Overview of Fugaku and its Future Perspectives

Satoshi Matsuoka, Director Riken R-CCS

ASCAC Presentation

29 July 2021[1]

What is a 'Exascale' Supercomputer?

1. FP64 Performance > 1 Exaflop (EF)

1.1. Achieve Rpeak (FP64) > 1 EF

1.2. Achieve Top500 – Linpack Rmax > 1 EF

- *Fugaku Rmax = 0.442 EF, Rpea = 0.537 EF-> NG*
- *However, very little correlation to real apps, symbolic*

2. Any floating point precision performance > 1 Exaflop

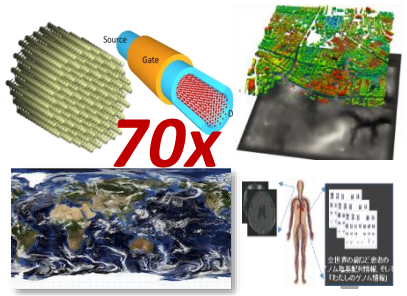
1.1. Peak FP performance > 1 EF

1.2. Measured performance in credible app or benchmark

- *Fugaku FP32, FP16 Peak, HPL-AI (2EF) > 1 Exaflop -> OK!*
- *However, ORNL Summit: FP16 Peak ~= 3 EF, GB2018 App ~= 2EF*

3. Real apps ~= 50~100x 2011~12 10~20PF SCs

- **Fugaku ~70x c.f. K (11PF Rmax) on 9 target apps**
- **"Applications First" -> The most important metric**



Fugaku: Largest & Fastest Supercomputer Ever

'Applications First' R&D Challenge--- High Risk "Moonshot" R&D

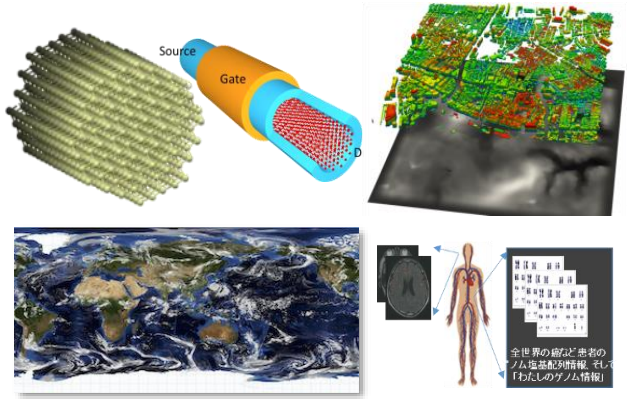
- **A new high performance & low power Arm A64FX CPU co-developed by Riken R-CCS & Fujitsu along with nationwide HPC researchers as a National Flagship 2020 project**

- 3x perf c.f. top CPU in HPC apps
- 3x power efficiency c.f. top CPU
- General purpose Arm CPU, runs same program as Smartphones
- Acceleration features for AI

"Moonshot" R&D Target

- **Fugaku x 2~3 = Entire annual IT in Japan**

	Smartphones		Servers (incl. IDC)		**Fugaku**		K Computer
Untis	20 million ~annual shipment in Japan	=	300,000 (~annual shipment in Japan	=	1 (160K nodes)		Max 120
Power (W)	10W×2,000万台= 200MW	=	600-700W×30万台= 200MW (incl cooling)	>>	30MW (very low)		15MW (less than 1/10 efficiency c.f. Fugaku)



- **Developed via extensive co-design**

"Science of Computing"

By Riken & Fujitsu & HPCI Centers, etc., Arm Ecosystem, Reflecting numerous research results

"Science by Computing"

"9 Priority Areas" to develop target applications to tackle important societal problems

“Applications First” Exascale R&D

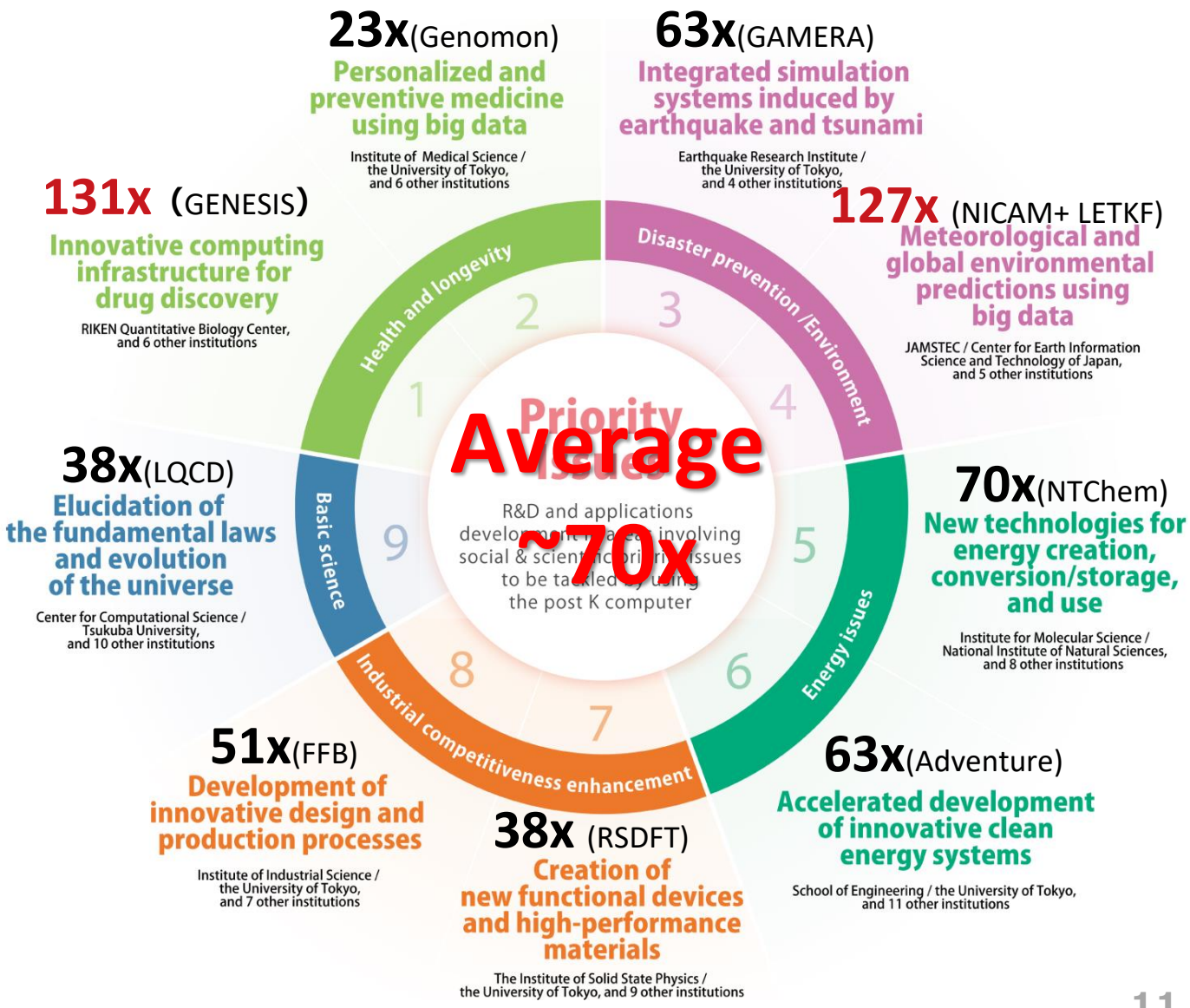
Fugaku Target Applications – Priority Research Areas

