Opening Address
From K to Post-K and Beyond

Satoshi Matsuoka
Director, Riken R-CCS
Riken R-CCS Symposium, Kobe
20190218

Post-K: The Game Changer

1. **Heritage of the K-Computer, HP in simulation via extensive Co-Design**

- High performance: up to x100 performance of K in real applications
- Multitudes of Scientific Breakthroughs via Post-K application programs
- Simultaneous high performance and ease-of-programming

2. New Technology Innovations of Post-K

- **High Performance, esp. via high memory BW**
  Performance boost by "factors" c.f. mainstream CPUs in many HPC & Society5.0 apps
- **Very Green e.g. extreme power efficiency**
  Ultra Power efficient design & various power control knobs
- **Arm Global Ecosystem & SVE contribution**
  Top CPU in ARM Ecosystem of 21 billion chips/year, SVE co-design and world's first implementation by Fujitsu
- **High Perf. on Society5.0 apps incl. AI**
  Architectural features for high perf on Society 5.0 apps based on Big Data, AI/ML, CAE/EDA, Blockchain security, etc.

Global leadership not just in the machine & apps, but as cutting edge IT

ARM: Massive ecosystem from embedded to HPC

Technology not just limited to Post-K, but into societal IT infrastructures e.g. Clouds

“Post-K” Chronology (part1)

- Jun 2006 Tokyo Tech. **TSUBAME1.0** becomes Top500#1 in Japan, first time based ***on general-purpose multi-core CPU***(*AMD Opteron*)
- Sep 2006 K Computer project officially launched (Kobe 2007)
- May 2009 Fujitsu unveils **Sparc64 fxVIII**, same day NEC drops out of K project, K Computer to become purely based ***on general-purpose multi-core CPU***
- Nov 2009 Govt. review almost cancels K => 2010 formulation of ***HPCI*** (Japan’s High Performance Computing Infrastructure consortium) based on **NAREGI** National Grid Project (2003-2007)
- Nov 2010 Tokyo Tech. **TSUBAME2.0**, becomes Top500#1 in Japan, ***first petascale and many-core SC in Japan*** (*NVIDIA Fermi GPU*)
- Apr 2011 **Riken AICS** (R-CCS predecessor) starts
- Jun 2011 K Computer becomes ***Top500#1 in the World***
- Nov 2011 ***ACM Gordon Bell Prizes*** for K Computer & Tsubame2.0

“Post-K” Chronology (part2)

- ~Apr 2011 SDHPC started, whitepaper on “Post-K” 2018~ SC
- Apr. 2012~2013 “Post-K Feasibility Study” officially starts
  - 3 architecture investigation teams, 1 application requirements team
- Apr 2014 “Post-K” project officially starts at Riken AICS, objective: **up to x100 speedup on benchmark apps** (NOT Exaflops on Linpack)
- Jun 2014 K Computer becomes Graph500#1 – Big Data convergence
- Jun 2016 K Computer becomes HPCG#1 – From FLOPS to BYTES
- Nov 2016 U-Tokyo – Tsukuba-U Oakforest PACS (Intel KNL) becomes Top500#1 in Japan – first general-purpose
- Aug 2017 HotChips announcement that “**Post-K” will adopt Arm ISA**
- Apr 2018 Riken AICS => Riken R-CCS (Riken Center for Computational Science), **Satoshi (me) becomes Director**
- Aug 2018 Fujitsu unveils Arm64fx @ Hotchips2018

“Post-K” Chronology (part3)

- *Oct 2018 Several basic research projects towards future architectures and AI&HPC convergence start at R-CCS*
- Nov 2018 **AIST ABCI** becomes Top500#1 in Japan – HPC & AI convergence becomes real based on **2017 TSUBAME3.0**
- Nov 2018 “Post-K” manufacturing and production officially approved by CSTI, Prime Minister’s Science and Technology Committee
- Nov 2018 MEXT committee on investigating the future usage of “Post-K” towards Society 5.0 starts – extreme broadening of SC usage
- Feb 2019 “Post-K” public naming commences, to be announced May~June 2019 – submit your ideas now(!)
- Mar 2019 “Post-K” Hardware specs to complete by Fujitsu and handed over to Riken R-CCS, machine sized at >150,000 nodes

“Post-K” Chronology (part4)

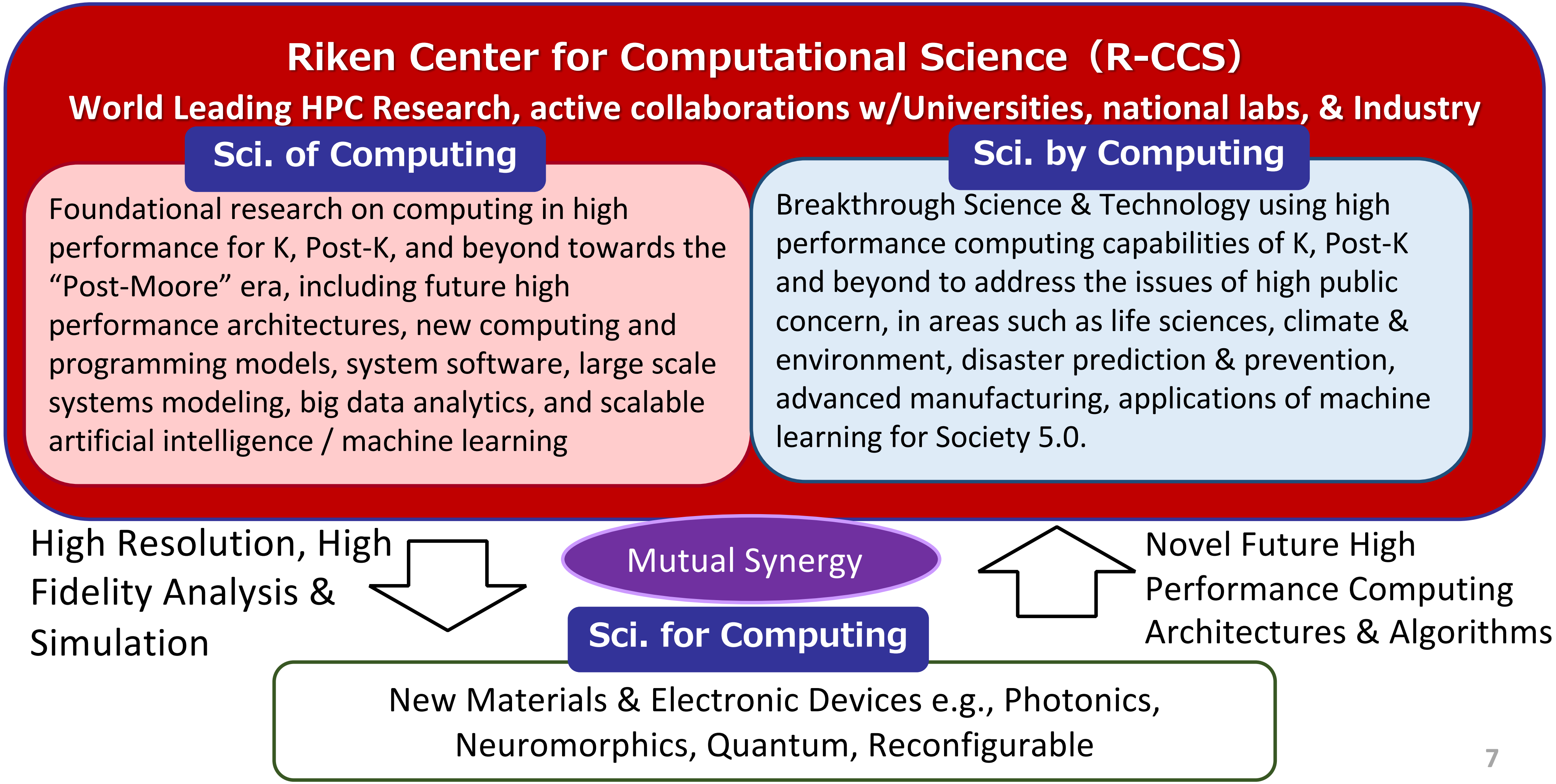
(Disclaimer: below includes speculative schedules and subject to change)

