500 #6 (#1 in JP), HPCG #3 (#2), Green500 #6 (#2)
- Jun. 2017: TOP500 #7 (#1 in JP), HPCG #5 (#2), Green500 #21
- Nov. 2017: TOP500 #9 (#2 in JP), HPCG #6 (#2), Green500 #22

- **JCAHPC: Joint Center for Advanced High Performance Computing)**
  - University of Tsukuba
  - University of Tokyo
    - The system was installed at Kashiwa-no-Ha (Leaf of Oak) Campus/U.Tokyo, which is located between Tokyo and Tsukuba
  - http://jcahpc.jp

JCAHPC

Computation node (Fujitsu PRIMERGY) with single chip Intel Xeon Phi 7250 (Knights Landing, 68 cores, 3+ TFLOPS) and Intel Omni-Path Architecture card (100Gbps)

Chassis with
8 nodes, 2U size

15 Chassis with 120 nodes per Rack

Spec:
4.24 MW (incl. Cooling)
3.44 MW (w/o Cooling)

Green500:
2.72 MW
13.55 PF => 4.98 GF/W
#6 (Nov.'16) => #21 => #22

Reedbush systems in ITC, U-Tokyo (not JCAHPC)

- Supercomputer System for Data Analyses & Scientific Simulations
  - HPE (ex-SGI)

	Reedbush-U	Reedbush-H	Reedbush-L
CPU	Intel Xeon E5-2695v4 (Broadwell-EP, 2.1GHz 18core) x 2		
Memory	256 GiB (153.6GB/sec)		
GPU	NA	NVIDIA Tesla P100 : (5.3TF, 720GB/sec, 16GiB, NVLink) x 2	NVIDIA Tesla P100 : (5.3TF, 720GB/sec, 16GiB, NVLink) x 4
Interconnect	IB EDR x 1ch	IB FDR x 2ch	IB EDR x 2ch
Topology	Full-bisection Fat-tree		
# of Nodes	420	120	64
FLOPS	508.0 TF	145.2 TF(CPU)+1.27 PF(GPU)= 1.42 PF	77.4 TF(CPU)+1.35 PF(GPU)= 1.43 PF
Operation	Jul. 2016	Mar. 2017	Oct. 2017