- **Advanced Applications Co-Design Program to Parallel Fugaku R&D**
- **Select one representative app from 9 priority areas**

SDGs Goals

- Health & Medicine
- Environment & Disaster
- Energy
- Materials & Manufacturing
- Basic Sciences

- **Up to 100x speedup c.f. K-Computer => achieved!**

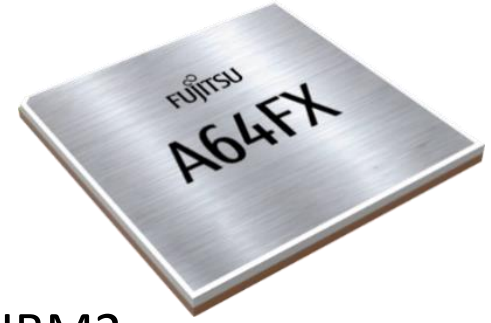


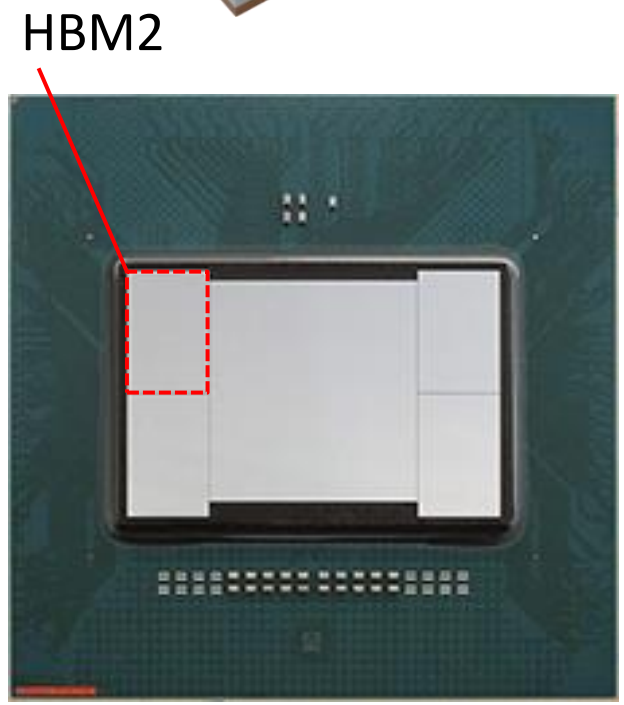
A64FX CPU for supercomputers

FUJITSU

- All-in-one 7nm SoC w/ low power consumption
  - Armv8.2-A, 512-bit SVE (Scalable Vector Extension)
  - Four HBM2, 32 GiB per package
  - Tofu Interconnect D integrated
  - HW inter-core barrier & sector cache
  - 48 compute cores &
    4 assistant cores for OS daemon & MPI offload

CPU core frequency	1.8	2.0	2.2	GHz
Peak DP perf (FP64)	2.7	3.0	3.3	TFLOPS
Peak SP perf (FP32)	5.5	6.1	6.7	TFLOPS
Peak HP perf (FP16)	11	12	13	TFLOPS
Memory peak bandwidth	1024			GB/s



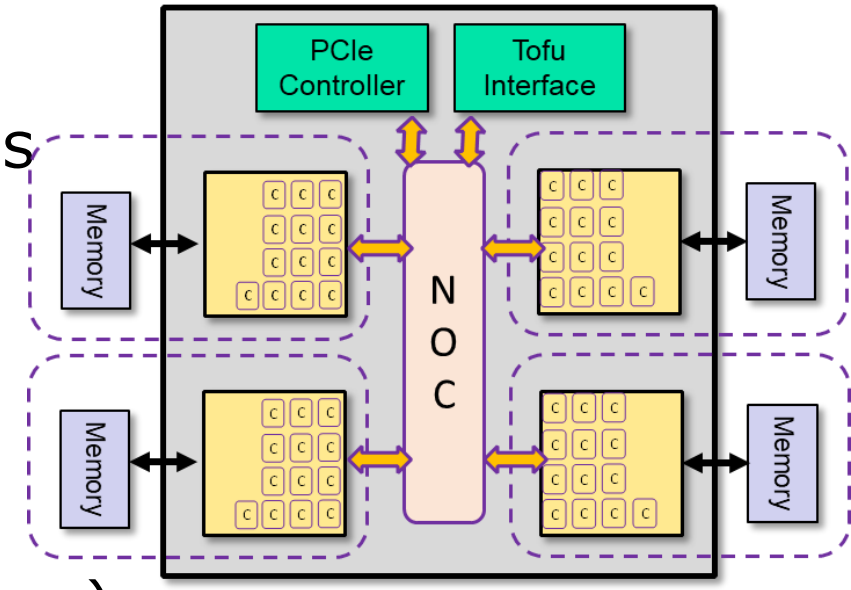


A64FX w/o LID

Fugaku: World Leading Advanced IT (not just SC)