- 1H2019 “Post-K” manufacturing budget approval by the Diet, actual manufacturing commences
- Apr 2019 R-CCS lead research activities on next-gen architectures will commence => whitepaper to be written by Winter
- Aug 2019 End of K-Computer operations
- 4Q2019~1Q2020 “Post-K” installation starts
- 1H2020 “Post-K” preproduction operation starts
- 2020~2021 “Post-K” production operation starts (hopefully)
- And of course we move on…

Watch for announcements on “Post-K” technology commercialization by Fujitsu and its partner vendors RSN

Apr 1 2018 Became Director of Riken-CCS:

Science, of Computing, by Computing, and for Computing



Riken Center for Computational Science（R-CCS)

World Leading HPC Research, active collaborations w/Universities, national labs, & Industry

Sci. of Computing

Foundational research on computing in high performance for K, Post-K, and beyond towards the “Post-Moore” era, including future high performance architectures, new computing and programming models, system software, large scale systems modeling, big data analytics, and scalable artificial intelligence / machine learning

Sci. by Computing

Breakthrough Science & Technology using high performance computing capabilities of K, Post-K and beyond to address the issues of high public concern, in areas such as life sciences, climate & environment, disaster prediction & prevention, advanced manufacturing, applications of machine learning for Society 5.0.

High Resolution, High Fidelity Analysis & Simulation

Mutual Synergy

Novel Future High Performance Computing Architectures & Algorithms

Sci. for Computing

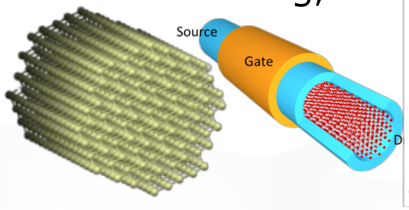
New Materials & Electronic Devices e.g., Photonics, Neuromorphics, Quantum, Reconfigurable

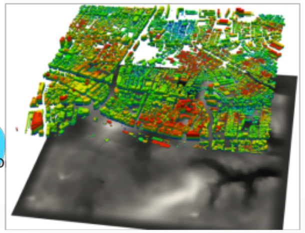
Co-Design Activities in Post-K

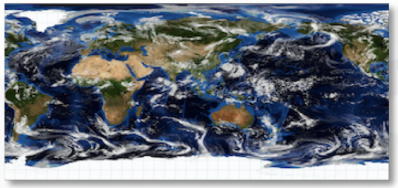
Multiple Activities since 2011

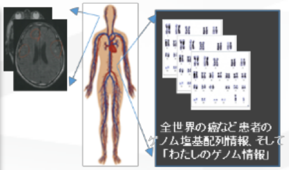
Science by Computing

- 9 Priority App Areas: High Concern to General Public: Medical/Pharma, Environment/Disaster, Energy, Manufacturing, …





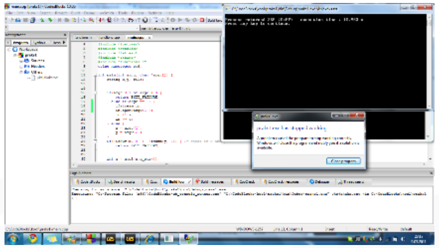




全世界の癌など患者のゲノム塩基配列情報、そして「わたしのゲノム情報」

Select representatives from 100s of applications signifying various computational characteristics

Science of Computing

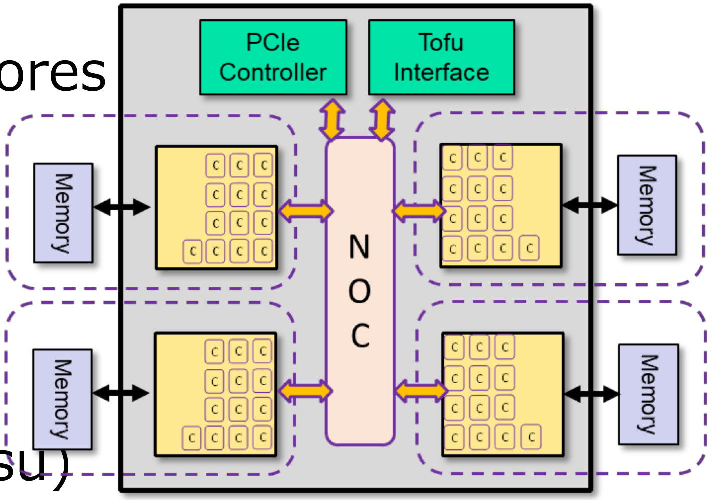


Design systems with parameters that consider various application characteristics

- **Extremely tight collabrations between the Co-Design apps centers, Riken, and Fujitsu, etc.**
- **Chose 9 representative apps as “target application” scenario**
- **Achieved up to x100 speedup c.f. K-Computer**
- **Also ease-of-programming, broad SW ecosystem, very low power, …**

“Post-K” Arm64fx Processor is…

- **an Many-Core ARM CPU…**
  - 48 compute cores + 2 or 4 assistant (OS) cores
  - Brand new core design by Fujitsu
  - Near Xeon-Class Integer performance core
  - ARM V8.2 --- 64bit ARM ecosystem



- **…but also a GPU-like processor**
  - SVE 512 bit vector extensions (ARM & Fujitsu)
    - Integer (1, 2, 4, 8 bytes) + Float (16, 32, 64 bytes)
  - Cache + access localization (sector cache) – similar to scratchpad
  - HBM2 OPM – Massive Mem BW (1TByte/s, Bytes/DPF ~0.4 same as K)
    - Streaming memory access, strided access, scatter/gather etc.
  - Intra-chip barrier synch. and other memory enhancing features
  - 40GByte/s Tofu-.D interconnect + PCIe 3

