BDEC Japan Update

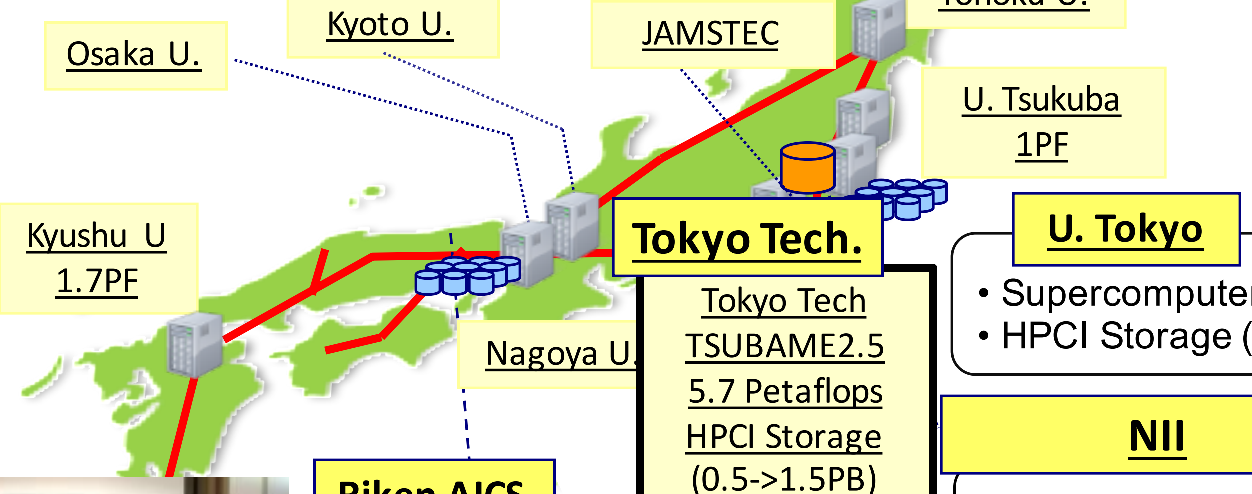
Satoshi Matsuoka (Tokyo Tech)
Yutaka Ishikawa (Riken AICS)

BDEC Meeting
2016/06/17
Frankfurt, Germany

Japan's High Performance Computing Infrastructure (HPCI) (Similar to PRACE)

HPCI: a nation-wide HPC infrastructure

- Supercomputers ~40 PFlops (2015 April)
- National Storage 22.5 PB HDDs + Tape (~14PB Used)
- Research Network (SINET4), 40+10GBps->SINET5 (2016) 200~400Gbps
- SSO (HPCI-ID), Distributed FS (Gfarm FS)
- National HPCI Allocation Process



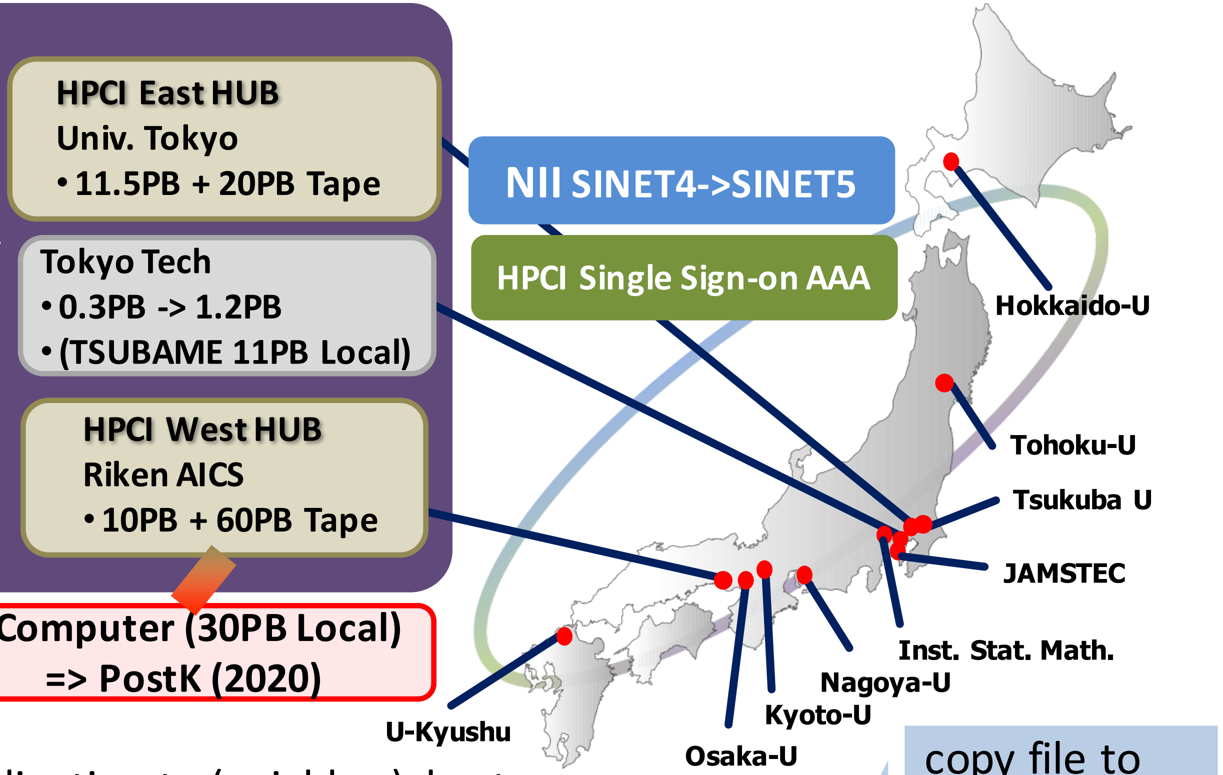
U. Tokyo
- Supercomputers 1.1PF
- HPCI Storage (12PB)

NII
- Management of SINET &Single sign-on

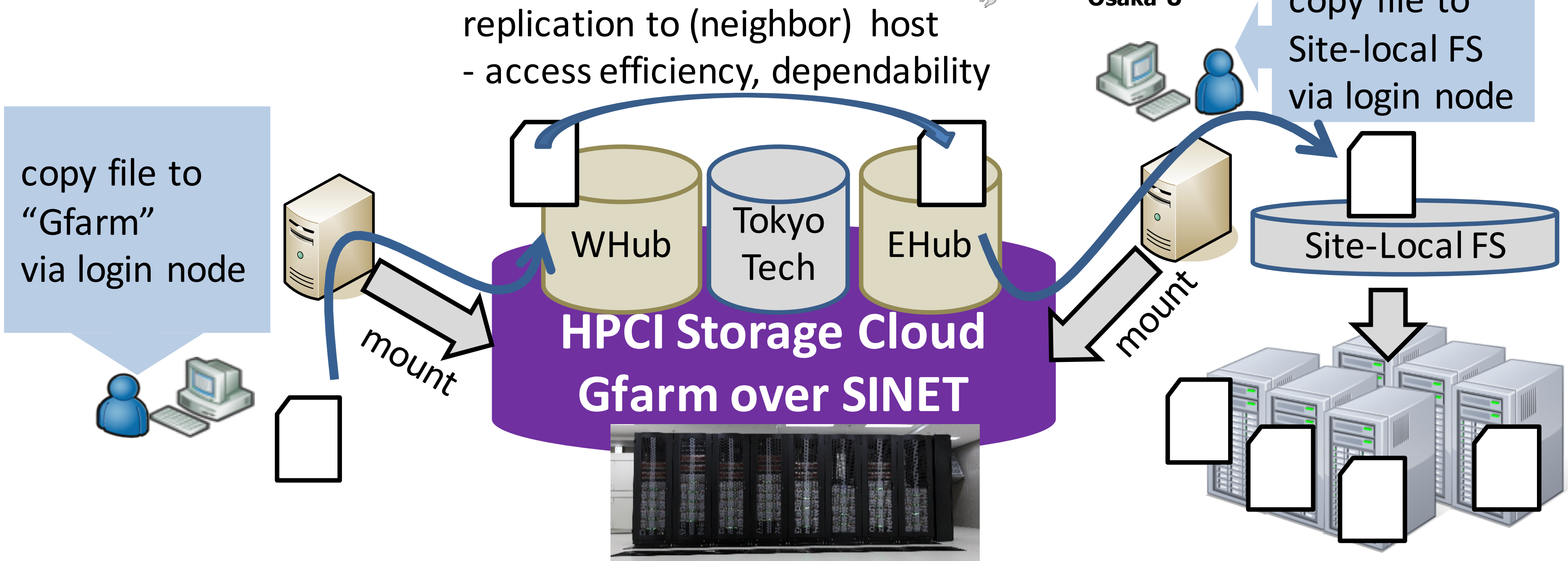
Riken AICS
- "K" computer 11Petaflops
- HPCI Storage (10PB)

HPCI Nationwide HPC Storage Cloud

- 21.8 PB (separate from local) *~70% full*
- High resiliency and availability
  - Redundant Servers · RAID6
  - Active Repair
- Multi-Tier Distributed Storage
  - Multi-vendor utility
  - ZABBIX, Ganglia
- Fault Detection & Information Sharing

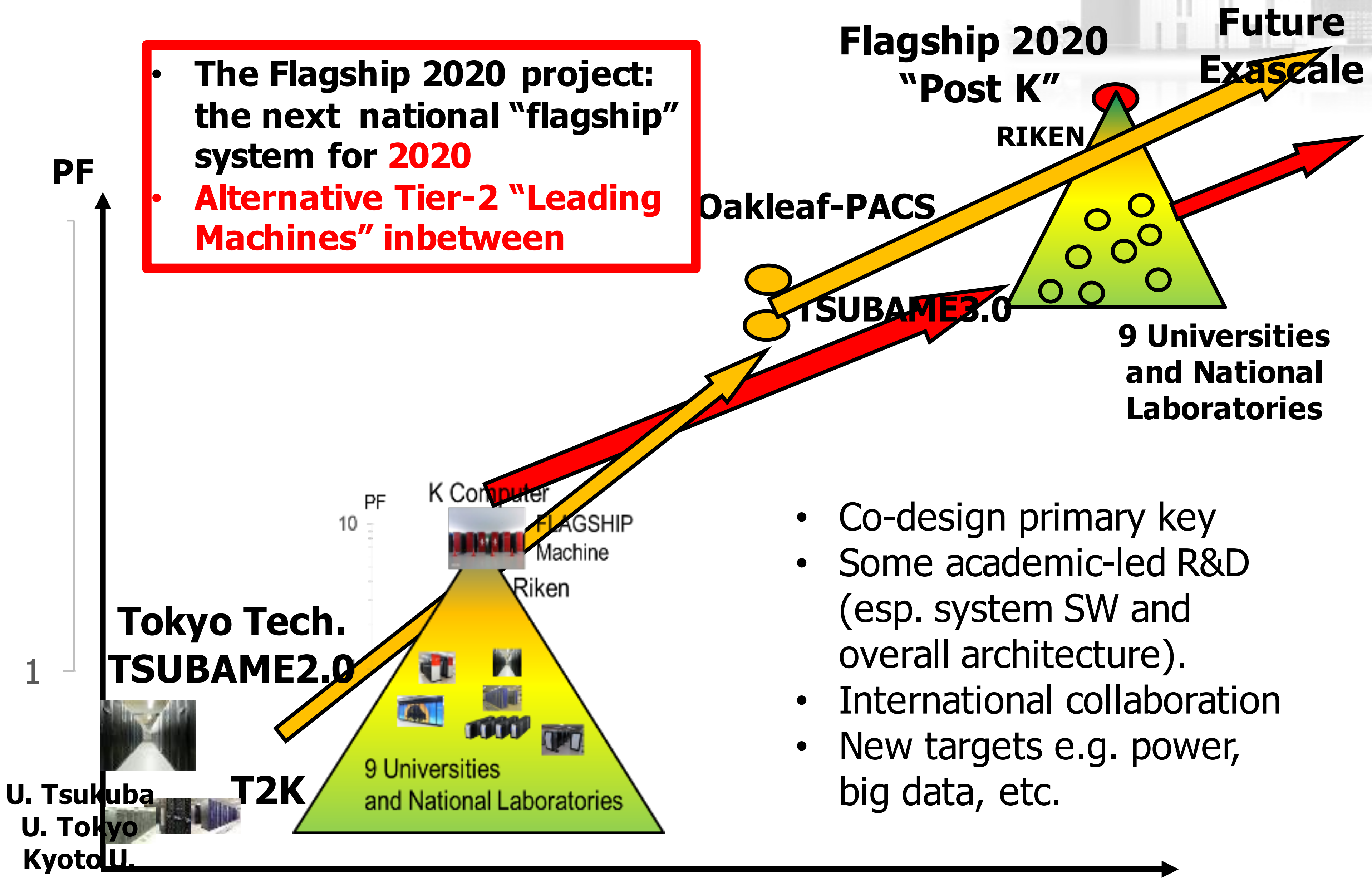


HPCI High Performance Computing Infrastructure



Towards the Next Flagship Machine & Beyond

- **The Flagship 2020 project: the next national "flagship" system for 2020**
- **Alternative Tier-2 "Leading Machines" inbetween**