- **CPU: Highest performing general purpose CPU for high-end computing**
  - First server CPU w/7nm process
  - 3x faster c.f. latest CPUs from US competitors w/SVE & HBM2, etc.
  - 3x power efficient -> GPU-class power efficiency
  - Arm v8.2 ISA compliant (own μ-architecture) => e.g. RHEL works out of the box
- **Network/Interconnect: highest bandwidth & lowest latency (Tofu-D)**
  - 400Gbps-class network/node, 0.5μs latency (c.f. IDC 10~100Gbps, 10~100μs latency)
  - First server CPU w/ on-die NIC & switch => 160K nodes interconnected w/o external switch, 1.6 million switch ports, > 100K AoC cables
  - ~6 PetaByte/s injection bandwidth => 10x aggregate GAFAM IDCs traffic
- **System Architecture => World's first ultra-scale disaggregated architecture**
  - CPU cores (esp. L2 Cache), memory (HBM2) and NIC all connected via on-chip network with multiple DMACs => any memory region in the system of 160K noes accessible by any CPU via RDMA and injected onto on-die L2 cache w/sub-μs latecy

Fugaku's FUjitsu A64fx Processor is…



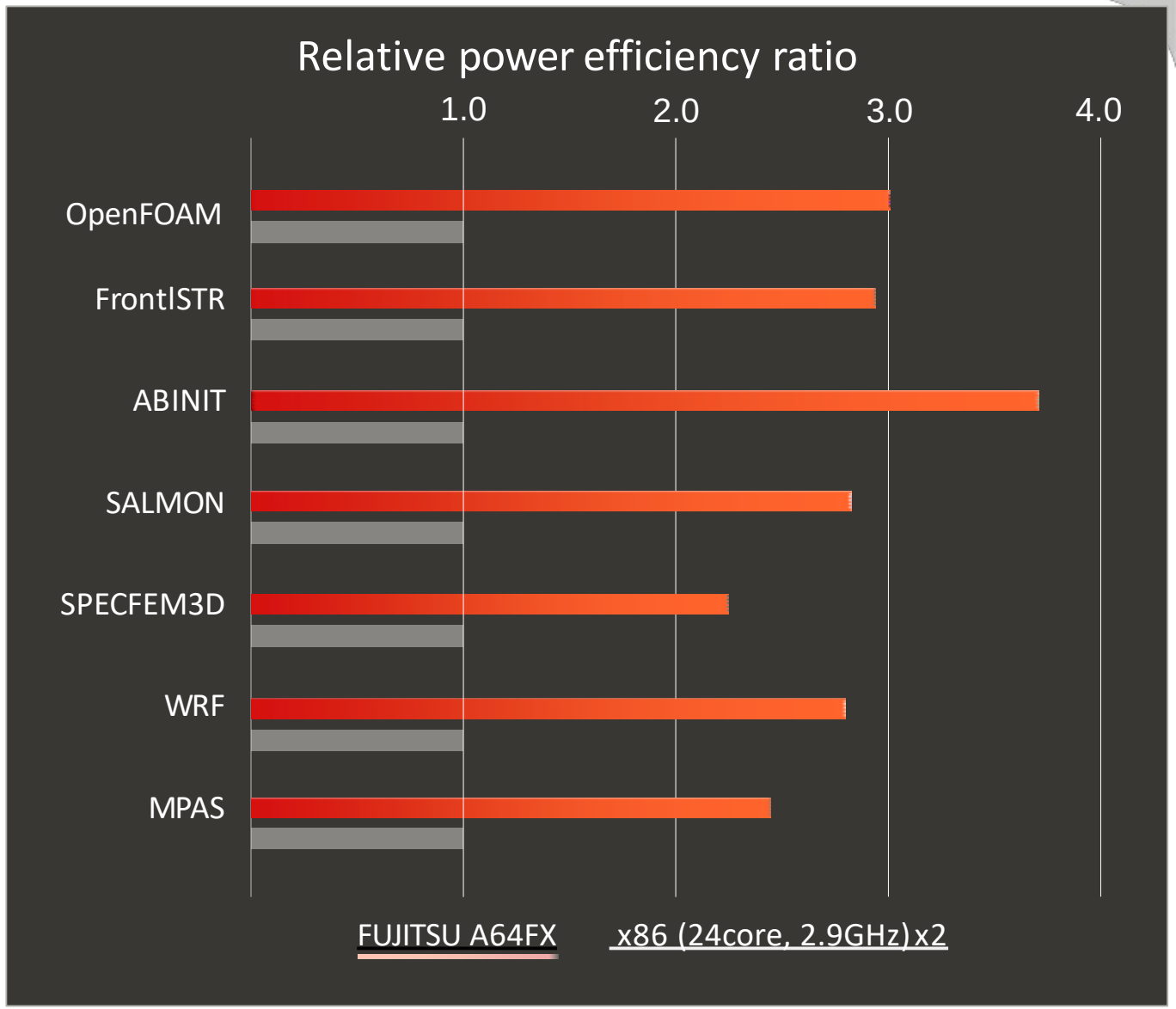
- **an Many-Core ARM CPU…**
  - 48 compute cores + 2 or 4 assistant (OS) cores
  - Brand new core design
  - Near Xeon-Class Integer performance core
  - ARM V8 --- 64bit ARM ecosystem
  - Tofu-D + PCIe 3 external connection
- **…but also an accelerated GPU-like processor**
  - SVE 512 bit x 2 vector extensions (ARM & Fujitsu)
    - Integer (1, 2, 4, 8 bytes) + Float (16, 32, 64 bytes)
  - Cache + memory localization (sector cache)
  - HBM2 on package memory – Massive Mem BW (Bytes/DPF ~0.4)
    - Streaming memory access, strided access, scatter/gather etc.
  - Intra-chip barrier synch. and other memory enhancing features

