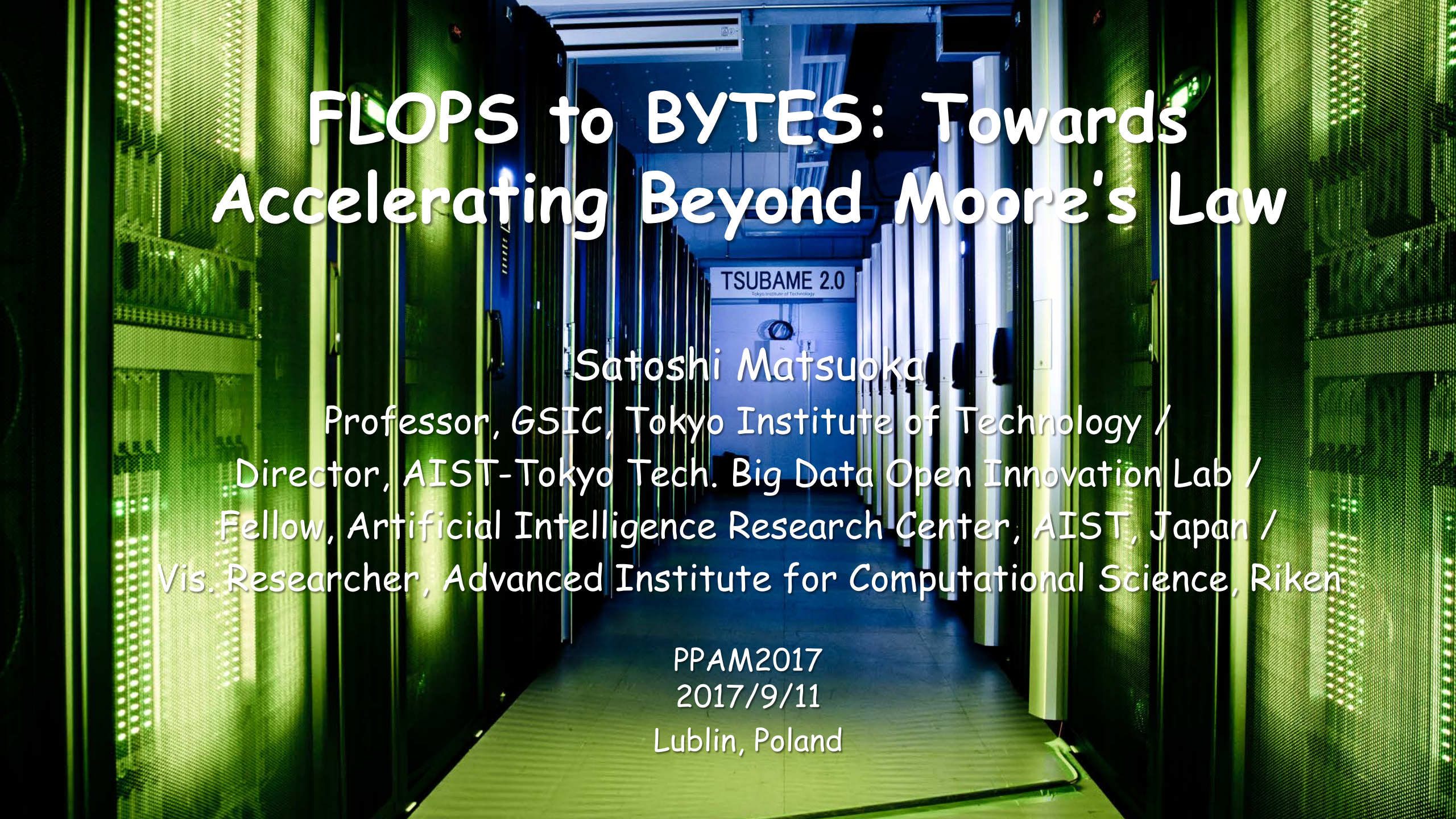


FLOPS to BYTES: Towards Accelerating Beyond Moore's Law

Satoshi Matsuoka
Professor, GSIC, Tokyo Institute of Technology /
Director, AIST-Tokyo Tech. Big Data Open Innovation Lab /
Fellow, Artificial Intelligence Research Center, AIST, Japan /
Vis. Researcher, Advanced Institute for Computational Science, Riken

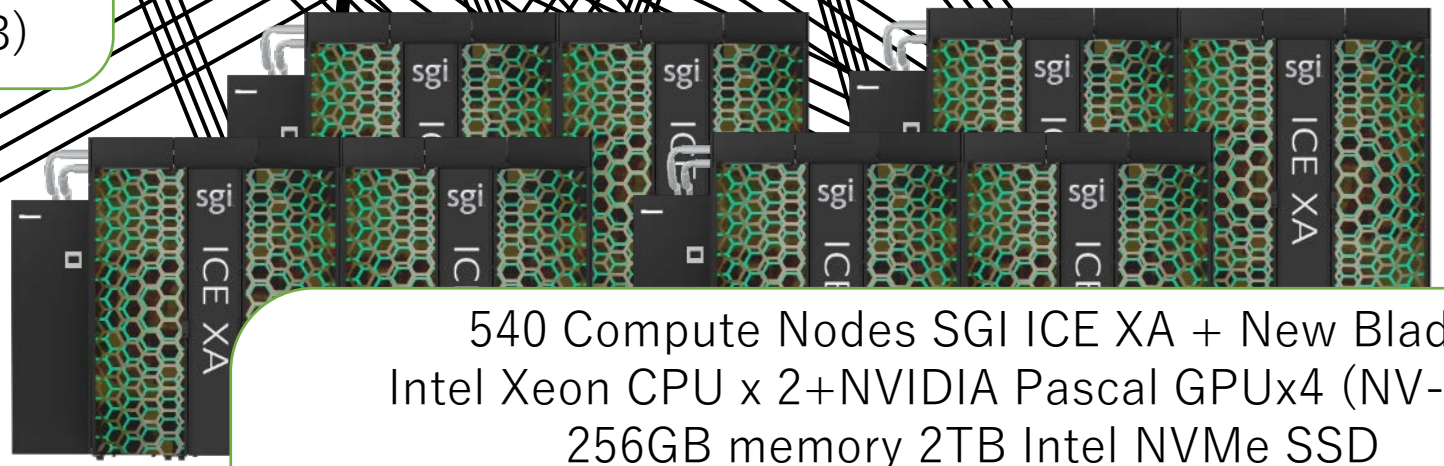
PPAM2017
2017/9/11
Lublin, Poland

Overview of TSUBAME3.0
BYTES-centric Architecture, Scalaibility to all 2160 GPUs, all nodes, the entire memory hiearchy

Full Operations
Aug. 2017

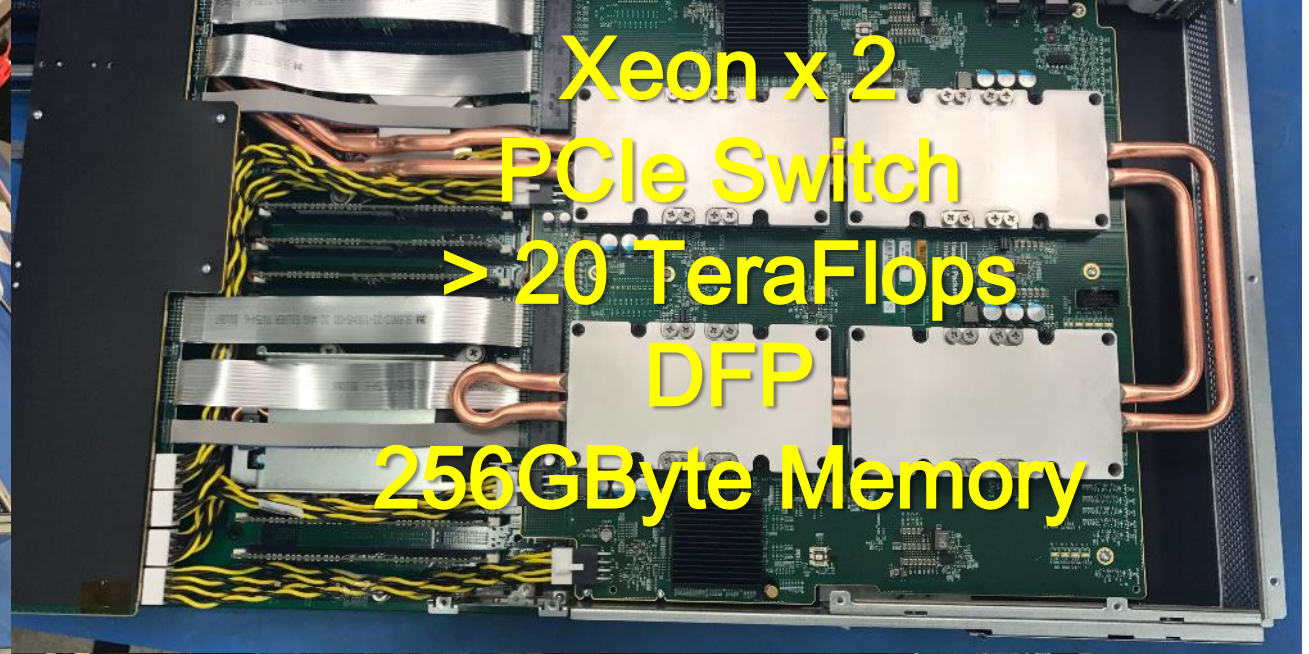
Full Bisection Bandwidgh
Intel Omni-Path Interconnect. 4 ports/node
Full Bisection / 432 Terabits/s bidirectional
~x2 BW of entire Internet backbone traffic

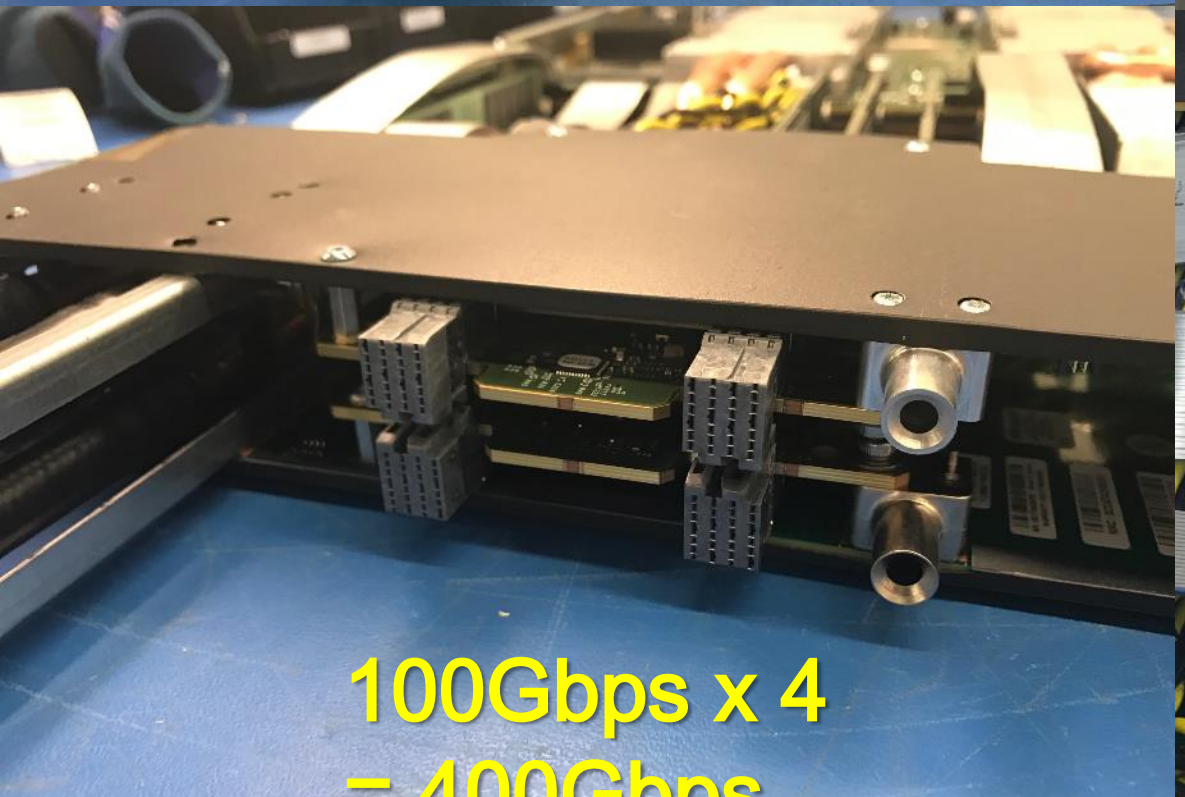
DDN Storage
(Lustre FS 15.9PB+Home 45TB)

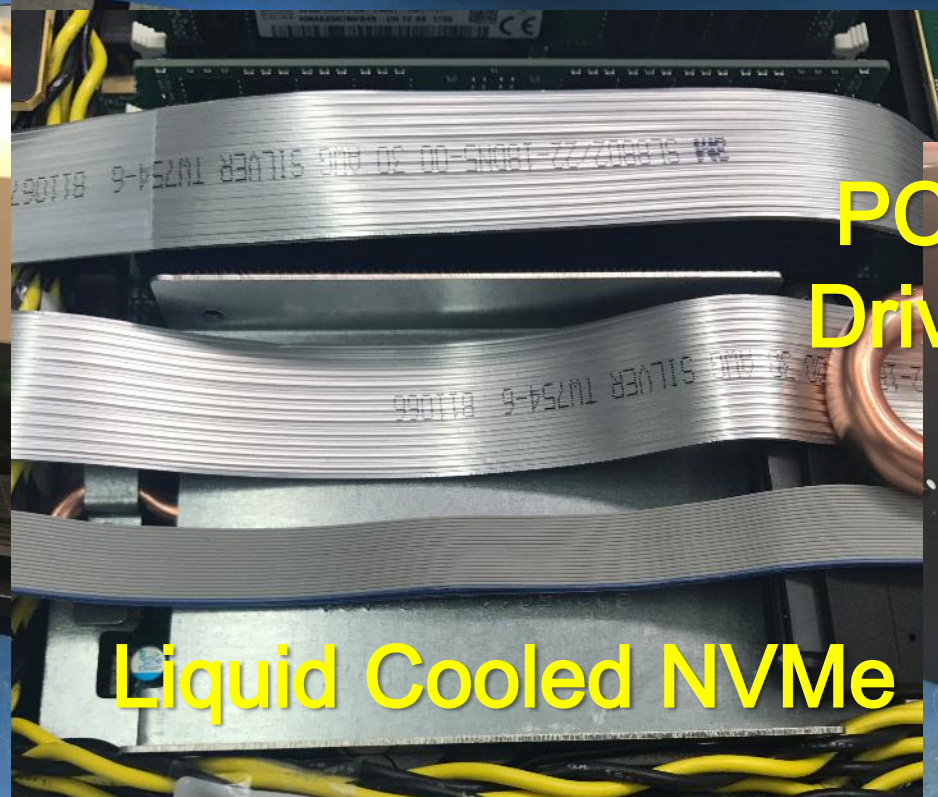


540 Compute Nodes SGI ICE XA + New Blade
Intel Xeon CPU x 2+NVIDIA Pascal GPUx4 (NV-Link)
256GB memory 2TB Intel NVMe SSD
47.2 AI-Petaflops, 12.1 Petaflops

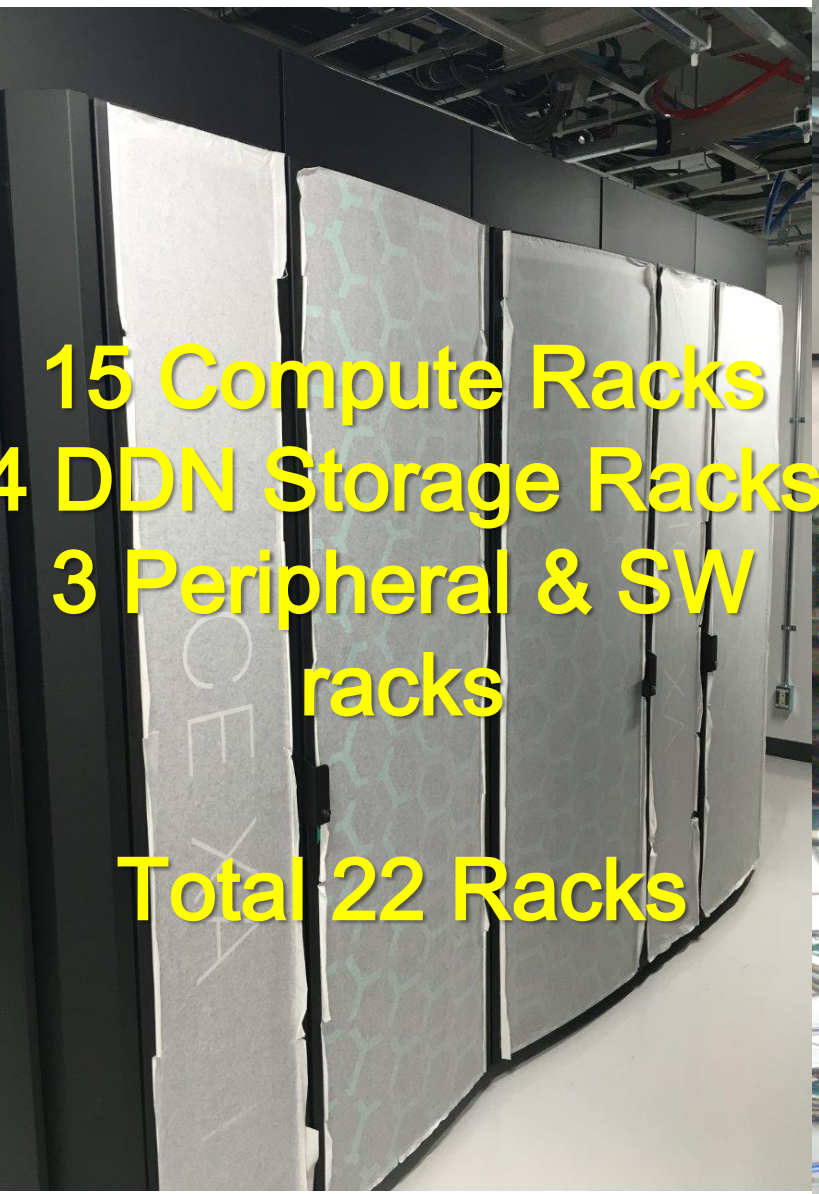

Liquid Cooled
“Hot Pluggable” ICE-XA Blade
Smaller than 1U server, no cables or pipes

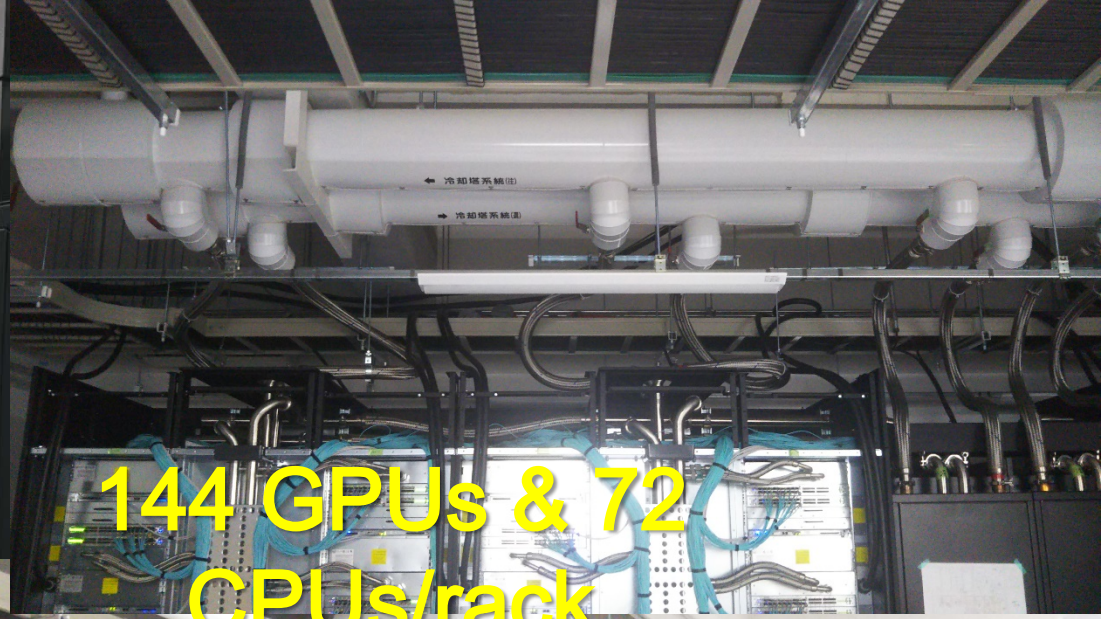

Xeon x 2
PCIe Switch
> 20 TeraFlops DFP
256GByte Memory

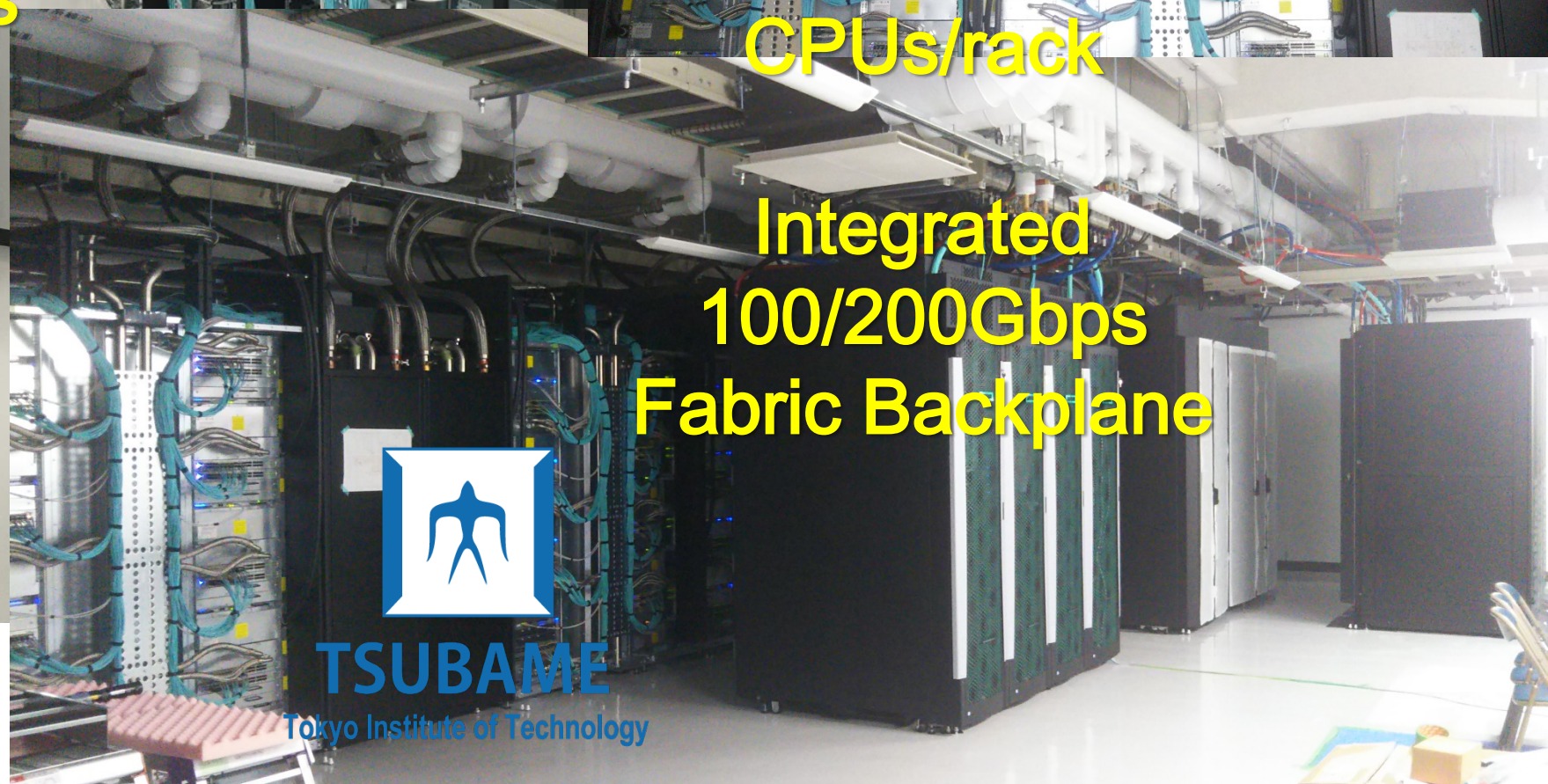

100Gbps x 4
= 400Gbps


Liquid Cooled NVMe


PCIe NVMe
Drive Bay x 4

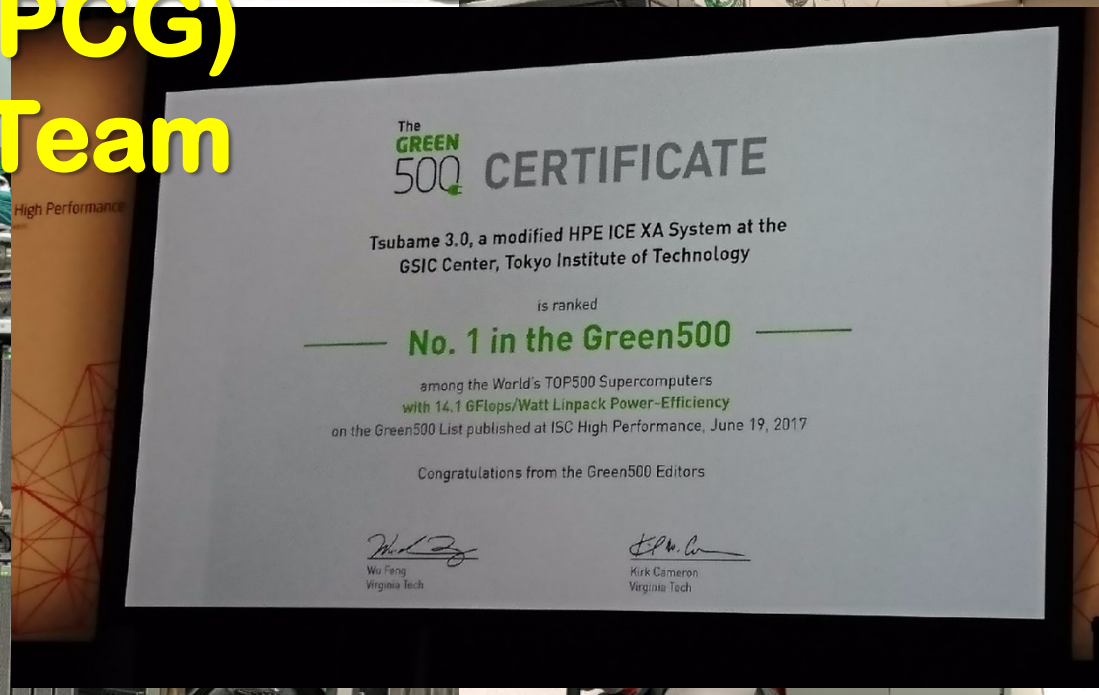

15 Compute Racks
4 DDN Storage Racks
3 Peripheral & SW racks
Total 22 Racks


144 GPUs & 72 CPUs/rack


Integrated 100/200Gbps Fabric Backplane
TSUBAME
Tokyo Institute of Technology

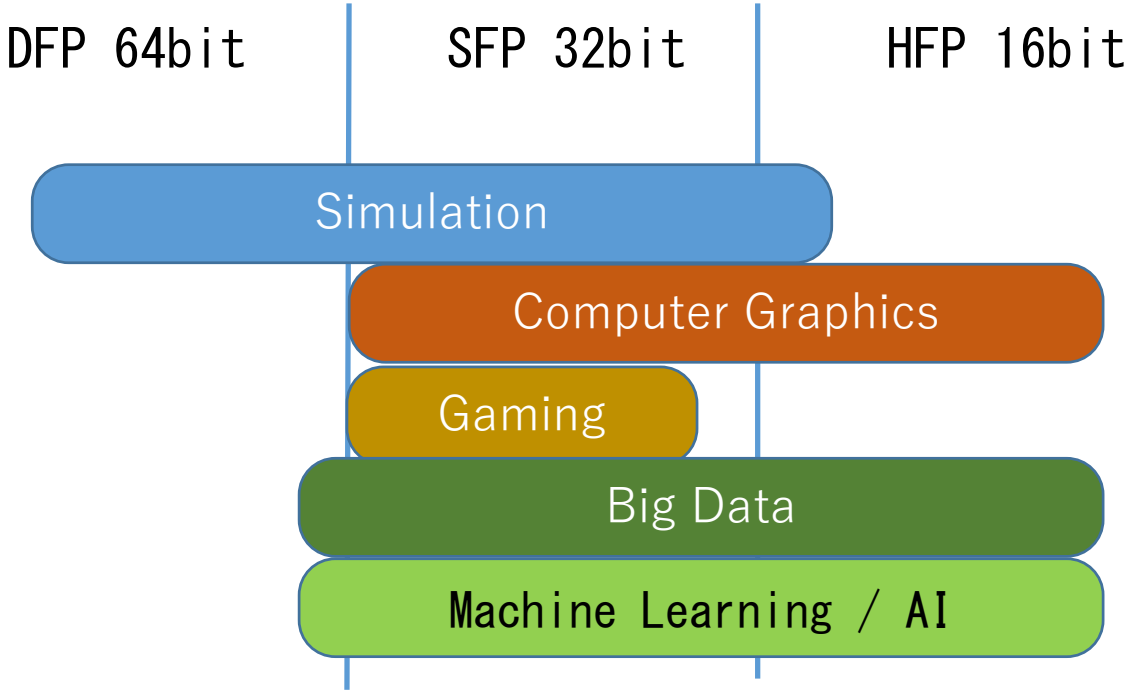
Running Benchmarks (HPL, HPCG)
Tokyo Tech + HPE/SGI Japan Team

Top 500 Submission =
1.x Petaflops @ 144 Nodes

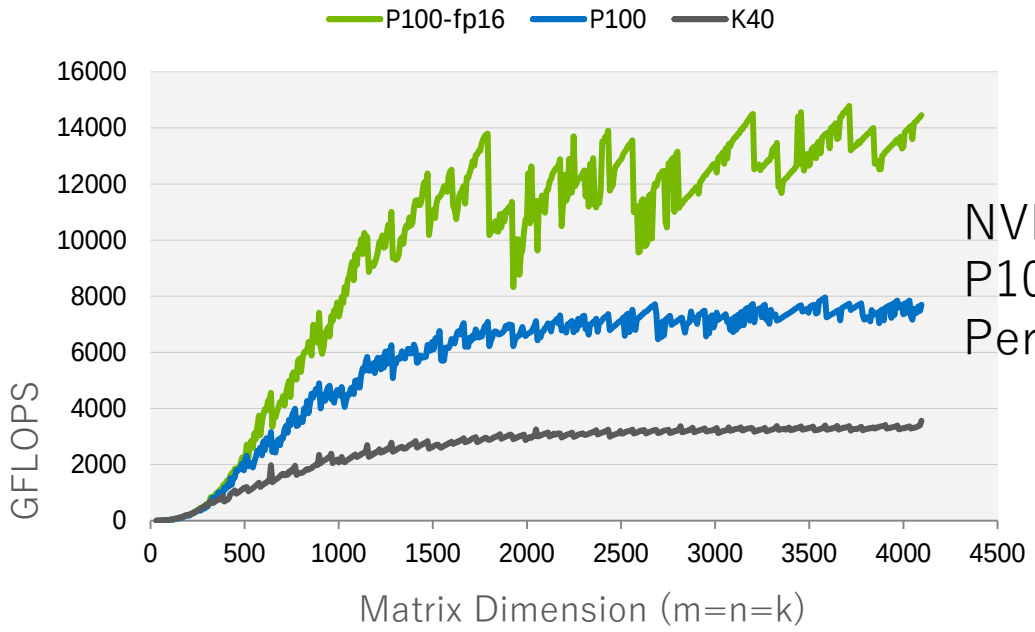


Green 500 Submission =
14.11 Gflops/W (Nov. 2016 #1 = 9.46 Gflops/W)
Announcement ISC17 Opening,
9AM Monday, 19th June World #1 !!!