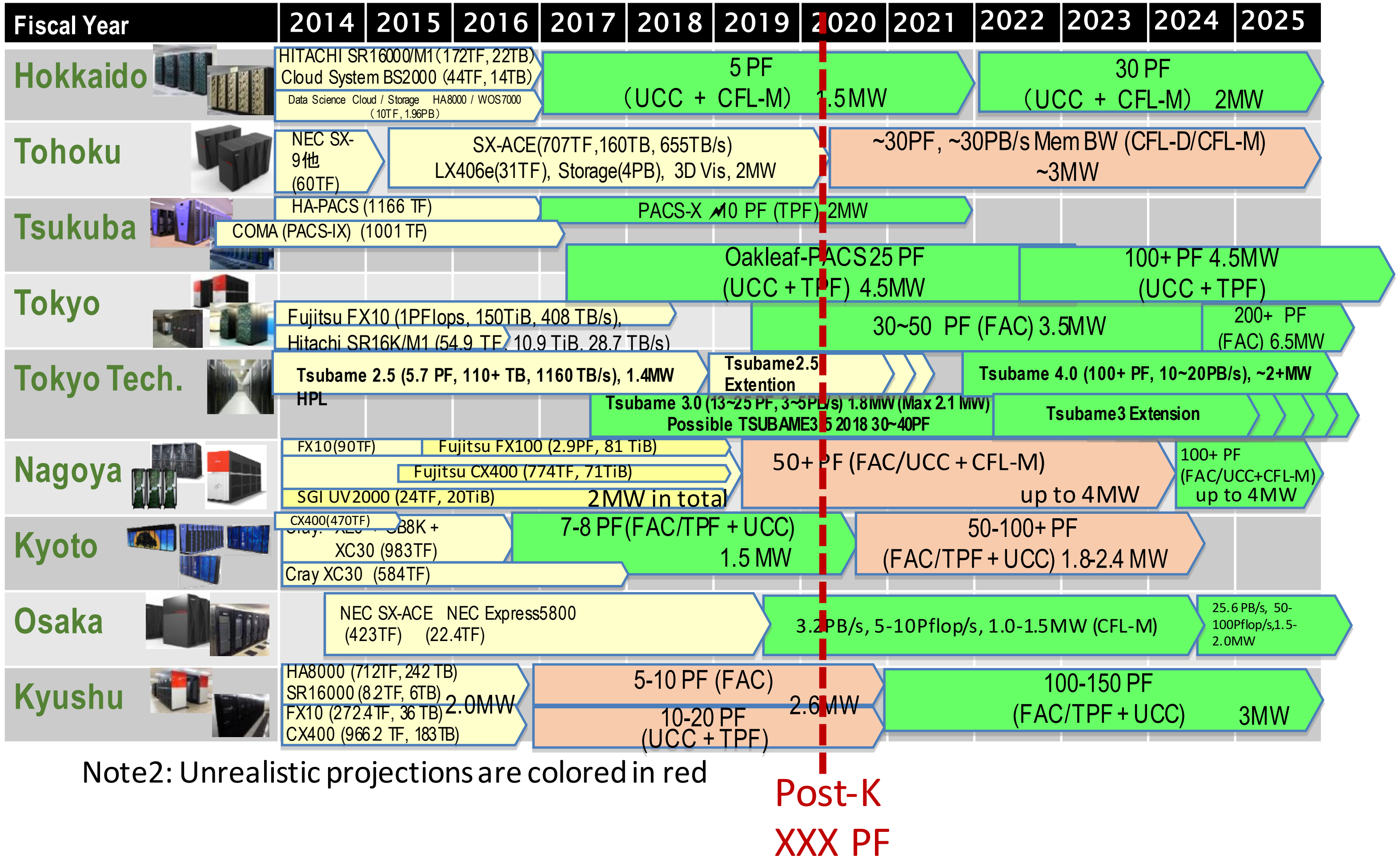


- Co-design primary key
- Some academic-led R&D (esp. system SW and overall architecture).
- International collaboration
- New targets e.g. power, big data, etc.

Japanese HPCI Centers Supercomputing Infrastructures Roadmap

(as of Mar. 2016, Tokyo Tech updated to Dec. 2015 plans)

NOTE: Year is Japanese fiscal year Apr~Mar. e.g. FY2014 is Apr 2014~Mar 2015



Note2: Unrealistic projections are colored in red

Flagship 2020 Project

- Dual mission
  - Develop the next Japanese flagship computer, tentatively called "post K"
  - Simultaneously develop a range of application codes, to run on the "post K", to help solve major societal and science issues

- Architecture: Many core processor
- Target performance: 100 times (maximum) of K by the capacity computing
  50 times (maximum) of K by the capability computing
- Power consumption of 30-40MW (cf. K computer: 12.7~20MW)

- Budget
  - 110 billion JPY (about 1.06 billion US$ if 1US$=104JPY) + Fujitsu 30 billion JPY + ~10 billion JPY/Year x 6 years
  - R&D + manufacturing of the post K system
  - Development of applications
  - Operations

Japan Flagship 2020 “Post K” Supercomputer

✓CPU
- A NEW many-core processor (NOT x86)
- Multi-hundred petaflops peak total
- Power Knob feature for saving power

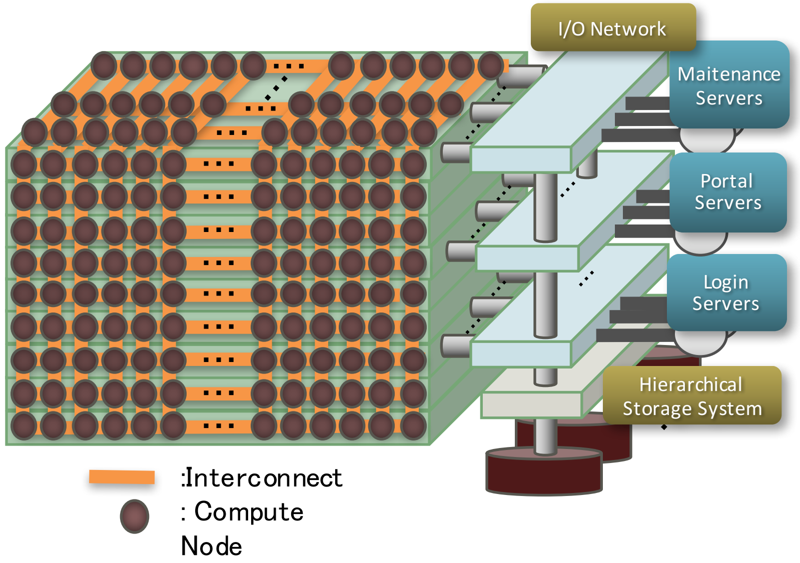
✓Memory
✓3-D stacked DRAM, Terabyte/s BW

✓Interconnect
- TOFU3 CPU-integrated 6-D torus network

- I/O acceleration
- 30MW+ Power
- Being designed and will be manufactured by Fujitsu

- Development Leaders: Yutaka Ishikawa, Mitsuhisa Sato (Riken)

Prime Minister Abe visiting K Computer 2013



Outline of system development

- **Science – driven System**
  - Basic design based on Priority Issues and Target applications
  - Application - System Co-design
- **Global Competitiveness**
  - Realize general purpose system which has ability to compete in oversea markets on the issues of computing performance, power-efficiency and cost
- **International Cooperation**
  - Strategic use of International Cooperation (e.g. system software)
- **Inheriting property of K computer**
  - Full use of Technologies, Human resources and Applications established by K computer, as a succession machine
- **Upgradable System**
  - Design which allows upgrade performance in response to progress of semiconductor technology after 2020

Schedule

Fiscal Year	2014	2015	2016	2017	2018	2019	2020
System	Basic Design		Trial Production Detailed Design		Manufacture	Installation Coordination	Operation
Application	Development and utilization of Application for Exascale Computing						

Application Co-Design Targets

- SPIRE (Strategic Programs for Innovative Research) Program for the K computer
  - The projects were organized around 2011.
- For Flagship2020,
  - A government committee (from academia and industry) is organized to identify "priority research areas" (9) and "frontier research areas"(5)
  - Accepted 9 proposals for priorty areas, about $2 million each / year
  - Frontier areas a few thousand $$$ / year, under review

Nine Priority Application Areas

①Innovative Drug Discovery

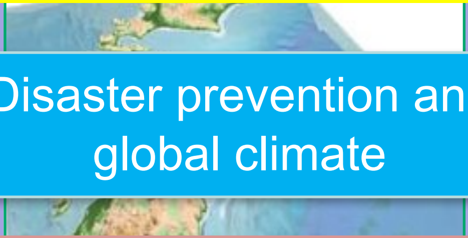
RIKEN Quant. Biology Center

②Personalized and Preventive Medicine

Inst. Medical Science, U. Tokyo

Society with health and longevity

③Hazard and Disaster induced by Earthquake and Tsunami



Earthquake Res. Inst., U. Tokyo

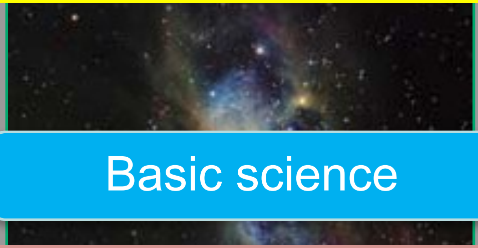
Disaster prevention and global climate

⑧ Innovative Design and Production Processes for the Manufacturing Industry in the Near Future

Inst. of Industrial Science, U. Tokyo

Industrial competitiveness

⑨Fundamental Laws and Evolution of the Universe



Cent. for Comp. Science, U. Tsukuba

Basic science

④Environmental Predictions with Observational Big Data

Center for Earth Info., JAMSTEC

⑦New Functional Devices and High-Performance

Inst. For Solid State Phys., U. Tokyo

⑥Innovative Clean Energy Systems

Grad. Sch. Engineering, U. Tokyo

⑤High-Efficiency Energy Creation, Conversion/Storage and Use

Inst. Molecular Science, NINS

Energy issues

Exploratory Application Areas – BDEC Affinity

Interactive Models of Socio-Economic Phenomena and their Applications

Frontiers of Basic Science - challenge to extremes -

Formation of exo-planets (second Earth) and Environmental Changes of Solar Planets

Mechanisms of Neural Circuits for Human Thoughts and Artificial Intelligence

New Application areas
➔ Use of other HPCI Resources such as TSUBAME and Oakleaf/U-Tokyo possible (esp. 2018- when K is decommissioned)