Green500: 199 kW, 459.8 TF
2.31 GF/W
#69 (Nov.'16) => #78 => n/a

Green500: 93.6 kW, 802.4 TF
8.57 GF/W
#11 (Jun. '17) => #16

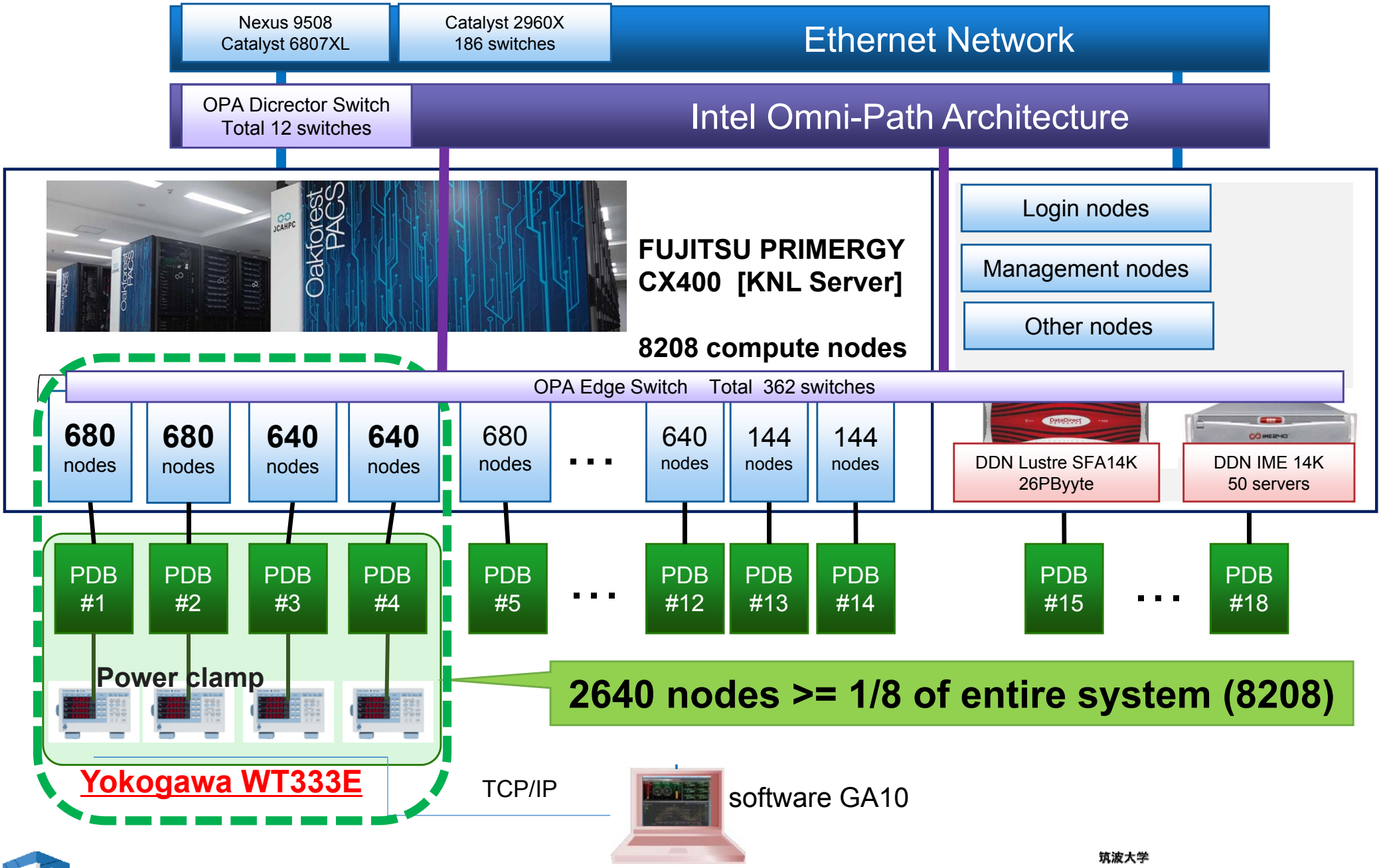
Green500: 79.0 kW, 805.6 TF
10.167 GF/W
#11 (Nov. '17)

Specification

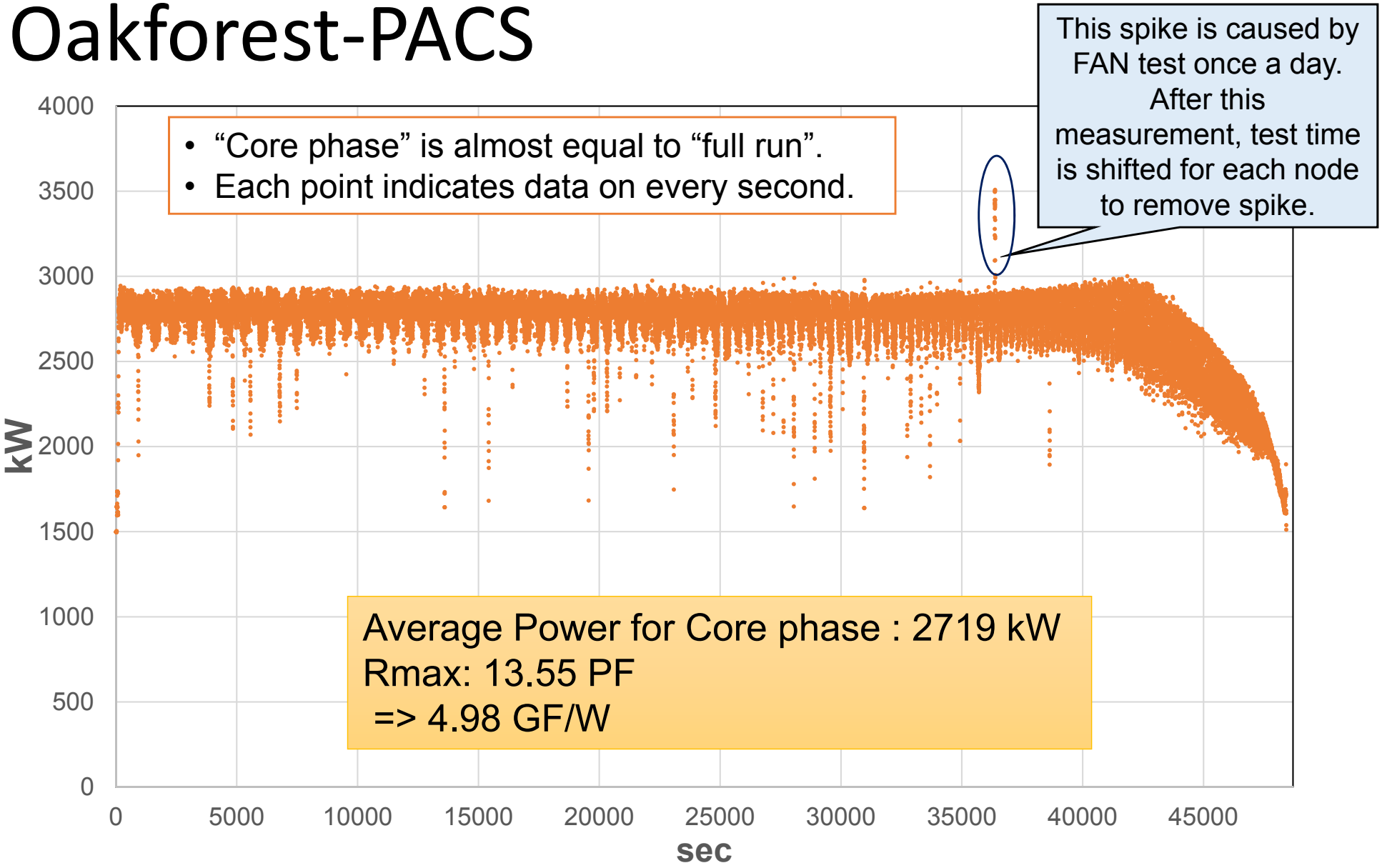
What we requested in the specification for procurement:

- To provide the functions which can monitor and record power consumption of entire system in real time.
- To enable Level 2/3 measurement based on the Energy Efficient High Performance Computing Power Measurement Methodology 2.0
  - Level 2 … Oakforest-PACS
  - Level 3 … Reedbush systems
- Actual measurement was done by vendors (Fujitsu for OFP, HPE for Reedbush)
- HPL program: by Fujitsu for OFP
  by NVIDIA for RB-H and RB-L

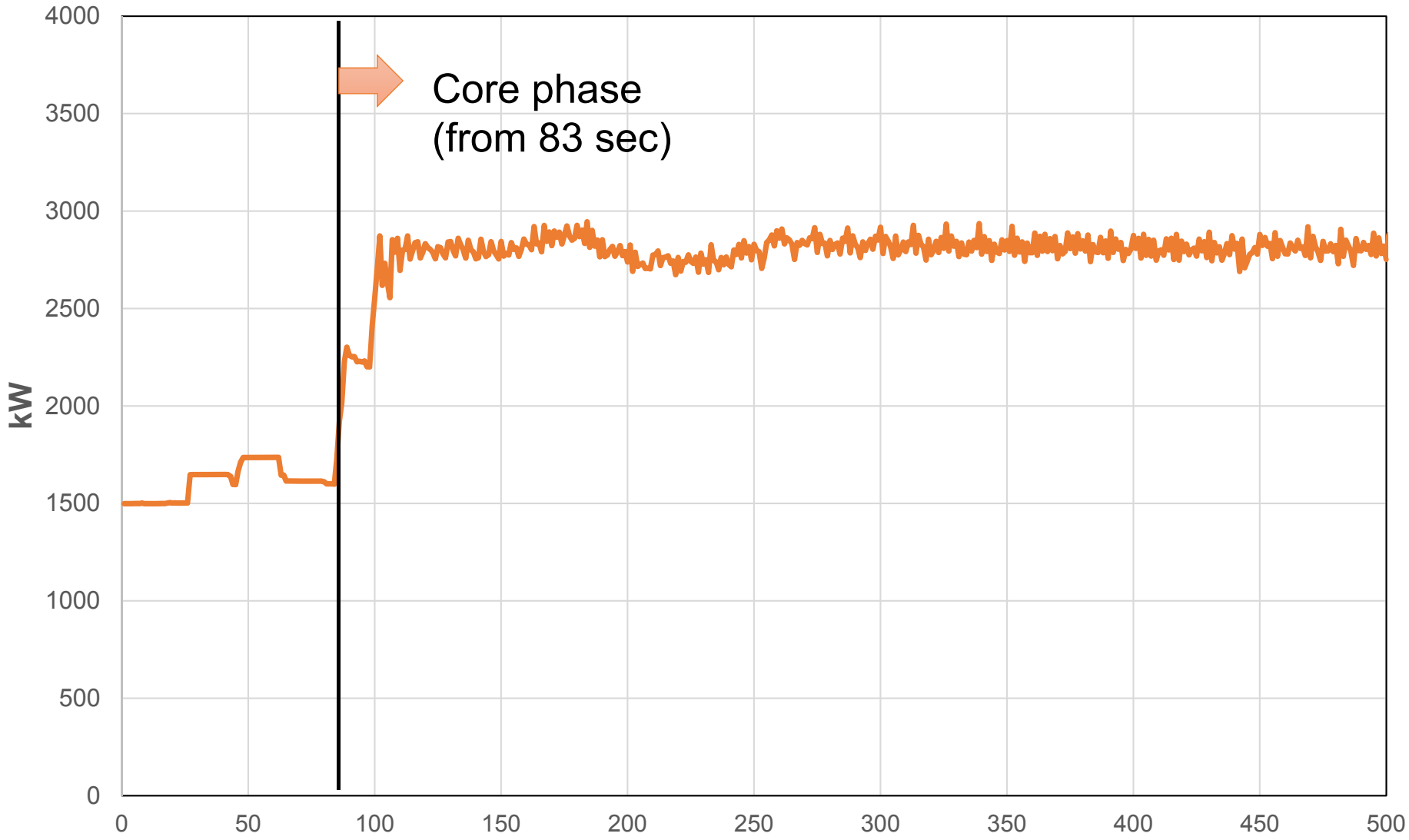
Power Monitoring for Oakforest-PACS



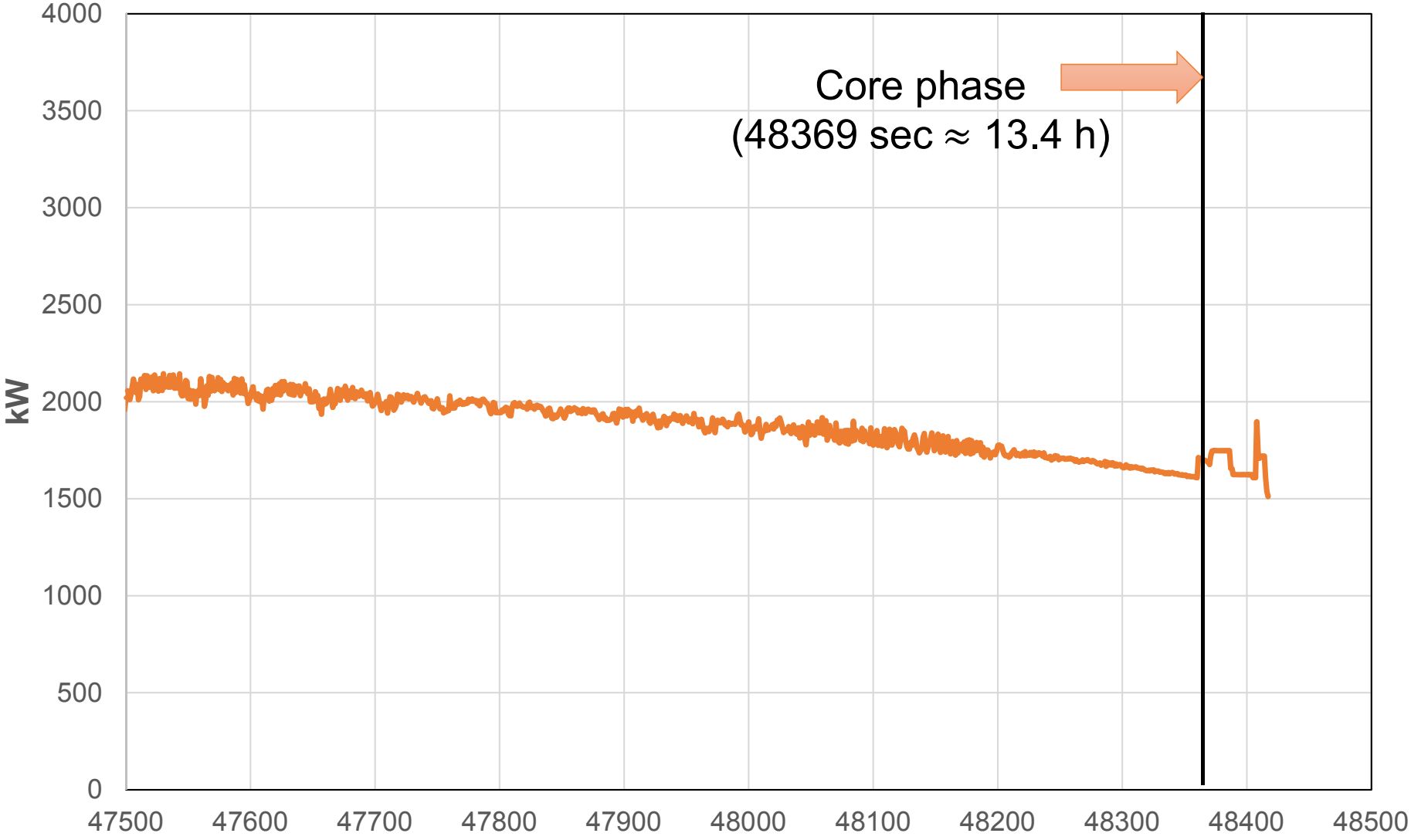
Transition of Power Consumption on Oakforest-PACS