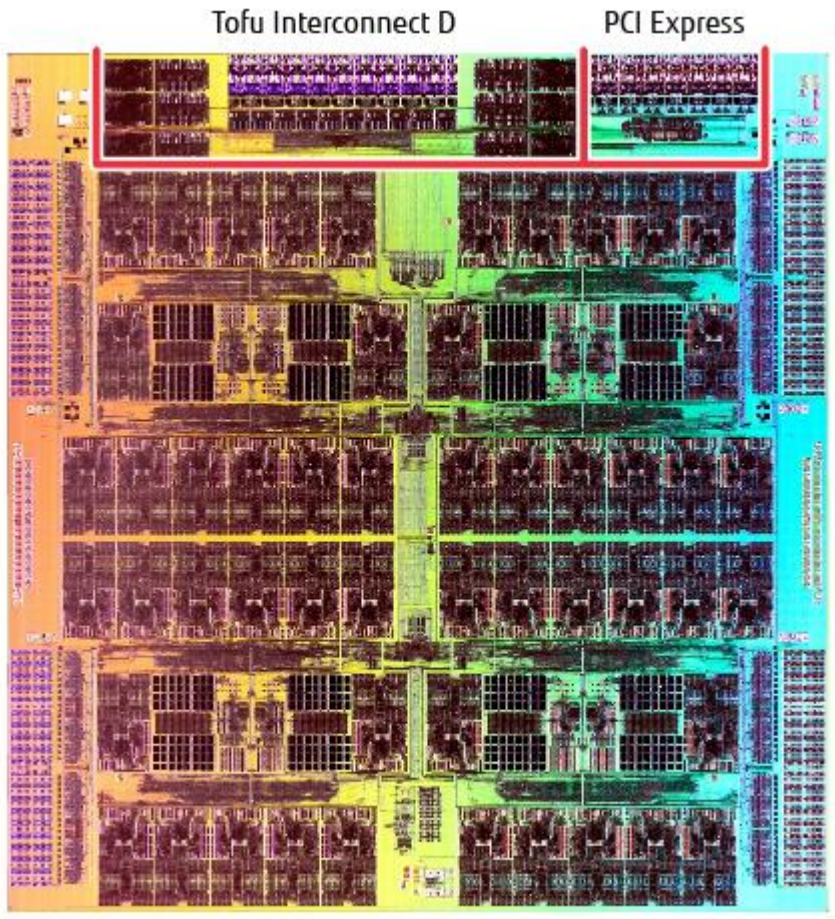
- **GPU-like performance in real-world HPC especially CFD-- Weather & Climate (even w/traditional Fortran code) + AI/Big Data**

A64FX CPU power efficiency for real apps

FUJITSU

- Performance /Energy consumption on an A64FX @ 2.2GHz
- Up to 3.7x more efficient over the latest x86 processor (24core, 2.9GHz) x2
- High efficiency is achieved by energy-conscious design and implementation





- **Tofu-D logic Embedded into CPU die**
- **25mm² die area (~6% of entire die)**
- **Power: 8~9W (incl. SerDes&AOC, very low power c.f. 100GbE, EDR/HDR IB @ 25-30W)**
  - Constant irrespective of state
  - ~ 4~5 % of entire node
- **Directly connected to on-chip torus network**
  - No I/O bus inbetween e.g. PCI-E
  - Direct DMAC access to L2 cache
- **6-D torus router switch + DMAC**
  - ~160,000 low dimension switch on Fugaku
  - ~1.6 million ports total
- **CPU, Memory, and Tofu-D directly connected to on-chip Xbar & NW => disaggregated architecture**

Fugaku Tofu-D Performance

FUJITSU

- 8B Put transfer between nodes on the same board

	Communication settings	Latency
Tofu1(K)	Descriptor on main memory	1.15 µs
	Direct Descriptor	0.91 µs
TofuD	**To/From far CMGs**	**0.54 µs**
	To/From near CMGs	**0.49 µs**

C.f. 100GbE in IDC
Latency 10~100µs

- Total Injection Bandwidth

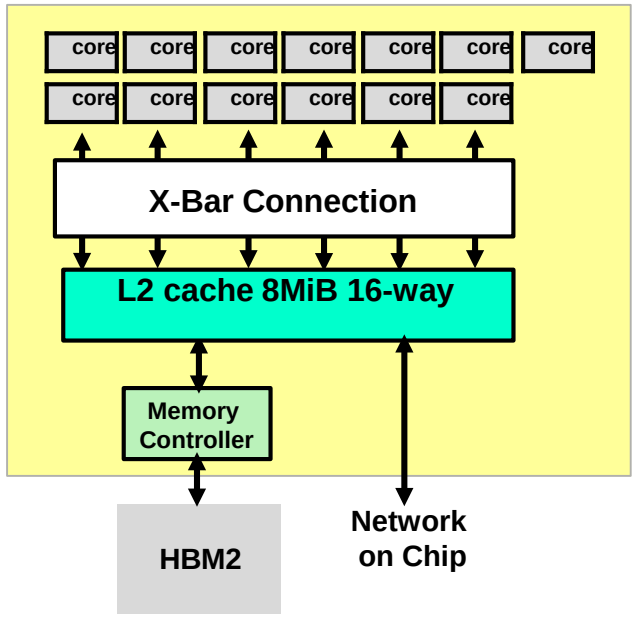
	Injection rate	Efficiency
Tofu1 (K)	15.0 GB/s	77 %
Tofu1 (FX10)	17.6 GB/s	88 %
TofuD	**38.1 GB/s**	**93 %**