Proposals for exploratory areas are currently under examination

Co-design in the Post K development

Nine social & scientific priority issues and their R&D organizations have been selected from the following point of view:

- High priority issues from a social and national viewpoint
- Promising creation of world-Leading achievement
- Promising strategic use of post K computer

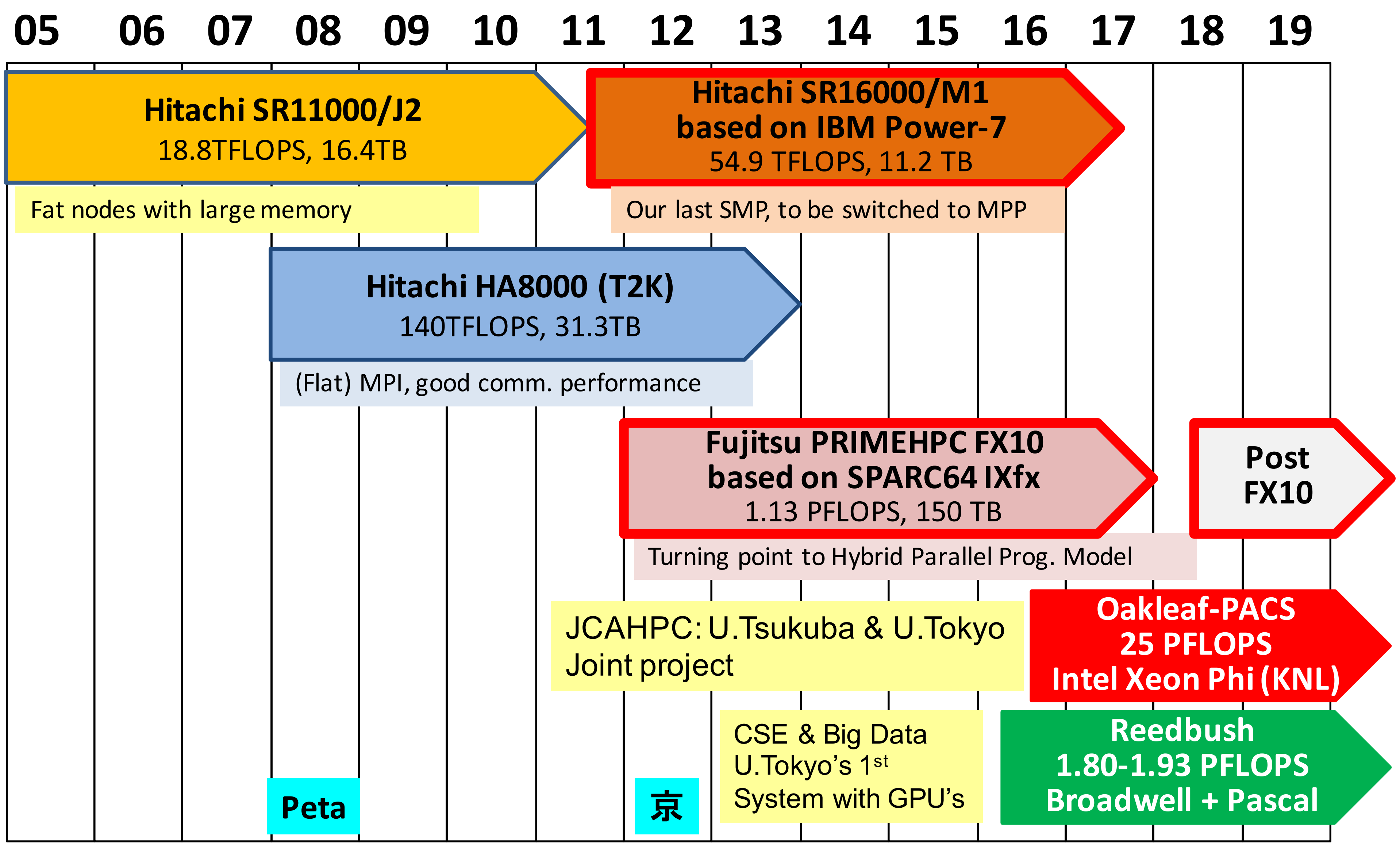
	Target Application	
	Program	Brief description
①	GENESIS	MD for proteins
②	Genomon	Genome processing (Genome alignment)
③	GAMERA	Earthquake simulator (FEM in unstructured & structured grid)
④	NICAM+LETK	Weather prediction system using Big data (structured grid stencil & ensemble Kalman filter)
⑤	NTChem	molecular electronic (structure calculation)
⑥	FFB	Large Eddy Simulation (unstructured grid)
⑦	RSDFT	an ab-initio program (density functional theory)
⑧	Adventure	Computational Mechanics System for Large Scale Analysis and Design (unstructured grid)
⑨	CCS-QCD	Lattice QCD simulation (structured grid Monte Carlo)

Basic Design Verification Review for the "post K"

- OVERVIEW
  - The verification review for "post K" concluded that its basic architectural design would make progress toward the state of the art system and its objectives.
  - Those are:
    - Solving major social/scientific problems
    - Archiving competitiveness internationally
- CO-DESIGN
  - In constant dialogue/discussion, Co-design has been working successfully. Having had mutual commitment between applications and architecture, the performance of "post K" would be enhanced
- REMARKS
  - Need improvements on:
    - Power consumption (GF/W)
    - Effective performance of the target applications (max seedup to exceed 100 times higher than the K computer's performance)

Supercomputers in ITC/U.Tokyo+Tsukuba U

2 big systems, 6 yr. cycle



JCAHPC

- Joint Center for Advanced High Performance Computing (http://jcahpc.jp)
（最先端共同HPC基盤施設）
- Very tight collaboration for “post-T2K” with two universities
  - For main supercomputer resources, *uniform specification* to ***single shared system***
  - Each university is financially responsible to introduce the machine and its operation
    -> unified procurement toward single system with ***largest scale in Japan***
  - To manage everything smoothly, a joint organization was established
    -> JCAHPC

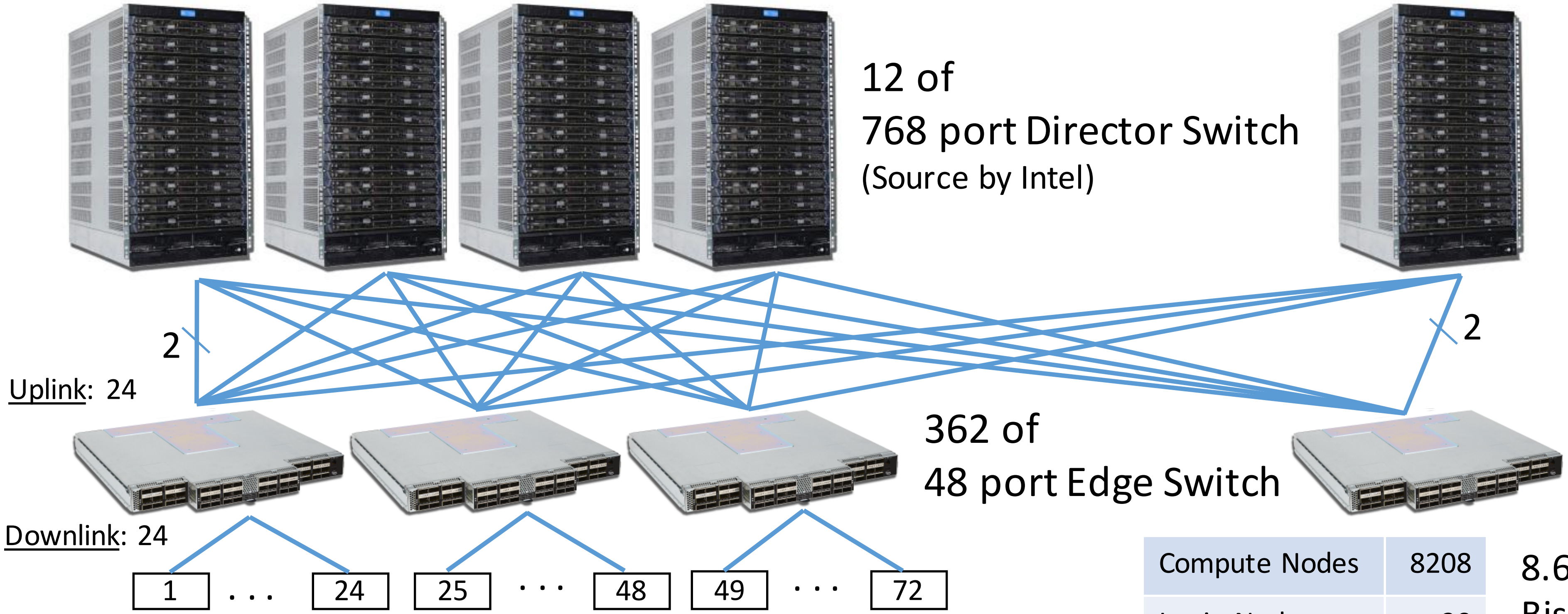
(pre) Photo of Oakleaf-PACS computation node

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

Chassis with 4 nodes, 2U size

Manufactured
by Fujitsu

Full bisection bandwidth Fat-tree by Intel® Omni-Path Architecture



Firstly, to reduce switches&cables, we considered :

- All the nodes into subgroups are connected with FBB Fat-tree
- Subgroups are connected with each other with >20% of FBB

But, HW quantity is not so different from globally FBB, and globally FBB is preferred for flexible job management.

Compute Nodes	8208
Login Nodes	20
Parallel FS	64
IME	300
Mgmt, etc.	8
Total	**8600**

8.6Tb Bisection BW

> CISCO projection on Global Intra-IDC BW circa 2016

JCAHPC

Specification of Oakforest-PACS system

Total peak performance			25 PFLOPS
Total number of compute nodes			8,208
Compute node	Product		Fujitsu Next-generation PRIMERGY server for HPC (under development)
	Processor		Next-generation of Intel® Xeon Phi™ (Code name: Knights Landing), >60 cores
	Memory	High BW	16 GB, > 400 GB/sec (MCDRAM, effective rate)
		Low BW	96 GB, 115.2 GB/sec (DDR4-2400 x 6ch, peak rate)
Inter-connect	Product		Intel® Omni-Path Architecture
	Link speed		100 Gbps
	Topology		Fat-tree with (completely) full-bisection bandwidth
Login node	Product		Fujitsu PRIMERGY RX2530 M2 server
	# of servers		20
	Processor		Intel Xeon E5-2690v4 (2.6 GHz 14 core x 2 socket)
	Memory		256 GB, 153 GB/sec (DDR4-2400 x 4ch x 2 socket)

Specification of Oakforest-PACS system (I/O)

Parallel File System	Type		Lustre File System
	Total Capacity		26.2 PB
	Meta data	Product	DataDirect Networks MDS server + SFA7700X
		# of MDS	4 servers x 3 set
		MDT	7.7 TB (SAS SSD) x 3 set
	Object storage	Product	DataDirect Networks SFA14KE
		# of OSS (Nodes)	10 (20)
		Aggregate BW	500 GB/sec
Fast File Cache System	Type		Burst Buffer, Infinite Memory Engine (by DDN)
	Total capacity		940 TB (NVMe SSD, including parity data by erasure coding)
	Product		DataDirect Networks IME14K
	# of servers (Nodes)		25 (50)
	Aggregate BW		1,560 GB/sec