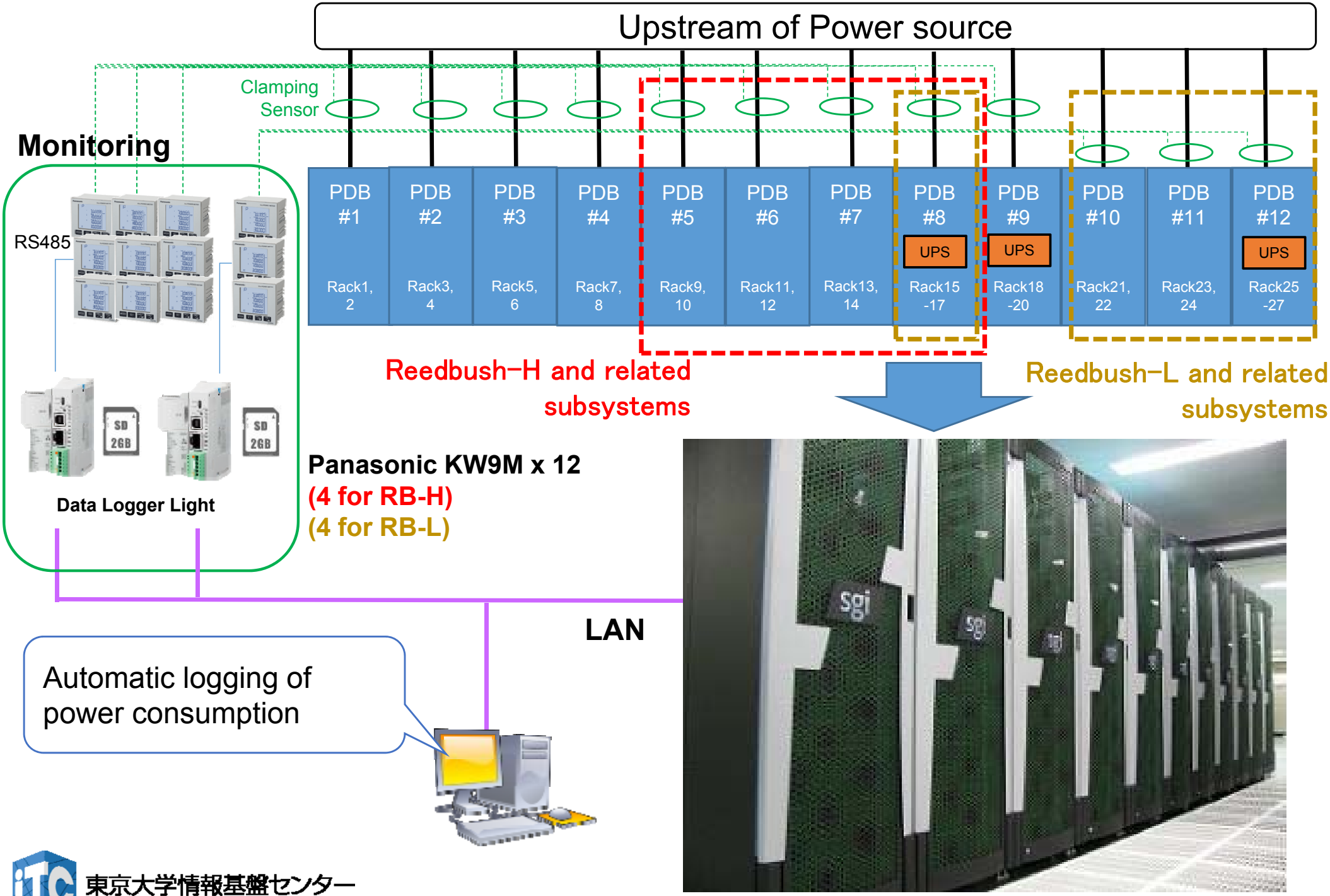
Detail on OFP (Startup stage)