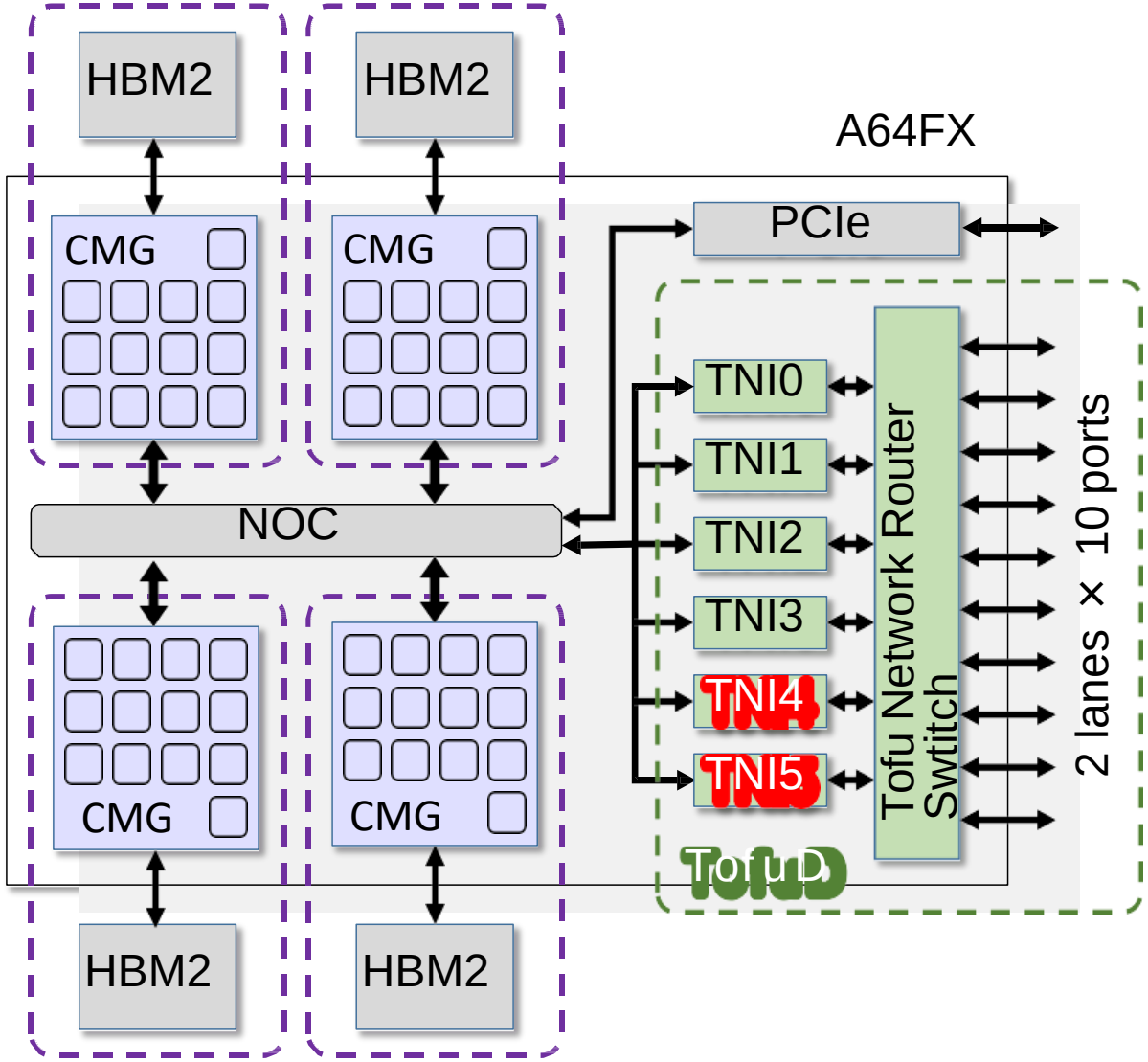
C.f. 100GbE in IDC
Bandwidth ~10GB/s

Disaggregated Architecture of A64FX

- Any CPU can access any memory in system via RDMA (TNI) to its L2
  - Entire 160K Fugaku Nodes
  - Sub microsecond latency
  - NOC + Tofu-D NW Switch on every node (on-die)

CMG Configuration (13 cores + L2 + MC=>HBM2)



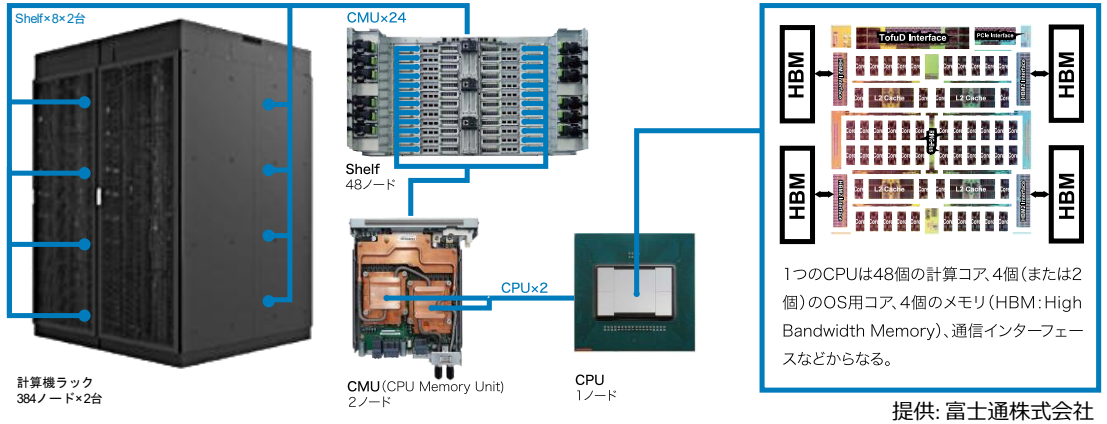


HBM2=>NoC=>TNI=>SW…AoC…SW=>TNI=>NoC=>L2&HBM2

Fugaku Total System Config & Performance

- **Total # Nodes: 158,976 nodes**
  - 384 nodes/rack x 396 (full) racks = 152,064 nodes
  - 192 nodes/rack x 36 (half) racks = 6,912 nodes