Benchmark team



Tokyo Tech GSIC leads Japan in aggregated AI-capable FLOPS TSUBAME3+2.5+KFC, in all Supercomuters and CloudsNV



NVIDIA Pascal P100 DGEMM Performane

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

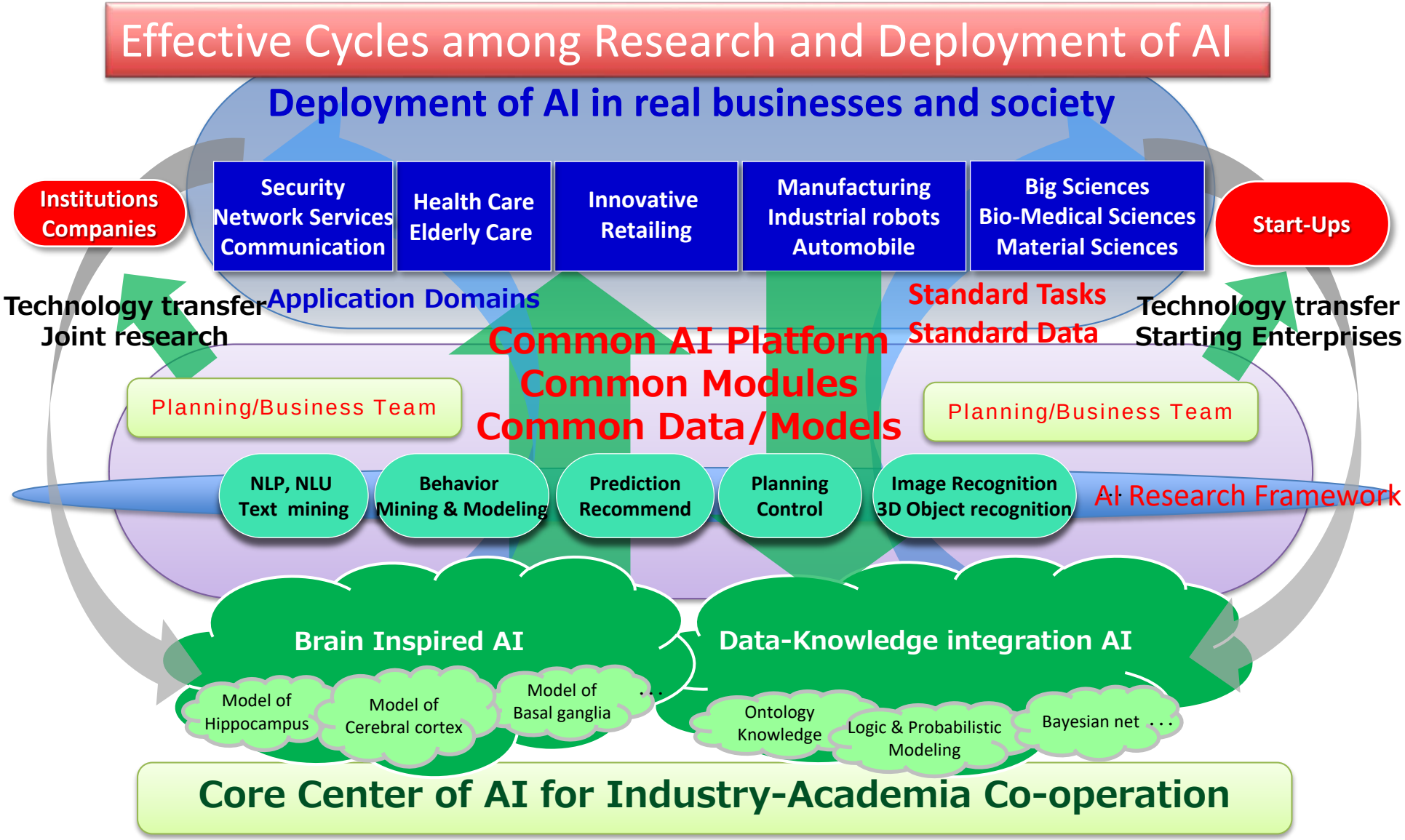
- Three national centers on Big Data and AI launched by three competing Ministries for FY 2016 (Apr 2015-)
  - METI – AIRC (Artificial Intelligence Research Center): AIST (AIST internal budget + > $200 million FY 2017), April 2015
    - Broad AI/BD/IoT, industry focus
  - MEXT – AIP (Artificial Intelligence Platform): Riken and other institutions ($~50 mil), April 2016
    - A separate Post-K related AI funding as well.
    - Narrowly focused on DNN
  - MOST – Universal Communication Lab: NICT ($50~55 mil)
    - Brain –related AI
  - $1 billion commitment on inter-ministry AI research over 10 years => Supplanting HPC activities?

Vice Minsiter Tsuchiya@MEXT Annoucing AIP estabishment

2015- AI Research Center (AIRC), AIST

Now > 400+ FTEs

Director:
Jun-ichi Tsujii



Matsuoka : Joint appointment as "Designated" Fellow since July 2017

National Institute for Advanced Industrial Science and Technology (AIST)

Joint Lab established Feb. 2017 to pursue BD/AI joint research using large-scale HPC BD/AI infrastructure

Tokyo Institute of Technology / GSIC

TSUBAME
Tokyo Institute of Technology

GSIC
Global Scientific Information and Computing Center

Tsubame 3.0/2.5 Big Data /AI resources

Ministry of Economics Trade and Industry (METI)

Resources and Acceleration of AI / Big Data, systems research

AIST Artificial Intelligence Research Center (AIRC)

Application Area
Natural Langauge Processing
Robotics
Security

Joint Research on AI / Big Data and applications

AIST-Tokyo Tech Real World Big-Data Computation Open Innovation Laboratory (RWBC-OIL)

Director: Satoshi Matsuoka

Basic Research in Big Data / AI algorithms and methodologies

ITCS Departments

Industrial Collaboration in data, applications

ABCI
AI Bridging Cloud Infrastructure

Industry

YAHOO! JAPAN IT LAB DENSO

DENSO IT LABORATORY, INC.

Other Big Data / AI research organizations and proposals
JST BigData CREST
JST AI CREST
Etc.

The current status of AI & Big Data in Japan

We need the triage of advanced algorithms/infrastructure/data but we lack the cutting edge infrastructure dedicated to AI & Big Data (c.f. HPC)

Tokyo Tech

AIRC

Joint RWBC Open Innov. Lab (OIL) (Director: Matsuoka)

AIST NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST) AIST-AIRC

東京工業大学 Tokyo Institute of Technology

Riken -AIP

NICT- UCRI

AI Venture Startups

Preferred Networks

UEI

AI/BD Centers & Labs in National Labs & Universities

R&D ML Algorithms & SW

Big Companies AI/BD R&D (also Science)

IT LAB DENSO IT LABORATORY, INC.

Panasonic

ABEJA

MIZUHO みずほ情報総研

DeNA

ZMP

Seeking Innovative Application of AI & Data

Use of Massive Scale Data now Wasted

DENSO Petabytes of Drive Recording Video

FANUC FA&ロボット&ロボマシン FA&Robots

TOYOTA

JARI 一般財団法人 日本自動車研究所

Web access and merchandice

Y! YAHOO! JAPAN

SoftBank

NTT

IoT Communication, location & other data

"Big" Data

Massive Rise in Computing Requirements (1 AI-PF/person?)

SAKURA Internet

Google Cloud Platform

Windows Azure

amazon web services

In HPC, Cloud continues to be insufficient for cutting edge research => dedicated SCs dominate & racing to Exascale

Over $1B Govt. AI investment over 10 years

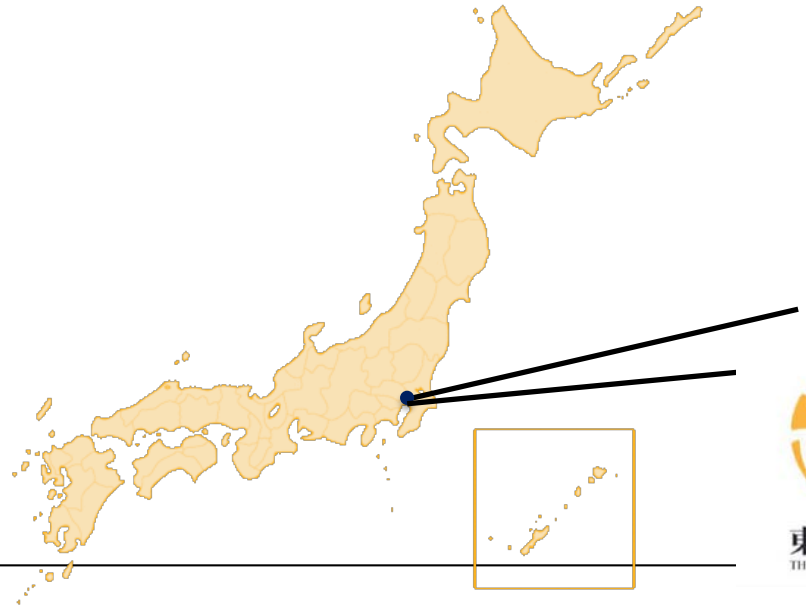
AI&Data Infrastructures

Massive "Big" Data in Training

METI AIST-AIRC ABCI

as the *worlds first large-scale OPEN AI Infrastructure*

- **ABCI**: AI Bridging Cloud Infrastructure
  - Top-Level SC compute & data capability for DNN (>130 AI-Petaflops)
  - Open Public & Dedicated infrastructure for AI & Big Data Algorithms, Software and Applications
  - Platform to accelerate joint academic-industry R&D for AI in Japan



- >130~ AI-Petaflops
- < 3MW Power
- < 1.1 Avg. PUE
- Operational 2017Q4 ~2018Q1

東京大学 THE UNIVERSITY OF TOKYO

AIST NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST)

Univ. Tokyo Kashiwa Campus

The “Real” ABCI – 2018Q1

- **Extreme computing power**
  - w/ >**130 AI-PFlops** for AI/ML especially DNN
  - **x1 million speedup** over high-end PC: 1 Day training for 3000-Year DNN training job
  - TSUBAME-KFC (1.4 AI-Pflops) x 90 users (T2 avg)
- **Big Data and HPC converged modern design**
  - For advanced data analytics (Big Data) and scientific simulation (HPC), etc.
  - Leverage Tokyo Tech’s “TSUBAME3” design, **but differences/enhancements being AI/BD centric**
- **Ultra high BW & Low latency memory, network, and storage**
  - For accelerating various AI/BD workloads
  - Data-centric architecture, optimizes data movement
- **Big Data/AI and HPC SW Stack Convergence**
  - **Incl. results from JST-CREST EBD**
  - **Wide contributions from the PC Cluster community desirable.**
- **Ultra-Green (PUE<1.1), High Thermal (60KW) Rack**
  - **Custom, warehouse-like IDC building and internal pods**
  - **Final “commoditization” of HPC technologies into Clouds**

ABCI Cloud Infrastructure

- **Ultra-dense IDC design from ground-up**
  - Custom inexpensive lightweight “warehouse” building w/ substantial earthquake tolerance
  - **x20 thermal density of standard IDC**
- **Extreme green**
  - Ambient warm liquid cooling, large Li-ion battery storage, and high-efficiency power supplies, etc.
  - **Commoditizing supercomputer cooling technologies to Clouds (60KW/rack)**
- **Cloud ecosystem**
  - Wide-ranging Big Data and HPC standard software stacks
- **Advanced cloud-based operation**
  - Incl. dynamic deployment, container-based virtualized provisioning, multitenant partitioning, and automatic failure recovery, etc.
  - Joining HPC and Cloud Software stack for real
- **Final piece in the commoditization of HPC (into IDC)**
- **Open Sourcing of Next-Gen IDC Architecture for AI**

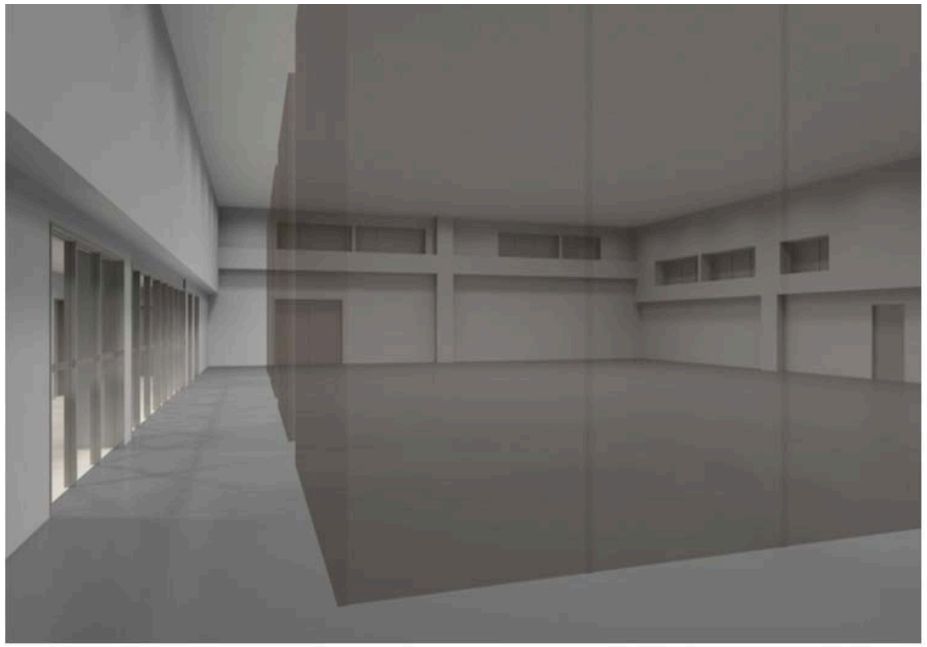
ABCI AI-IDC CG Image

イメージスケッ

Reference Image

引用元: NEC導入事例

ABCI Datacenter ~$10 million (Just broke ground, to be completed late 2017)



Cutting Edge Research AI Infrastructures in Japan Accelerating BD/AI with HPC

(and my effort to design & build them)

Tokyo Tech

In Production

Oct. 2015
TSUBAME-KFC/DL
(Tokyo Tech.)
1.4 AI-PF(Petaflops)

x5.8

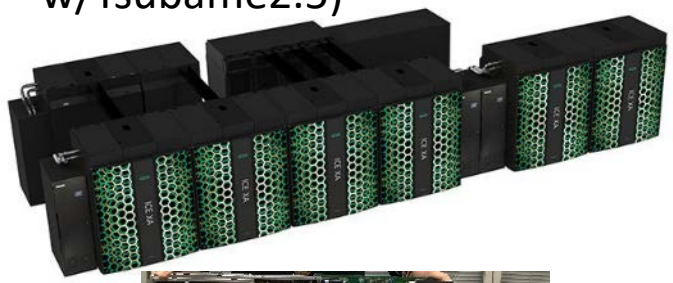
Under Acceptance

Mar. 2017
AIST AI Cloud
(AIST-AIRC)
8.2 AI-PF

x5.8

Being Manufactured

Aug. 2017
TSUBAME3.0 (Tokyo Tech.)
47.2 AI-PF (65.8 AI-PF w/Tsubame2.5)



x2.8~4.2

Mar. 2018
ABCI (AIST-AIRC)
130-200 AI-PF

Draft RFC out
IDC under construction

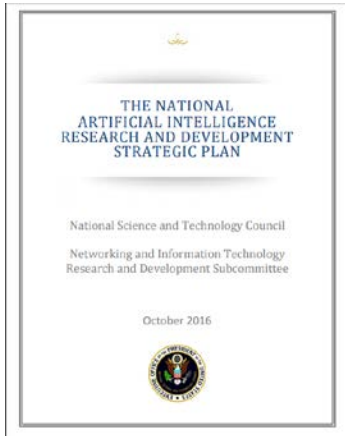
x5.0~7.7

Mar. 2019?
"ExaAI"
~1 AI-ExaFlop
Undergoing Engineering Study

~x1000 in 3.5 years

Mar. 2017
AI Supercomputer
Riken AIP
4.1 AI-PF

R&D Investments into world leading AI/BD HW & SW & Algorithms and their co-design for cutting edge Infrastructure absolutely necessary (just as is with Japan Post-K and US ECP in HPC)


THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN

National Science and Technology Council

Networking and Information Technology Research and Development Subcommittee

October 2016

ORNL Summit

- ~200 Petaflops FP64, ~3 Exaflop FP16 by 1H2018

Peak FP64 Exaflop TSUBAME in 2020 – Just getting Flops/W is within reach

- 7nm+ post Volta GPU (Pascal P100 16nm)
- ~10,000 CUDA Cores (P100 3840), 12.5 Teraflops/Chip (P100 5.3TF) w/matrix engine
- 80,000 chips => 80 million small cores
- 4 GPUs/node => 20,000 nodes (x40 TSUBAME3, 500~600 racks)
- Scalable High-Dimensional Torus or Hypercube topology (Tsubame3 : Full Fattree)
- x3 power efficiency 50GF/W (x1.9 via process, x1.6 via arch) (TSUBAME3.0 14.1GF/W)
- 1 Exa DFP Peak, ~600 PF Linpack, 12MW Facility Power
- So the DARPA Exascale report projection turned out to be fairly accurate
- But is just getting FLOPS all that valuable?

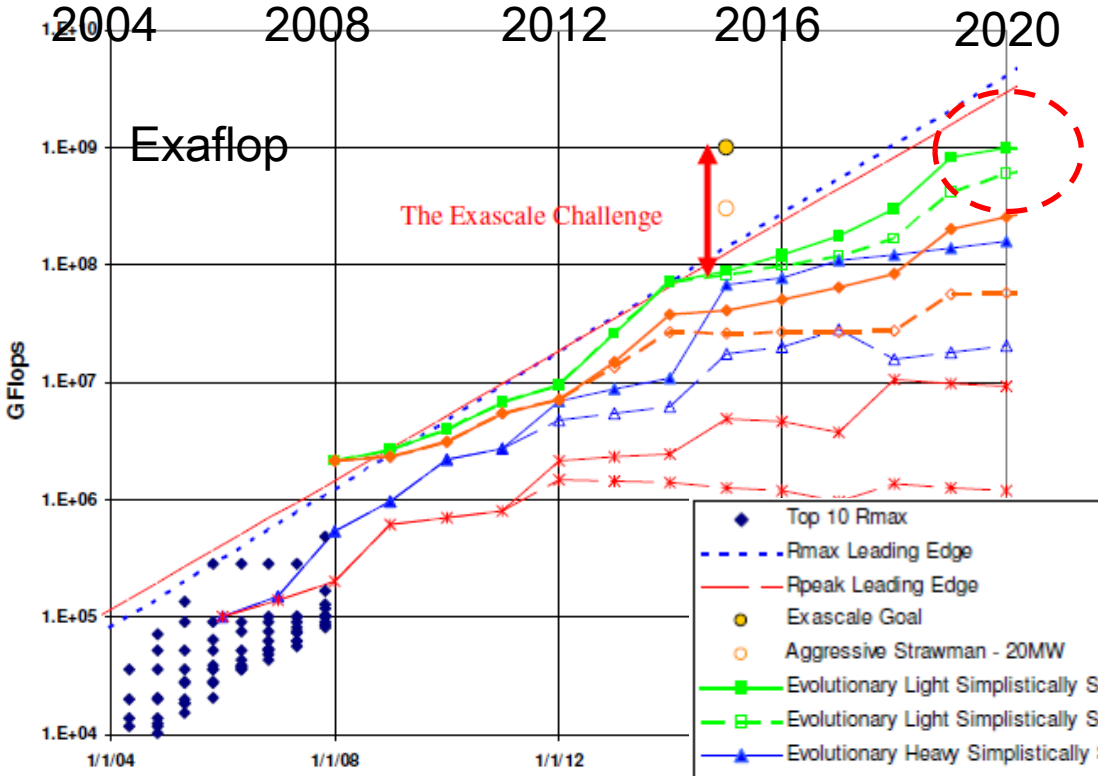
ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems

Peter Kogge, Editor & Study Lead
Keren Bergman
Shekhar Borkar
Dan Campbell
William Carlson
William Dally
Monty Denneau
Paul Franzon
William Harrod
Kerry Hill
Jon Hiller
Sherman Karp
Stephen Keckler
Dean Klein
Robert Lucas
Mark Richards
Al Scarpelli
Steven Scott
Allan Snavely
Thomas Sterling
R. Stanley Williams
Katherine Yelick

DARPA ExaScale Report 2018

September 28, 2008

This work was sponsored by DARPA IPTO in the ExaScale Computing Study with Dr. William Harrod

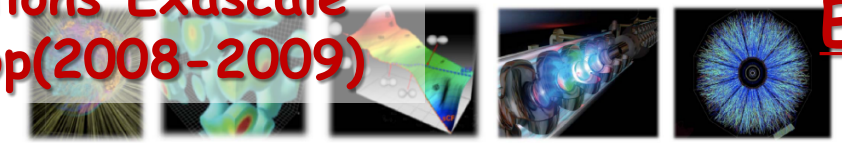


Exa-scale Computational Resources

(slide courtesy Martin Savage)

- Meeting structured around present Nuclear Physics areas of effort

Applications Exascale Workshop(2008-2009)

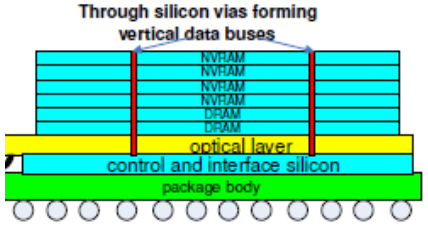


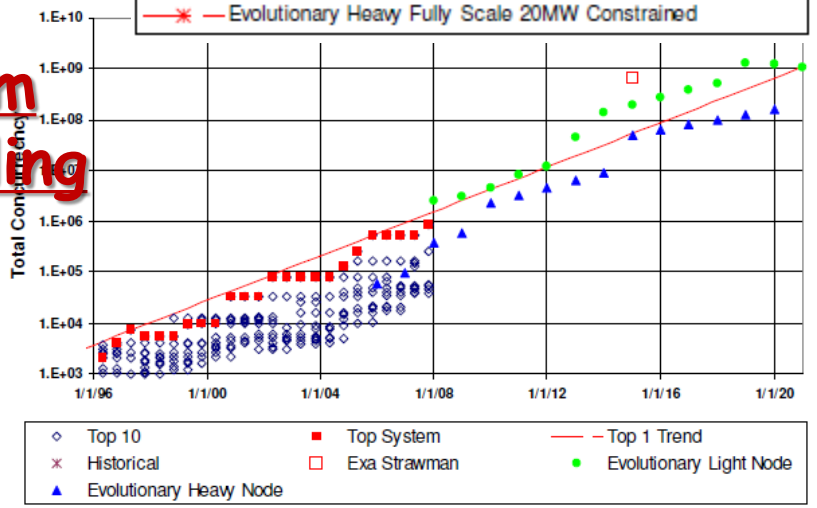
Nuclear Astrophysics | Cold QCD and Nuclear Forces | Nuclear Structure and Reactions | Accelerator Physics | Hot and Dense QCD

- Exa-scale computing is REQUIRED to accomplish the Nuclear Physics mission in each area
- Staging to Exa-flops is crucial :
  - 1 Pflop-yr to 10 Pflop-yrs to 100 Pflop-yrs to 1 Exa-flop-yr (sustained)