- **GPU-like High performance in HPC, AI/Big Data, Auto Driving…**

Post-K A64fx A0 (ES) performance

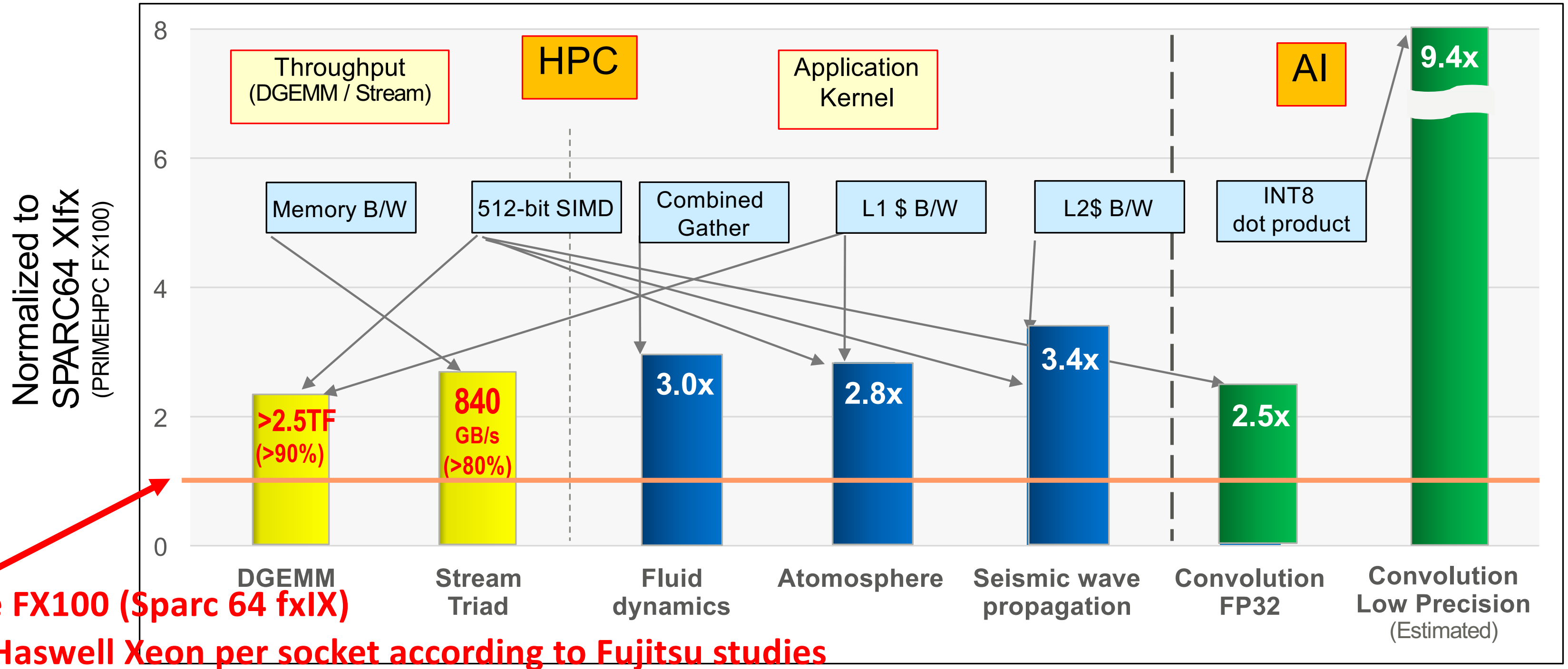
	Performance / CPU					Machine Performance (HPC)		
	Peak TF (DFP)	Peak Mem. BW	Stream Triad	Theoretical B/F	DGEMM Efficiency	Linpack Efficiency	GF/W	Network BW Per Chip
Post-K A64fx (A0 Eng. Sample)	2.764/ 3.072	1024GB/s	840GB/s	0.37/ 0.33	94 %	87.7 %	>15	TOFU-D 40.8GB/s (6.8x 6)
Intel KNL	3.0464	600GB/s	490GB/s	0.20	66%	54.4 %	4.9	12.5 GB/s
Intel Skylake	1.6128	127.8GB/s	97 GB/s	0.08	80 %	66.7 %	4.5	6.2GB/s
NVIDIA V100 (DGX-2)	7.8	900 GB/s	855GB/s	0.12		76 %	15.113	160GB/s 6.2GB/s

Performance

FUJITSU

- A64FX boosts performance up by microarchitectural enhancements, 512-bit wide SIMD, HBM2 and process technology
  - > 2.5x faster in HPC/AI benchmarks than SPARC64 XIfx (Fujitsu's previous HPC CPU)
    - The results are based on the Fujitsu compiler optimized for our microarchitecture and SVE

A64FX Benchmark Kernel Performance (Preliminary results)



Baseline FX100 (Sparc 64 fxIX)

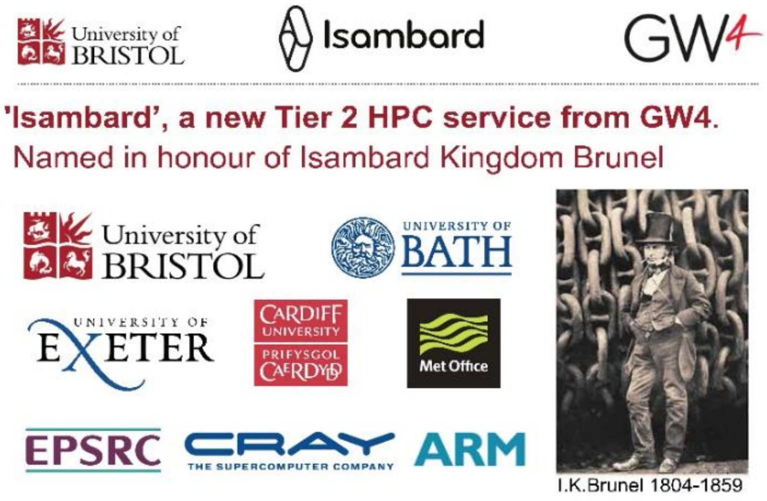
~x2 c.f. Haswell Xeon per socket according to Fujitsu studies

https://www.ssken.gr.jp/MAINSITE/event/2015/20151028-sci/lecture-04/SSKEN_sci2015_miyoshi_presentation.pdf

Baseline: SPARC64 XIfx (PRIMEHPC FX100)

ARM HPC ECOSYSTEM
(slides courtesy Prof. Simon Mcintosh-Smith, U-Bristol)

- **“Isambard” Cavium TX2 HPC Cluster**
- **Various Portings and Benchmarking**
  - Practically all x86 codes work “out-of-the-box”
  - Compiler dependency more crucial c.f. ISA
  - Performance competitive due to most applications being memory BW dependent, and Cavium BW 33% superior

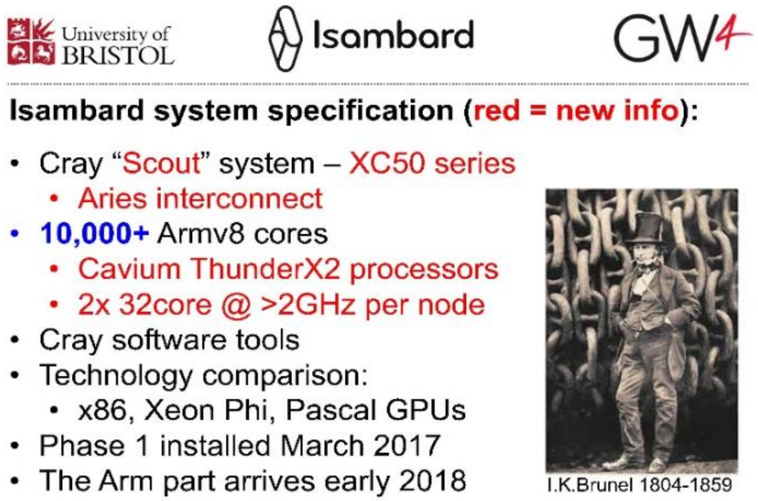


University of BRISTOL Isambard GW4

'Isambard', a new Tier 2 HPC service from GW4.
Named in honour of Isambard Kingdom Brunel

University of BRISTOL
University of BATH
University of EXETER
CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD
Met Office
EPSRC
CRAY THE SUPERCOMPUTER COMPANY
ARM

I.K.Brunel 1804-1859

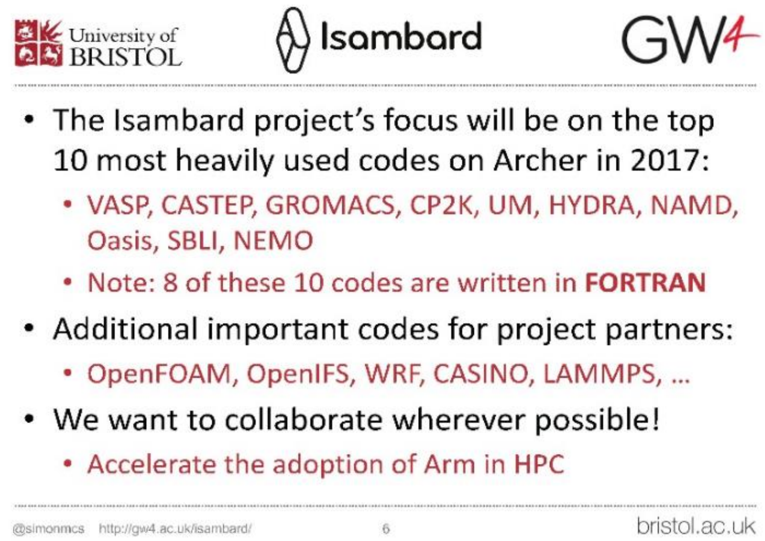


University of BRISTOL Isambard GW4

Isambard system specification (red = new info):

- Cray “Scout” system – XC50 series
  - Aries interconnect
- 10,000+ Armv8 cores
  - Cavium ThunderX2 processors
  - 2x 32core @ >2GHz per node
- Cray software tools
- Technology comparison:
  - x86, Xeon Phi, Pascal GPUs
- Phase 1 installed March 2017
- The Arm part arrives early 2018