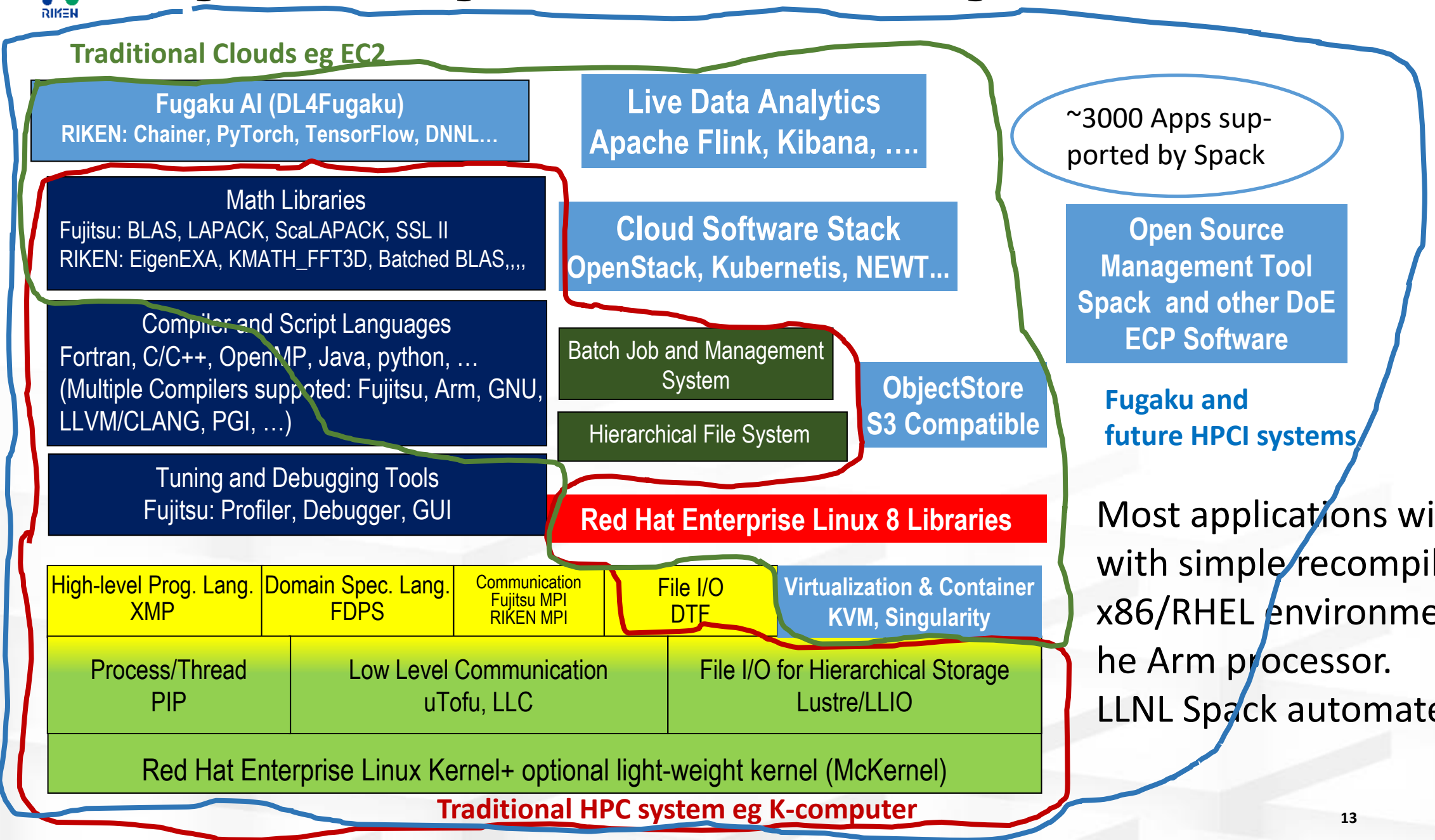
  c.f. K Computer 88,128 nodes
- **Theoretical Peak Compute Performances**
  - Normal Mode (CPU Frequency 2GHz)
    - 64 bit Double Precision FP: 488 Petaflops
    - 32 bit Single Precision FP: 977 Petaflops
    - 16 bit Half Precision FP (AI training): 1.95 Exaflops
    - 8 bit Integer (AI Inference): 3.90 Exaops
  - Boost Mode (CPU Frequency 2.2GHz)
    - 64 bit Double Precision FP: 537 Petaflops
    - 32 bit Single Precision FP: 1.07 Exaflops
    - 16 bit Half Precision FP (AI training): 2.15 Exaflops
    - 8 bit Integer (AI Inference): 4.30 Exaops
- **Theoretical Peak Memory Bandwidth: 163 Petabytes/s**



提供: 富士通株式会社

- <u>C.f. K Computer performance comparison (Boost)</u>
  - 64 bit Double Precision FP: 48x
  - 32 bit Single Precision: 95x
  - 16 bit Half Precision (AI training): 190x
    - K Computer Theoretical Peak: 11.28 PF for all precisions
  - 8 bit Integer (AI Inference): > 1,500x
    - K Computer Theoretical Peak: 2.82 Petaops (64 bits)
  - Theoretical Peak Memory Bandwidth: 29x
    - K Computer Theoretical Peak: 5.64 Petabytes/s

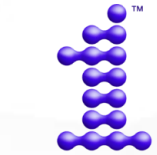
Fugaku HPC+Big Data+AI+Cloud 'Converged' Software Stack



Most applications will work with simple recompile from x86/RHEL environment to t he Arm processor.
LLNL Spack automates this.

Standard Software Ecosystem & OSS Contributions

- Arm v8.2 + SVE and other server standards fully compliant
- Standard Linux distributions work out of the box, most Cloud, HPC, BD OSSs as well
- Standardized configurations via frameworks (e.g., OneAPI, Spack), VMs, Containers
- High Performance AI being developed w/OneDNN & others)



Most Software on x86 HPC Clusters & Clouds Simply Work on Fugaku

Cloud Service Providers Partnership

https://www.r-ccs.riken.jp/library/topics/200213.html (in Japanese)

HPC SYSTEMS

Altair

XTREME-D

NIMBIX

FOCUS 公益財団法人 計算科学振興財団

exa EXA CORPORATION

Agency for Science, Technology and Research SINGAPORE

×

ORACLE

RIKEN R-CCS RIKEN Center fo Computa...