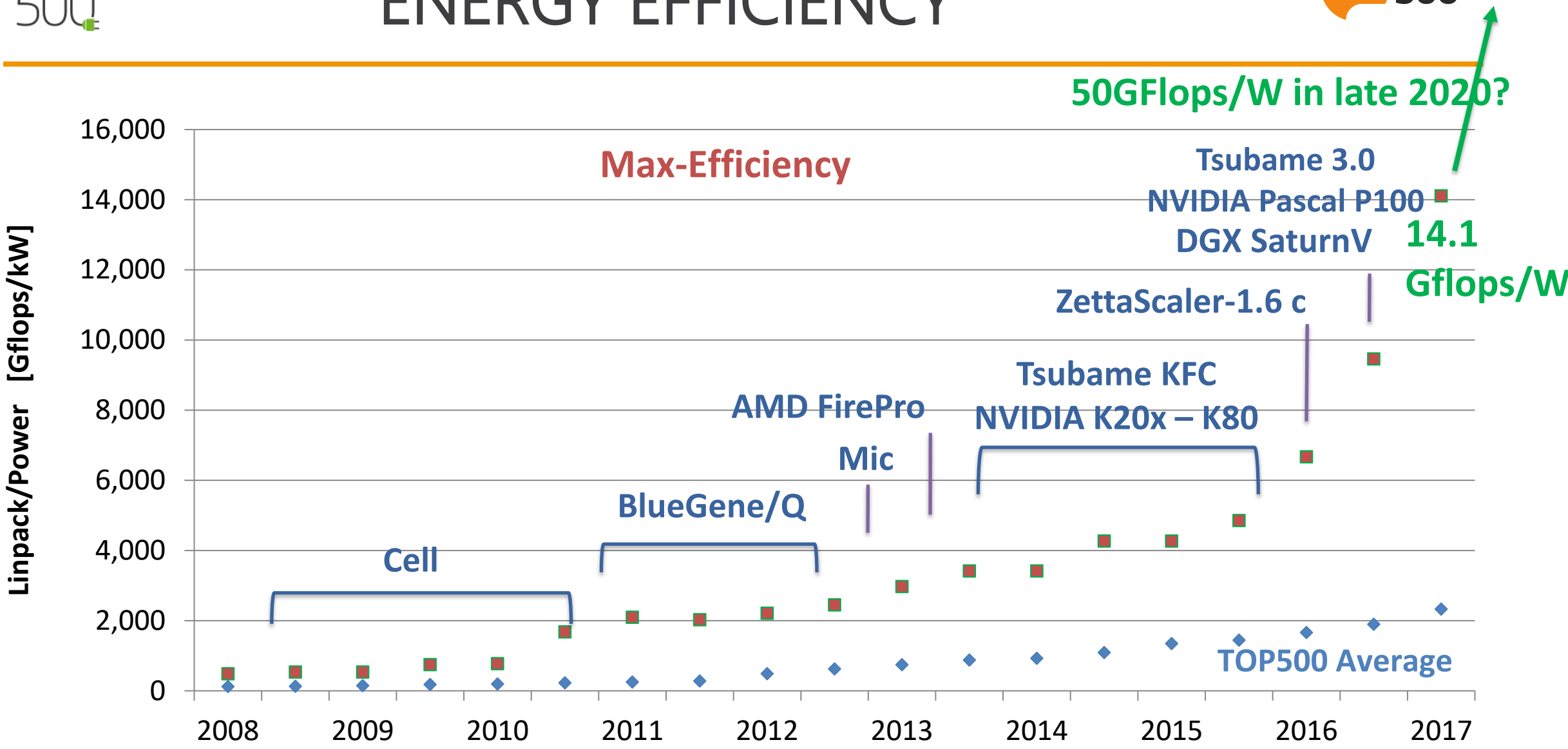
Paul Messina June 28, 2009

LW CPU core 2020
10^8~10^9 pararelism
End of Denard Scaling

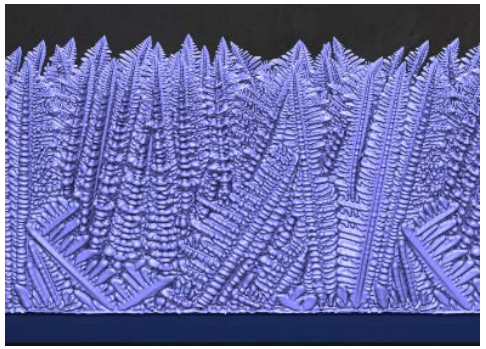


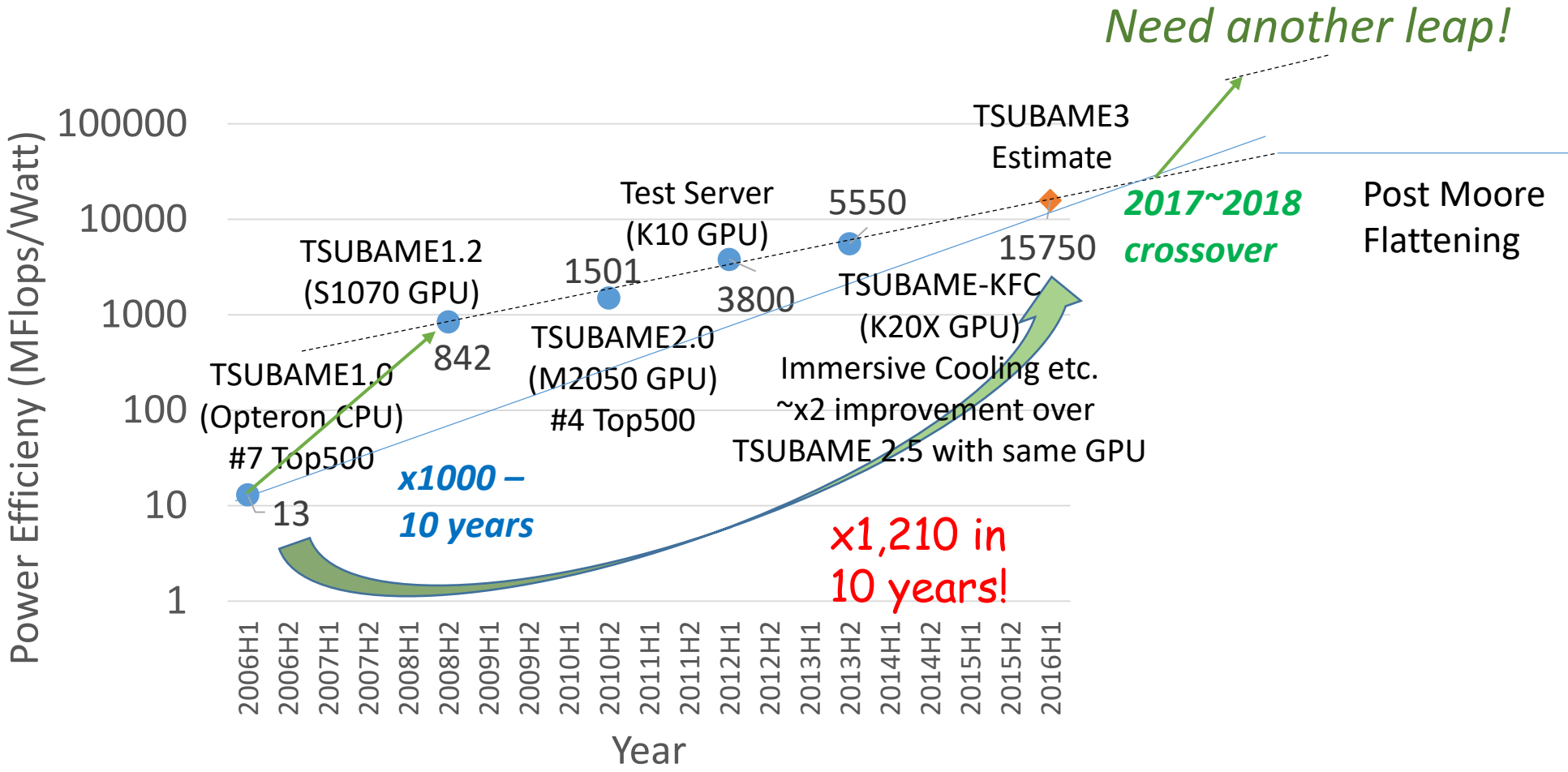


ENERGY EFFICIENCY



Many core was a good step but we already used it once, and cannot use it again for boosting





Measured for the 2011 Gordon Bell Award Dendritic Solidification App
Flop/s/W = Total #Flops / J = energy to solution given same problem

What is worse: Moore's Law will end in the 2020's

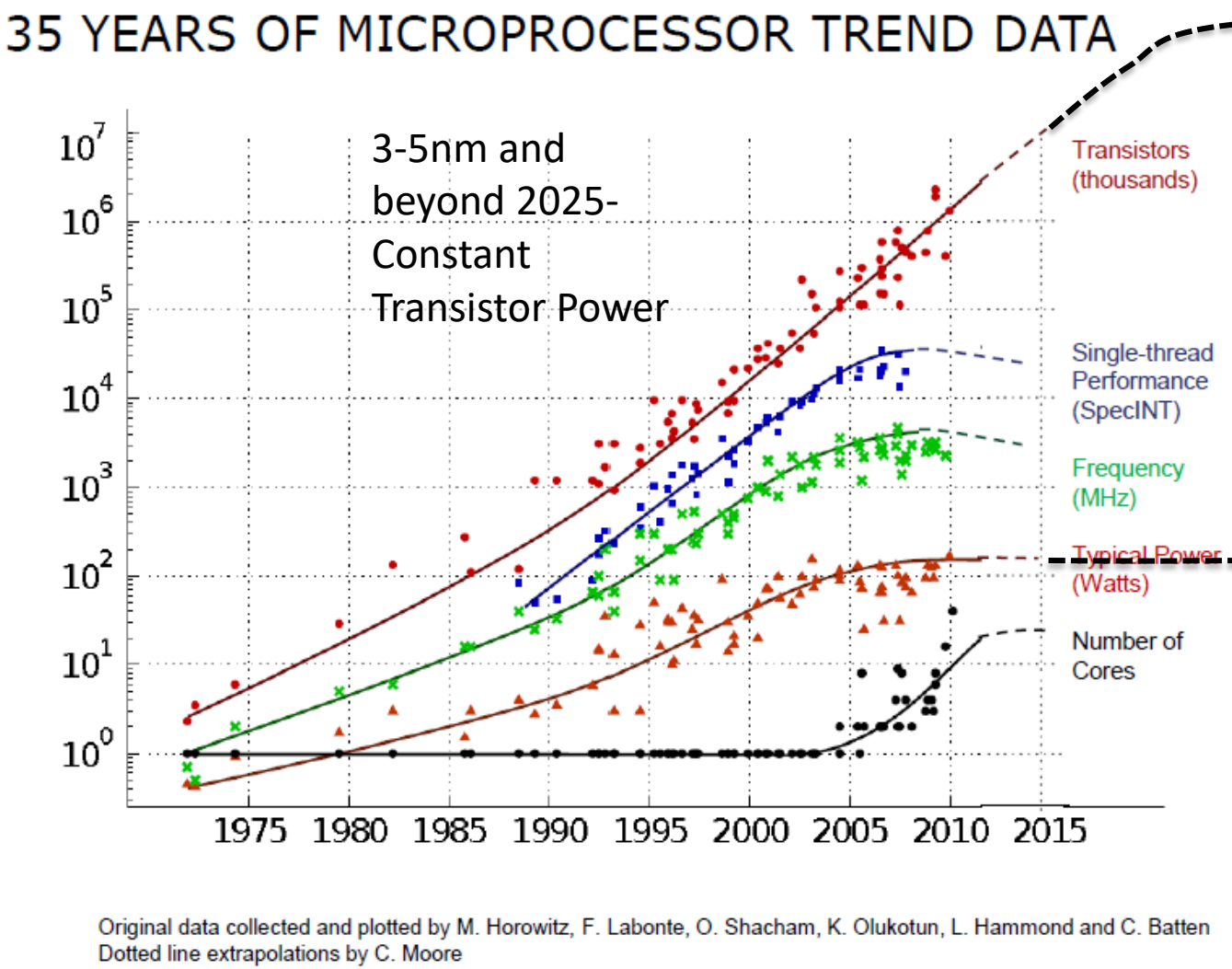
- Much of underlying IT performance growth due to Moore's law
  - "LSI: x2 transistors in 1~1.5 years"
  - Causing qualitative "leaps" in IT and societal innovations
  - The main reason we have supercomputers and Google…
- But this is slowing down & ending, by mid 2020s…!!!
  - End of Lithography shrinks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

Gordon Moore

The curse of constant transistor power shall soon be upon us

- How do we *sustain* "performance growth" beyond the "end of Moore"?
  - Not just one-time speed bumps
  - ***Will affect all aspects of IT, including BD/AI/ML/IoT, not just HPC***
  - ***End of IT as we know it***

20 year Eras towards of End of Moore's Law



20-year Moore-Dennard Single Core ILP-Vector Killer-Micro Era

- 1980s~2004 Dennard scaling, perf+ = single thread+ = transistor & freq+ = power+

20 year Post-Dennard Many-Core Era

- 2004~2015 feature scaling, perf+ = transistor+ = core#+, constant power
- 2015~2025 all above gets harder

20-year Next-Gen Post-Moore era

- 2025~ post-Moore, **constant feature&power = flat performance**

Need to realize the next 20-year era of supercomputing

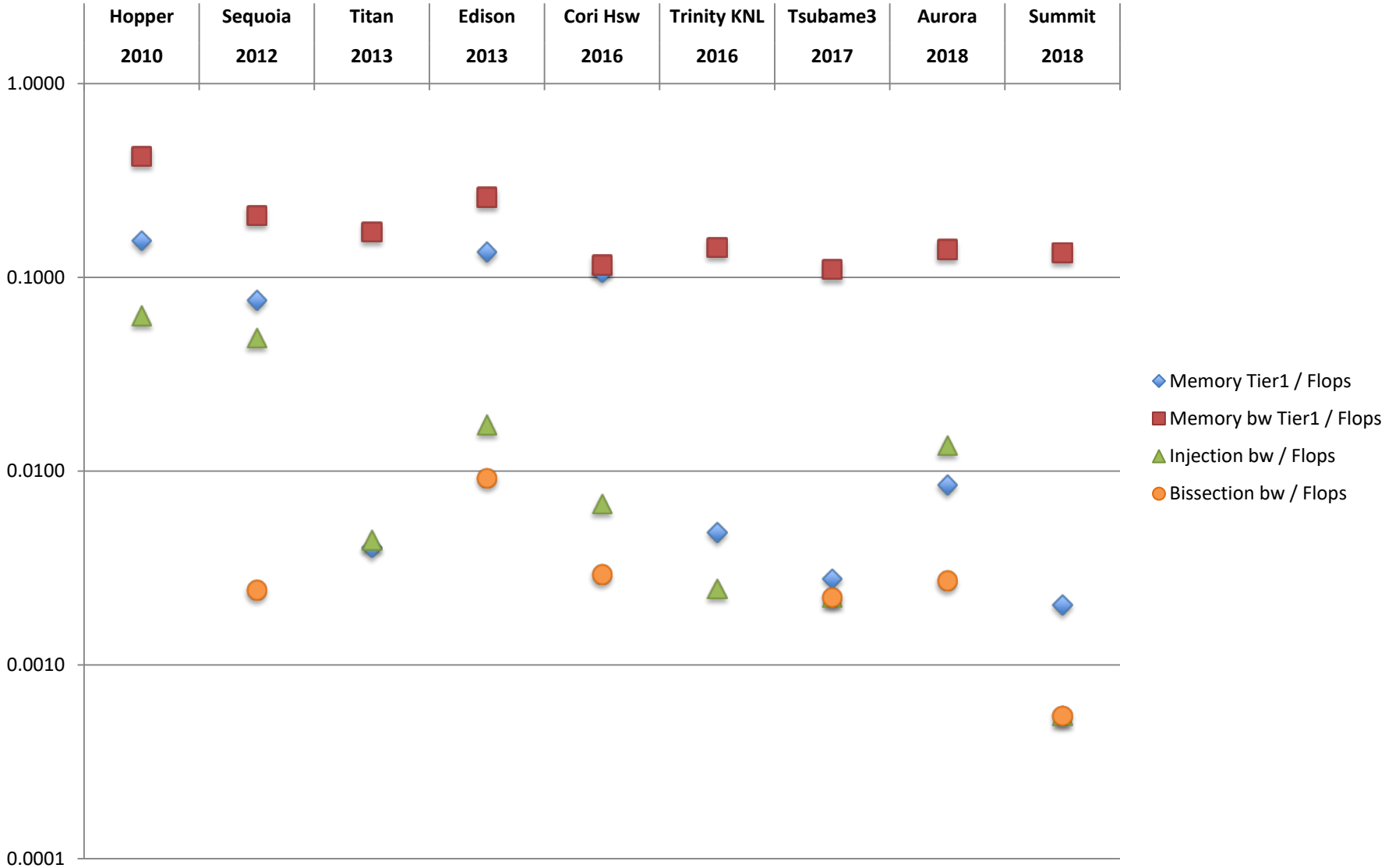
The “curse of constant transistor power”
- Ignorance of this is like ignoring global warming -

- Systems people have been telling the algorithm people that “FLOPS will be free, bandwidth is important, so devise algorithms under that assumption”
- This will certainly be true until exascale in 2020…
- But when Moore’s Law ends in 2025-2030, constant transistor power (esp. for logic) = FLOPS will no longer be free!
- So algorithms that simply increase arithmetic intensity will no longer scale beyond that point
- Like countering global warming – need disruptive change in computing – in HW-SW-Alg-Apps etc. for the next 20 year era

Performance growth via ***data-centric computing:***
"From FLOPS to BYTES"

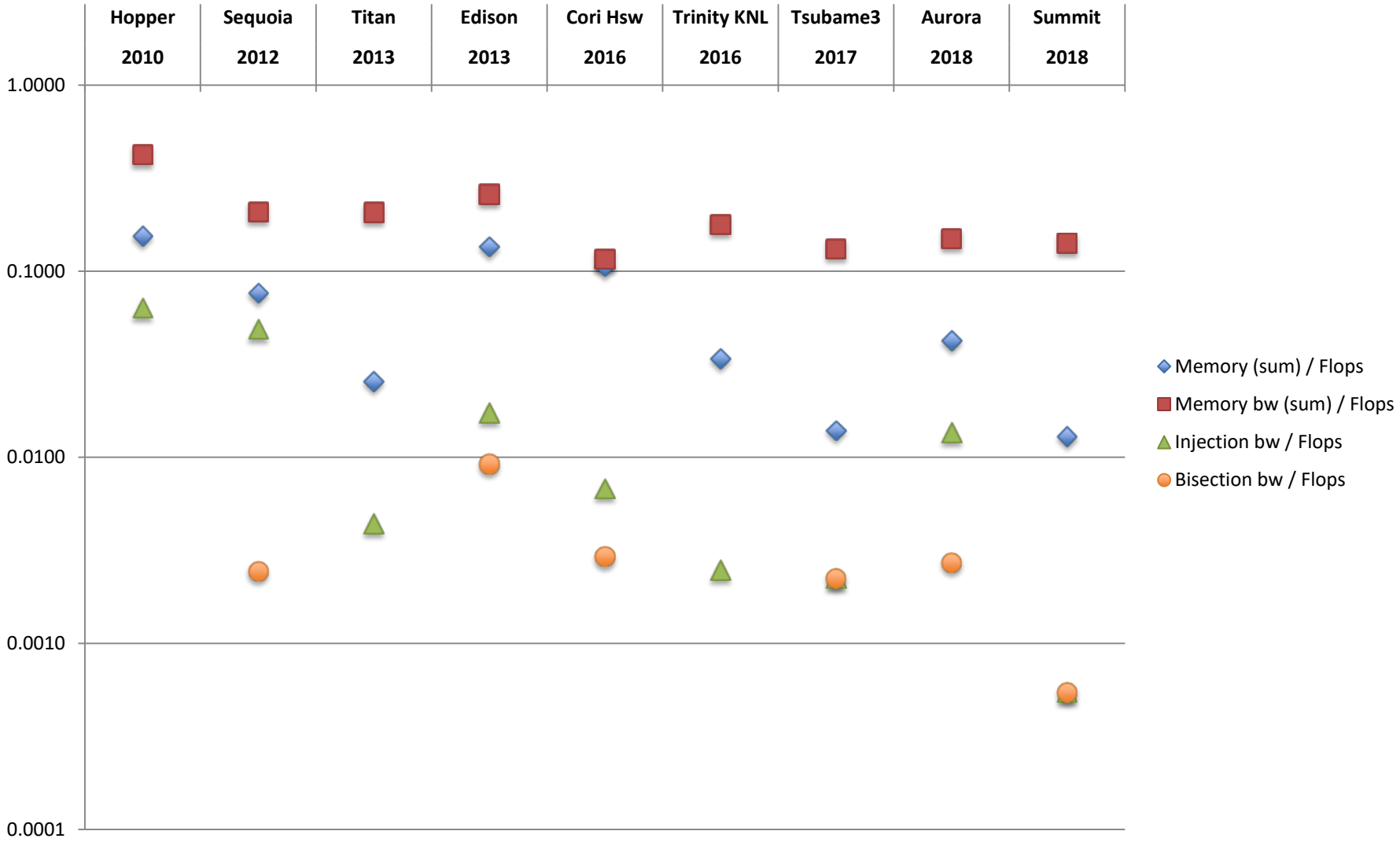
- *Identify the new parameter(s) for scaling over time*
- Because data-related parameters (e.g. capacity and bandwidth) *will still likely continue to grow towards 2040s*
- Can grow transistor# for compute, but CANNOT use them AT THE SAME TIME(Dark Silicon) => **multiple computing units specialized to type of data**
- **Continued capacity growth**: 3D stacking (esp. direct silicon layering) and low power NVM (e.g. ReRAM)
- **Continued BW growth:** Data movement energy will be **capped constant** by dense 3D design and advanced optics from silicon photonics technologies
- Almost back to the old "vector" days(?), but no free lunch – latency still problem, locality still important, *need* ***general algorithmic acceleration thru data capacity and bandwidth,*** *not FLOPS*

Leadership Systems Ratios (Tier1 memory)(slide courtesy Nic Dube@HPE)



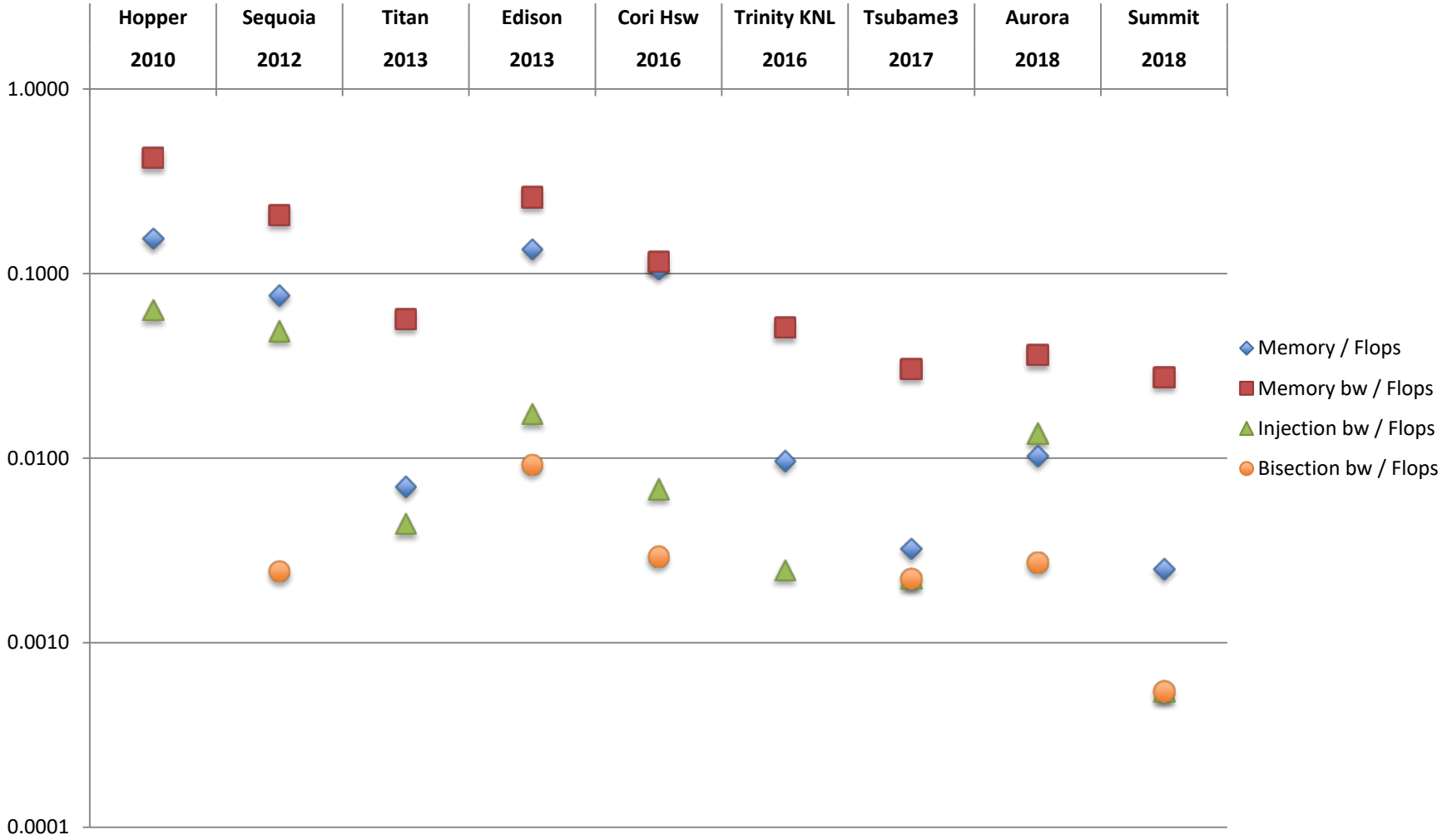
Hewlett Packard Enterprise

Leadership Systems Ratios (sum)) (slide courtesy Nic Dube@HPE)



Hewlett Packard Enterprise

Leadership Systems Ratios (wAvg) (slide courtesy Nic Dube@HPE)



Hewlett Packard Enterprise

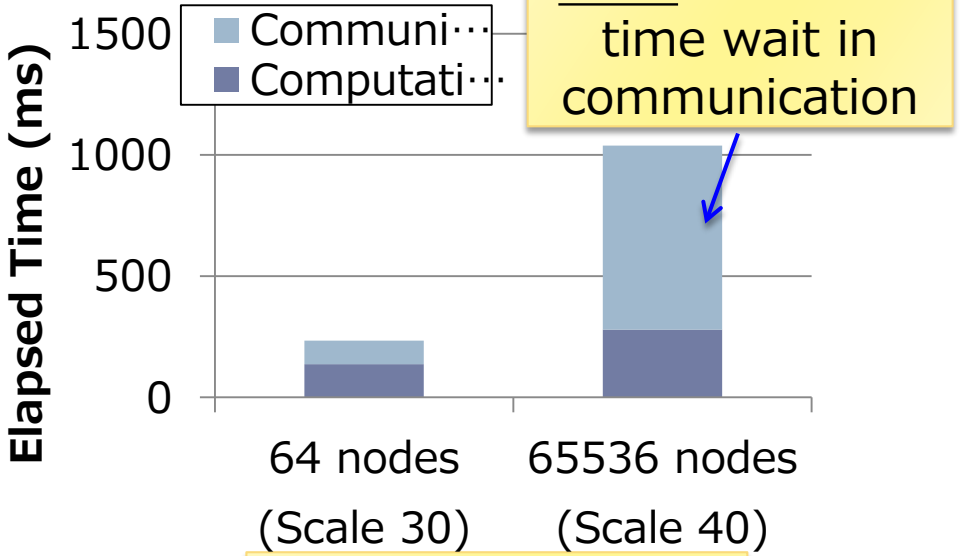
HPCG Top 10 ranking June 2017

#	T	Site	Manufacturer	Computer	Country	HPCG [Pflop/s]	Rmax [Pflop/s]	HPCG/ Peak	HPCG/ HPL
1	8	RIKEN Advanced Institute for Computational Science	Fujitsu	K Computer SPARC64 VIIIfx 2.0GHz, Tofu Interconnect	Japan	0.6027	10.5	5.3%	5.7%
2	2	National University of Defense Technology	NUDT	Tianhe-2 NUDT TH-IVB-FEP, Xeon 12C 2.2GHz, IntelXeon Phi	China	0.5801	33.9	1.1%	1.7%
3	3	Swiss National Supercomputing Centre (CSCS)	Cray	Piz Daint Cray XC50, Xeon E5 12C 2.6GHz, Aries, NVIDIA Tesla P100	Switzerland	0.4700	19.6	1.9%	2.4%
4	7	JCAHPC Joint Center for Advanced HPC	Fujitsu	Oakforest-PACS PRIMERGY CX1640 M1, Intel Xeons Phi 7250 68C 1.4 GHz, OmniPath	Japan	0.3855	13.6	1.5%	2.8%
5	1	National Supercomputing Center in Wuxi	NRCPC	Sunway TaihuLight NRCPC Sunway SW26010, 260C 1.45GHz	China	0.3712	93.0	0.3%	0.4%
6	6	Lawrence Berkeley National Laboratory	Cray	Cori Cray XC40, Intel Xeons Phi 7250 68C 1.4 GHz, Aries	USA	0.3554	14.0	1.3%	2.5%
7	5	Lawrence Livermore National Laboratory	IBM	Sequoia BlueGene/Q, Power BQC 16C 1.6GHz, Custom	USA	0.3304	17.2	1.6%	1.9%
8	4	Oak Ridge National Laboratory	Cray	Titan Cray XK7, Opteron 16C 2.2GHz, Gemini, NVIDIA K20x	USA	0.3223	17.6	1.2%	1.8%
9	10	Los Alamos NL / Sandia NL	Cray	Trinity Cray XC40, Xeon E5 16C 2.3GHz, Aries	USA	0.1826	8.10	1.6%	2.3%
10	15	NASA/ Ames Research Center/NAS	HPE	Pleiades SGI ICE X, Xeon E5 10C 2.4 2.8GHz, Infiniband FDR	USA	0.1750	5.95	2.5%	2.9%

Sparse BYTES: The Graph500 – 2015~2016 – world #1 x 4

K Computer #1 Tokyo Tech[Matsuoka EBD CREST] Univ.