U-Tokyo Reedbush

- SGI was awarded (Mar. 22, 2016)
- Compute Nodes (CPU only): Reedbush-U
  - Intel Xeon E5-2695v4 (Broadwell-EP, 2.1GHz 18core,) x 2socket (1.210 TF), 256 GiB (153.6GB/sec)
  - InfiniBand EDR, Full bisection Fat-tree
  - Total System: 420 nodes, 508.0 TF
- Compute Nodes (with Accelerators): Reedbush-H
  - Intel Xeon E5-2695v4 (Broadwell-EP, 2.1GHz 18core) x 2socket, 256 GiB (153.6GB/sec)
  - NVIDIA Pascal GPU (Tesla P100)
    - (4TF, 1TB/sec, 16GiB) x 2 / node
  - InfiniBand FDR x 2ch (for ea. GPU), Full bisection Fat-tree
  - 120 nodes, 145.2 TF(CPU)+960 TF(GPU)= 1.1 PF

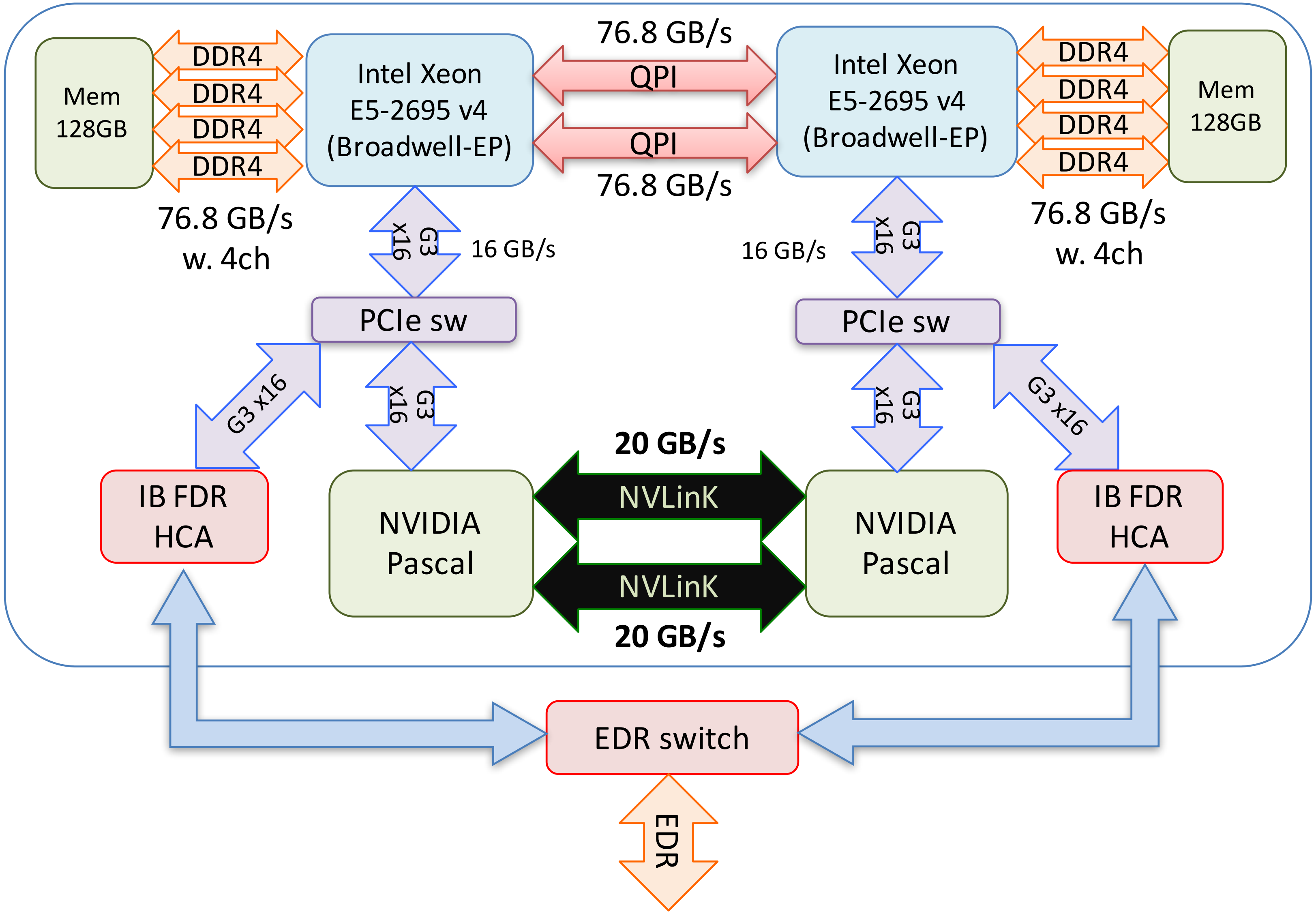
Why “Reedbush” ?

Blaise Pascal
(1623-1662)

- L'homme est un roseau pensant.
- Man is a thinking reed.
- 人間は考える葦である

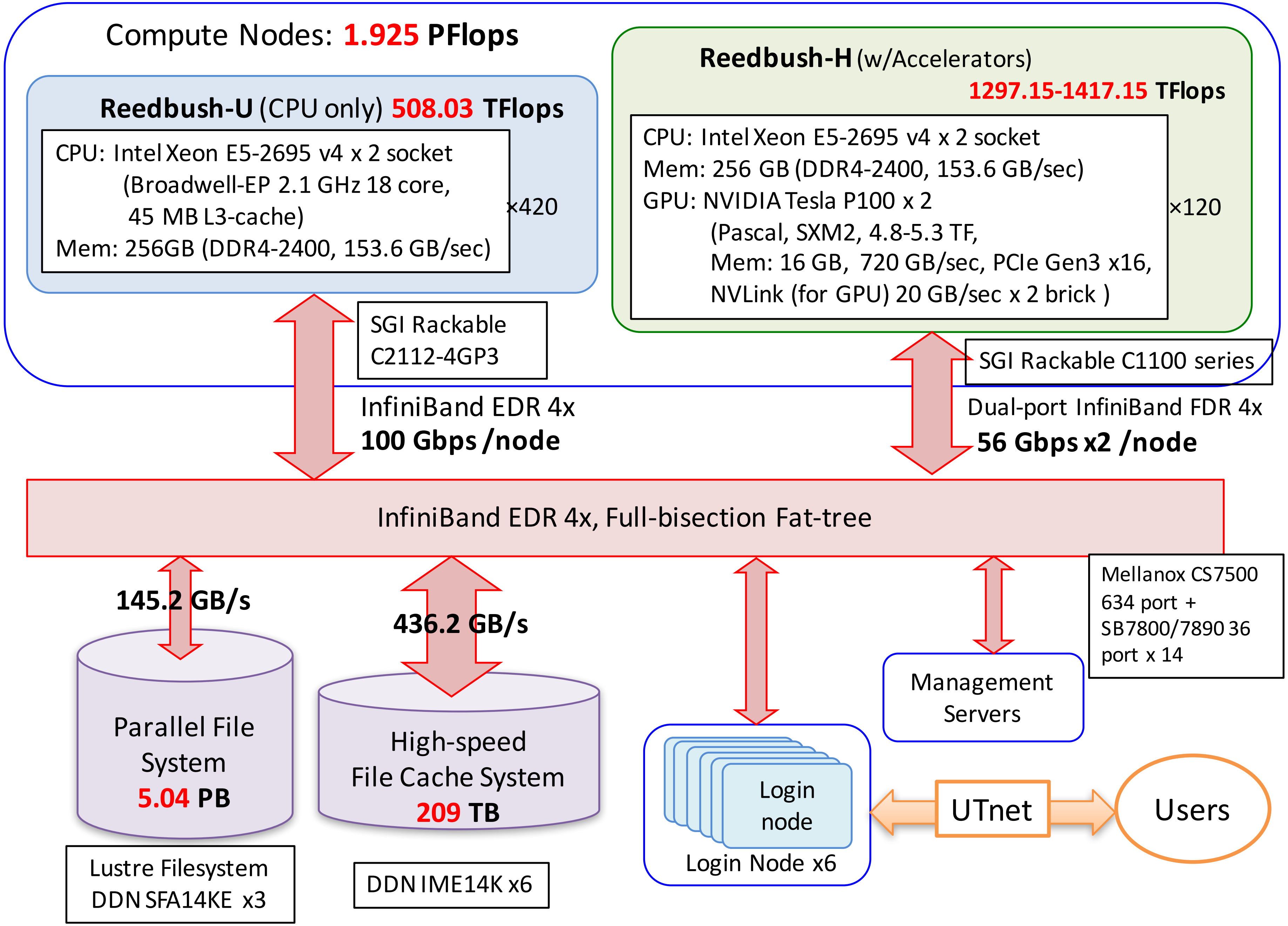
Pensées (Blaise Pascal)

Configuration of Each Compute Node of Reedbush-H



Reedbush (Mini PostT2K) (2/2)

- Storage/File Systems
  - Shared Parallel File-system (Lustre)
    - 5.04 PB, 145.2 GB/sec
  - Fast File Cache System: Burst Buffer (DDN IME (Infinite Memory Engine))
    - SSD: 209.5 TB, 450 GB/sec
- Power, Cooling, Space
  - Air cooling only, < 500 kVA (without A/C): 378 kVA
  - < 90 m^2
- Software & Toolkit for Data Analysis, Deep Learning …
  - OpenCV, Theano, Anaconda, ROOT, TensorFlow
  - Torch, Caffe, Cheiner, GEANT4



Compute Nodes: 1.925 PFlops
Reedbush-U (CPU only) 508.03 TFlops
CPU: Intel Xeon E5-2695 v4 x 2 socket
(Broadwell-EP 2.1 GHz 18 core,
45 MB L3-cache)
Mem: 256GB (DDR4-2400, 153.6 GB/sec)
×420
Reedbush-H (w/Accelerators)
1297.15-1417.15 TFlops
CPU: Intel Xeon E5-2695 v4 x 2 socket
Mem: 256 GB (DDR4-2400, 153.6 GB/sec)
GPU: NVIDIA Tesla P100 x 2
(Pascal, SXM2, 4.8-5.3 TF,
Mem: 16 GB, 720 GB/sec, PCIe Gen3 x16,
NVLink (for GPU) 20 GB/sec x 2 brick)
×120
SGI Rackable C2112-4GP3
InfiniBand EDR 4x
100 Gbps /node
SGI Rackable C1100 series
Dual-port InfiniBand FDR 4x
56 Gbps x2 /node
InfiniBand EDR 4x, Full-bisection Fat-tree
145.2 GB/s
436.2 GB/s
Mellanox CS7500 634 port + SB7800/7890 36 port x 14
Parallel File System
5.04 PB
High-speed File Cache System
209 TB
Management Servers
Login node
Login Node x6
UTnet
Users
Lustre Filesystem DDN SFA14KE x3
DDN IME14K x6

Tsubame current & future plans

- TSUBAME 2.5 (Production) Sep. 2013 – Mar 2019 (and beyond)
  - TSUBAME2.0 Nov. 2010-Sep. 2013, upgrade M2050 GPU -> K20X
  - 1424 nodes / 4224 GPUs, to be reduced to ~1300 nodes upon TSUBAME3 deployment
  - 5.7Petaflops (DFP), 17.1Petaflops (SFP)
- TSUBAME-KFC/DL (experimental, T3 Proto) – Oct 2013 – Sep 2018
  - Upgrade to KFC/DL Oct. 2015 K20X GPU -> K80 GPU
  - 42 nodes / 336 GPU chips, 0.5/1.5 PF DFP/SFP
  - Oil immersion, ambient cooling, PUE < 1.09
- TSUBAME 3.0 (Production) beginning of Q3 2017 ~2021 (and beyond)
  - 13~25 Petaflops DFP depending on funding
  - Parallel production to TSUBAME2.5
  - Focus on BD / AI workloads, not just traditional HPC => ~100PF max for AI combined with 2.5
- New IDC space construction for Tsubame3 and staggered operations beyond (T3+T4)
  - Power (4MW) + ambient cooling + storage (up to 100PB HDD) + high floor load (> 1 Ton / m^2)
  - To be completed March 2017
  - Power/Energy minimization for joint op in development