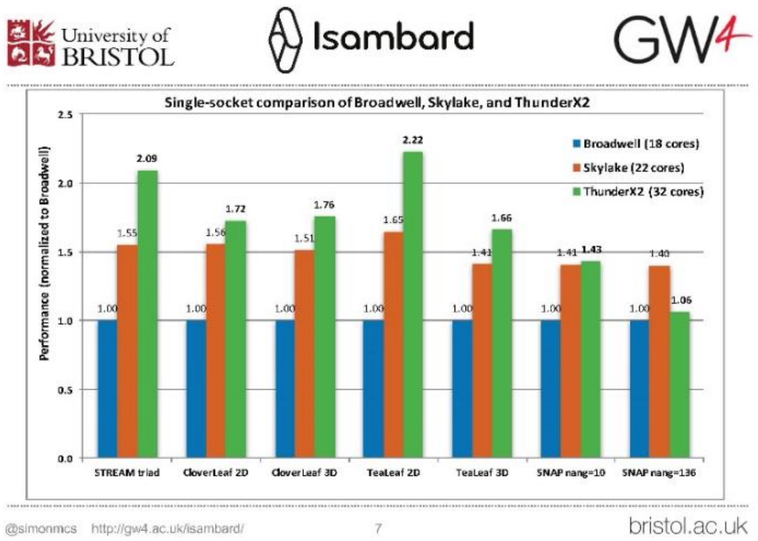
I.K.Brunel 1804-1859



University of BRISTOL Isambard GW4

- The Isambard project’s focus will be on the top 10 most heavily used codes on Archer in 2017:
  - VASP, CASTEP, GROMACS, CP2K, UM, HYDRA, NAMD, Oasis, SBLI, NEMO
  - Note: 8 of these 10 codes are written in **FORTRAN**
- Additional important codes for project partners:
  - OpenFOAM, OpenIFS, WRF, CASINO, LAMMPS, …
- We want to collaborate wherever possible!
  - Accelerate the adoption of Arm in HPC

@simonmcs http://gw4.ac.uk/isambard/ 6 bristol.ac.uk

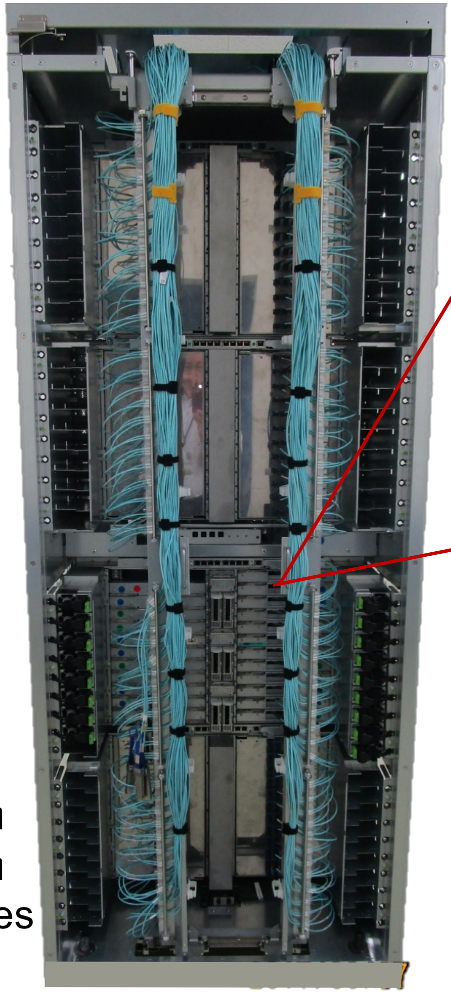


University of BRISTOL Isambard GW4

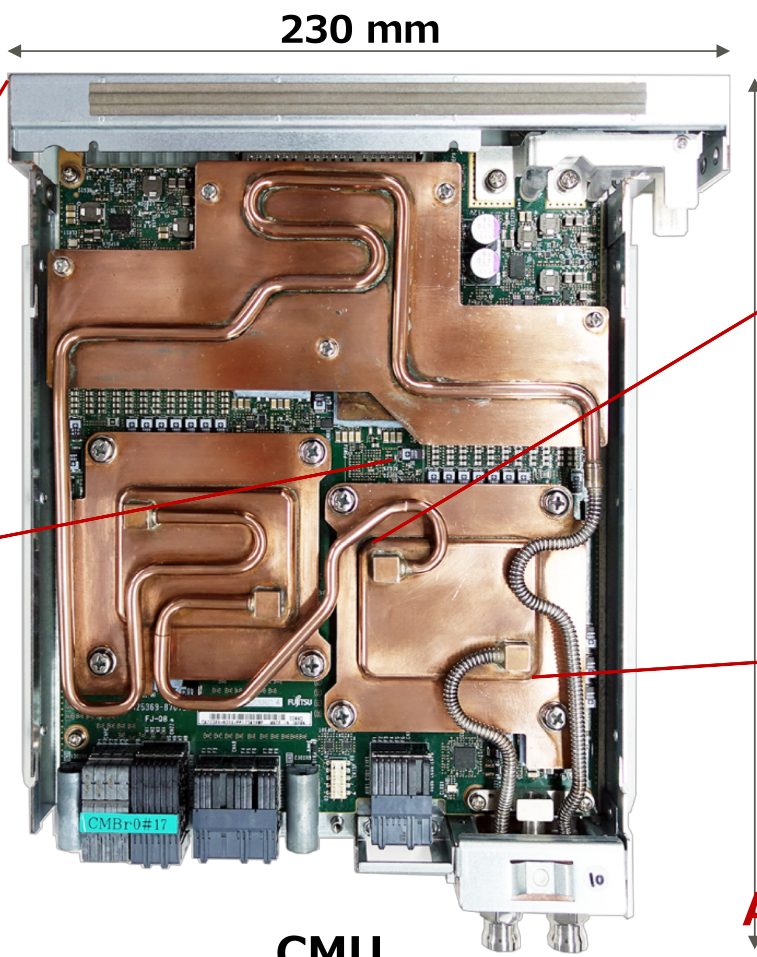
Single-socket comparison of Broadwell, Skylake, and ThunderX2

@simonmcs http://gw4.ac.uk/isambard/ 7 bristol.ac.uk

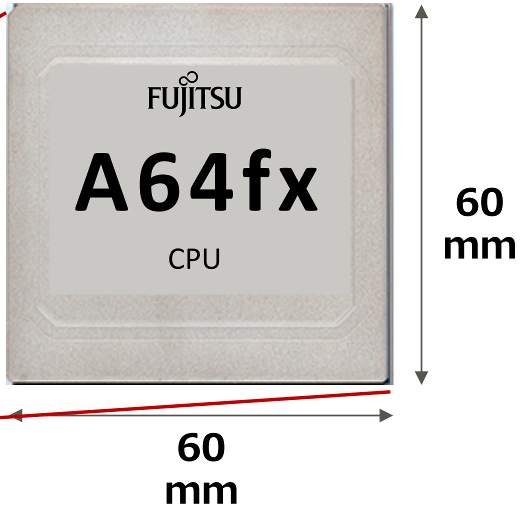
Post-K Chassis, PCB (w/DLC), and A64fx CPU Package