Kyushu [Fujisawa Graph CREST], Riken AICS, Fujitsu



73% total exec time wait in communication

88,000 nodes,
660,000 CPU Cores
1.3 Petabyte mem
20GB/s Tofu NW

K computer

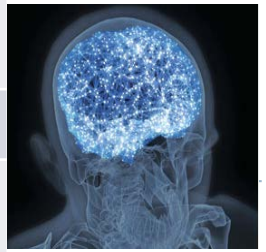
Effective x13 performance c.f. Linpack

TaihuLight
10 million CPUs
1.3 Petabyte mem

LLNL-IBM Sequoia
1.6 million CPUs
1.6 Petabyte mem

BYTES Rich Machine + Superior BYTES algoithm

List	Rank	GTEPS	Implementat
November 2013	4	5524.12	Top-down o
June 2014	1	17977.05	Efficient hybrid
November 2014	2	19585.2	Efficient hybrid
June, Nov 2015 June Nov 2016	**1**	**38621.4**	Hybrid + Node Compression



BYTES, not FLOPS!

K computer “Still the best” for Bandwidth (Data-centric) workloads (It’s the Bandwidth!) And TSUBAME3, too

FUJITSU

	2011	2012	2013	2014	2015	2016	2017
1.TOP500 List		2	4	4	4	7	8
2. HPCG					2		
3. Gordon Bell Prize						Finalist	
3. HPC Challenge Awards (HPC、Random Access、STREAM、FFT)							
4. Graph500	4	2					
5. Green500(TSUBAME)					2		

Characteristics of Big Data and AI Computing

As BD / AI

Graph Analytics e.g. Social Networks
Sort, Hash, e.g. DB, log analysis
Symbolic Processing: Traditional AI

As HPC Task

Integer Ops & Sparse Matrices
Data Movement, Large Memory
Sparse and Random Data, Low Locality

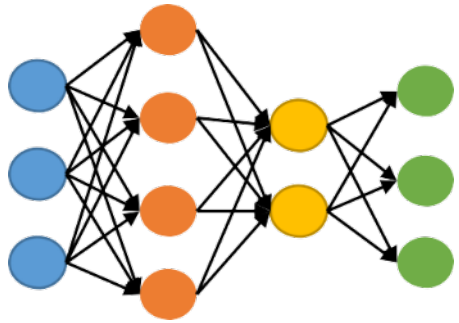
Acceleration, Scaling

Opposite ends of HPC computing spectrum, but HPC simulation apps can also be categorized likewise

Acceleration via Supercomputers adapted to AI/BD

As BD / AI

Dense LA: DNN
Inference, Training, Generation



As HPC Task

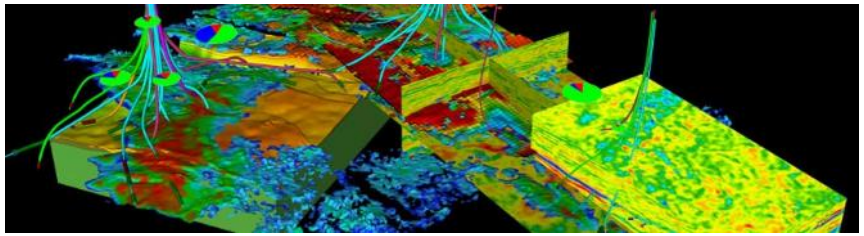
Dense Matrices, Reduced Precision
Dense and well organized neworks and Data

Acceleration, Scaling

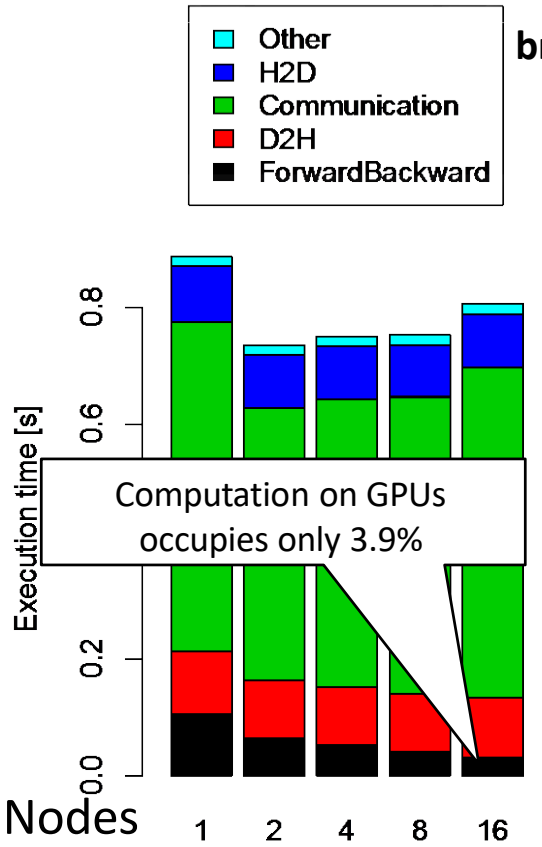
(Big Data) BYTES capabilities, in bandwidth and capacity, unilaterally important but often missing from modern HPC machines in their pursuit of FLOPS…

- **Need <u>BOTH bandwidth and capacity</u> (BYTES) in a HPC-BD/AI machine:**
  - Obvious for lefthand sparse ,bandwidth-dominated apps
  - But also for righthand DNN: Strong scaling, large networks and datasets, in particular for future 3D dataset analysis such as CT-scans, seismic simu. vs. analysis…)

(Source: https://www.spineuniverse.com/image-library/anterior-3d-ct-scan-progressive-kyphoscoliosis)



(Source: http://www.dgi.com/images/cvmain_overview/CV4DOverview_Model_001.jpg)

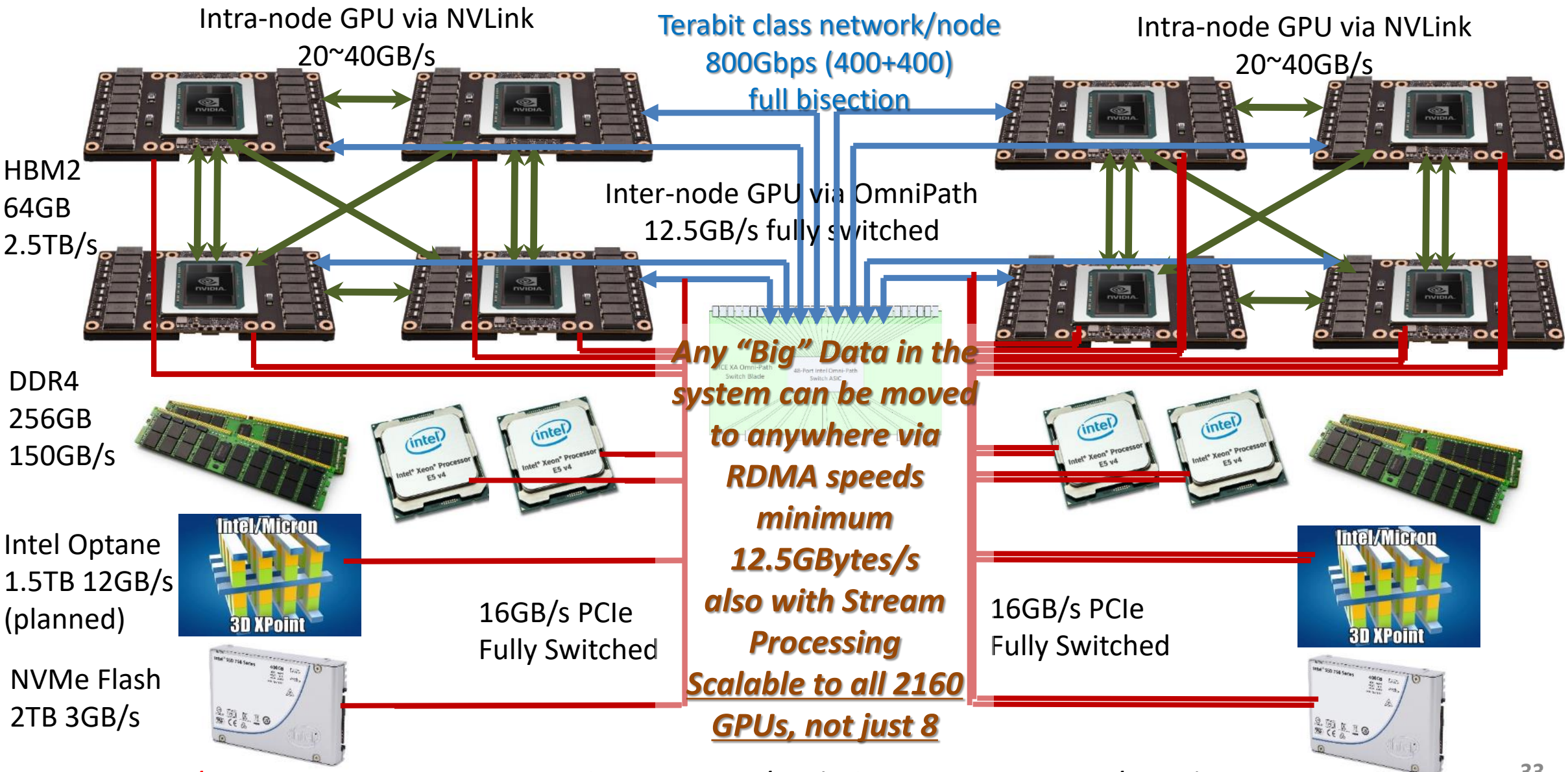


Our measurement on breakdown of one iteration of CaffeNet training on TSUBAME-KFC/DL (Mini-batch size of 256)

Proper arch. to support large memory cap. and BW, network latency and BW important

Number of GPUs = 8 per node

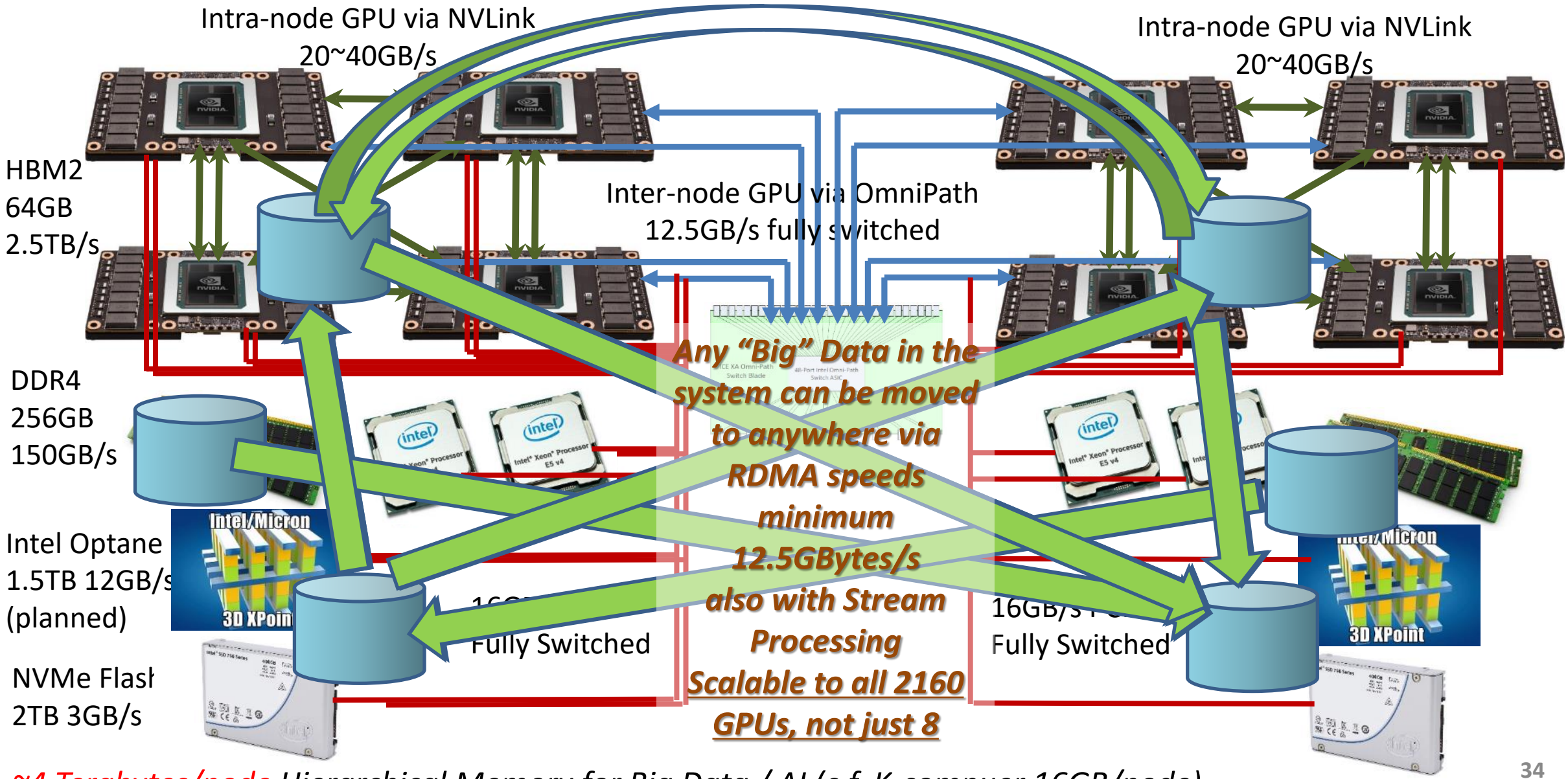
TSUBAME3: *A Massively BYTES Centric Architecture* for Converged BD/AI and HPC



~4 Terabytes/node Hierarchical Memory for Big Data / AI (c.f. K-compuer 16GB/node)

➔ *Over 2 Petabytes in TSUBAME3, Can be moved at 54 Terabyte/s or 1.7 Zetabytes / year*

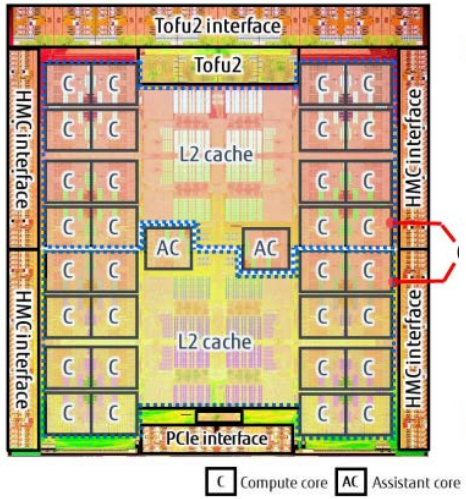
TSUBAME3: *A Massively BYTES Centric Architecture* for Converged BD/AI and HPC

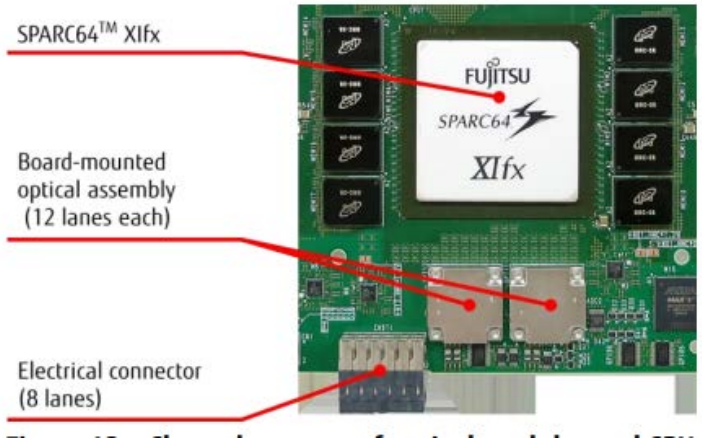


~4 Terabytes/node Hierarchical Memory for Big Data / AI (c.f. K-compuer 16GB/node)

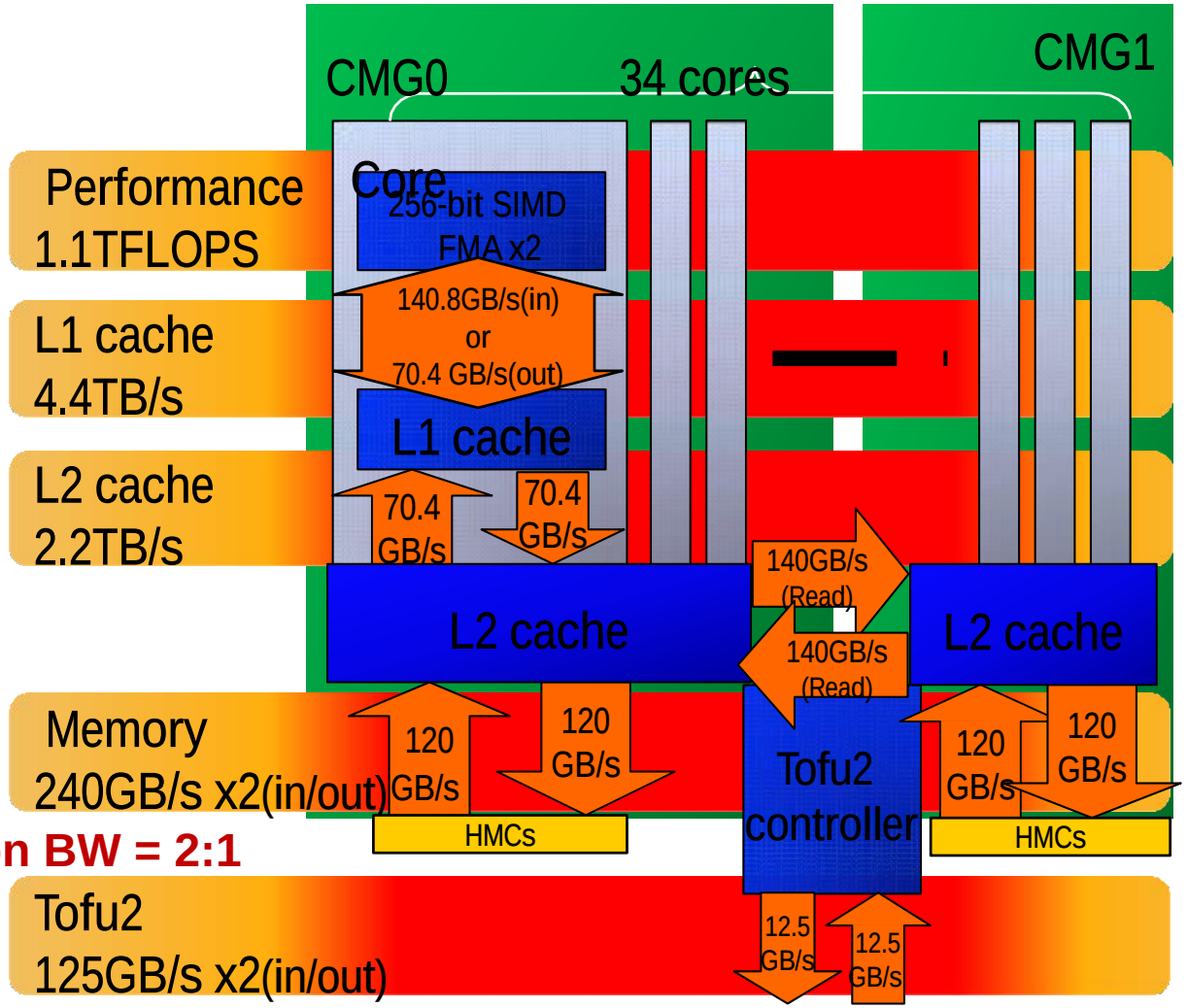
➔ *Over 2 Petabytes in TSUBAME3, Can be moved at 54 Terabyte/s or 1.7 Zetabytes / year*

Fujitsu SPARC64™ Xifx (2015)





Memory BW : Injection BW = 2:1
(TSUBAME3: 40:1)



Japan Flagship 2020 “Post K” Supercomputer

✓CPU
- ARM v8 + 512 bit SVE extensions
- Multi-hundred petaflops peak total
- Power Knob feature for saving power

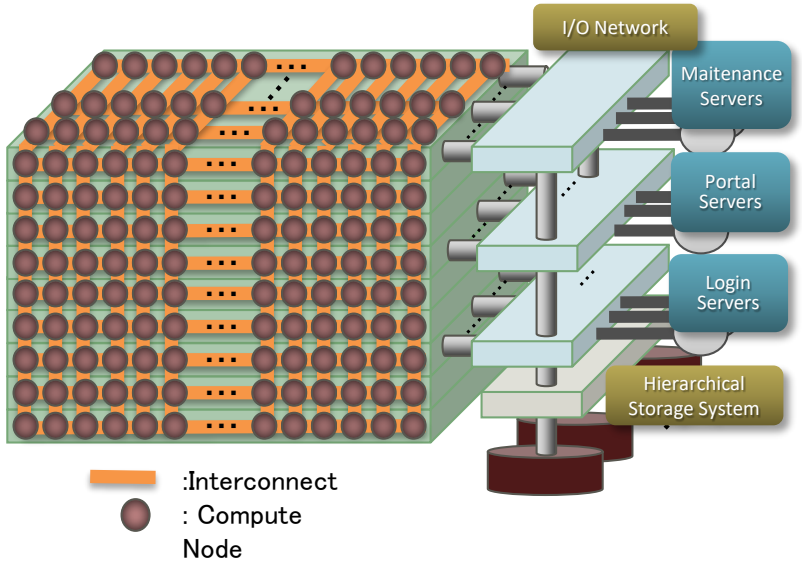
✓Memory
- ✓3-D stacked DRAM, Terabyte/s BW

✓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)

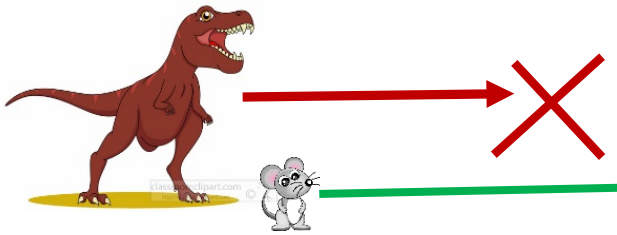
Prime Minister Abe visiting K Computer 2013



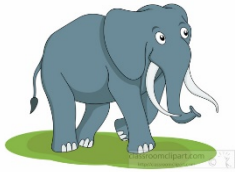
Post-K will be/have

- Continuum of K
- Pretty good DFP FLOPS
- Very Good low precision FLOPS
- Awesome Memory Bandwidth
- (but) relatively low memory capacity
- Awesome Network Injection Bandwidth
- Very Good Network Bisection Bandwidth
- Relatively low I/O speed

Many Core Era



Post Moore Era

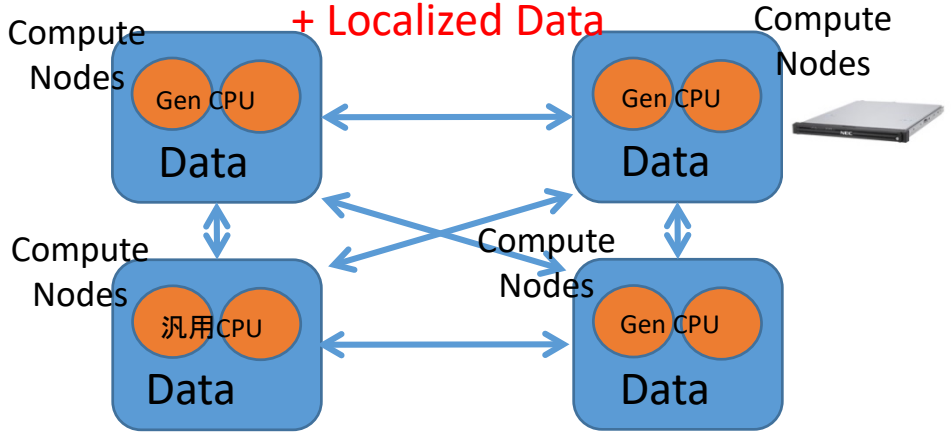


Flops-Centric Algorithms and Apps

Flops-Centric System Software

Hardware/Software System APIs
Flops-Centric Massively Parallel Architecture

Homogeneous General Purpose Nodes
+ Localized Data



Loosely Coupled with Electronic Interconnect

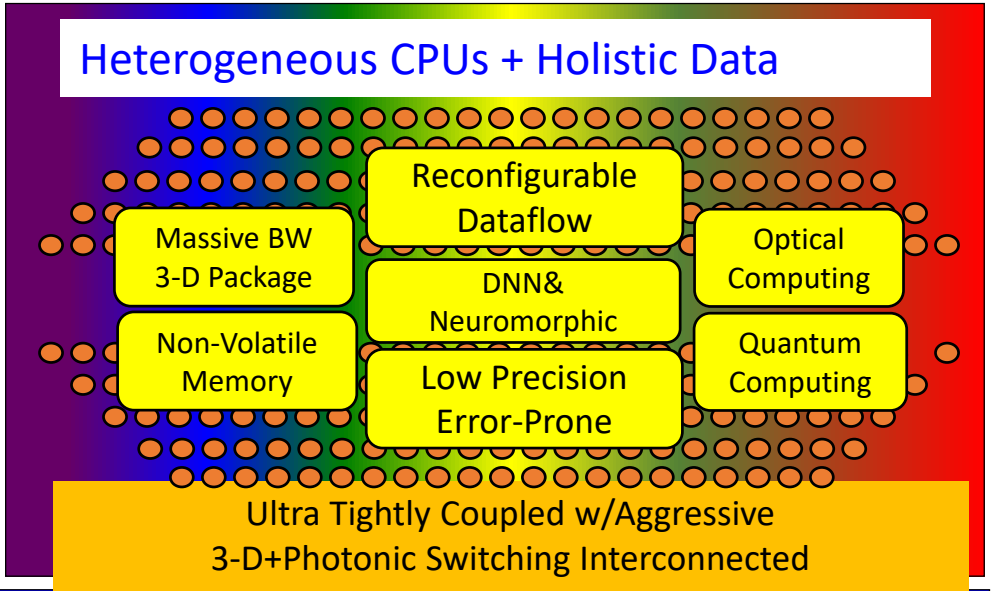
Transistor Lithography Scaling
(CMOS Logic Circuits, DRAM/SRAM)

~2025
M-P Extinction Event

Bytes-Centric Algorithms and Apps

Bytes-Centric System Software

Hardware/Software System APIs
Data-Centric Heterogeneous Architecture

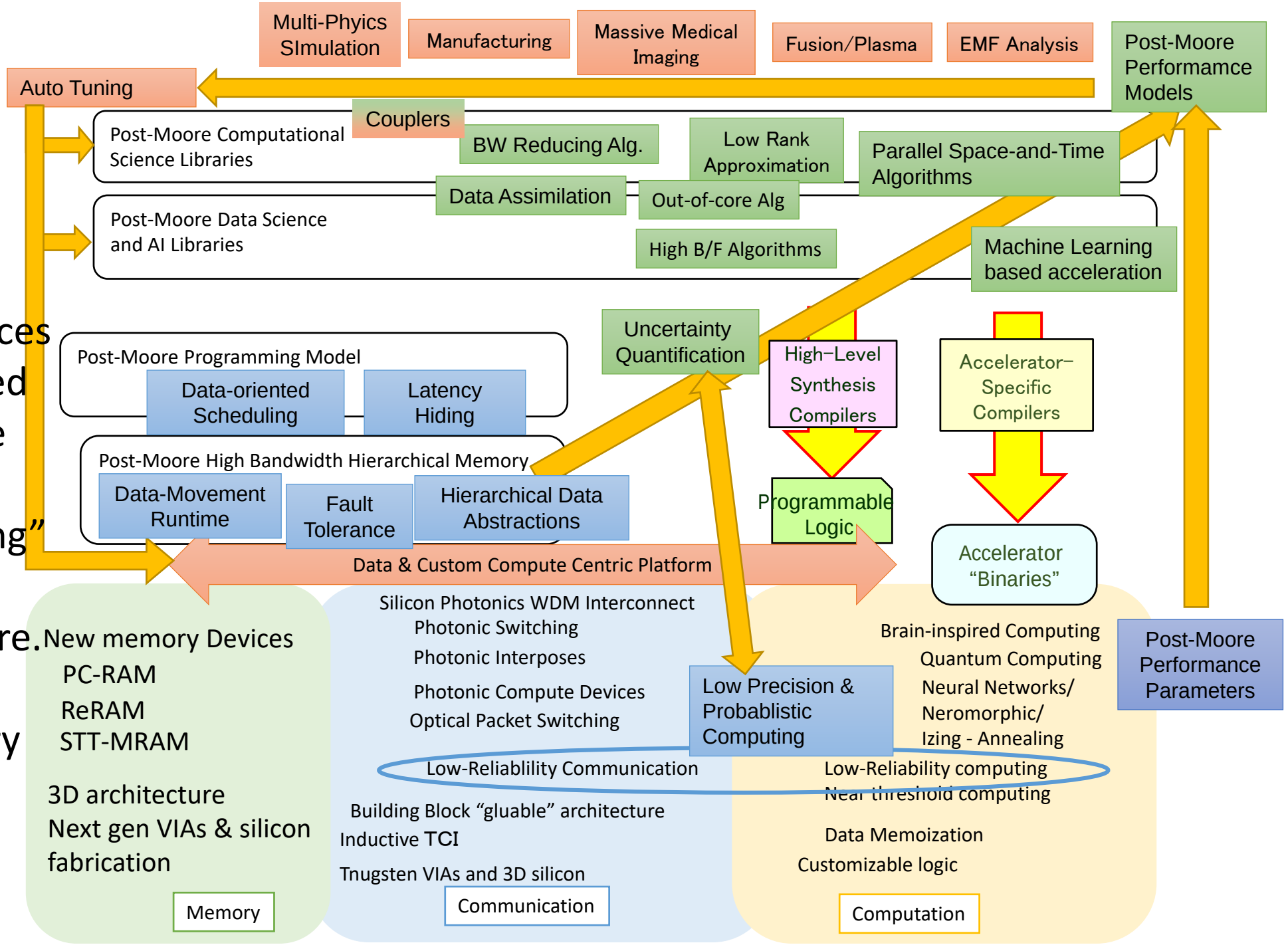


Novel Devices + CMOS (Dark Silicon)
(Nanophotonics, Non-Volatile Devices etc.)