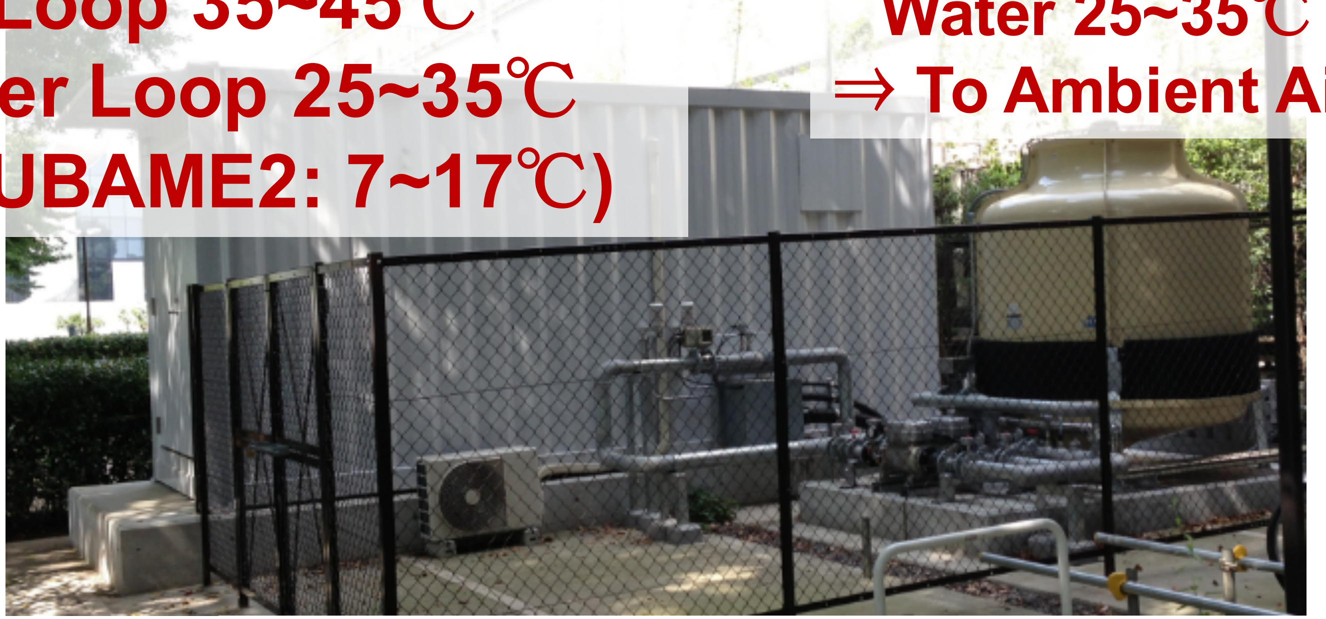
25

TSUBAME-KFC/DL: TSUBAME3 Prototype [ICPADS2014]

Oil Immersive Cooling＋ Hot Water Cooling + High Density Packaging + Fine-Grained Power Monitoring and Control, *upgrade to /DL Oct. 2015*

High Temperature Cooling
Oil Loop 35~45℃
⇒ Water Loop 25~35℃
(c.f. TSUBAME2: 7~17℃)

Cooling Tower:
Water 25~35℃
⇒ To Ambient Air



Single Rack High Density Oil Immersion
168 NVIDIA K80 GPUs + Xeon
413+TFlops (DFP)
1.5PFlops (SFP)

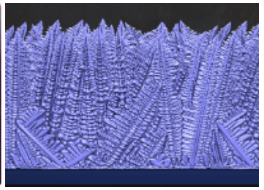
Experimental Container Facility
20 feet container (16m²)
Fully Unmanned Operation

Mid-2017 TSUBAME3.0 Towards Exa & Big Data

1. "Everybody's Supercomputer" – High Performance (15~20 Petaflops, ~4PB/s Mem, ~1Pbit/s NW), innovative high cost/performance packaging & design, in mere 100m^2…
2. "Extreme Green" – 9~10GFlops/W power-efficient architecture, system-wide power control, advanced cooling, future energy reservoir load leveling & energy recovery
3. "Big Data/AI Convergence" – Extreme high BW &capacity, deep memory hierarchy, extreme I/O acceleration, Big Data SW Stack, focus on AI/ML /DNN, graph processing, …
4. "Cloud SC" – dynamic deployment, container-based node co-location & dynamic configuration, resourc elasticity, assimilation of public clouds…
5. "Transparency" - full monitoring & user visibility of machine & job state, accountability via reproducibility

2006 TSUBAME1.0
80 Teraflops, #1 Asia #7 World
"Everybody's Supercomputer"

2010 TSUBAME2.0
2.4 Petaflops #4 World
"Greenest Production SC"

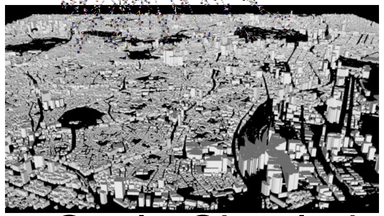


2011 ACM Gordon Bell Prize

2013
TSUBAME2.5
upgrade
5.7PF DFP
/17.1PF SFP
20% power
reduction

2013 TSUBAME-KFC
#1 Green 500

2017 TSUBAME3.0
13~25PF(DFP) 2~4PB/s Mem BW
9~10GFlops/W power efficiency
Big Data & AI Convergence



Large Scale Simulation
Big Data Analytics
Industrial Apps

27

Tremendous Recent Rise in Interest by the Japanese Government on Big Data, DL, AI, and IoT

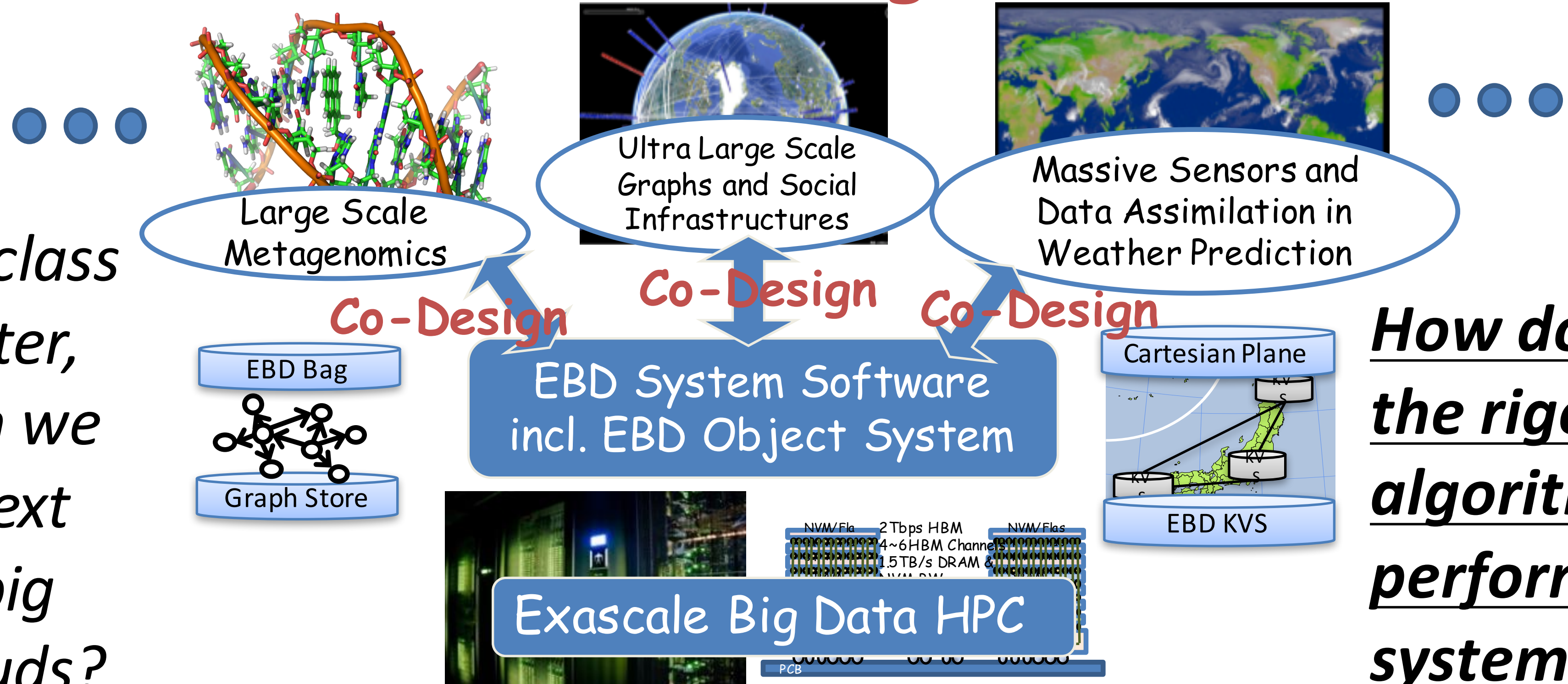
- Three projects and centers on Big Data and AI launched by three competing Ministries for FY 2016 (Apr 2016-)
  - MEXT – AIP (Artificial Intelligence Platform): Riken and other institutions ($~50 mil)
    - A separate Post-K related AI funding as well.
  - METI – AIRC (Artificial Intelligence Research Center): AIST (AIST internal budget + $~8 mil)
  - MOST – Universal Communication Lab: NICT ($50~55 mil)
  - $1 billion commitment on inter-ministry AI research over 10 years
- However, lack of massive platform and expertise in parallel computing c.f. Google, FB, Baidu…
  - MEXT attempts to suggest use of K computer
    -> community revolt "we want to use lots of GPUs like Google!"
  - MEXT Vice Minister Sadayuki Tsuchiya himself visits Matsuoka at Tokyo Tech Feb 1st, 2016.
    - "What is GPU and why is it so good for DL/AI?"
    - "Can you and TSUBAME can contribute to the MEXT projects directly over multiple years, with appropriate funding?"
  - Similar talks with METI & AIRC
    - "Can TSUBAME be utilized to cover the necessary research workload at AIRC?" --- Satoshi Sekiguchi, Director of Informatics, AIST

28

JST-CREST “Extreme Big Data” Project (2013-2018)

Future Non-Silo Extreme Big Data Scientific Apps

Given a top-class supercomputer, how fast can we accelerate next generation big data c.f. Clouds?