W 800mm
D1400mm
H2000mm
384 nodes



CMU



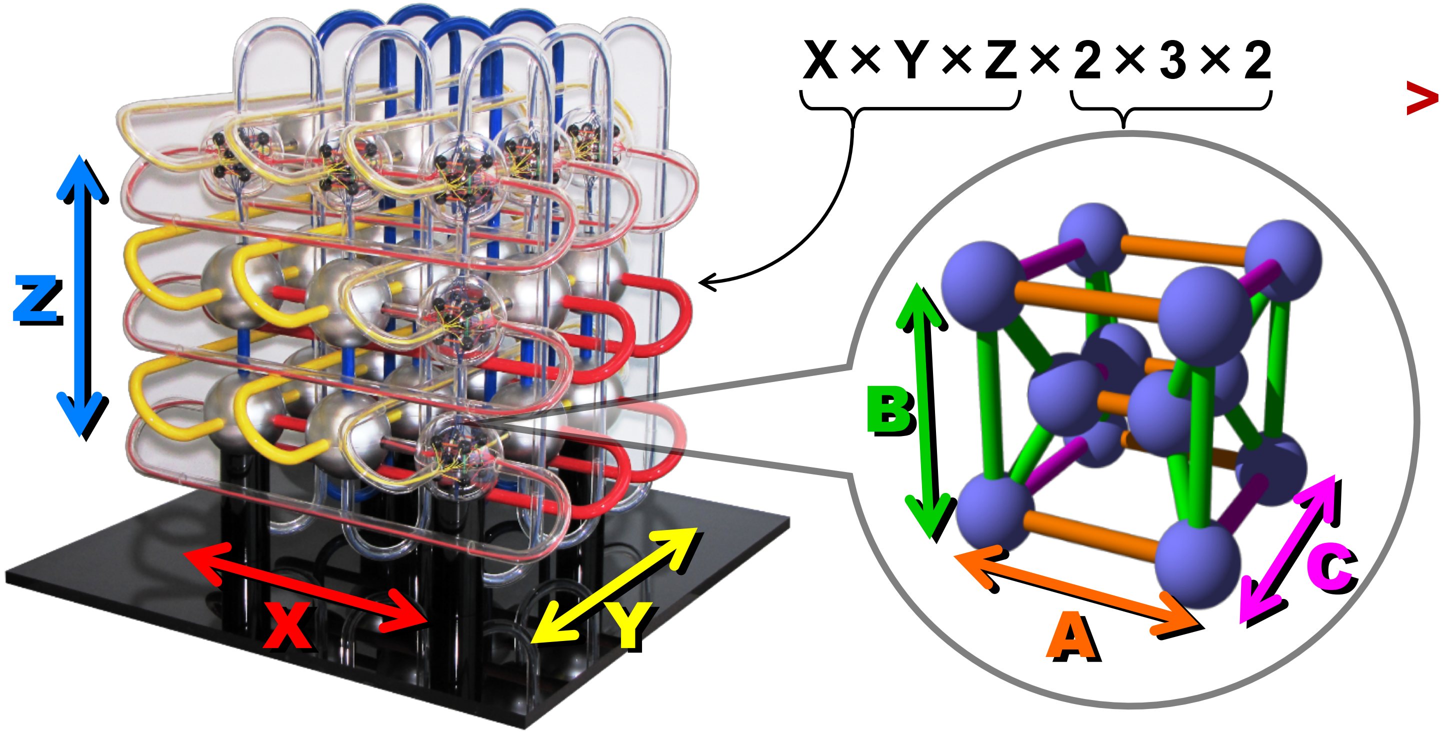
CPU Package

A0 Chip Booted in June
Undergoing Tests
B0 underway

TOFU-D 6D Mesh/Torus Network

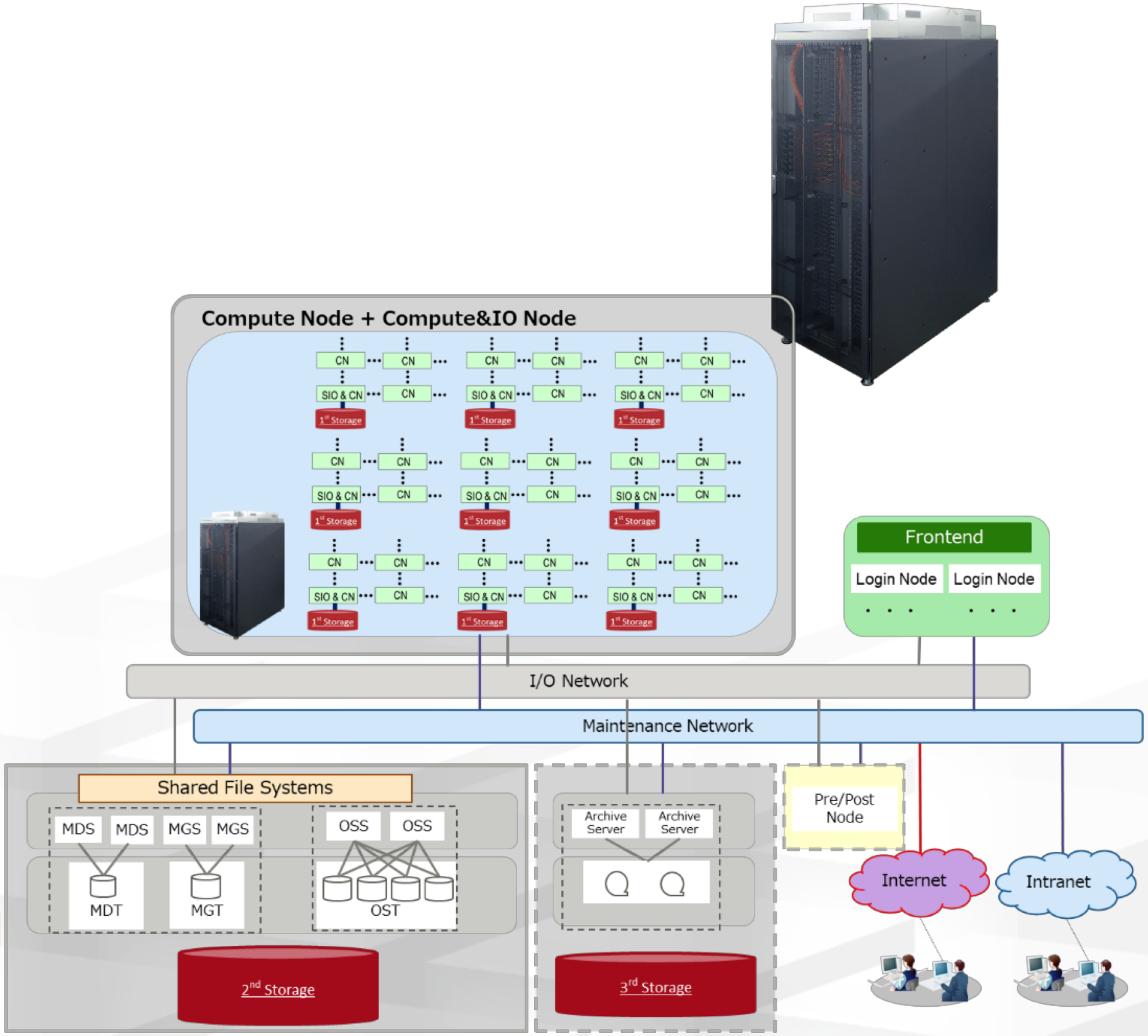
- Six coordinate axes: X, Y, Z, A, B, C
  - X, Y, Z: the size varies according to the system configuration
  - A, B, C: the size is fixed to 2 × 3 × 2
- Tofu stands for “torus fusion”: (X, Y, Z) × (A, B, C)

Embedded on-Chip
0.49 µs latency
38.1GByte/s throughput
Scalable to
> 100,000 nodes



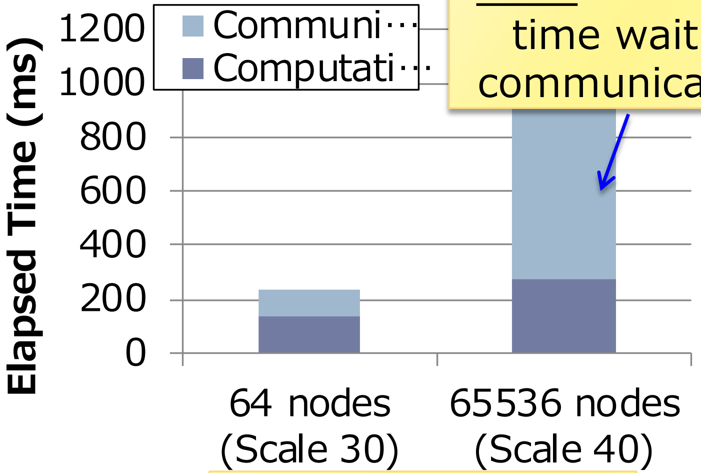
Overview of Post-K System

- **Compute Node, Compute + I/O Node connected by TOFU-D**
- **3-level hierarchical storage**
  - 1st Layer: GFS Cache + Temp FS
  - 2nd Layer: Lustre-based GFS
  - 3rd Layer: Off-site Cloud Storage
- **Full Machine Spec**
  - **>150,000 nodes, ~8 million High Perf. Arm v8.2 Cores**
  - **> 400 racks**
  - **~40 MegaWatts Machine+IDC PUE ~ 1.1 High Pressure DLC**
  - **~= 15~30 million state-of-the art competing CPU Cores for HPC workloads (both dense and sparse problems)**