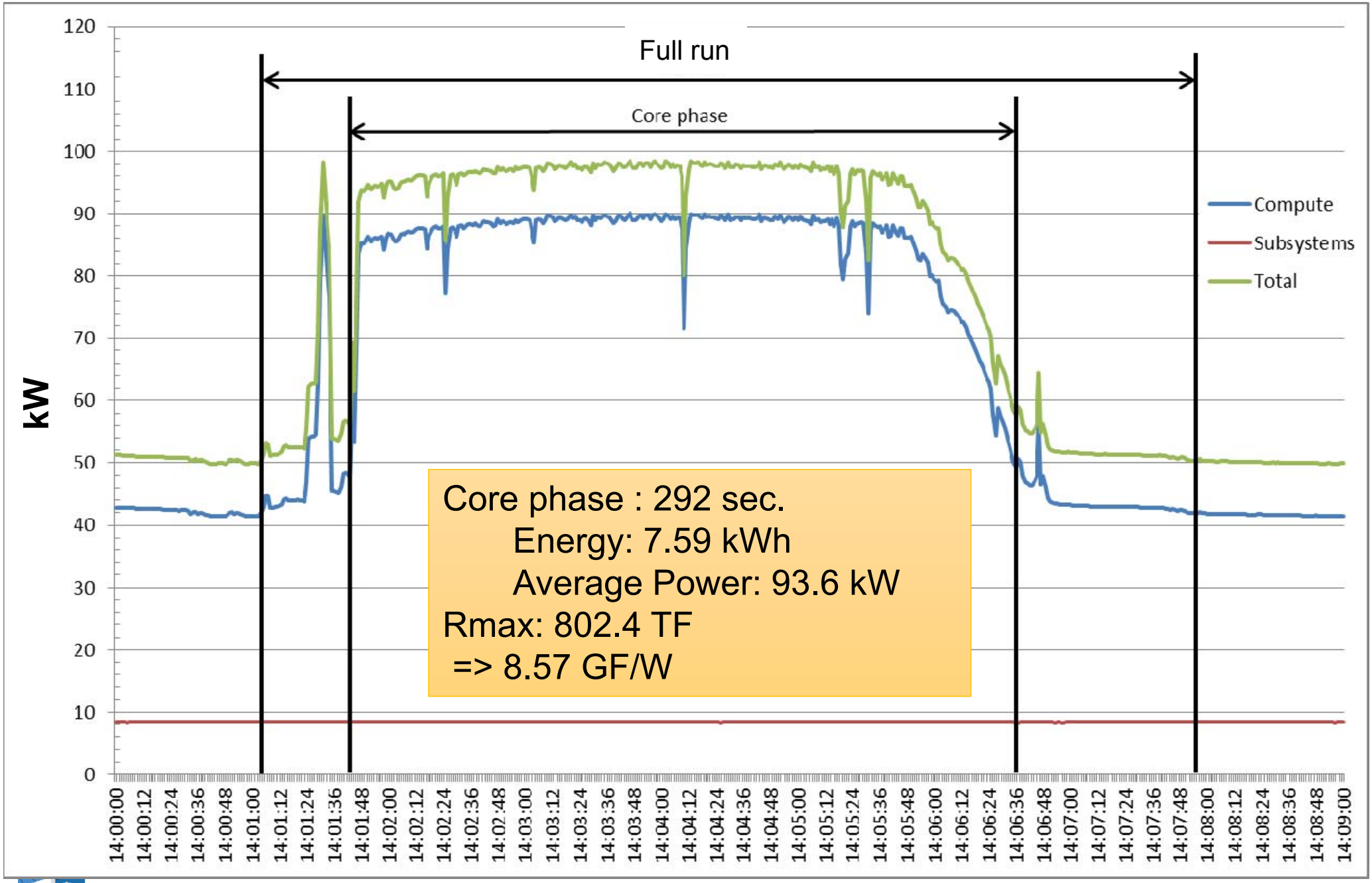
Detail on OFP (Final stage)