Post-Moore is NOT a More-Moore device as a panacea

Device & arch. advances improving data-related parameters over time

"Rebooting Computing" in terms of devices, architectures, software. Algorithms, and applications necessary
=> Co-Design even more important
c.f. Exascale



Problem Specific Architectures to exploit dark silicone "What are they good for?" – c.f. Berkeley Dwarfs

- Deep Neural Network Accelerator (Many, incl. Google)
- Spiking Neuromorphic Architecture (Manchester SpiNNaker, IBM TrueNorth, Heidelberg BrainScaleS)
- Ising Model optimization architecture (Hitachi)
- Automata Processor (Micron)
- Advanced FPGAs (Alltera, Xilinx)
- Network & I/O accelerator (Mellanox)
- …
- And of course Quantum Annealing and Computing (D-Wave)

Fujitsu Deep Learning Processor (DLU™)

FUJITSU

京 K computer

Supercomputer K technologies

FY2018～

DLU™

(Deep Learning Unit)

FUJITSU DLU™ Deep Learning Unit

DLU™ features

- Architecture designed for Deep Learning
- High performance HBM2 memory
- Low power design

→ Goal: 10x Performance/Watt compared to others

- Massively parallel： Apply supercomputer interconnect technology

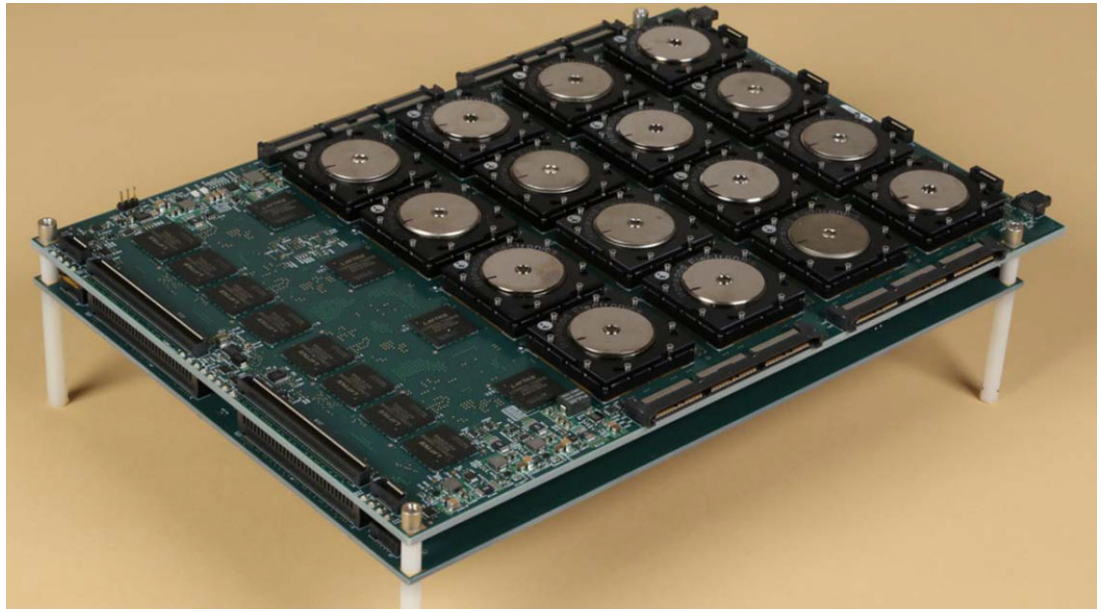
→ Ability to handle large scale neural networks

→ TOFU Network derivative for massive scaling

Designed for Scalable Learning, technically superior to Google TPU2

"Exascale" AI possible in 1H2019

Neuromorphic Architectures
(Not to be confused with DNN Accelerators)



IBM TrueNorth

SpiNNaker
Biologically Inspired Massively Parallel Architectures

MANCHESTER 1824

Distributed Neuronal Computation "on chip"

Asynchronous Spiking Neural Computation Hardware for low-power real-time operation in Closed-Loop Systems

... to simulate 1 Billion Spiking Neurons in real-time

882 Core "Rack Version"

72 Core Evaluation

18 Core Stand-Alone Prototype

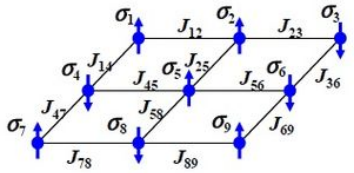
- Multi-channel spiking input and output
- Stand-alone spiking computing system
- Simulates ~20.000 neurons in real time
- Small (~20x20mm); low power (~600mW)
- Flexibly configurable, extendable, stackable

Manchester SpiNNaker
(ARM Based)

- Hitachi@ISSCC2015
  "An 1800-Times-Higher Power-Efficient 20k-spin Ising Chip for Combinational Optimization Problem with CMOS Annealing"
- Competitive to Quantum Annealing, room temperature, easy to scale
- Could be applicable to deep learning?

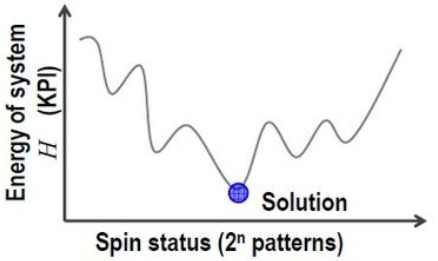
Computing with Ising model

- Ising model: expressing behavior of magnetic spins
- Using Ising model as natural phenomenon to map problems



$$H = -\sum_{\langle i,j \rangle} J_{ij}\sigma_i\sigma_j - \sum_j h_j\sigma_j$$

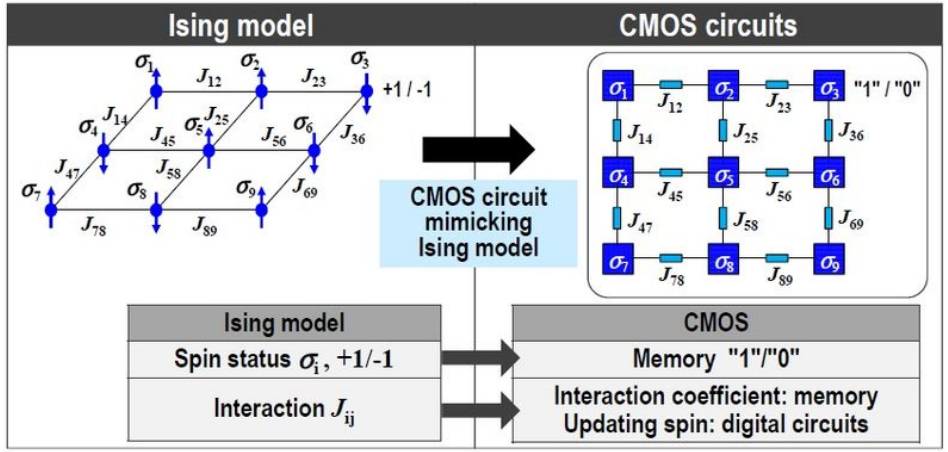
H: Energy of system
σ_i: Spin status (+1/-1)
J_{ij}: Interaction coefficient



n: number of spins

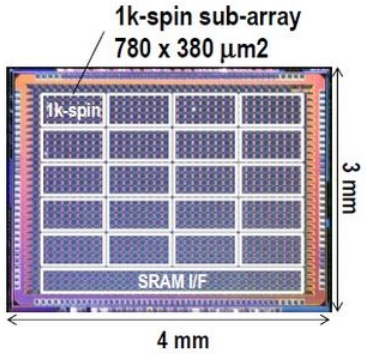
CMOS Ising computing

- Mimicking Ising model with CMOS circuits
- Easy to manufacture, easy to use, good scalability



Ising model	CMOS
Spin status σ_i, +1/-1	Memory "1"/"0"
Interaction J_{ij}	Interaction coefficient: memory Updating spin: digital circuits

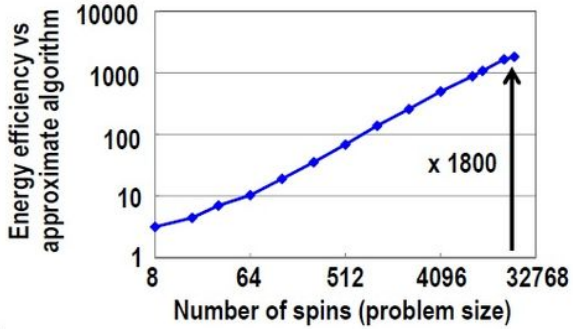
Fabrication results



Items	Value
Number of spins	20k (80 x 256)
Process	65 nm
Chip area	4x3=12 mm²
Area of spin	11.27 x 23.94 =270 µm²
Number of SRAM cells	260k bits Spin value: 1 bit Interaction factor: 2 bit x 5=10 bits External magnetic coefficient: 2 bits
Memory IF	100 MHz
Interaction speed	100 MHz
Operating current of core circuits (1.1 V) Do not include IO	Write: 2.0 mA Read: 6.0 mA Interaction: 44.6 mA

Measurement results of energy

- 1,800 times higher energy efficiency than conventional approximation algorithm on CPU



Conditions:
Randomly generated problems, energy for same preciseness solution
Ising chip: VDD=1.1 V, 100-MHz interaction, best solution among 10-times trial is selected.
Approximation algorithm: SG3(*) is operated on Core i5, 1.87 GHz, 10 W/core.

(*): Sera Kahruman et al., "On Greedy Construction Heuristics for the Max-Cut Problem," International Journal on Computational Science and Engineering, Volume 3, Number 3/2007, pp. 211-218, 2007.

Towards Understanding the Performance of FPGAs using OpenCL Benchmarks [HiPEAC Reconfigurable Computing WS 2015 Extended version to appear SC16]

Hamid Zohouri (Tokyo Tech), Naoya Maruyama (Riken AICS), Satoshi Matsuoka (Tokyo Tech), Motohiko Matsuda (RIKEN AICS)

In collaboration with:
Aaron Smith (Microsoft Research),

Supported by Altera

TSUBAME
Tokyo Institute of Technology

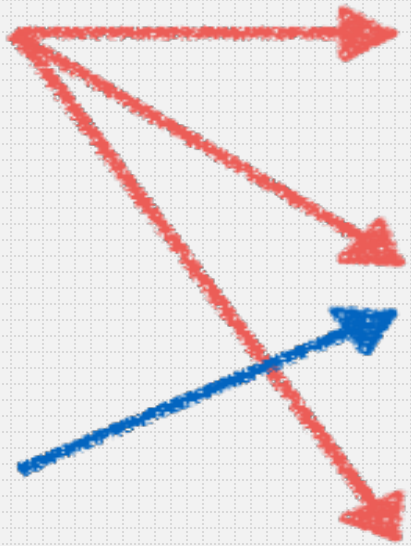
Microsoft Research

ALTERA

Parallelism in Altera OpenCL

Explicit:
Thread/SIMD parallelism

Implicit:
Pipeline parallelism



- **Inter pipelines**
  - Configurable number of duplicated pipelines
- **Intra pipeline**
- **SIMD**
  - Instantiate SIMD units base on user direction (attribute `num_simd_work_items`)

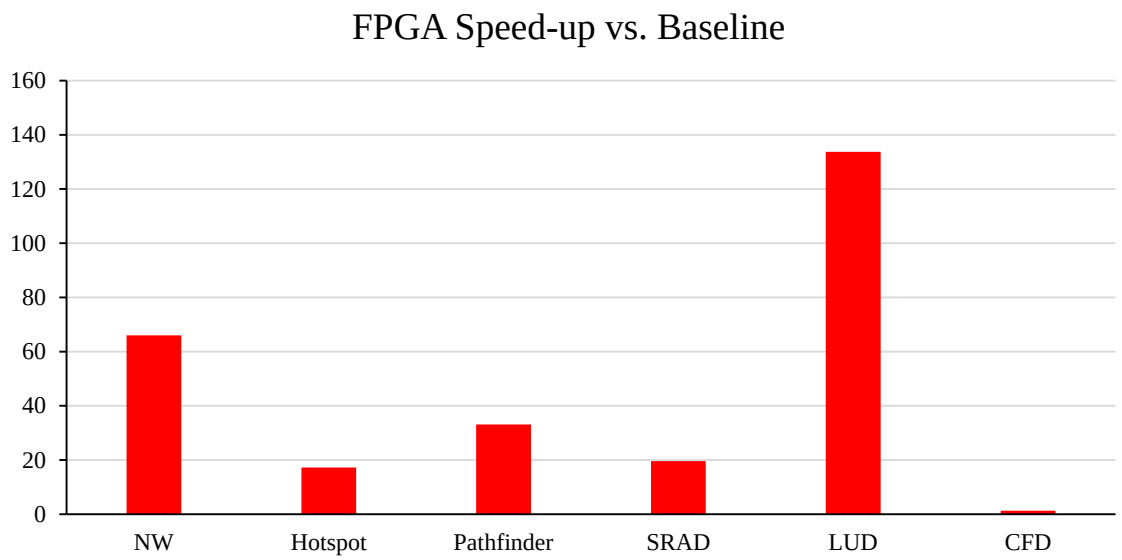
Optimization Effects

Type	Optimization	F_{max} (MHz)	Run Time (ms)	Power Dissipation (Watt)	Power Usage (J)
MT	None	277.2	16574	12.01	199.1
Pipeline	None	243.4	117523	10.59	1245.2
MT	Basic	194.7	2445	16.94	41.4
Pipeline	Basic	249.1	116457	9.93	1156.7
Pipeline	Advanced	148.0	251	15.44	3.8

Sliding window is 66x faster than baseline

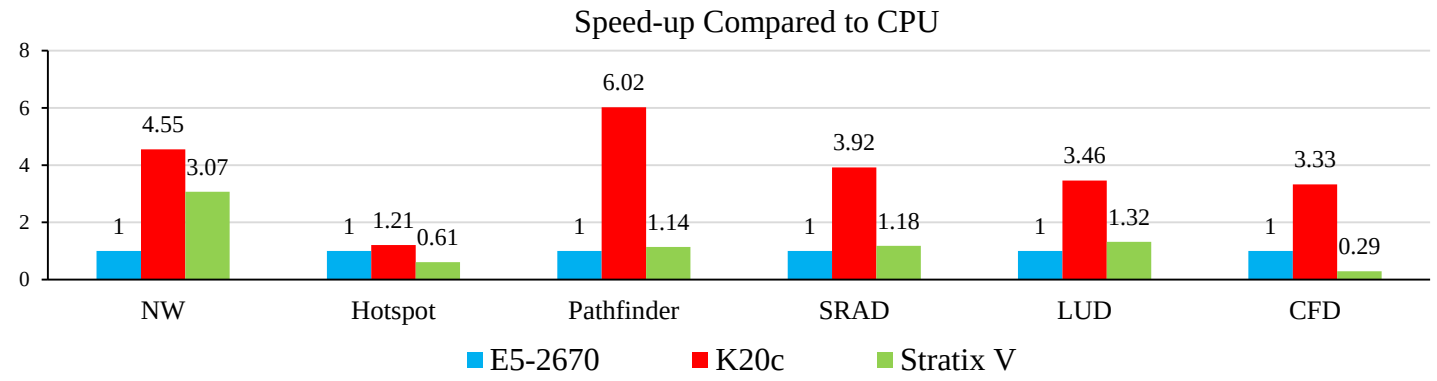
Stratix V Speed-up vs. Baseline

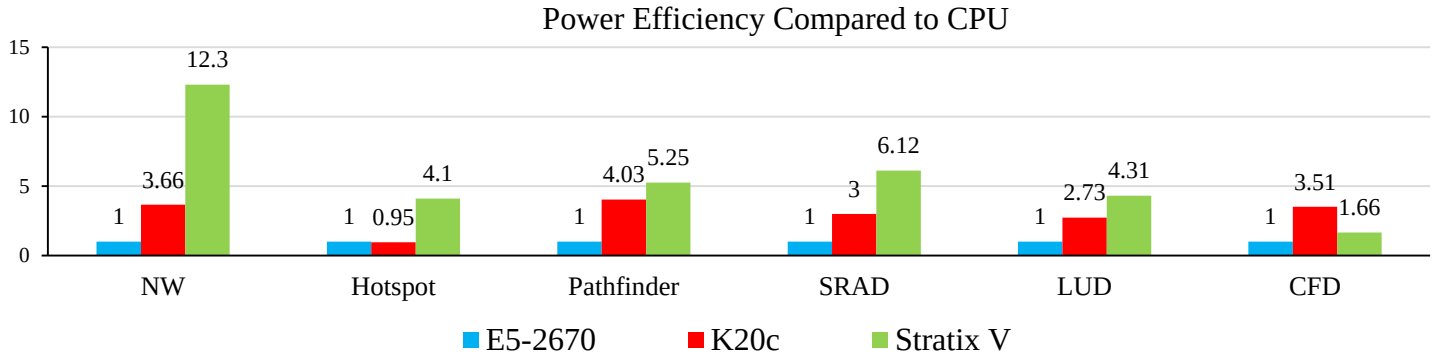
- Up to 133x speed-up
- CFD speed-up is minimal due to lack of area for optimization



Stratix V vs. E5-2670 & K20c

- CFD only has basic optimization
  - Performance is not good
- Stratix V:
  - Beats CPU performance in all benchmarks except Hotspot
  - Beats CPU power efficiency in all benchmarks
  - Cannot beat GPU performance in any benchmark
  - Beats GPU power efficiency in all benchmarks up to 3.4x





Portfolio of Accelerators – What are they good for – (1) general purpose

- Vector (SIMD) HPC accelerators => General purpose HPC esp. SIMD
  - GPUs, Xeon Phi, Shenwei SW26010
  , Post-K ARM-SVE
- Macro Dataflow Processors => asynch threads, functional programming etc.
  - ETL EM-4, Wave computing, …
- High memory bandwidth accelerators (FLOPS increase = Moore's law ending) => memory capacity and BW increase, =>
  - NEC SX-Aurora Vector processor BYTES/FLOPS ~= 1
  - Future 3-D die stacked architectures, NVM including NV-DIMM
- FPGAs – programmable HLS languages e.g. OpenCL
  - Intel Stratix 10, Xilinx Vertex 9
- Superconducting Accelerators
  - Massive single thread performance to accelerate serial bottleneck

Portfolio of Accelerators – – What are they good for – (2) ML / AI

- DNN accelerators – accelerating tensor operations => many ML apps
  - Small Matrix-Tensor Engine (NVIDIA Volta TensorCore, Intel LakeCrest?)
  - Systoric Array (Google TPU2)
  - Small Vector processor arrays (Fujitsu DLU)
- Neuromorphic accelerators –Spiking Neural Networks => brain simulation, more power efficient ML?
  - IBM TrueNorth
  - Manchester U Spinnaker
  - Heidelberg U BrainScales
  - Many others (DoE, U-Tokyo – NEC, SIngapore, …)
- Symbolic Computing accelerators => string searches. E.g. Genomics?
  - Micron Automata Processor

Portfolio of Accelerators – – What are they good for – (3) Quantum and Pseudoquantum

- Quantum Annealers (theory invented at Tokyo Tech.) => ML?
  - D-Wave
  - Others in the lab
- Pseudoquantum (CMOS) Annealers => ML?
  - Hitachi Ising chip (ISSCC 2015)
  - Fujitsu (Pseudo)quantum annealing chip
- Quantum Gate processors => Quantum simulation, Cryptography
  - Many ongoing work

I/O and Data accelerators – What are they good for (4) I/O

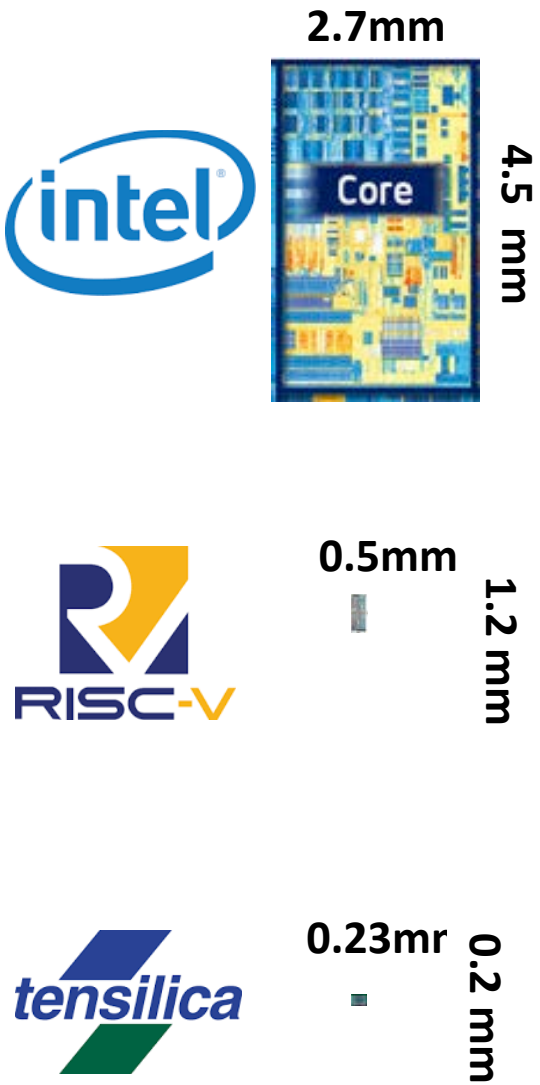
- Cray MTA – graph operations
- Burst Buffers – I/O intensive ops e.g. checkpoints
- Database accelerators – classic

Non-Volatile Memory and 3-D Stacking

- Many devices
- Various stacking technologies

- Results: Massive capacity, extreme bandwidth, low power
- Exploits Z-direction locality
- New breed of “in memory computing”
- Could persist as a trajectory for the next 20 years

When does data movement dominate? (Original Slide Courtecy John Shalf@LBNL)



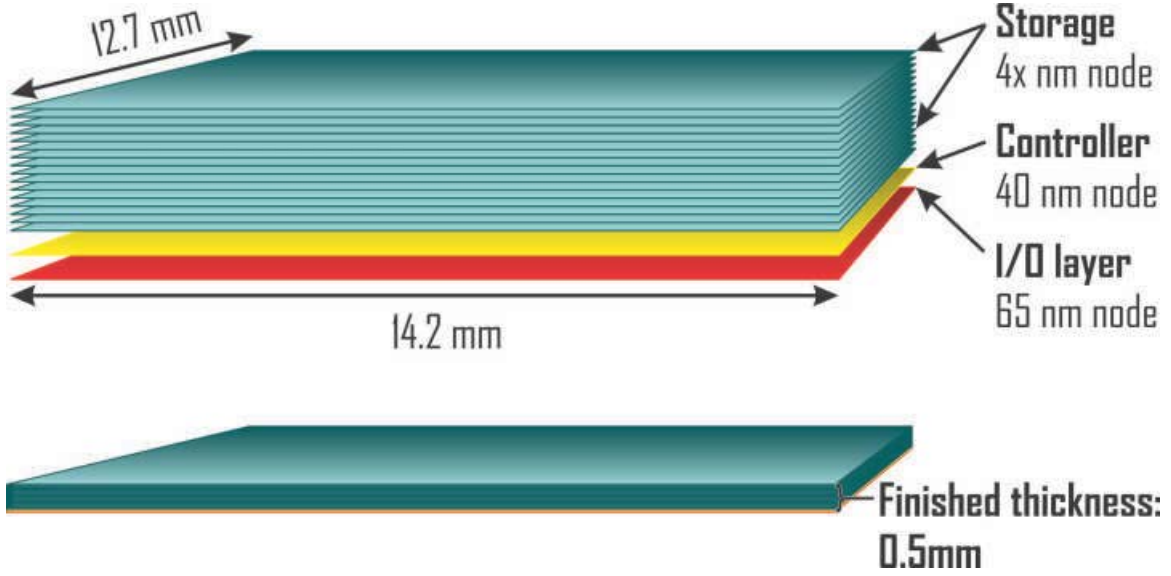
Core Energy/Area est.	Data Movement Cost
Area: 12.25 mm^2 Power: 2.5W Clock: 2.4 GHz E/op: 651 pj	*Compute Op == data movement Energy @* **108mm** *Energy Ratio for 20mm* **0.2x**
Area: 0.6 mm^2 Power: 0.3W (<0.2W) Clock: 1.3 GHz E/op: 150 (75) pj	*Compute Op == data movement Energy @* **12mm** *Energy Ratio for 20mm* **1.6x**
Area: 0.046 mm^2 Power: 0.025W Clock: 1.0 GHz E/op: 22 pj	*Compute Op == data movement Energy @* **3.6mm** *Energy Ratio for 20mm* **5.5x**