Sparse BYTES: The Graph500 – 2015~2018 – world #1 x 7

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



73% total exec time wait in communication

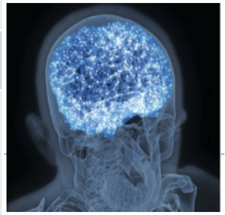
88,000 nodes,
660,000 CPU Cores
1.3 Petabyte mem
20GB/s Tofu NW

K computer

#1 38621.4 GTEPS
(#7 10.51PF Top500)

List	Rank	GTEPS	Implementati
November 2013	4	5524.12	Top-down or
June 2014	1	17977.05	Efficient hybrid
November 2014	2	19585.2	Efficient hybrid
June 2015, June 2016 ~ Nov 2018	**1**	**38621.4**	**Hybrid + Node Compression**

BYTES Rich Machine + Superior BYTES algoithm



LLNL-IBM Sequoia
1.6 million CPUs
1.6 Petabyte mem

≫

Effective x13 performance c.f. Linpack

TaihuLight
10 million CPUs
1.3 Petabyte mem

#3 23751 GTEPS
(#4 17.17PF Top500)

#2 23755.7 GTEPS
(#1 93.01PF Top500)

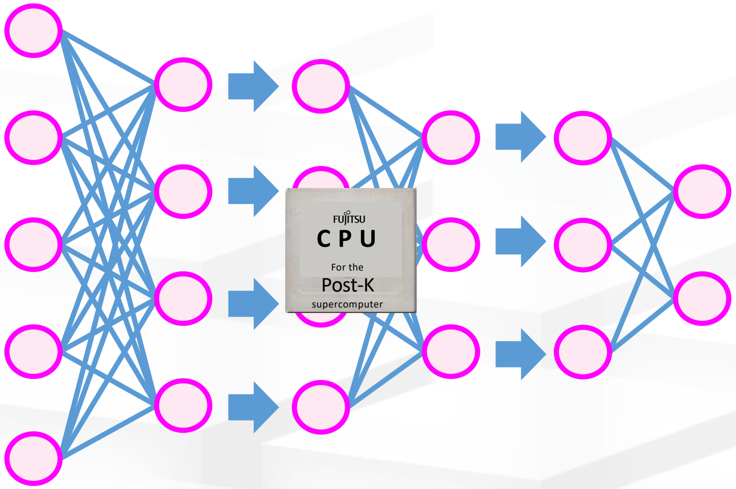
BYTES, not FLOPS!

Massive Scale Deep Learning on Post-K

Post-K Processor

- ◆High perf FP16&Int8
- ◆**High mem BW for convolution**
- ◆**Built-in scalable Tofu network**

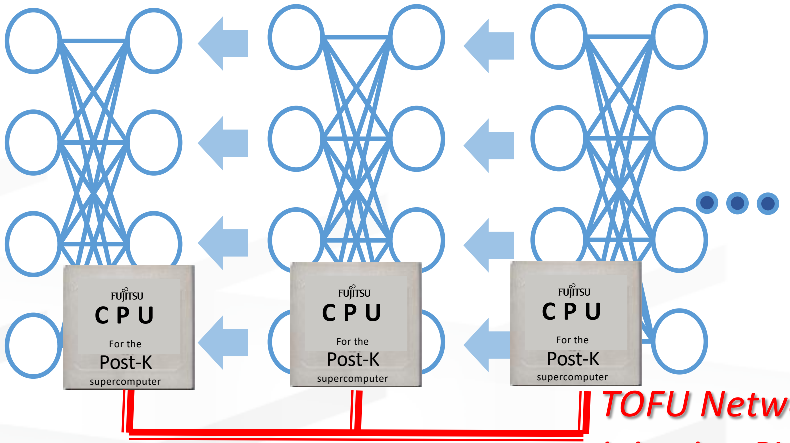
High Performance DNN Convolution



Low Precision ALU + High Memory Bandwidth + Advanced Combining of Convolution Algorithms (FFT+Winograd+GEMM)

Unprecedened DL scalability

High Performance and Ultra-Scalable Network for massive scaling model & data parallelism



TOFU Network w/high injection BW for fast reduction

Unprecedented Scalability of Data/

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

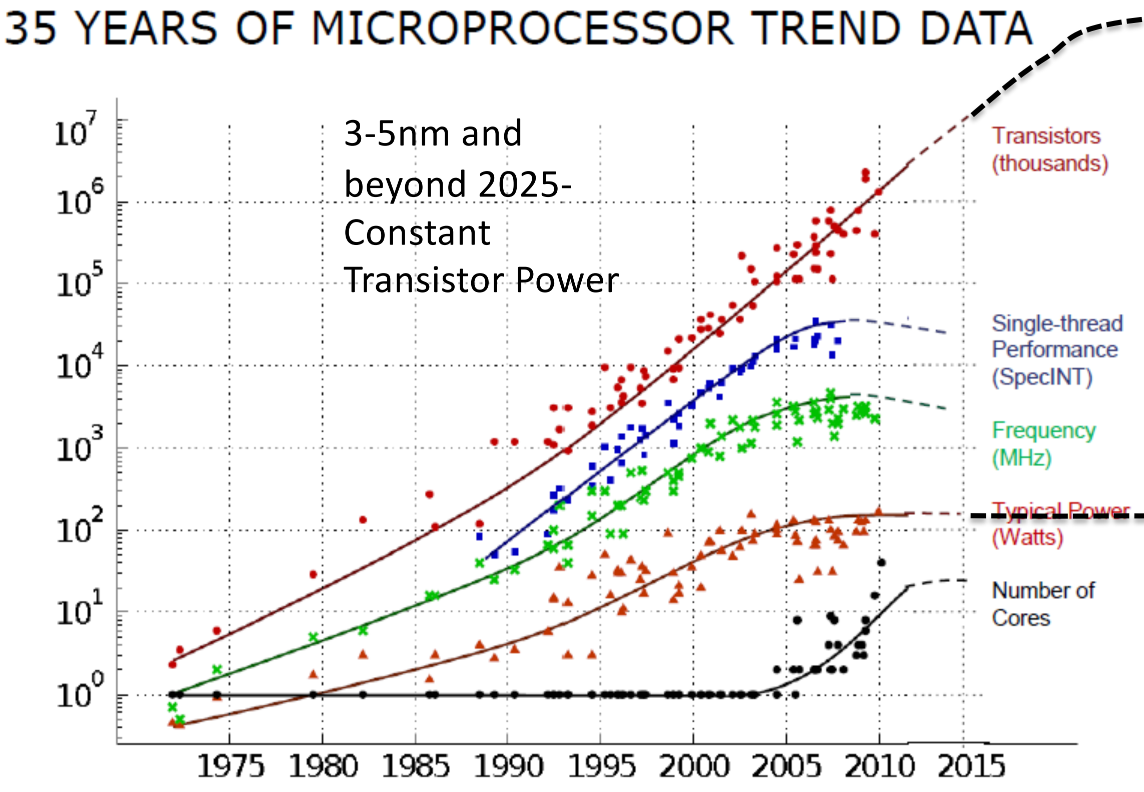
- Much of underlying IT performance growth due to Moore's law
  - "LSI: x2 transistors in 1~1.5 years"
  - Causing qualitative "leaps" in IT and societal innovations
  - The main reason we have supercomputers and Google...
- But this is slowing down & ending, by mid 2020s…!!!
  - End of Lithography shrinks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