<u>How do we bring the rigor of HPC algorithms, performance and systems research into Big Data / AI?</u>

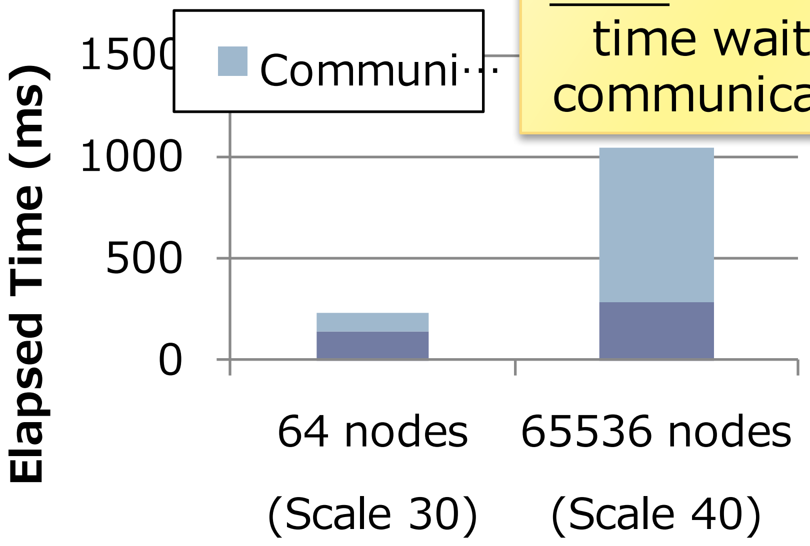
Convergent Architecture (Phases 1~4)
Large Capacity NVM, High-Bisection NW

Cloud IDC
Very low BW & Efficiency
Highly available, resilient

Supercomputers
Compute&Batch-Oriented
More fragile

The Graph500 – June 2014 and June 2015

K Computer #1 Tokyo Tech[EBD CREST] Univ. Kyushu [Fujisawa Graph CREST], Riken AICS, Fujitsu



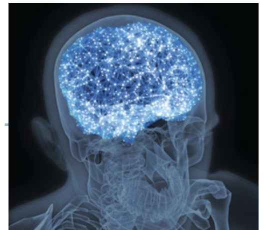
73% total exec time wait in communication

88,000 nodes,
700,000 CPU Cores
1.6 Petabyte mem
20GB/s Tofu NW

LLNL-IBM Sequoia
1.6 million CPUs
1.6 Petabyte mem

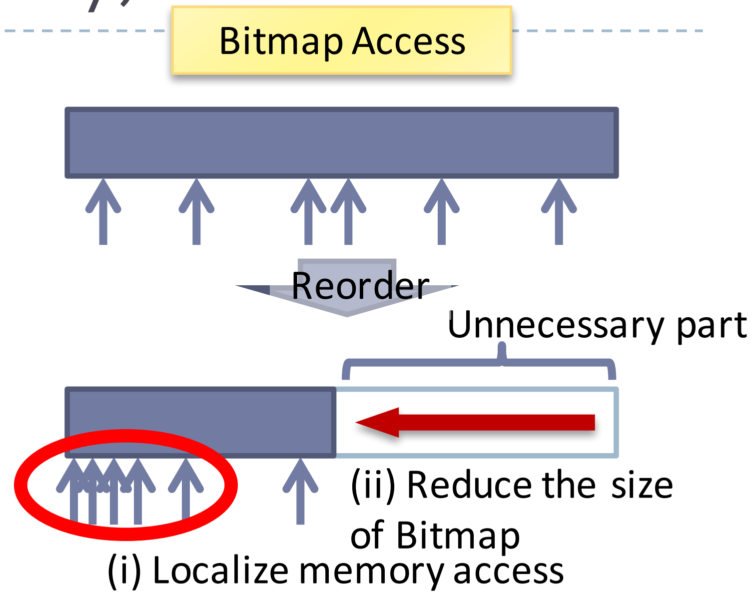
List	Rank	GTEPS	Implementation
November 2013	4	5524.12	Top-down only
June 2014	1	17977.05	**Efficient hybrid**
November 2014	2		**Efficient hybrid**
June 2015	**1**	**38621.4**	**Hybrid + Node Compression**

*Problem size is weak scaling "Brain-class" graph



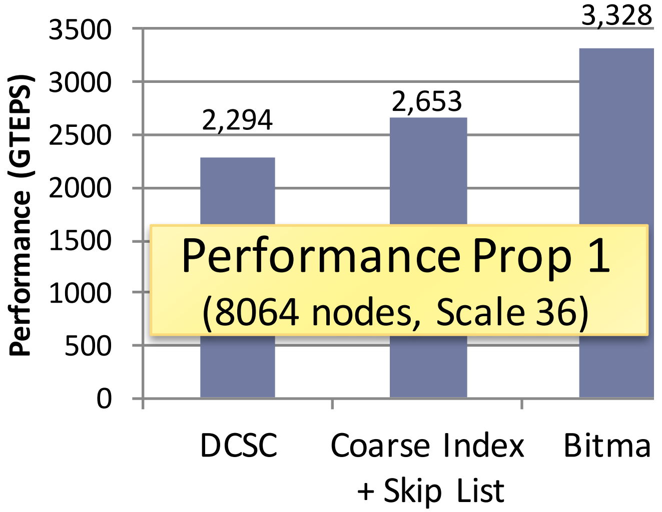
Optimized Graph500 program– Bandwidth Reducing Algorithm (Concept from HPC, e.g. by Jim Demmel @ UC Berkeley)

- **Problem:** Partitioned graph: hyper sparse matrix =>Traditional sparse matrix representation inefficient
- Proposal1: a new bigmap-based Sparse Matrix Representation
  - Enables **compression of row indexes&fast access to each row**.
- Proposal2: Vertex Reordering for Bitmap Optimization
  - Reordered vertex number by sorting vertices by degree.
  - Use reordered # for bitmap access and original # for other processing.
  - Result: 16% speedup by reduction of bitmap data, 28% speedup by localized memory access, and 49% speedup in total. (8064 nodes)

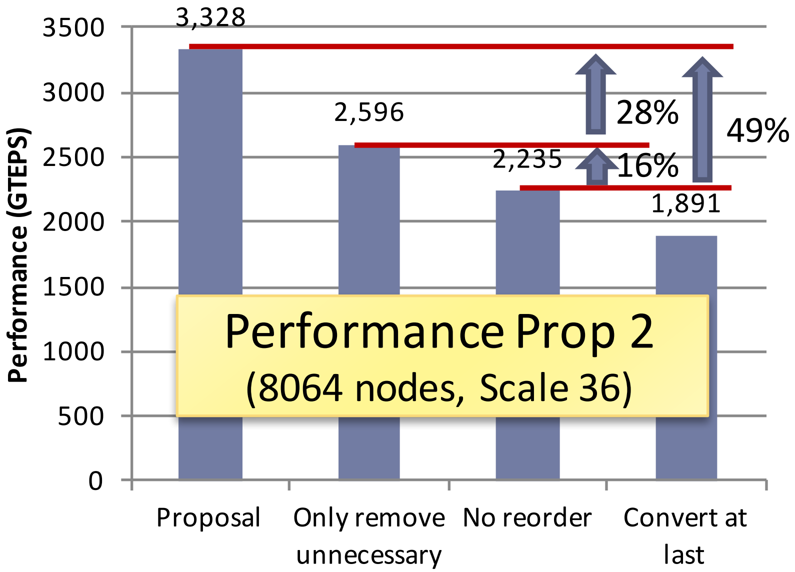


CSR (Compressed Sparse Row)	1806
DCSC	861
Coarse Index + Skip List	309
Bitmap (Proposal)	**337**

Data size of row index (MB/node)
(8064 nodes, Scale 36)



Performance Prop 1
(8064 nodes, Scale 36)



Performance Prop 2
(8064 nodes, Scale 36)

Estimated Compute Resource Requirements for Deep Learning [Source: Preferred Network Japan Inc.]

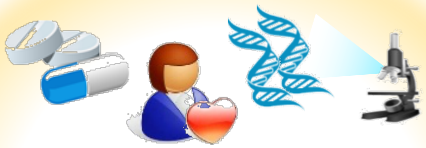
To complete the learning phase in one day

P:Peta
E:Exa
F:Flops

Image/Video Recognition

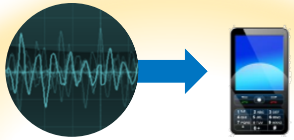
10P（Image）～ 10E（Video） Flops
学習データ：1億枚の画像 10000クラス分類
数千ノードで6ヶ月 [Google 2015]

Bio / Healthcare



100P ～ 1E Flops
一人あたりゲノム解析で約10M個のSNPs
100万人で100PFlops、1億人で1EFlops

Image Recognition



10P～ Flops
1万人の5000時間分の音声データ
人工的に生成された10万時間の
音声データを基に学習 [Baidu 2015]

Auto Driving

1E～100E Flops
自動運転車１台あたり1日 1TB
10台～1000台, 100日分の走行データの学習

Robots / Drones

1E～100E Flops
1台あたり年間1TB
100万台～1億台から得られた
データで学習する場合

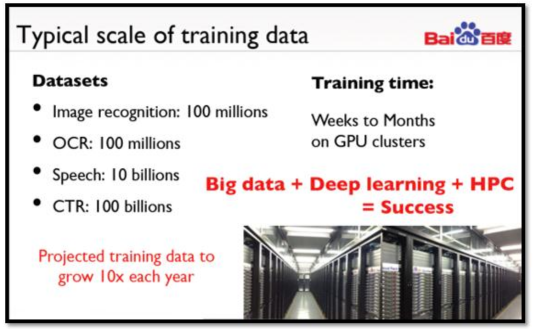
機械学習、深層学習は学習データが大きいほど高精度になる
現在は人が生み出したデータが対象だが、今後は機械が生み出すデータが対象となる

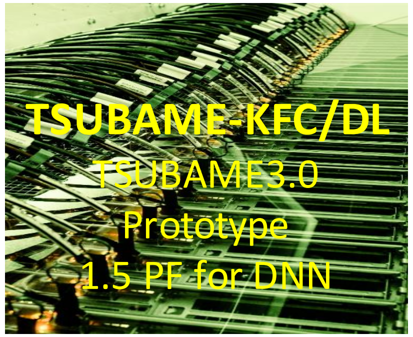
各種推定値は1GBの学習データに対して1日で学習するためには
1TFlops必要だとして計算

10PF 100PF 1EF 10EF 100EF
2015 2020 2025 2030

Research on Advanced Deep Learning Applications (Part of JST Extreme Big Data Project 2013-2018)

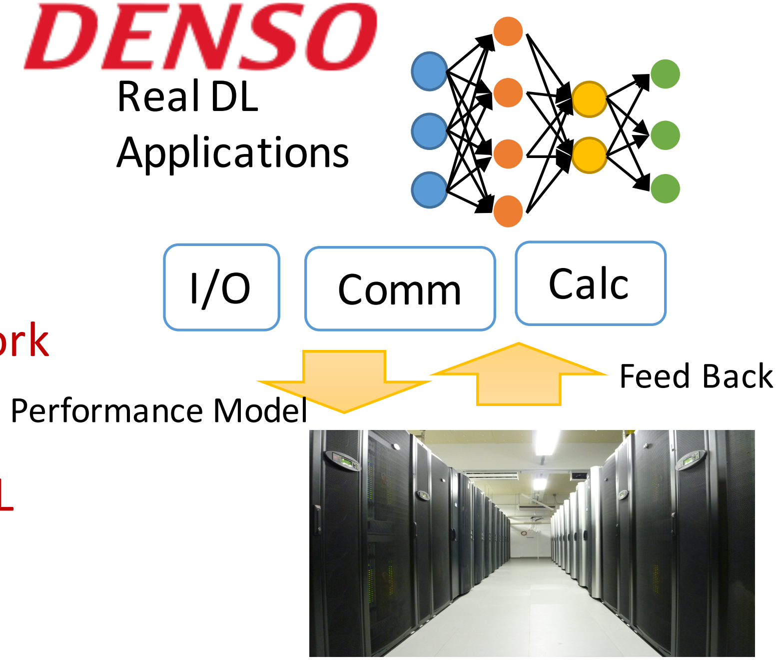
- Deep Learning IS HPC!
  - Training models – mostly dense MatVec
  - Data Access for training target data sets
  - Sharing updated training parameters in neural networks
- Goals
  - Accelerate DL applications in EBD architectures ?
    - Extreme-scale Parallelization, Fast Interconnects, Storage I/O, etc.
  - Performance bottlenecks of multi-node parallel DL algorithms on current HPC systems ?
- Current Status
  - Official Collaboration w/DENSO IT Lab signed November
  - Profiling based bottleneck identification and performance modeling & optimization of a real DL application on TSUBAME
    - Great result, joint paper being prepared for submission
  - > 100 million images, 1500 GPUs (6 Pflops) 1 week grand challenge run
  - Compete w/Google, MS, Baidu etc. in ILSVRC in ImageNet with shallow network
    - To fit within smaller platforms e.g. Jetson
    - Got reasonable results, about 10% accuracy with 15-layer CNN
  - Denso Lab continues to run workloads on TSUBAME2.5 and TSUBAME-KFC/DL
  - In talks with other companies, e.g. Yahoo! Japan