Could be reduced by orders of magnitude by 3D, as Z-direction movement is under 1mm

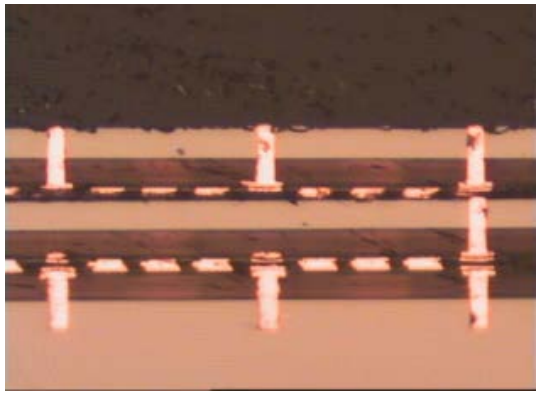
Capacity by dense NVM w/DRAM cache

Example Innovation: Tungsten TSV at 2um ultra fine pitch with die thinning by Tezzaron Semiconductor

- Suppose 4TF SFP @ 7nm, 16TB/s internal chip BW vs. 200GB/s external chip mem BW => 80 times speedup!
- High-density, high-signaling TSV challenge
  - Wide I/O 2 1024 bits 1 Ghz -> 2~3 Ghz
  - We need 128,000 bits @ 1Ghz !
  - 10 micron TSV estimation
    - 400 x 400 TSVs on 20mx20m chip -> 50 micron spacing
    - With tungsten TSVs the chip area is negligible



Many-layer stacking via aggressive wafer thinning and self-diagnostics



Source: Tezzaron website
http://www.tezzaron.com

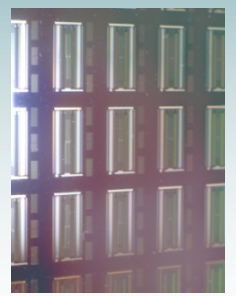
DiRAM4 Stack Overview

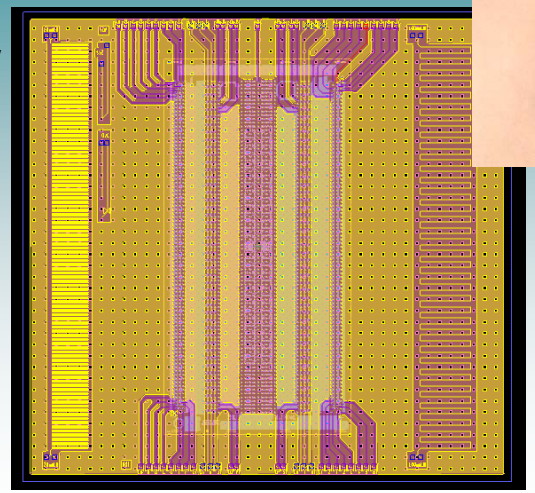
(Tezzaron slides taken from http://www.tezzaron.com/media/Tezzaron-Presentation-EPS-100814-dist-.pptx)

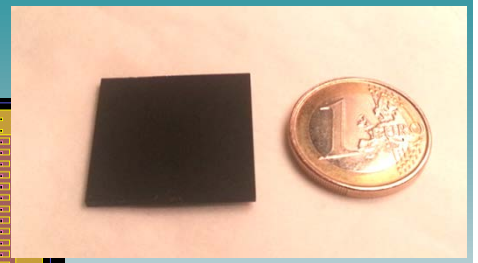
- **64 Gb** of Memory in 175 mm^2
- **256** fully independent RAMs
- **16 Banks** per RAM
- **64 bit** Sep I/O Data per RAM
- 7ns Access Time (Closed page to data)
- 12ns tRC (Page Open to Page Open in a Bank)
- 16 Tb/s Data Bandwidth
- Competitive Manufacturing Cost

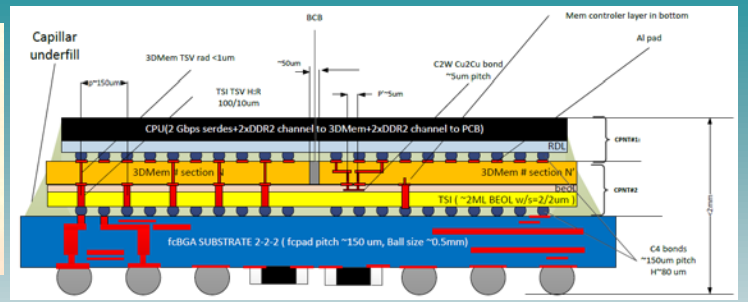
2.5/3D Circuits

IME A*STAR / Tezzaron Collaboration

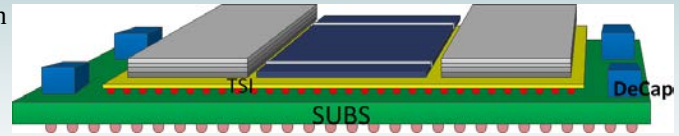


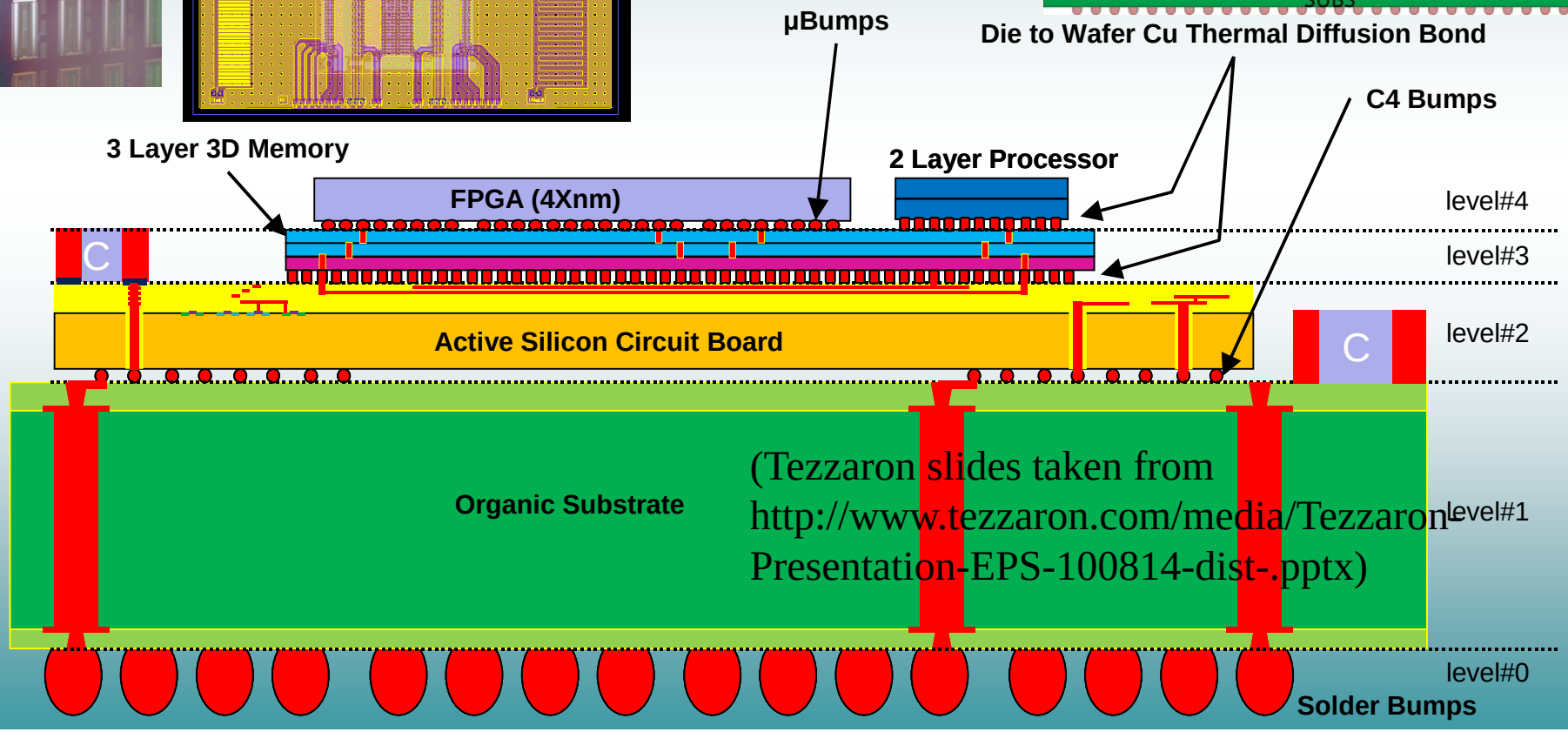






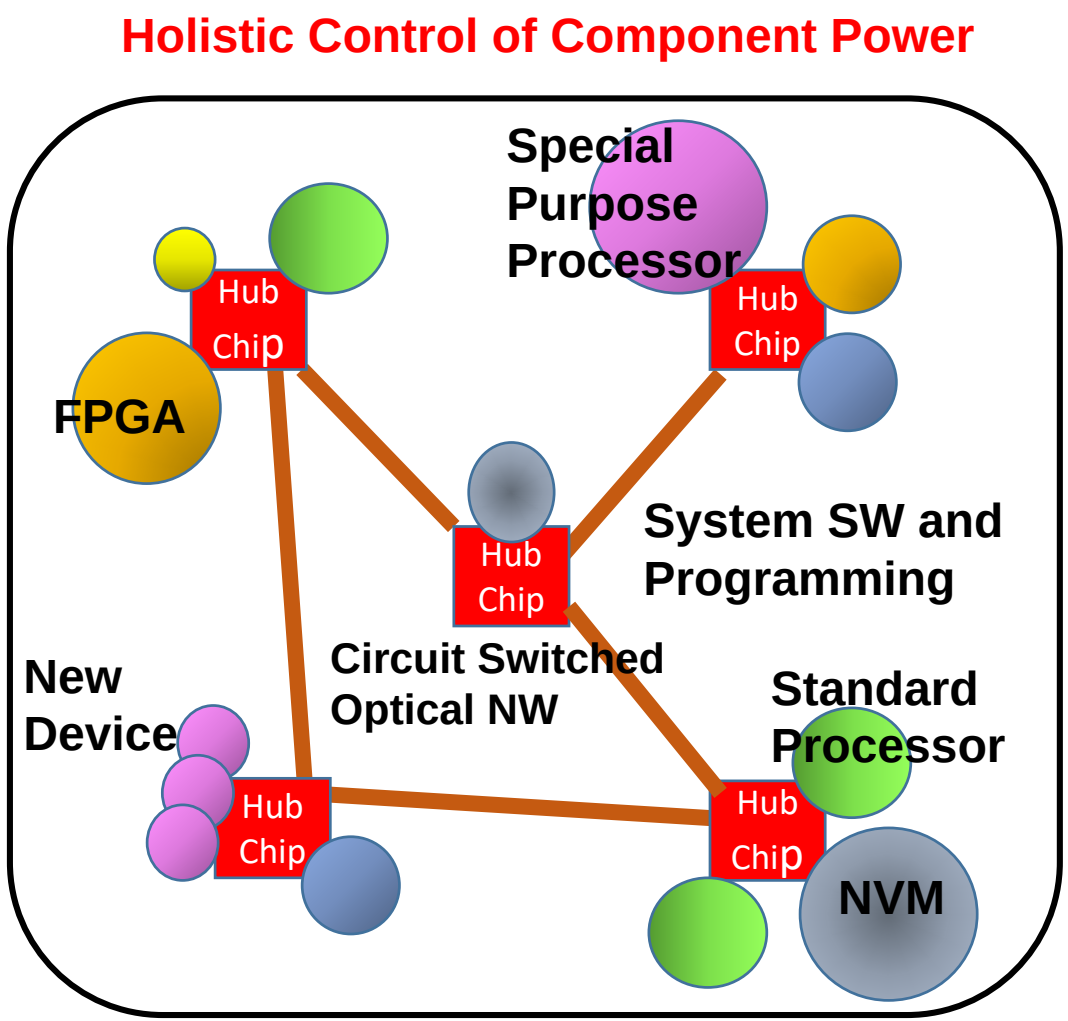
IME A*STAR / Tezzaron Collaboration





(Tezzaron slides taken from http://www.tezzaron.com/media/Tezzaron Presentation-EPS-100814-dist-.pptx)

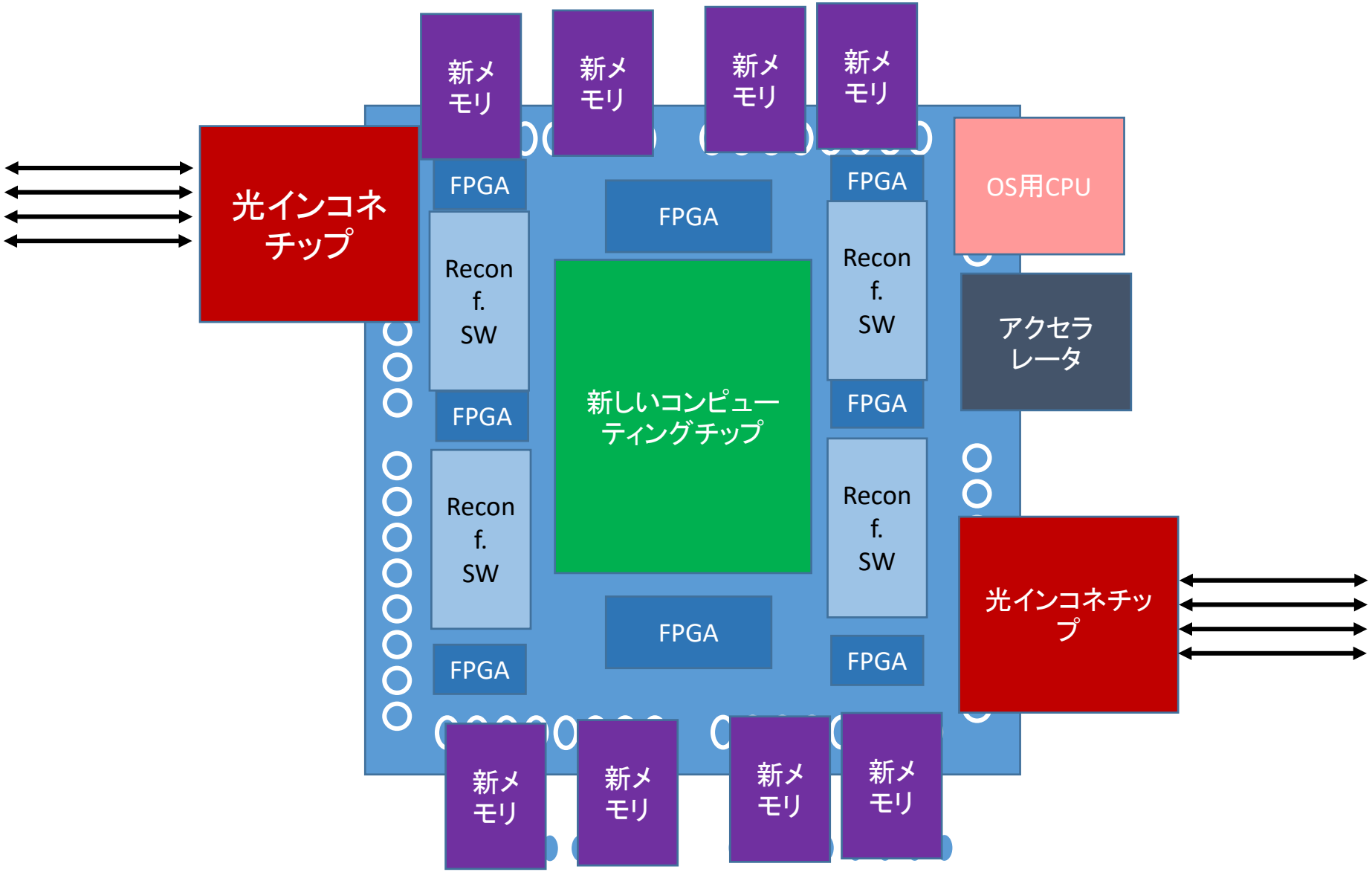
Super Building Block Archiecture (Amano, Keio U)



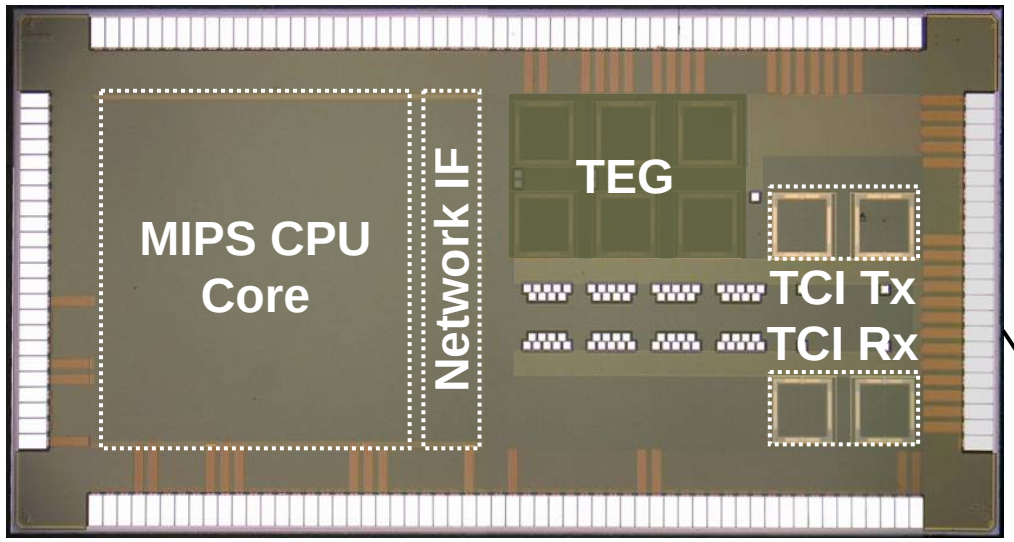
System View on "Post-Moore" Architecture

Not just a new device, but focus on how they are interconnected, and integrated as a system controlling their power

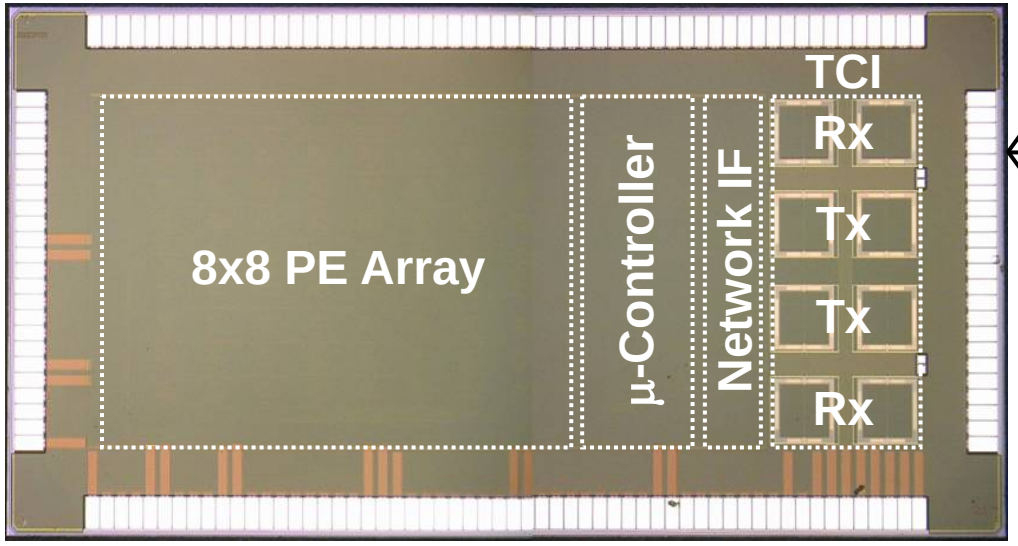
A Hub architecture that employs Inductive (3D) TCI and programmable FPGA+Switch



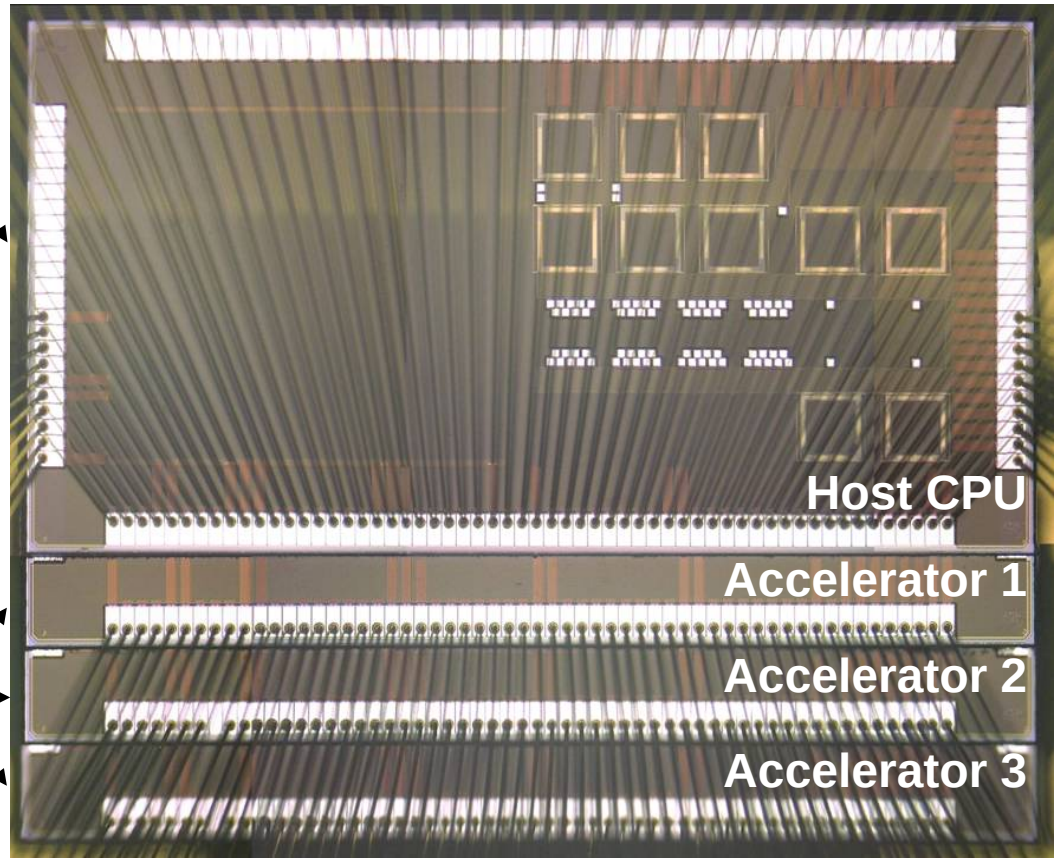
ドーターチップ接続のイメージ



Host CPU Chip



Accelerator Chip



Host CPU + Accelerator x3 Chip Stack
Fabricated in 65nm CMOS

Microphotograph of stacked test chips.

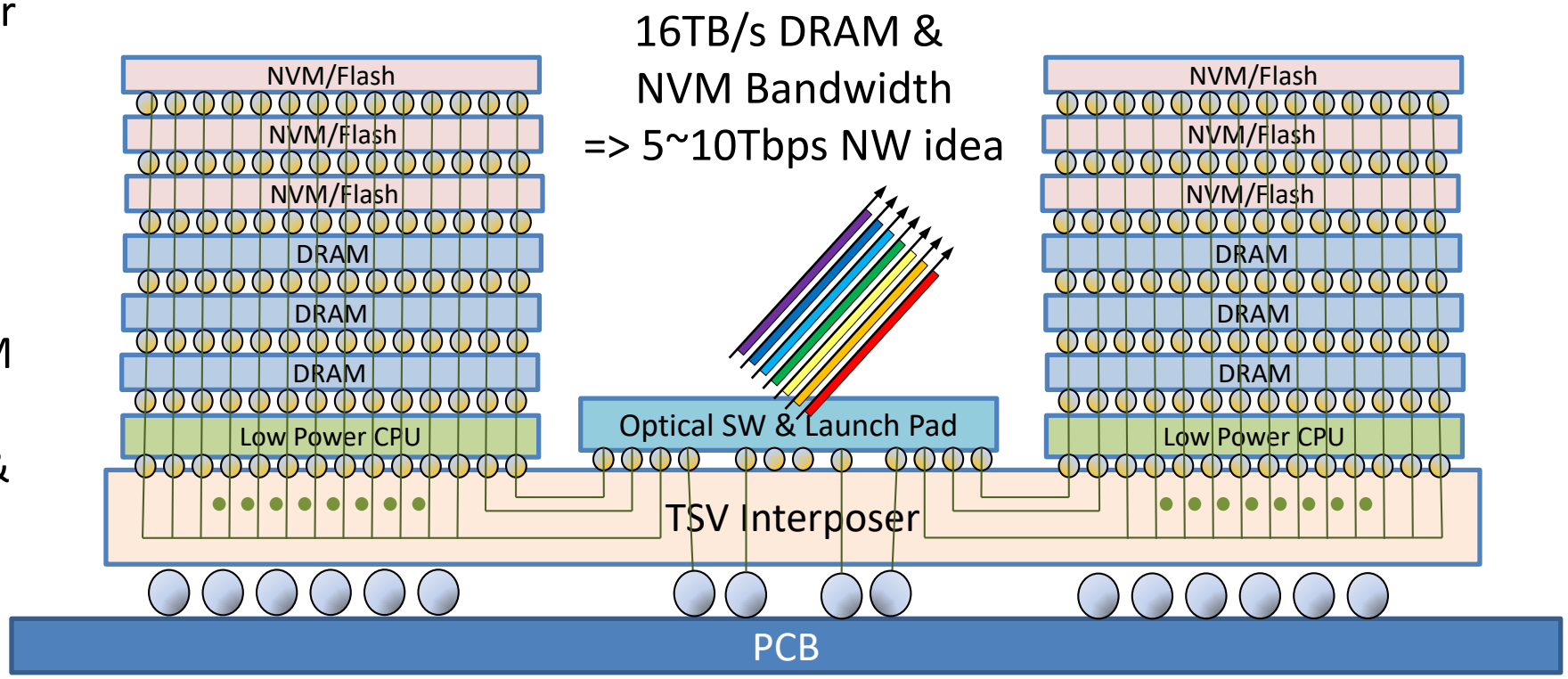
Strawman BYTES-Oriented Post-Moore Architecture

Low voltage & power CPU for direct stacking and large silicon area

Domain-specific hetero- and customizable processor configurations, including PIM

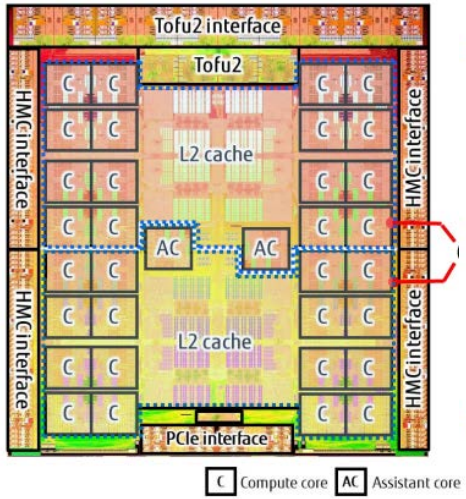
Extreme multi-layer DRAM & NVRAM stacking via high density tungsten TSV

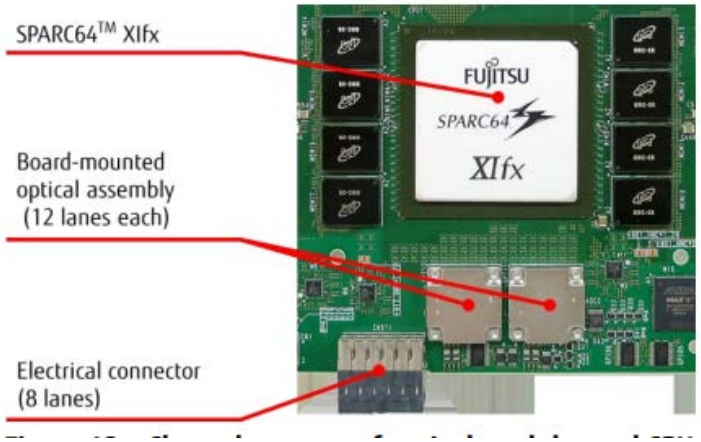
Direct WDM optics onto Interposer

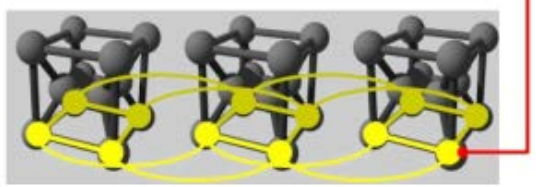


Direct Chip-Chip Interconnect with DWDM optics
Low Power Processor allows Direct 3D Stacking
Configurable Low-power CPU