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
  - ***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



3-5nm and beyond 2025- Constant Transistor Power

Original data collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond and C. Batten
Dotted line extrapolations by C. Moore

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

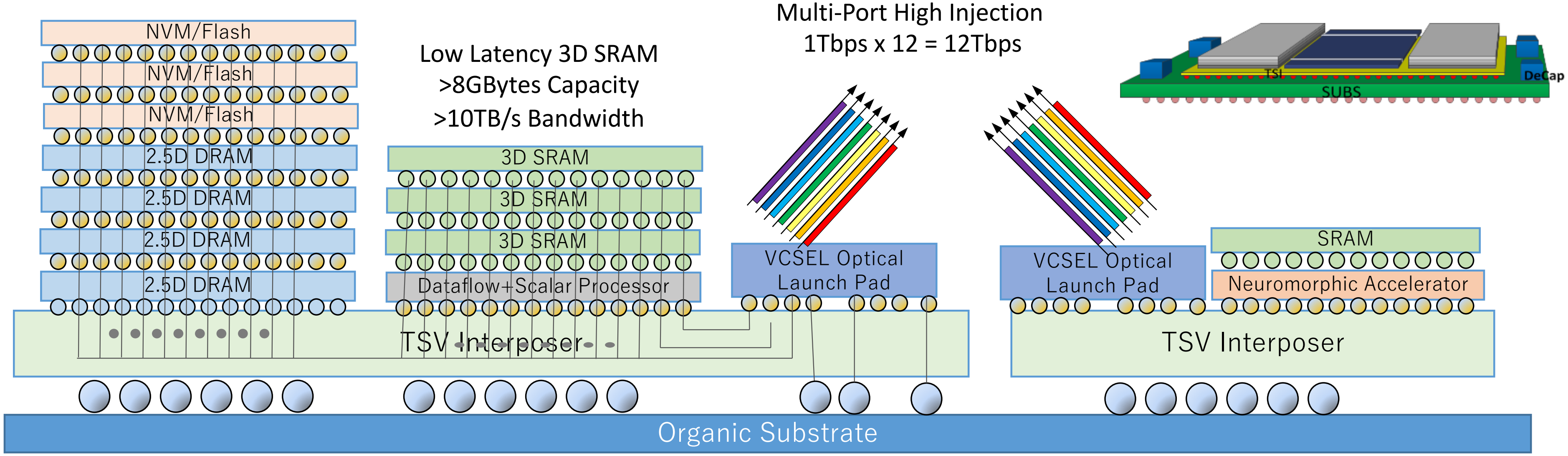
- 2004~2015 feature scaling, perf+ = transistor+ = core#+, constant power
- 2015~2025 all above gets harder

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

2025-2028 Post-Moore FLOPS-to-BYTES x100 Speedup Architecture



Medium Bandwith
2.5D DRAM
>64GBytes Capacity
~3TB/s Bandwidth

High Capacity
Flash NVM
>1 TBytes Capacity

Low Latency 3D SRAM
>8GBytes Capacity
>10TB/s Bandwidth

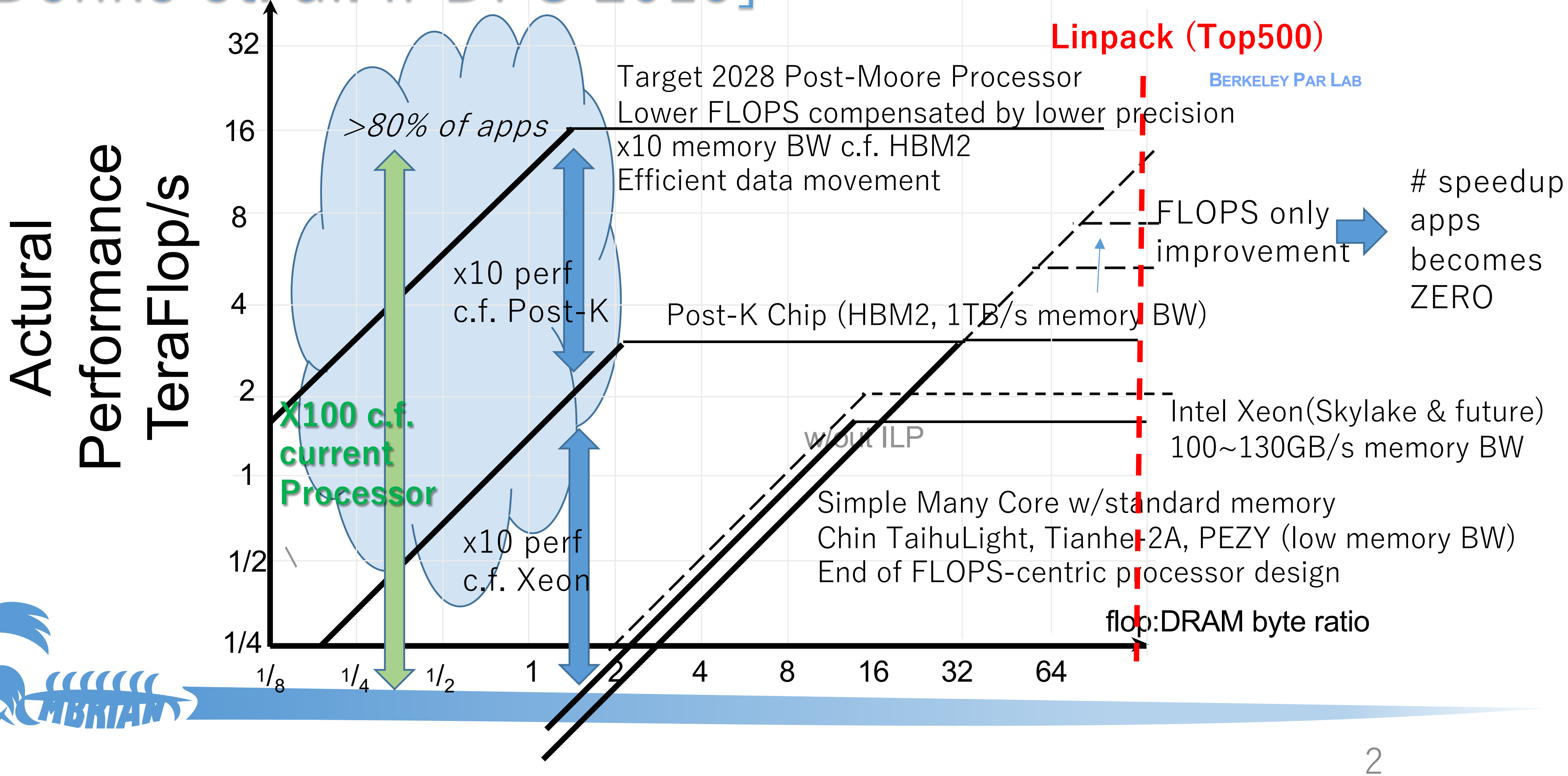
Photonic Network
VCSEL-based
Multi-Port High Injection
1Tbps x 12 = 12Tbps

3 nm UV fabrication

- General purpose processor: Heterogeneous reconfigurable dataflow + scalar many-core processor, 200 Teraflops SFP, 20TeraFlops DFP
- Accelerators: Neural/Neuromorphic, Ising, Graph, etc.