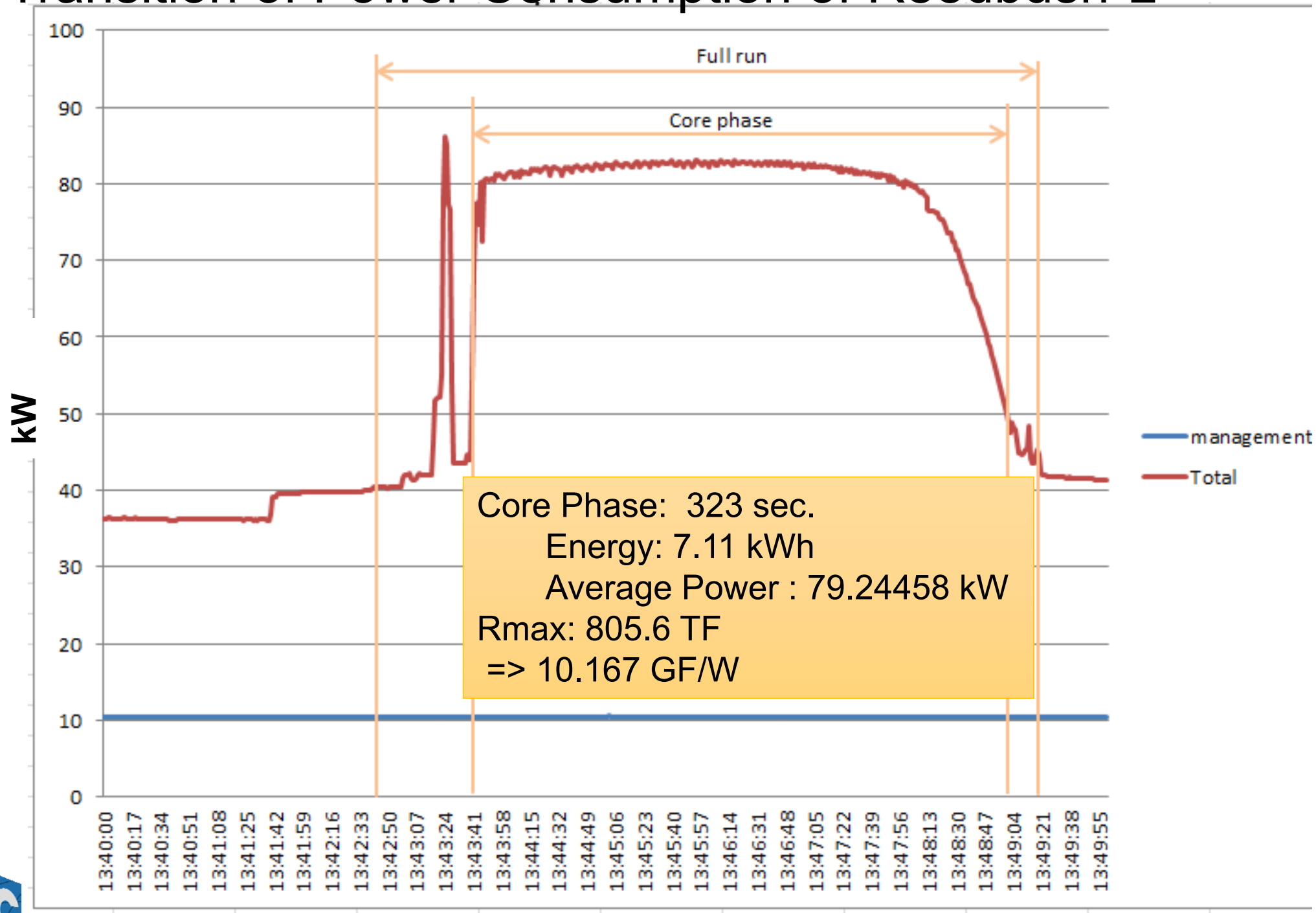
Power Monitoring for Reedbush-U, -H, and -L