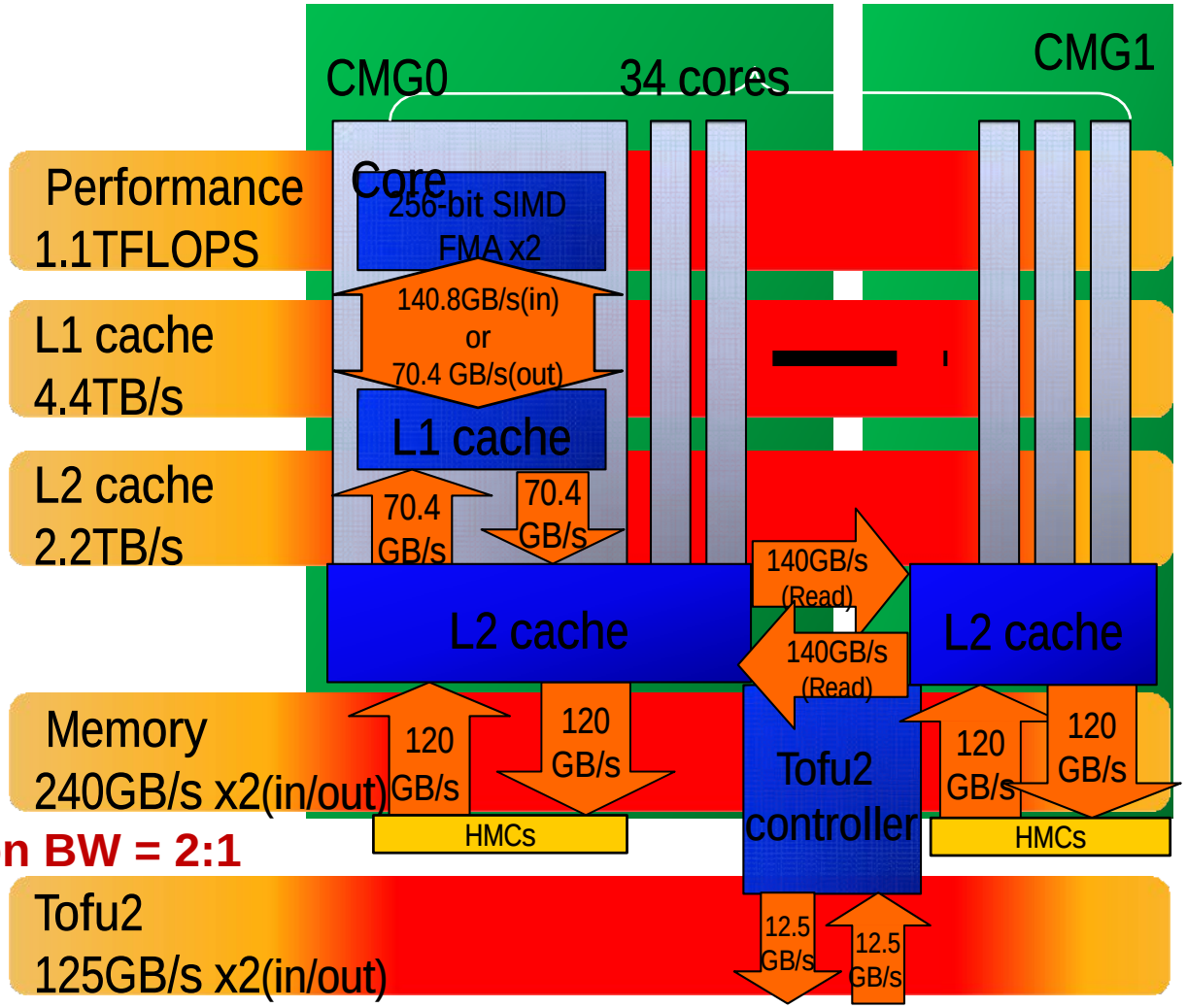
Fujitsu SPARC64™ Xifx (2015)





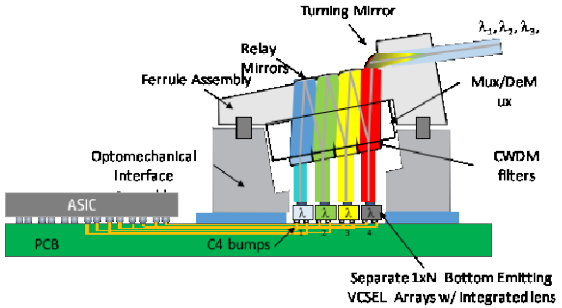


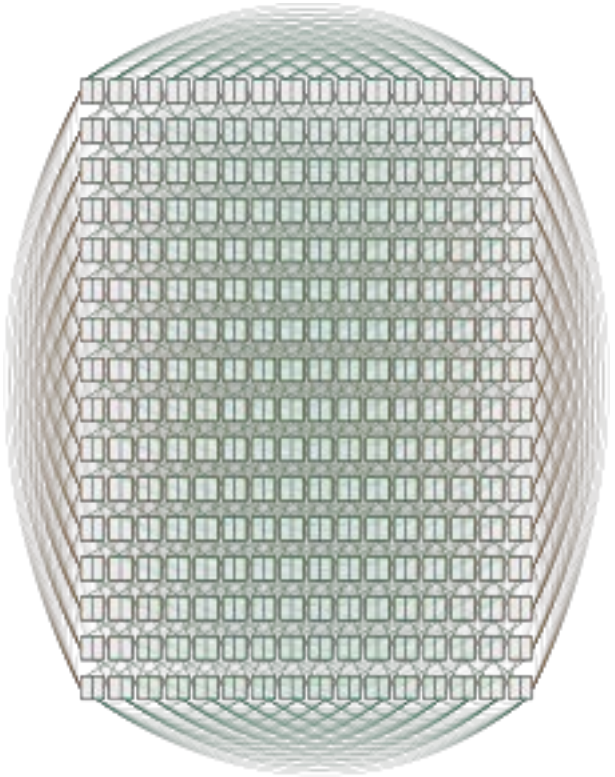
Memory BW : Injection BW = 2:1
(TSUBAME3: 40:1)

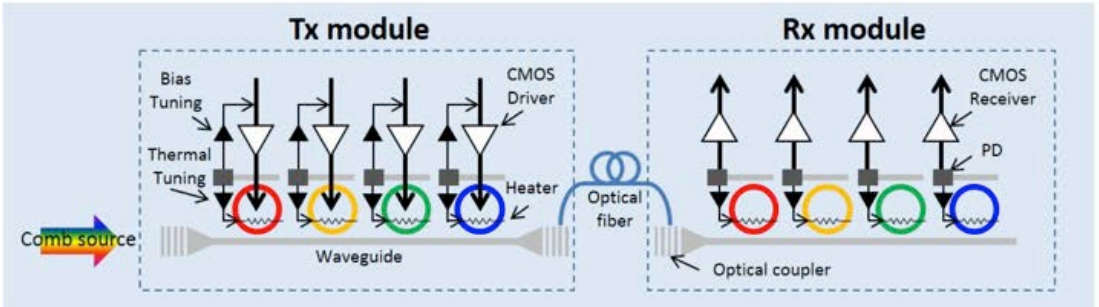


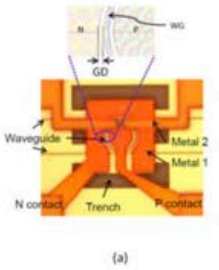
Interconnect Shortcomings

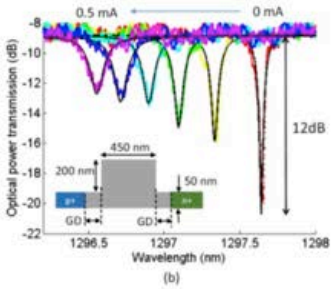
– Current technology:
  – 10$ / Gbps and 50 pJ per bit, per link
  – 1 exaflops -> 10 PB/s injection bw
  – O (1B$) and O(5MW) (node link only)
– First stepping stone: mid-board optics - vcsels
  – Advanced Development program at HPE
  – Cheaper, more efficient, can be water-cooled
– Exascale technology target: silicon photonics - ring resonators
  – 10 cents per Gbps, 5 pJ per bit
– Enabling enhanced topologies like the Hyper-X will require new “widgetry”











Optical Network Technology for Future Supercomputers & IDC

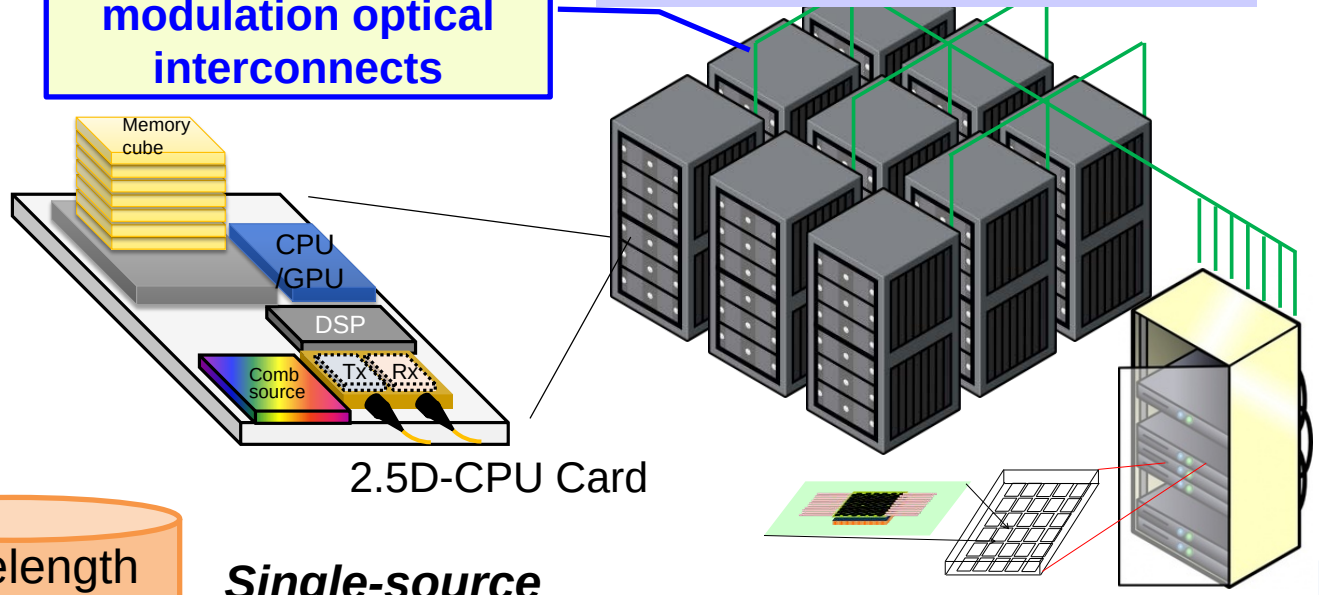
AIST IMPULSE

- Large-scale silicon photonics based cluster switches
- DWDM, multi-level modulation, highly integrated "elastic" optical interconnects
- Ultra-low energy consumption network by making use of optical switches

SC14 AIST Booth #2531

DWDM, multi-level modulation optical interconnects

Datacenter server racks



2.5D-CPU Card

Silicon photonics cluster switches

- Ultra-compact switches based on silicon photonics
- 3D integration by amorphous silicon
- A new server architecture

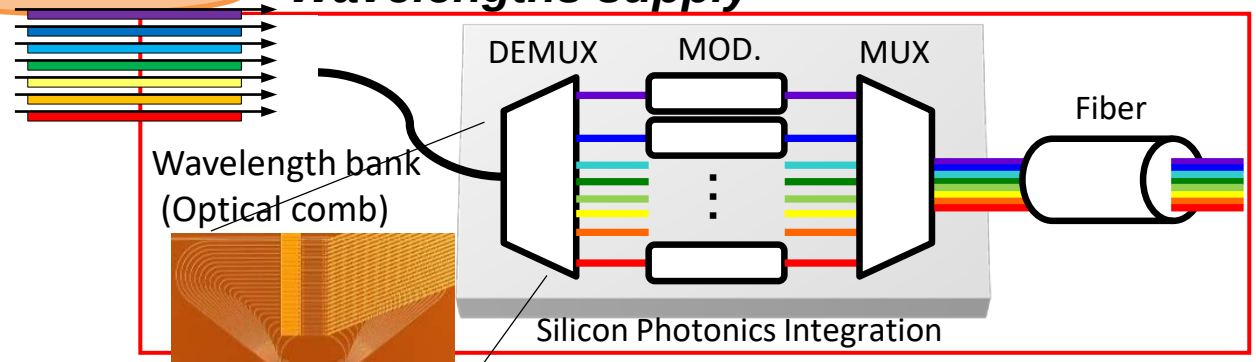
Current electrical switches:
~1pbps
↓
~500Pbps

No of λs	Order of mod.	Bit rate
1	1	20 Gbps
4	8	640 Gbps
32	8	5.12 Tbps

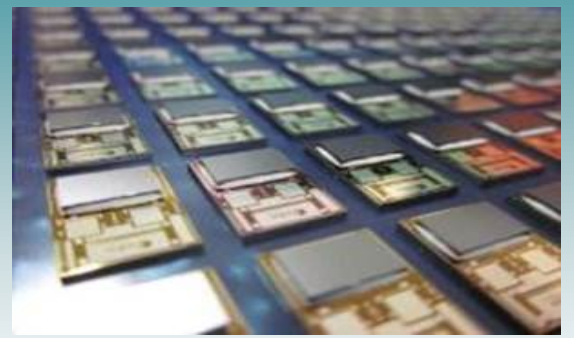
Current state-of-the-art Tx
100Gbps
↓
~ 5.12Tbps

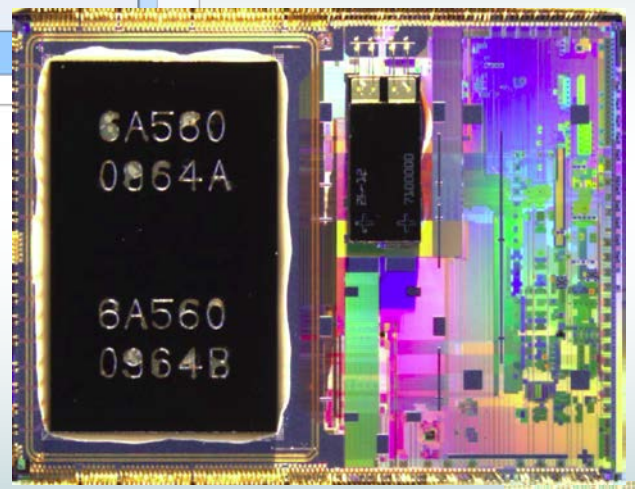
Wavelength Bank

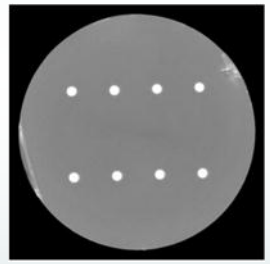
Single-source Wavelengths supply

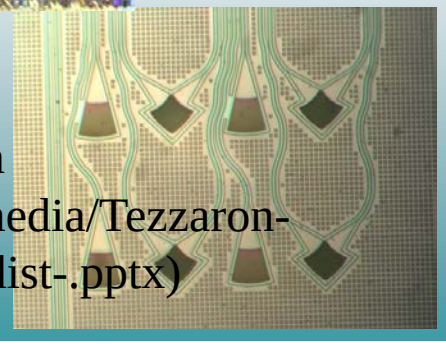


Luxtera 2.5D Photonic Data Pump









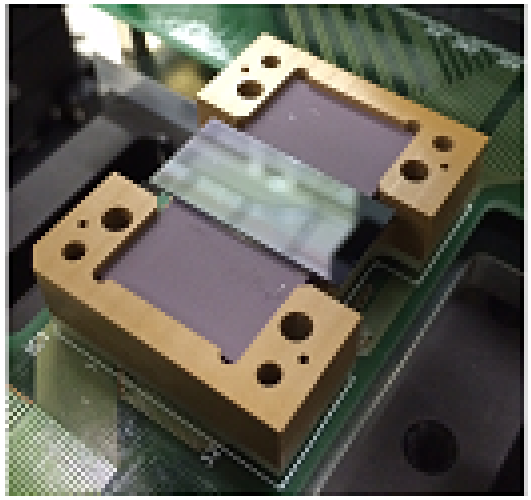
- 2.5pJ/bit power
- Bare metal protocol
  - Ultra low latency
  - Protocol agnostic
- 8 core Fiber
- 25Gb SERDES or 3.125Gb interface
- Self-calibrating self-tuning
- >1.6Tb/s payload

(Tezzaron slides taken from http://www.tezzaron.com/media/Tezzaron-Presentation-EPS-100814-dist-.pptx)

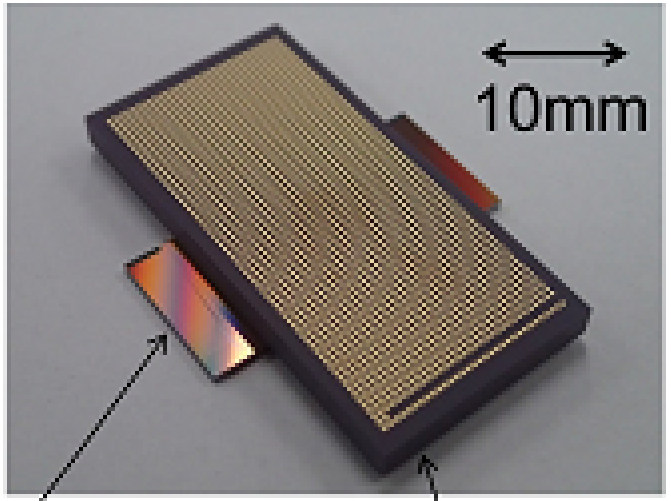
32 x 32 Optical Circuit Switch (Courtesy NTTAIST)

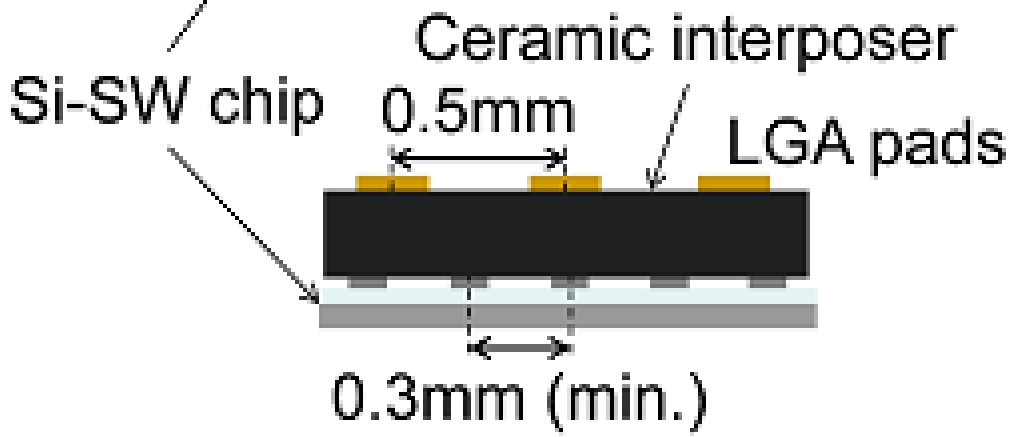
- Ceramic LGA interposer with 0.5-mm pitch
- Flip-chip bonding with Au bumps and non-conductive paste
- LGA socket to contact PCB

LGA socket



After FC bonding





Problem: heavy optical loss

Fast Optical Crossbar Swtch (EECS, UCB)

Seok et. al. "Large-scale broadband digital silicon photonic switches with vertical adiabatic couplers" Optica, 3-1, 2016

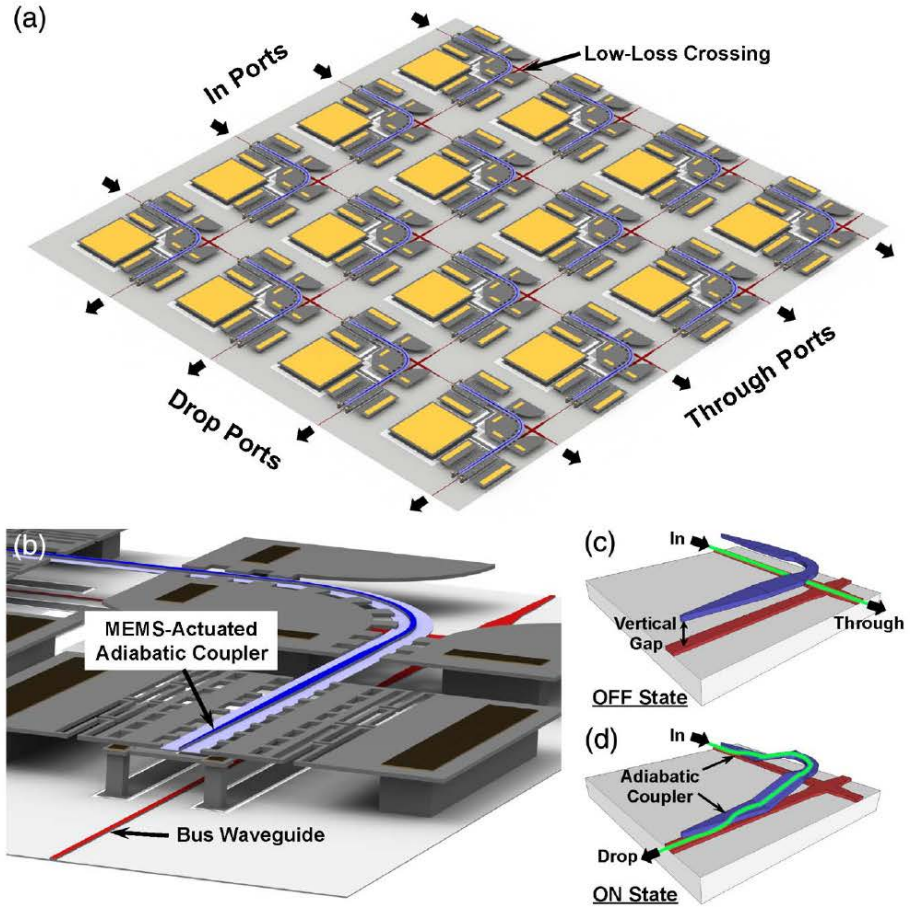


Fig. 1. Schematics of silicon photonic MEMS switches. (a) Matrix architecture of silicon photonic MEMS switch, (b) close-up view of a MEMS-actuated adiabatic coupler, (c) switch unit cell in the OFF state, and (d) switch unit cell in the ON state. The adiabatic coupler is precisely positioned at the optimum distance to the bus waveguide.

- Array of 64x64 MEMS optical crossbar switch
- 3.7db on-chip insertion loss
- 0.91microsecond switching time

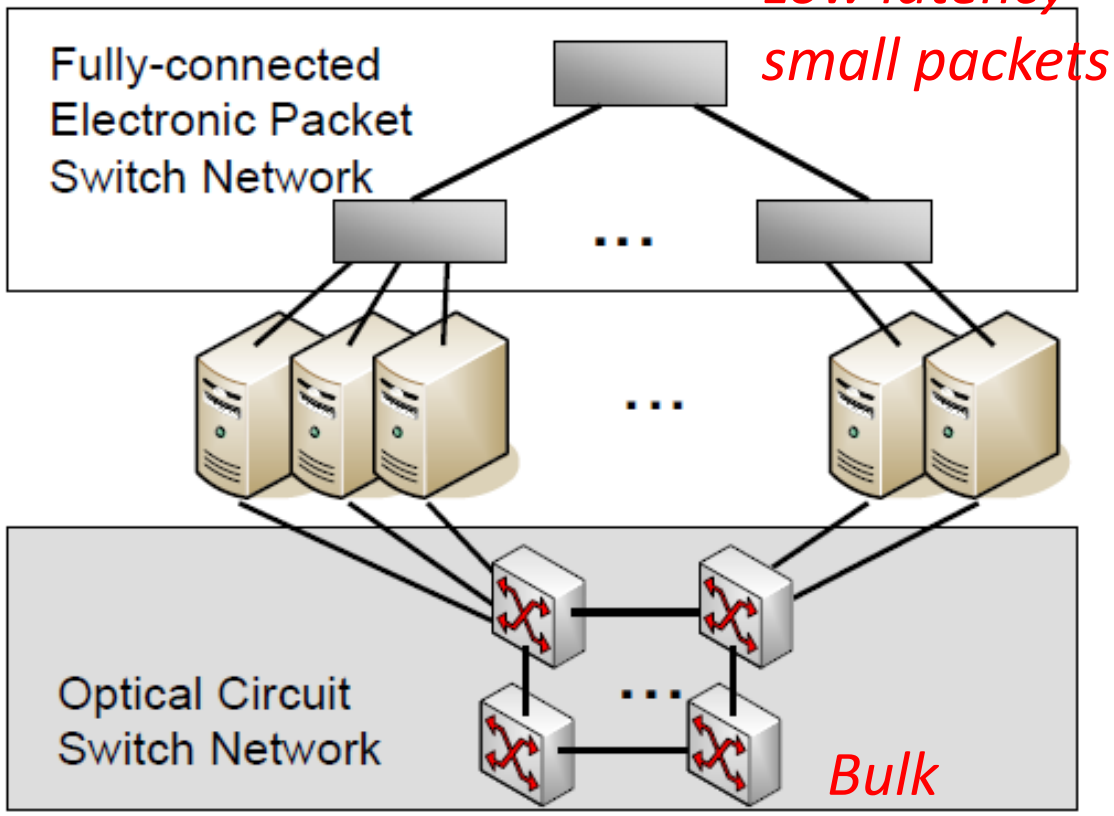
- At 100,000 ports – 9 hop network
  - 33db+ loss
  - 8.2 microsecond switching time => 1Tb 800Kbyte BW x Delay

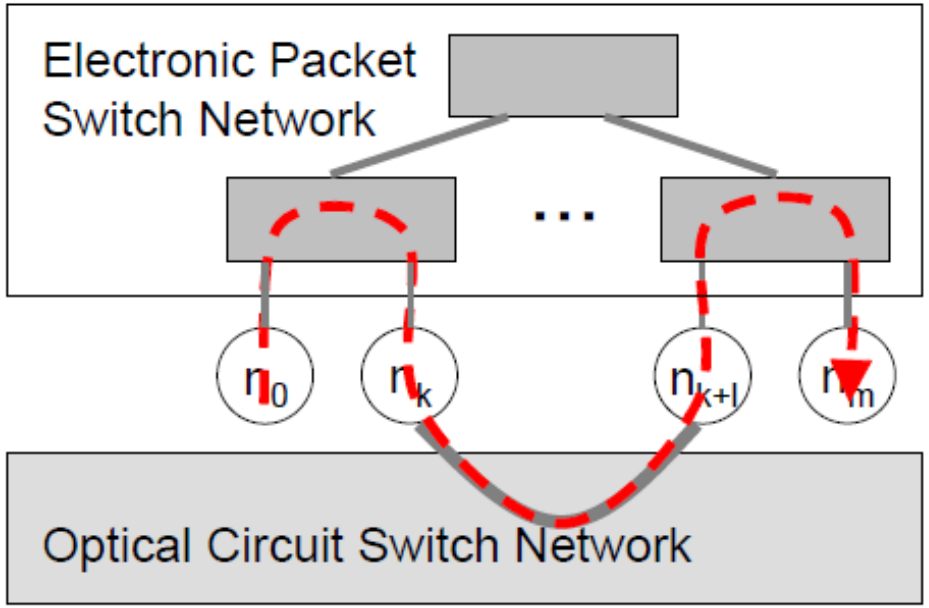
Solution: Hybrid EO Network

- Idea1: use low (latency/diameter, bandwidth, power) electrical network for low latency messages, and use optical circuits for high bandwidth and fixed topology messages
- Idea2: merge the electronic switch and optical MEMS switch, and use the latter as the control plane of the optical MEMs circuit
  - Thus the electronic switches become the optical speculative "buffer"