- Direct Chip-Chip Interconnect with DWDM VCSEL micro-optics, 12Tbps injection bandwidth
- Low arity switches for multi-dimensional torus, multi-channel network injection ports

FLOPS to Bytes Data Centric Processor [Donke et. al. IPDPS 2019]



Target 2028 Post-Moore Processor
Lower FLOPS compensated by lower precision
x10 memory BW c.f. HBM2
Efficient data movement

FLOPS only improvement

speedup apps becomes ZERO

Post-K Chip (HBM2, 1TB/s memory BW)

Intel Xeon(Skylake & future)
100~130GB/s memory BW

Simple Many Core w/standard memory
Chin TaihuLight, Tianhe-2A, PEZY (low memory BW)
End of FLOPS-centric processor design

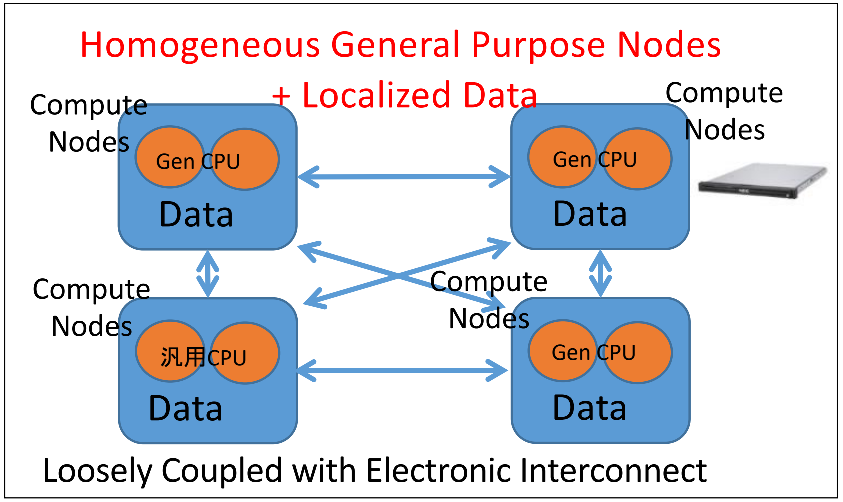
Many Core Era

Post Moore
Cambrian Era

Flops-Centric Monolithic Algorithms and Apps

Flops-Centric Monolithic System Software

Hardware/Software System APIs
Flops-Centric Massively Parallel Architecture



Homogeneous General Purpose Nodes
+ Localized Data

Compute Nodes

Gen CPU

Data

Compute Nodes

Gen CPU

Data

Compute Nodes

汎用CPU

Data

Compute Nodes

Gen CPU

Data

Loosely Coupled with Electronic Interconnect

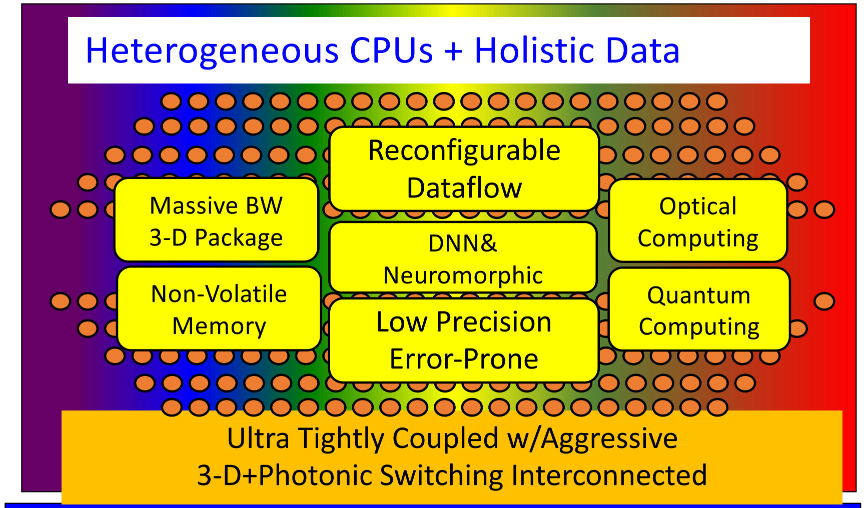
Transistor Lithography Scaling
(CMOS Logic Circuits, DRAM/SRAM)

~2025
M-P Extinction
Event

Cambrian Heterogeneous Algorithms and Apps

Cambrian Heterogeneous System Software

Hardware/Software System APIs
“Cambrian” Heterogeneous Architecture



Heterogeneous CPUs + Holistic Data

Reconfigurable Dataflow

Massive BW 3-D Package

Optical Computing

DNN& Neuromorphic

Non-Volatile Memory

Quantum Computing

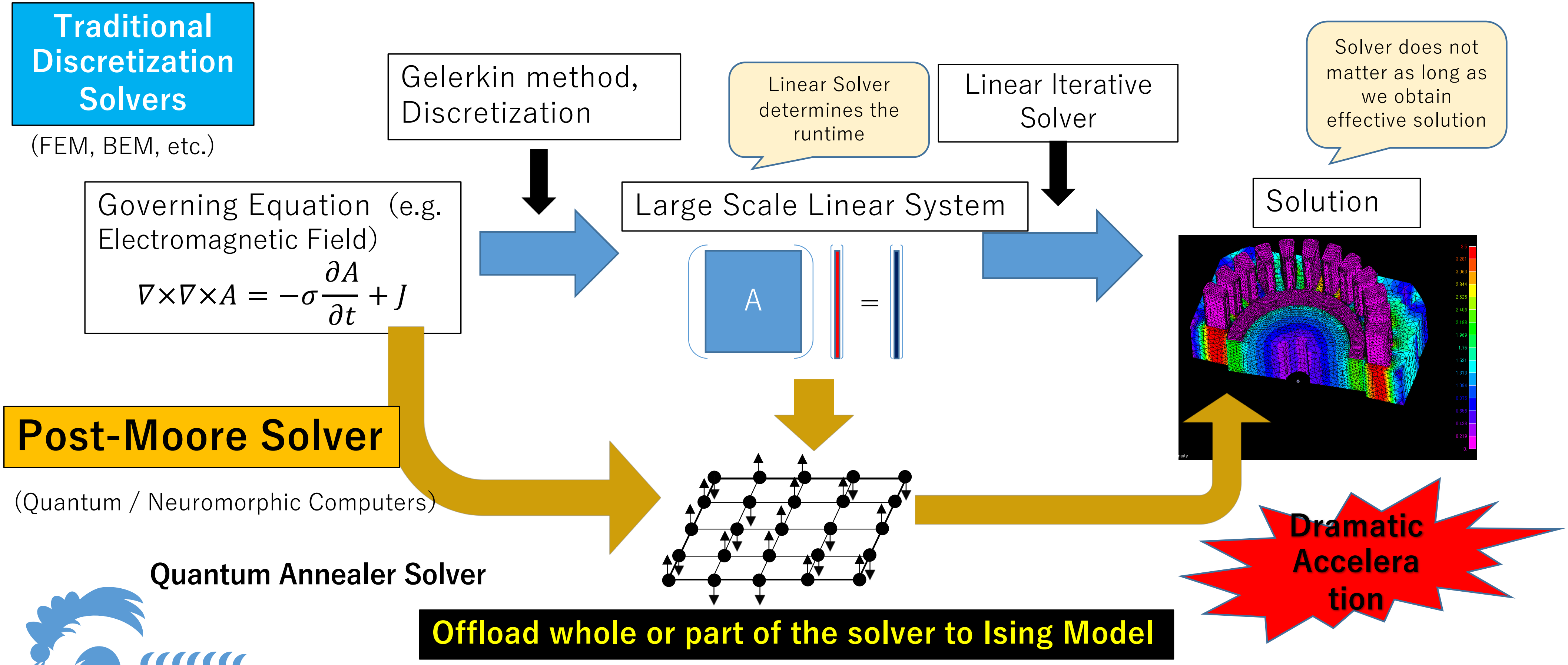
Low Precision Error-Prone

Ultra Tightly Coupled w/Aggressive
3-D+Photonic Switching Interconnected

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

Iwashita (Hokkaido U)

Re-thinking of Solvers in the Post-Moore Architecture



Traditional Discretization Solvers

(FEM, BEM, etc.)

Governing Equation (e.g. Electromagnetic Field)

$$\nabla\times\nabla\times A = -\sigma\frac{\partial A}{\partial t} + J$$

Gelerkin method, Discretization

Large Scale Linear System

Linear Solver determines the runtime

Linear Iterative Solver

Solution

Solver does not matter as long as we obtain effective solution

Post-Moore Solver

(Quantum / Neuromorphic Computers)

Quantum Annealer Solver

Offload whole or part of the solver to Ising Model

Dramatic Acceleration