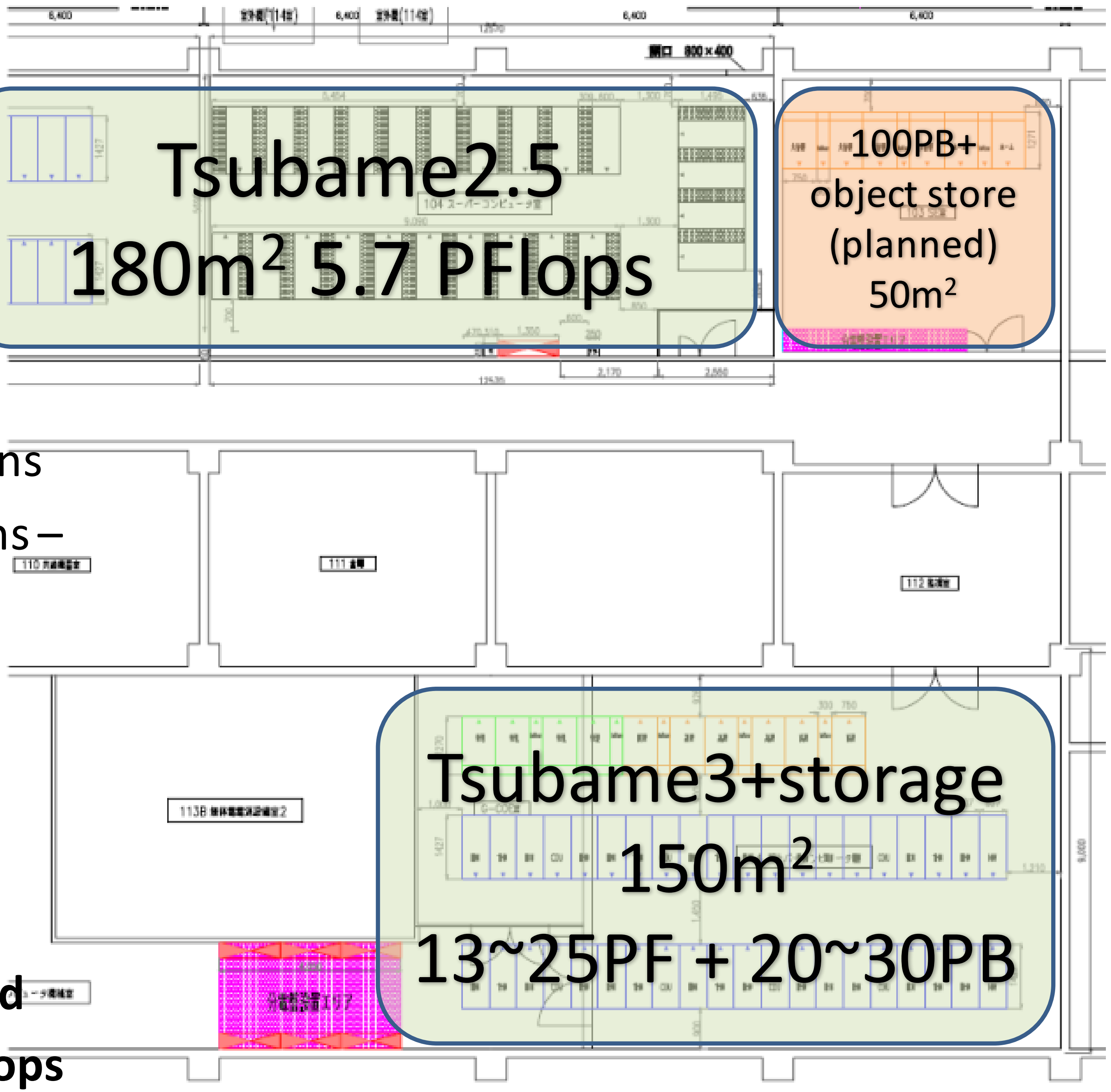
Many companies (ex. Baidu, etc.) employ GPU-based Cluster Architectures, similar to TSUBAME2 & KFC

DENSO



TSUBAME2&3 Joint Operation Plan

- New dedicated datacenter space for Tsubame3 => retain TSUBAME2
- Joint operation 2017~2019
  - TSUBAME3: mainline HPC operations
  - TSUBAME2.5: specialized operations – industry jobs, long running, AI/BD.
- Power capped not to exceed power & cooling limits (4MW)
- **Total ~8000 GPUs, 100Pflops for AI**
  - **Storage enhanced to cope w/capacity**
  - **<u>Pending budgetary allocation</u>**
- **Construction on new IDC space started**
- **Future: TSUBAME3+TSUBAME4 joint ops**



Comparison of Machine Learning / AI Capabilities of TSUBAME3+2.5 and K-Computer

東京工業大学
Tokyo Institute of Technology
GSIC
Global Scientific Information and Computing Center

独立行政法人理化学研究所
計算科学研究機構
RIKEN
RIKEN Advanced Institute for Computational Science

TSUBAME2.5(2013)
+TSUBAME3.0(2017) 7-8000GPUs

Deep Learning / AI Capabilities
FP16+FP32 up to ~100 Petaflops
+ up to 100PB online storage

X7~10

>>

(effectively more due to optimized DL SW Stack on GPUs)

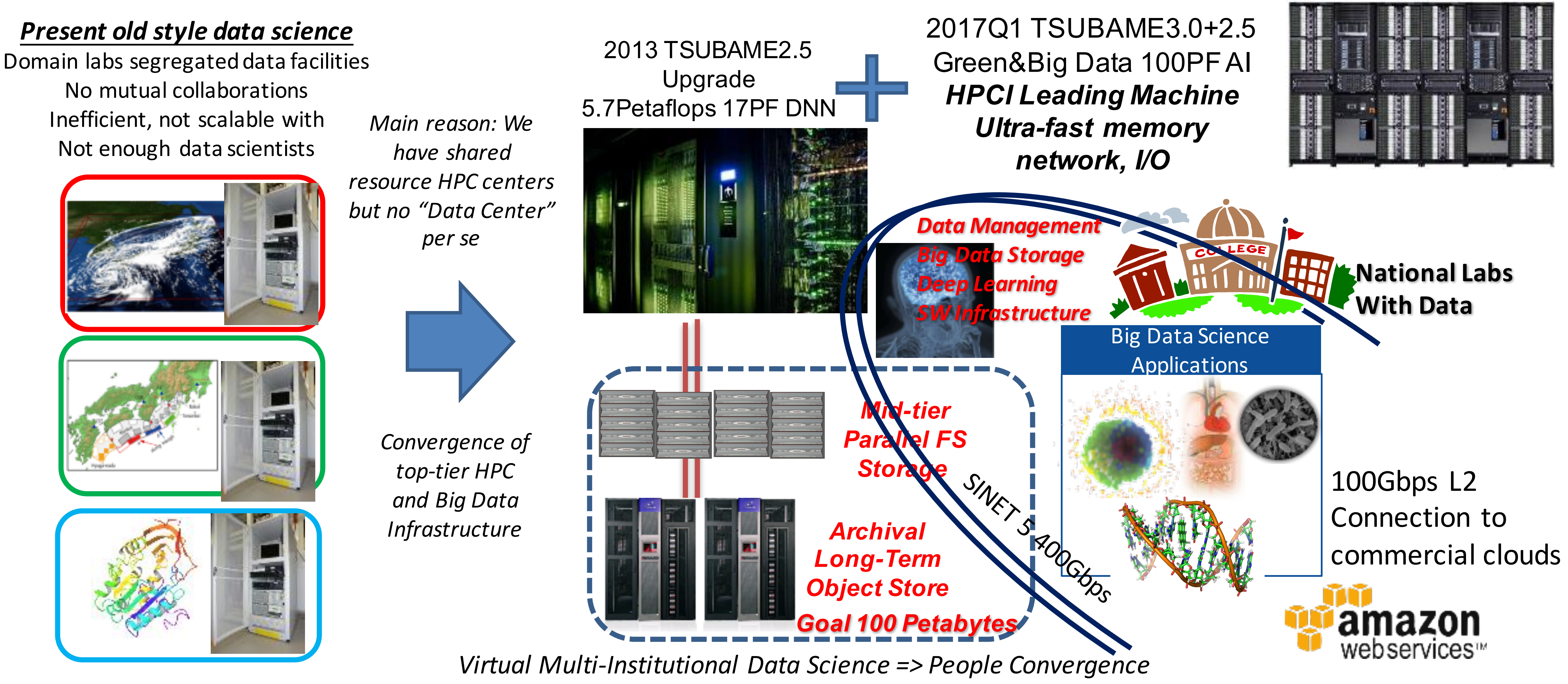
K Computer (2011)

Deep Learning
FP32 11.4 Petaflops

BG/Q Sequoia (2011)
22 Petaflops SFP/DFP

2015 Proposal to MEXT – Big Data and HPC Convergent Infrastructure
=>Big Data and AI supercomputer （Tokyo Tech GSIC)

- "Big Data" currently processed managed by domain laboratories => No longer scalable
- HPCI HPC Center => Converged HPC and Big Data Science Center
- People convergence: domain scientists + data scientists + CS/Infrastructure => Big data & AI center
- Data services, ML/DNN/AI services…



Domestic & International BDEC Joint Labs/Centers and Research Effort

International

- **JLESC** – Joint Labs on Exascale Computing
  - NCSA/UIUC, INRIA, ANL, BSC, Juelich SC, Riken AICS
- **ADAC** – Accelerated Data And Computing Institute
  - ORNL, ETH/CSCS, Tokyo Tech GSIC (MOU signed May 2016)
- **US DoE – Japan MEXT** collab. on Exascale System Software
- **SPPEXA** German DFG - French ANR-Japanese JST Software for Exascale

Domestic

- **HPCI** – High Performance Computing Infrastructure
- **JCAHPC** – U-Tokyo & TSUKUBA HPC centers
- **OIL-** AIST-Tokyo Tech Open Innovation Lab on AI research

National Institute for Advanced Industrial Science and Technology (AIST)

経済産業省
Ministry of Economy, Trade and Industry

Ministry of Economics Trade and Industry (METI)

Matsuoka will be appointed 15% to AIST AI-OIL starting summer

ラボ長（産総研研究職 or 東工大教員/クロスアポ）
- 副ラボ長（産総研研究職）
- 副ラボ長（産総研事務職）
- ラボ研究主幹（産総研研究職）
- ラボ構成員

GSIC

Tsubame 3.0/2.5
Big Data /AI resources

Resources and Acceleration of AI / Big Data, systems research

Joing Organization@Odaiba

AIST-TokyoTech AI/Big Data Open Innovation Laboratory (OIL)

AIST Artificial Intelligence Research Center (AIRC)

Joint Research on AI / Big Data and applications

Application Area
Natural Langauge Processing
Robotics
Security

Industrial Collaboration in data, applications

Basic Research in Big Data / AI algorithms and methodologies

“Smart AI Technology Research Organization”

Other Big Data / AI research organizations and proposals

Industry

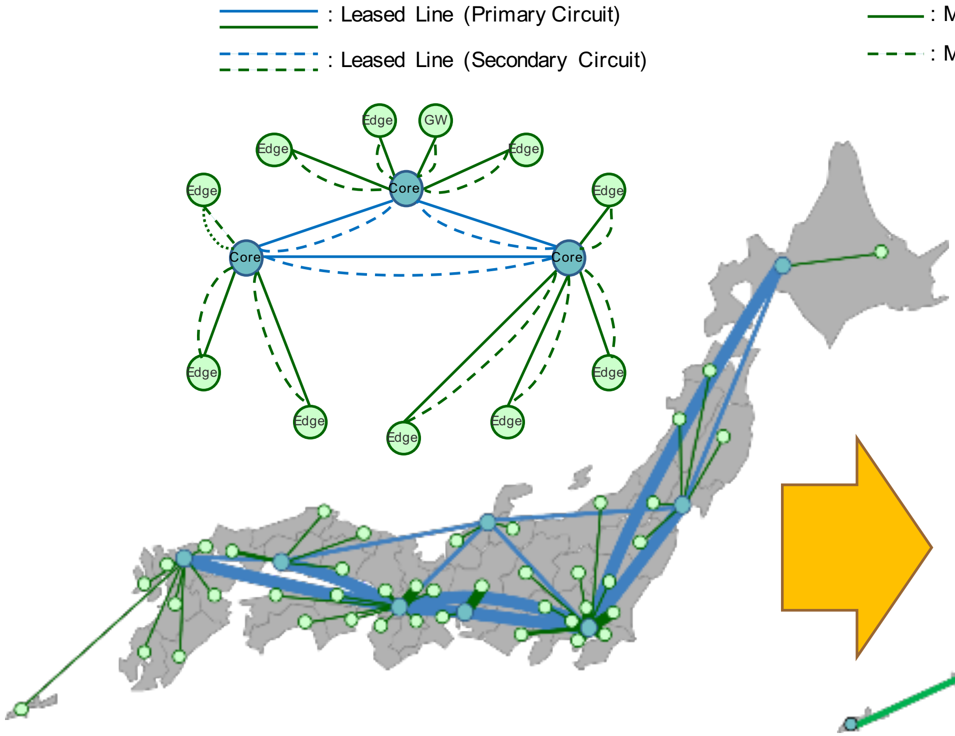
DENSO IT LABORATORY, INC.