RIKEN

<u>Action Items</u>

- Cool Project name and logo!
- Trial methods to provide computing resources of Fugaku to end-users via service providers
- Evaluate the effectiveness of the methods quantitatively as possible and organize the issues
- The knowledges gained will be feedbacked to scheme design of Fugaku by the government

Fugaku HPL-AI Results Comparisons (update Jun 2021)

- Compute units utilized (FP16)
  - A64FX: 32-element vector FP16 & FP64 mixed precision
  - GPUs: FP16 Matrix Engine (Tensor Core) & FP64 mixed precision
- FP16 vast difference in efficiency, while FP64 efficiency similar
- See our latest paper “Matrix Engines for High Performance Computing: A Paragon of Performance or Grasping at Straws?” [IEEE IPDPS 2021] https://arxiv.org/abs/2010.14373
- We will also release our code as OSS RSN to become a standard like HPL

	Main Processor	**HPL-AI Measured Performance**	**FP16 Peak Performance (full machine)**	**Efficiency**	**HPL-AI Performance /Chip**	Top500 /Linpack FP64 Measured Performance	FP64 Peak Performance	Efficiency
1. Fugaku	Fujitsu A64FX	**2.00 EF**	**2.14 EF**	**93.2%**	**12.6TF**	442.01 PF	537.21 PF	82.3%
2. Summit	NVIDIA V100	**1.15 EF**	**3.46 EF**	**33.2%**	**42.6TF**	148.60PF	200.79 PF	74.0%
3. Selene	NVIDIA A100	**0.63 EF**	**1.40 EF**	**45.0%**	**140.6TF**	63.46 PF	79.22 PF	80.1%

Note: Selene node count based on prerelease info

Development of DL software stack for Arm SVE

FUJITSU

Fugaku AI project
Signed on Nov. 25, 2019

FUJITSU

Framework & oneDNN porting & tuning

Naoki Shinjo, Akira Asato, Atsushi Ike, Koutarou Okazaki, **Yoshihiko Oguchi**, Masahiro Doteguchi, **Jin Takahashi**, Kazutoshi Akao, Masaya Kato, Takashi Sawada, **Naoto Fukumoto**, Kentaro Kawakami, Naoki Sueyasu, Kouji Kurihara, Masafumi Yamazaki, Takumi Honda

Fugaku AI project

RIKEN

Tuning for Fugaku

Satoshi Matsuoka, High Performance Artificial Intelligence Systems Research Team Leader
Kento Sato, High Performance Big Data Research Team Leader
Kazuo Minami, Application Tuning Development Unit Leader
Akiyoshi Kuroda, Application Tuning Development Unit

arm Linaro

CybozuLabs

Technical support

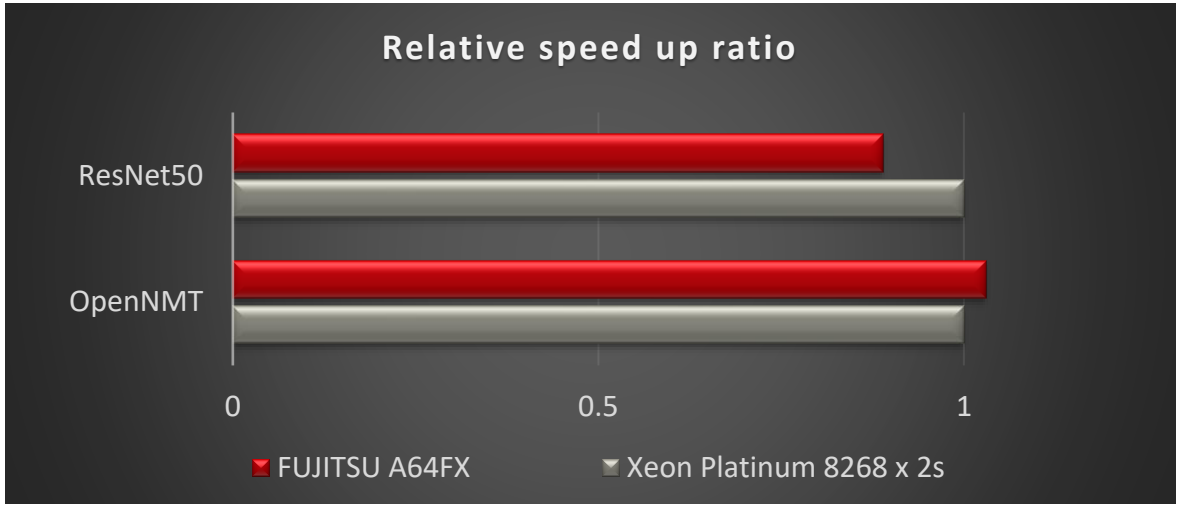
Shigeo Mitsunari

A64FX preliminary results for Deep Learning

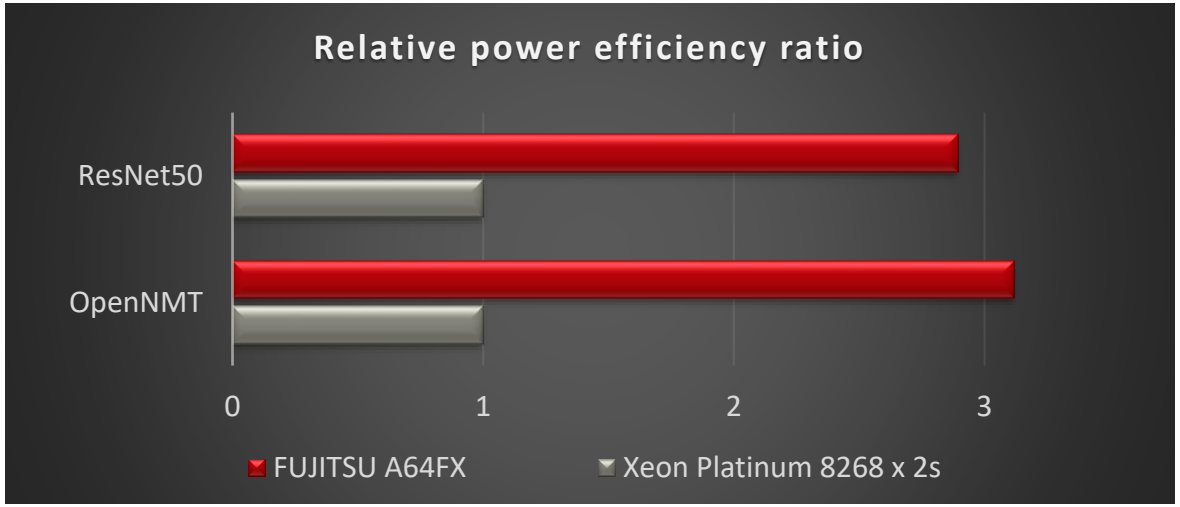
FUJITSU

- Setup
  - Using the same number of CPU cores
    - FX1000 single node (A64FX 2.2 GHz) vs. Xeon Platinum 8268 (24 core, 2.9GHz) x2
  - ResNet50 (image classification)
  - OpenNMT (natural lang. processing)
- Results
  - Performance:
    - Almost the same performance as Xeon
  - Energy efficiency:
    - Up to 2.8x more efficient over Xeon

FX1000



Training using fp32, PyTorch v1.5.0, OneDNN_aarch64, batch size 75 x 4proc.