Hybrid Electro-Optical Network w/shortcuts [Takizawa&Matsuoka LSPP07]

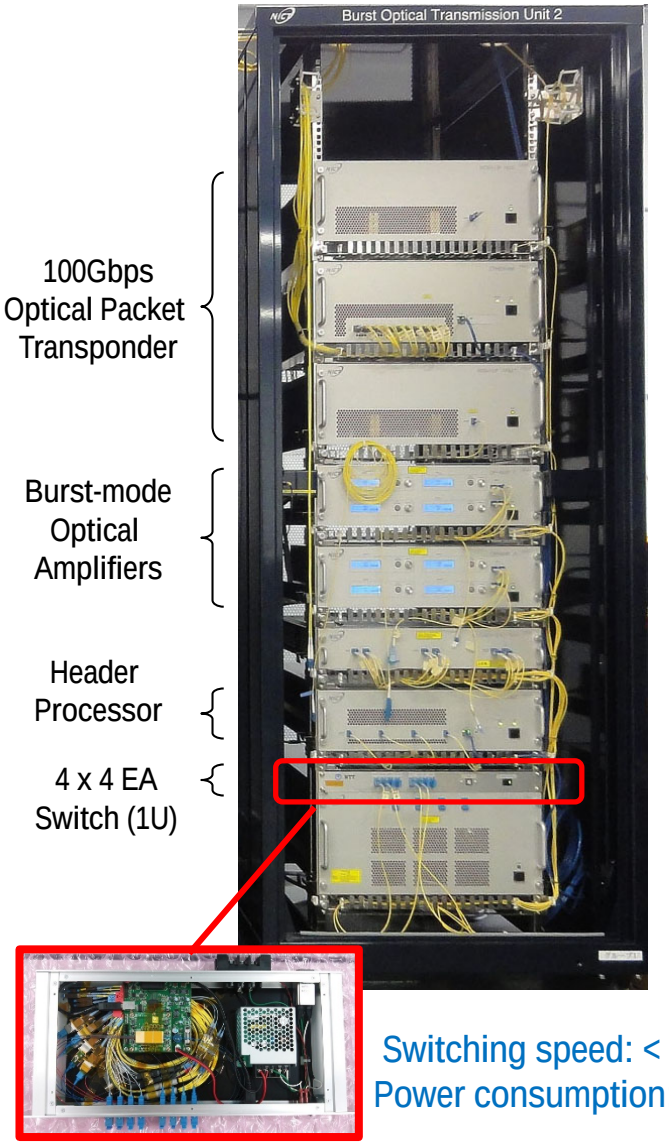
"Locality Aware MPI Communication on a Commodity Opto-Electronic Hybrid Network"





EO-OE Shortcut Forwarding of Messages Across Switches

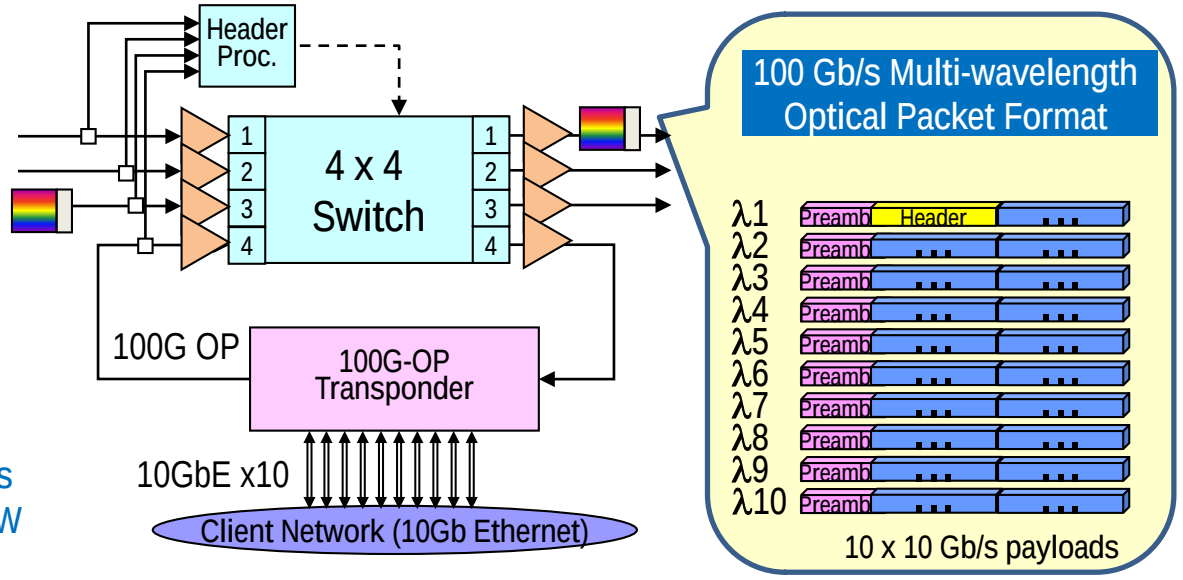
NICT Optical Packet Switch Node (Slides courtesy NICT)



Switching speed: < 8 ns
Power consumption: 3 W

Y. Muranaka, et.al, Photonics in Switching 2015.

- 4 x 4 OPS node with optical packet (OP) transponder
- 100Gb/s OPS port, 10GbE x 10 Client ports
- Stability: Tolerance for environmental disturbance (Polarization, Power fluctuation)
- Total throughput : 800 Gb/s
- Total power consumption: 141 W (w/o Transponder)
- 10-node hopping, 450 km fiber transmission



H. Furukawa, et.al, no.P.4.16, ECOC2015.

HIDEAKI FURUKAWA furukawa@nict.go.jp

Applications & Algorithms

Slides by Kengo Nakajima

Information Technology Center
The University of Tokyo

New Frontiers of Computer & Computational Science towards Post Moore Era

December 22, 2015, Takeda Auditorium, The University of Tokyo

Assumptions & Expectations towards Post-Moore Era

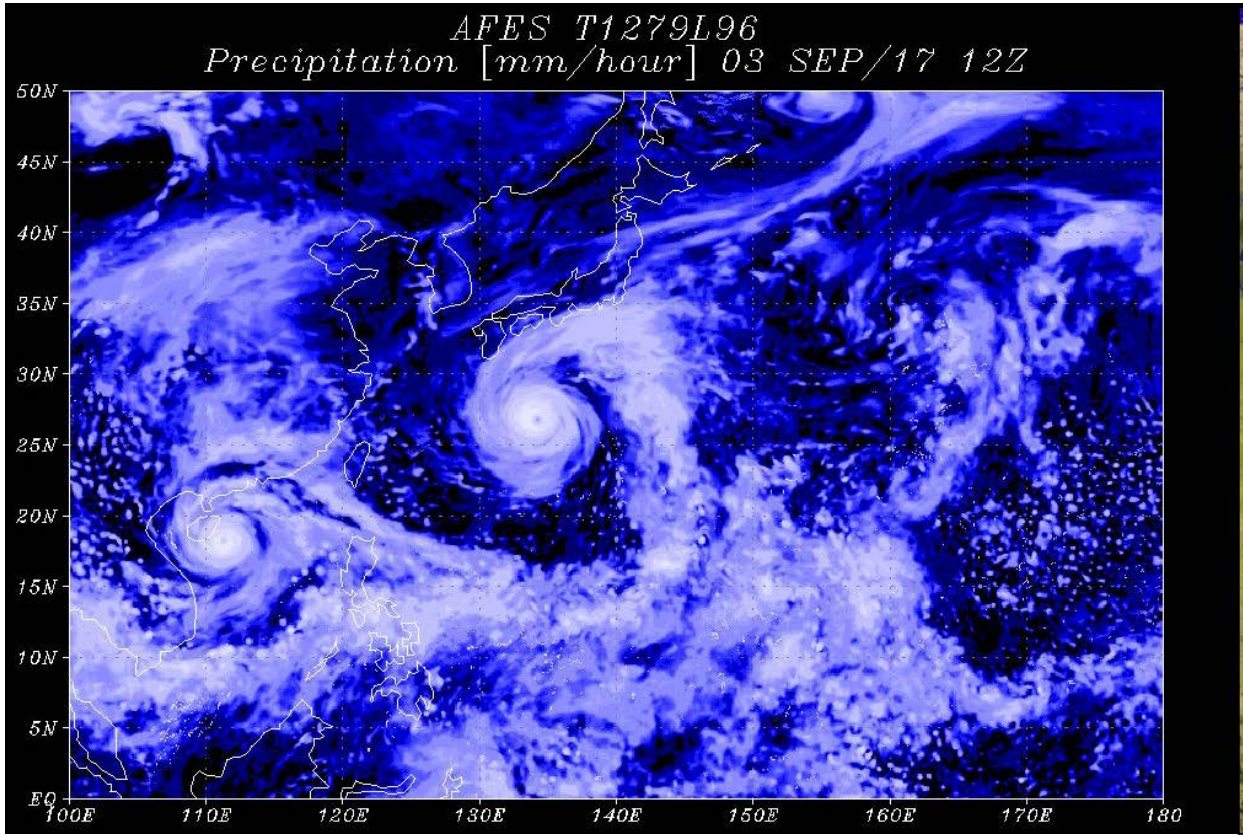
- **Higher Bandwidth**, Larger & Heterogeneous Latency
  - Memory: 3D Stacked Memory
  - Network: Optical Communication
  - Both of Memory & Network will be more hierarchical
- Larger Size of Memory & Cache
- Transaction/Transactional Memory

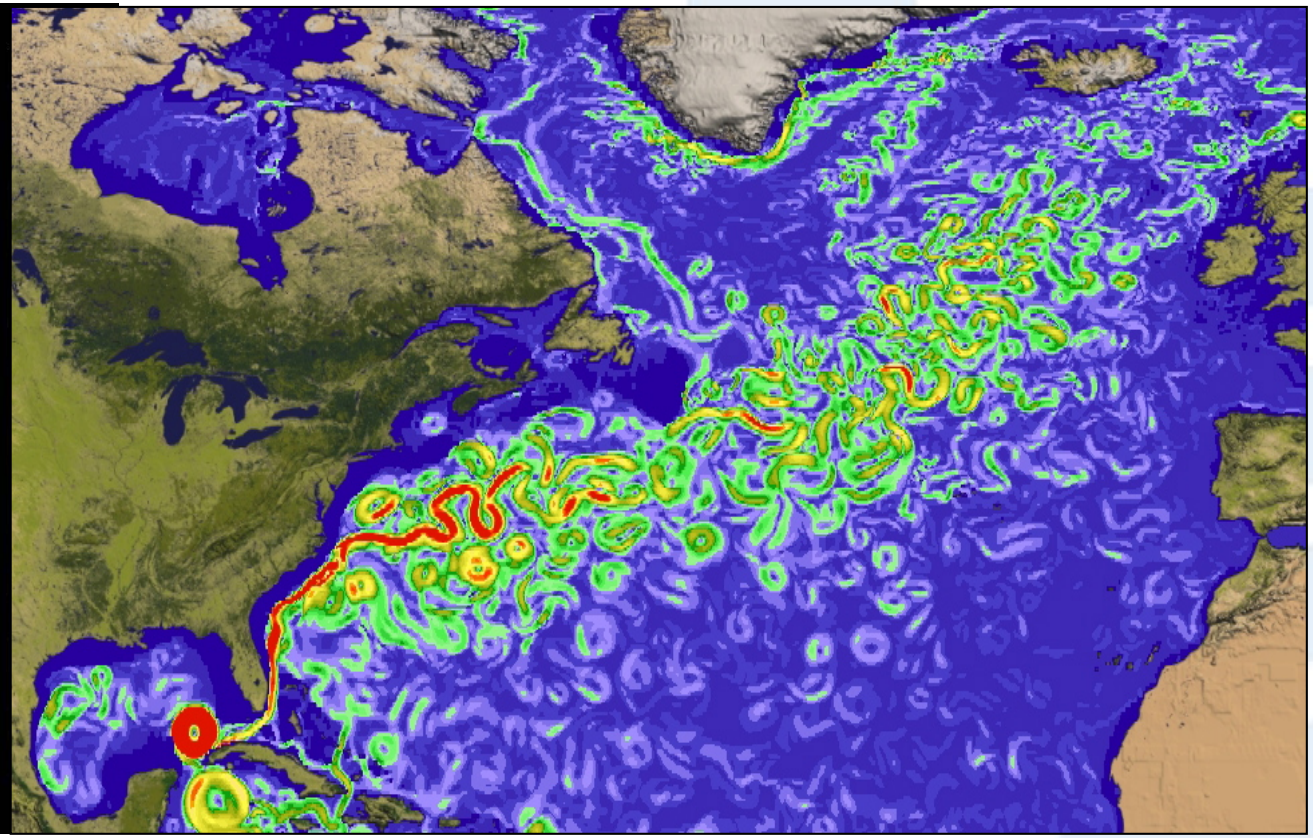
- Application-Customized Hardware, FPGA
- Large Number of Nodes/Number of Cores per Node
  - under certain constraints (e.g. power, space …)

Applications & Algorithms in Post-Moore Era (1/2)

- Compute Intensity ⇒ Data Movement Intensity
  - Non-Blocking Method, Out-of-Core Algorithm
- **Implicit scheme strikes back !**
  - I believe it was never defeated
  - Improvement of performance on sparse matrix computations
  - Big change and advancement are expected in all research areas related to algorithms for sparse matrices including preconditioning
  - **Everything might be easier… but don't relax too much!**
  - **Other Compute to Data Algorithms: H-Matrices**

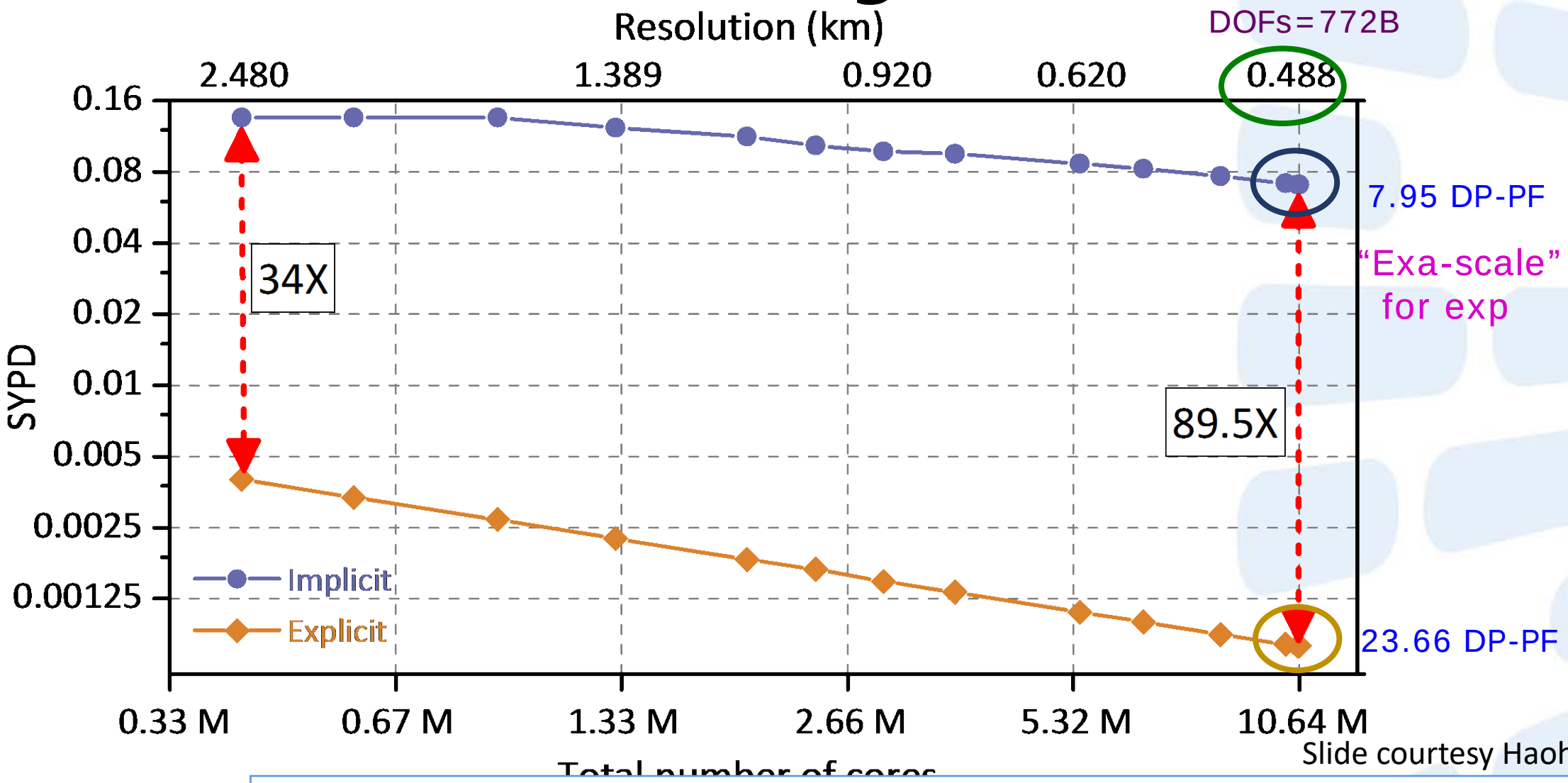
Highly-Scalable Atmospheric Simulation Framework (ACM Gordon Bell Prize 2016)





Slide courtesy Haohuan Fu

Weak-scaling results



Slide courtesy Haohuan Fu

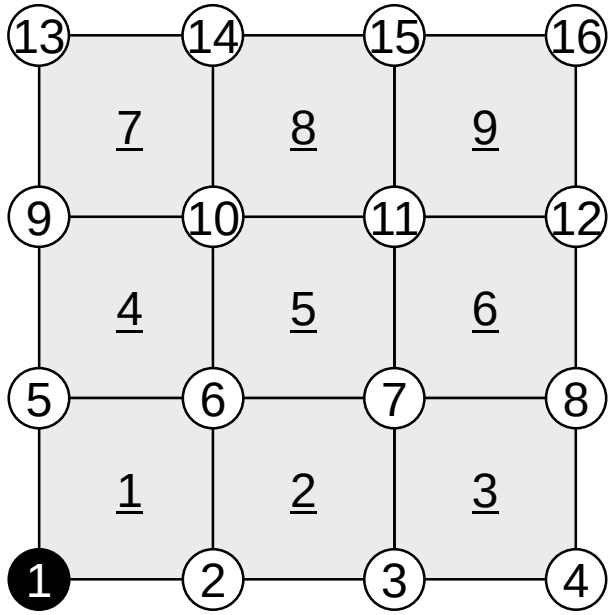
The 488-m res run: 0.07 SYPD, 10.6M cores, dt=240s, 89.5X speedup over explicit

GeoFEM Benchmark: ICCG for FEM Performance of a Node: Flat MPI

	SR11K/J2 Power5+	T2K AMD	FX10	K	Earth Sim 1
Core #/Node	16	16	16	8	8
Peak Performance (GFLOPS)	147.2	147.2	236.5	128.0	64.0
STREAM Triad (GB/s)	101.0	20.0	64.7	43.3	256.0
B/F	0.686	0.136	0.274	0.338	4.00
GeoFEM (GFLOPS)	19.0	4.69	16.0	11.0	25.6
% to Peak	12.9	3.18	6.77	8.59	40.0
LLC/core (MB)	18.0	2.00	0.75	0.75	-

Sparse Solver：Memory-Bound

Improvement of performance on sparse matrix computations due to higher memory bandwidth



Sparse Matrices:
- **FEM**
- **Indirect Memory Access**
- **Memory-Bound**

$$
\begin{bmatrix}
D & X & & & X & X & & & & & & & & & & \\
X & D & X & & X & X & X & & & & & & & & & \\
 & X & D & X & & X & X & X & & & & & & & & \\
 & & X & D & & & X & X & & & & & & & & \\
X & X & & & D & X & & & X & X & & & & & & \\
X & X & X & & X & D & X & & X & X & X & & & & & \\
 & X & X & X & & X & D & X & & X & X & X & & & & \\
 & & & & & & & & & & & & & & X & D
\end{bmatrix}
\begin{Bmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \\ \Phi_4 \\ \Phi_5 \\ \Phi_6 \\ \Phi_7 \\ \Phi_8 \\ \Phi_9 \\ \Phi_{10} \\ \Phi_{11} \\ \Phi_{12} \\ \Phi_{13} \\ \Phi_{14} \\ \Phi_{15} \\ \Phi_{16} \end{Bmatrix}
=
\begin{Bmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5 \\ F_6 \\ F_7 \\ F_8 \\ F_9 \\ F_{10} \\ F_{11} \\ F_{12} \\ F_{13} \\ F_{14} \\ F_{15} \\ F_{16} \end{Bmatrix}
$$

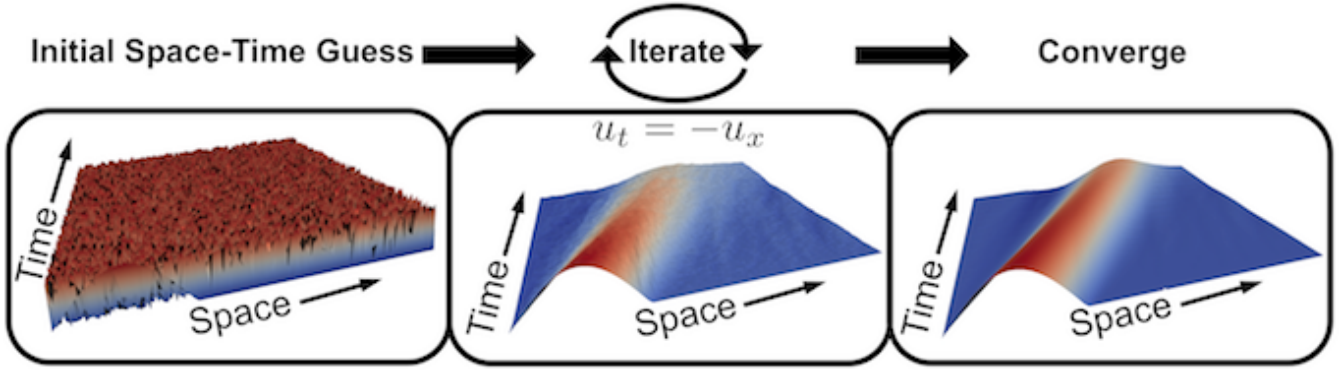
```
{Y}= [A]{X}

do i= 1, N
  Y(i)= D(i)*X(i)
  do k= INDEX(i-1)+1, INDEX(i)
    Y(i)= Y(i) + AMAT(k)*X(ITEM(k))
  enddo
enddo
```

Assumptions & Expectations towards Post-K/Post-Moore Era

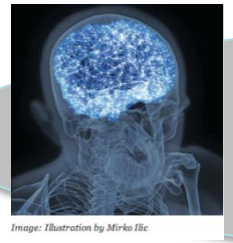
- Post-K (-2020)
  - Memory Wall
  - Hierarchical Memory (e.g. KNL: MCDRAM-DDR)
- **Post-Moore (-2025? -2029?)**
  - **Larger Size of Memory & Cache**
  - **Higher Bandwidth, Larger & Heterogeneous Latency**
    - **3D Stacked Memory, Optical Network**
    - **Both of Memory & Network will be more hierarchical**
  - **Application-Customized Hardware, FPGA**
- **Common Issues**
  - **Hierarchy, Latency (Memory, Network etc.)**
  - **Large Number of Nodes/Number of Cores per Node**
    - **under certain constraints (e.g. power, space …)**

Parallel-in-Space/Time (PiST)

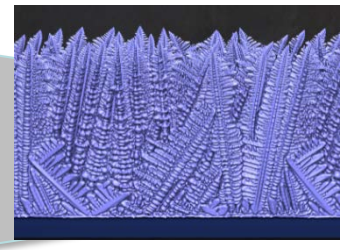
- MG is scalable, but improvement of performance is limited by parallelization only in space direction
- Time-Dependent Problems: Concurrency in Time Dir.
- Multigrid in (Space+Time) Direction
  - ✓ Traditional time-dependent method: Point-Wise Gauss Seidel
  - ✓ XBraid : Lawrence Livermore National Laboratory
    - ☐ Application to nonlinear problems (Transient Navier-Stokes Eqn's)
- MS with 3 sessions in SIAM PP16 (April 2016)
- PiST approach is suitable for the Post-Moore Systems with a complex and deeply hierarchical network that causes large latency.



Co-Designing Post-Moore HPC System Architecture



FLOPS-Oriented => BYTES-Oriented Numerical Applications and Algorithms



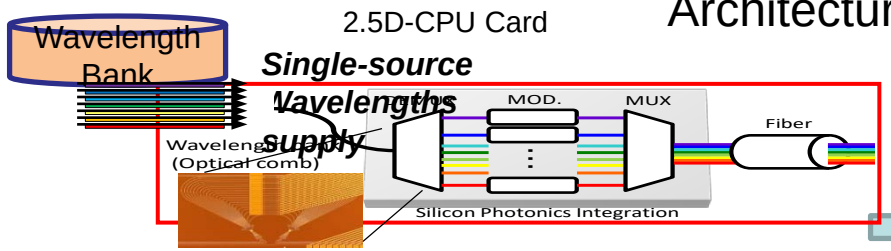
Programming Models and Abstraction?

System SW and Comm MW for Exabit Optical Inteconnect

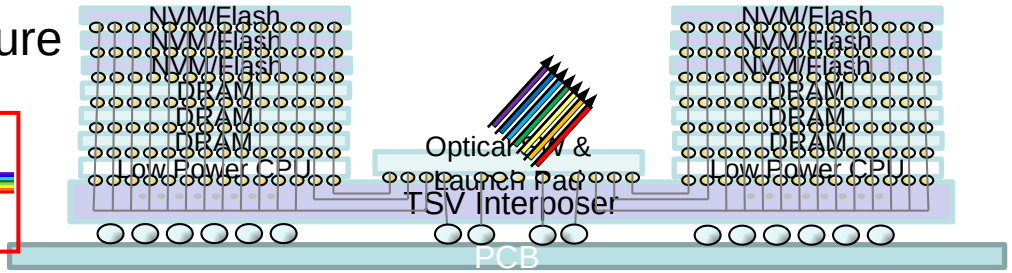
Programming & Perf Modeling of Super Building Block Architecture

System SW and & Programming of Deep and High Bandwidth Memory Hiearchy

Next Gen Exabit-class Optically Switched Interconnect



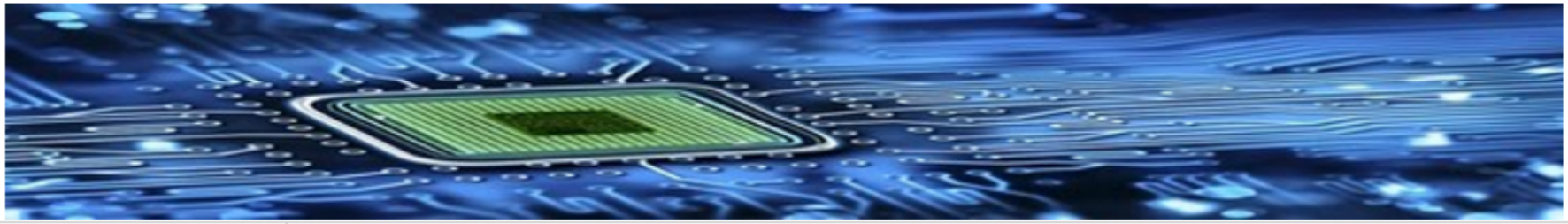
Specialized/Integrated/Reconfigurable Super Building Block Architecture



Advanced 3-D stacked Non-Volatile Memory >Tbytes, >10TByte/s

Post Moore Era Supercomputing Workshop @ SC17

- https://sites.google.com/site/2017pmes/
- 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