SINET5: Nationwide Academic Network: operational Apr 2016

- ◆ 2016 SINET5 connects all the SINET nodes in a fully-meshed topology and minimizes the latency between every pair of the nodes using nationwide dark fiber, 400Gbps, future 1Tbps
- ◆ MPLS-TP devices connect a pair of the nodes by primary and secondary MPLS-TP P2P paths.

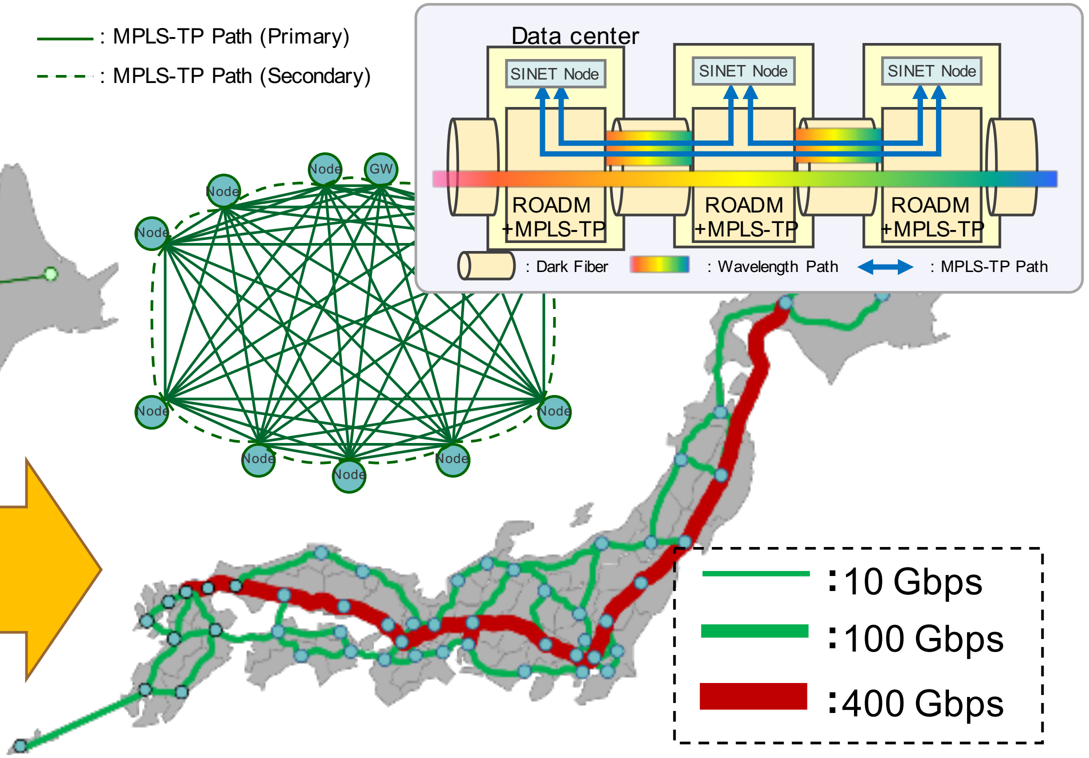
SINET4 present

- Connects nodes in a star-like topology
- Secondary circuits of leased lines need dedicated resources



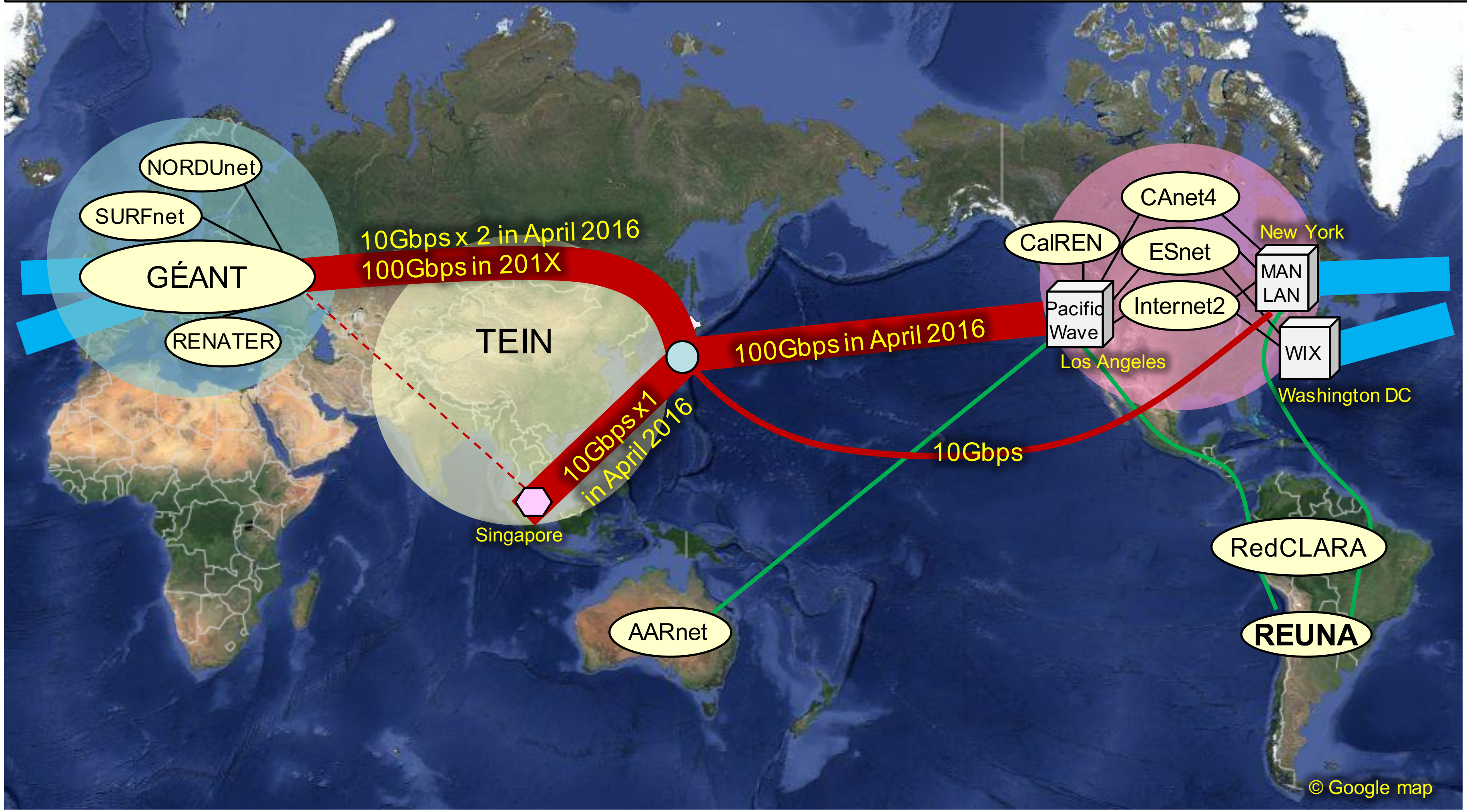
SINET5 2016

- Connects all the nodes in a fully-meshed topology with redundant paths
- Secondary paths do not consume resources



International Lines of SINET5

- 100-Gbps line to U.S. West Coast and will keep a 10-Gbps line to U.S. East Coast.
- Two direct 10-Gbps lines to Europe in April 2016, possibility of a 100-Gbps in the near future.
- SINET will keep a 10-Gbps line to Singapore in April 2016.



Towards TSUBAME4 and 5: Moore's Law will end in the 2020's

- Much of underlying IT performance growth due to Moore's law
  - "LSI: x2 transistors in 1~1.5 years"
  - Causing qualitative "leaps" in IT and societal innovations
  - The main reason we have supercomputers and Google...
- But this is slowing down & ending, by mid 2020s…!!!
  - End of Lithography shrinks
  - End of Dennard scaling
  - End of Fab Economics

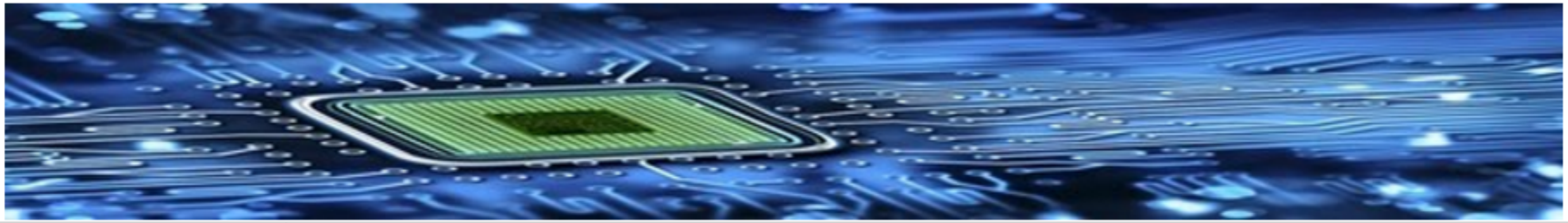
The curse of constant transistor power shall soon be upon us

Gordon Moore

- How do we *sustain* "performance growth" beyond the "end of Moore"?
  - Not just one-time speed bumps
  - Or do we give up and so something else?

Post Moore Era Supercomputing Workshop @ SC16

- https://sites.google.com/site/2016pmes/
- Jeff Vetter (ORNL), Satoshi Matsuoka (Tokyo Tech) et. al.



Search this sit

2016 Post-Moore's Era Supercomputing (PMES) Workshop Home

News

Call For Position Papers - Submission Deadline - June 17

Invited Speakers

Photos

Program

Resources

Workshop Venue

Sitemap

188

2016 Post-Moore's Era Supercomputing (PMES) Workshop Home

Co-located with SC16 in Salt Lake City
Monday, 14 November 2016

Workshop URL: http://j.mp/pmes2016
CFP URL: http://j.mp/pmes2016cfp
Submission URL (EasyChair): http://j.mp/pmes2016submissions
Submission questions: pmes16@easychair.org

This interdisciplinary workshop is organized to explore the scientific issues, challenges, and opportunities for supercomputing beyond the scaling limits of

News

PMES Submission Site Now Open!
PMES Workshop Confirmed for SC16!
Submissions open for PMES Position Papers on April 17

Important Dates

- Submission Site Opens: 17 April 2016