Transition of Power Consumption on Reedbush-H



Transition of Power Consumption of Reedbush-L



Comments from Fujitsu

- Preparation
  - HW screening is important to avoid performance degradation
    - Performance check with small scale HPL among subgroups of OFP
- During measurement
  - Measurement takes 14 hours
  - Support systems are established for the rapid recovery
  - Power monitoring makes it easy to find out problems
    - Sudden power degradation indicates something wrong happens

Comments from HPE

- Configuration of power meters and monitoring system are important to reduce the number of meters and to balance the system load for each power distribution board
- HW screening is important to remove low performance components
- Parameter tuning for HPL is very complicated because it depends on the settings of HW, drivers, parameters of MPI
- Total measurement time was 30 hours, while one trial was up to 10 min using 120 nodes
- Many subsystems such as Infiniband-EDR switch are shared by Reedbush-U and –H
  -> Isolation of these subsystems was required for measurement but cumbersome

Why we (JCAHPC and U-Tokyo) used Level 2/3?

- They reflect more realistic system potentials
  - Typically Level 1 number becomes better than L2/L3
    - On level 1, power of interconnect can be estimated, and power of subsystem can be omitted
- Our philosophy:
  - Measure the system performance under the same condition, configuration, and assumptions as the product run phase as much as we can
- Comments to measurement methodology of Green500:
  - About the ranking of Reedbush-L
    - Perf/W in Top500: #10, but #11 in Green500
  - "Power optimized number" is allowed to submit for Green500 but no one knows this fact without seeing Excel sheet
    - ex.) Power optimized run on Reedbush-L
      => 770.0TFLops, 10615.92 Mflops/W, can be 8th in current Green500
  - Only 2 systems are measured by L2/L3 in Top 10 of Green500 but only few knows…

Thank you !!

Extra Slides

Amendments to the The GREEN 500 Run Rules

Natalie Bates, Wu Feng, and Erich Strohmaier

VirginiaTech
Invent the Future

SyNeRG
synergy.cs.vt.edu

Amendments to Level 1 (version 1.2 → 2.0)

- Requires a measurement of the full core phase
  (V1.2: Only 20%)
- Requires measurement or estimation of the network
  (V1.2: Only compute subsystem was measured)
- Requires a larger fraction of the system to be measured
  - The largest of the following: minimum of 2 kW of power (V1.2: 1kW); 1/10 of the system (V1.2: 1/64); or 15 nodes (V1 did not include a node minimum) *OR*
  - A measurement of at least 40 kW
    (V1.2: Did not include a minimum power limit) *OR*
  - *The entire system*
    (V1.2: Did not include entire system)

Amendments to Level 2 (version 1.2 → 2.0)

- Requires a larger fraction of the system to be measured
  - The largest of the following: minimum of 10 kW of power, 1/8 of the system, or 15 nodes *OR*
  - The entire system

Cross-Cutting Requirement: Power-Meter Accuracy

- Minimum guaranteed accuracies for power meters (1% for L3, 2% for L2, 5% for L1, relaxed constraints if multiple power meters are used)
- Synchronization between power measurement & core phase (L3 requires no more than 10 seconds to be not covered, L1 and L2 allow up to 10%)
- Minimum sampling rate for L3 AC raised to 5 kHz (was 120 Hz before)

Alternate Performance & Power Numbers

- Submission of alternate performance and power numbers is *allowed* to the Green500 but with the following constraints:
  - The same ***full machine*** that was used for the TOP500 run is used for the Green500 run.
  - The same ***problem size*** that was used for the TOP500 run is used for the Green500 run.