Training using fp32, PyTorch v1.6.0, OneDNN_aarch64, batch size 3850 x 2proc.

Press releases
2020
2019
2018
2017
2016
2015
2014
2013
2012
2011
2010
2009
2008
2007

Fujitsu, AIST, and RIKEN Achieve Unparalleled Speed on the MLPerf HPC Machine Learning Processing Benchmark Leveraging Leading Japanese Supercomputer Systems

Fujitsu Limited,National Institute of Advanced Industrial Science and Technology,RIKEN

Tokyo, November 19, 2020

Fujitsu, the National Institute of Advanced Industrial Science and Technology (AIST), and RIKEN today announced a performance milestone in supercomputing, achieving the highest performance and claiming the ranking positions on the MLPerf HPC benchmark(1). The MLPerf HPC benchmark measures large-scale machine learning processing on a level requiring supercomputers, and the parties achieved these outcomes leveraging approximately half of the "AI-Bridging Cloud Infrastructure" ("ABCI") supercomputer system, operated by AIST, and about 1/10 of the resources of the supercomputer Fugaku, which is currently under joint development by RIKEN and Fujitsu.

Utilizing about half the computing resources of its system, ABCI achieved processing speeds 20 times faster than other GPU-type systems. That is the highest performance among supercomputers based on GPUs, computing devices specialized in deep learning. Similarly, about 1/10 of Fugaku was utilized to set a record for CPU-type supercomputers consisting of general-purpose computing devices only, achieving a processing speed 14 times faster than that of other CPU-type systems.

The results were presented as MLPerf HPC v0.7 on November 18th (November 19th Japan Time) at the 2020 International Conference for High Performance Computing, Networking, Storage, and Analysis (SC20) event, which is currently being held online.

FY2020-2022 Fugaku Research Promotion Program (1)

- **Area I Challenges for Solving Universal Problems of Mankind or Pioneering the Future**
  - Toward a unified view of the universe: from large scale structures to planets
  - Exploration of burning plasma confinement physics
  - Simulation for basic science: from fundamental laws of particles to creation of nuclei
  - Basic science for emergence and functionality of quantum matter
  - Biomolecular dynamics and function in a living cell using atomistic and coarse-grained simulations
  - Unravelling origin of cancer and diversity by large-scale data analysis and artificial intelligence technology
  - Human-scale whole brain simulation with connectome analysis and structure-function estimation
- **Area II Reinforcement of Efforts on Protecting People's Life and Property**
  - Overcoming heart failure pandemic with innovative integration of multi-scale heart simulator and large-scale clinical data
  - Large-scale numerical simulation of earthquake generation, wave propagation and soil amplification
  - Large Ensemble Atmospheric and Environmental Prediction for Disaster Prevention and Mitigation
  - Promotion of innovative drug discovery infrastructure for acceleration of precision medicine

Follow-on to the Priority Areas Program --- Real S&T Research Expected to Promote Fugaku

FY2020-2022 Fugaku Research Promotion Program (2)

- **Area III Enhancement of Industrial Competitiveness**
  - Environment-Compatible Chemical Substances
  - Multiscale simulations based on quantum theory toward the developments of energy-saving next-generation semiconductor devices
  - Digital Twins of Real World's Clean Energy Systems with Integrated Utilization of Super-simulation and AI
  - Development of high-performance permanent magnets by large-scale simulation and data-driven approach
  - Computational and Data Science Study for ET Revolution by Development of Next-Generation Battery and Fuel Cell
  - R&D of innovative fluid-dynamics simulations for aerodynamical/hydrodynamical performance predictions by using Fugaku (R&D of a turbomachinery design simulation system)
- **Area II Reinforcement of Efforts on Protecting People's Life and Property**
  - R&D of innovative fluid-dynamics simulations for aerodynamical/hydrodynamical performance predictions by using Fugaku (R&D of a vehicle design simulation system)
  - Leading research on innovative aircraft design technologies to replace flight test
- **Area IV Research Infrastructure**
  - Development of personalized medical support technology based on simulation data science of whole brain blood circulation

US DoE Labs Fugaku Usage via DoE-MEXT Partnership

- As a new phase of DoE – MEXT collaboration, DoE Labs and ECP were given opportunities to port their code to Fugaku for evaluation, with support from Riken and Fujitsu since January 2021
- Despite only 3 month in a brand new (and huge) environment, some teams were very successful in obtaining excellent performance and scalability results.
- Some groups naturally suffered primarily performance issues due to compilers, libraries, etc. Collaboration will continue till at least March 2022 to work on such problems to demonstrate Arm/SVE viability.
- Many thanks to Doug Kothe, Thuc Hoang, Mike Heroux, Lori Diachin, and all the members from the DoE labs and their collaborators that are taking part!

- **ECP groups:**
  - E3SM-MMF
  - CEED
  - ExaSky
  - ExaGraph
  - Kokkos
  - SLATE & HeFFTe & xSDK & PaPSEC & PAPI
  - PETSc/TAO

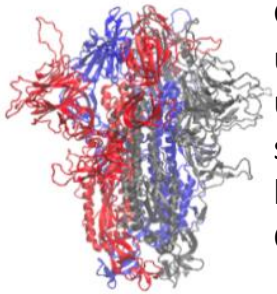
- **ASC and other DOE-MEXT collabo:**
  - StonyBrook (HPE/Cray Ookami)
  - LANL App Performance, FleCSALE
  - LLNL LBANN, MFEM, Spack, SW4
  - SNL App Performance, ATDM Kokkos, SuperContainers, Trilinos
  - ORNL Jeff Vetter's group

MEXT Fugaku Program: Fight Against COVID19

Fugaku resources made available a year ahead of general production (more research topics under international solicitation, also joined US-lead COVID-19 High Performance Computing Consortium)

Medical-Pharma

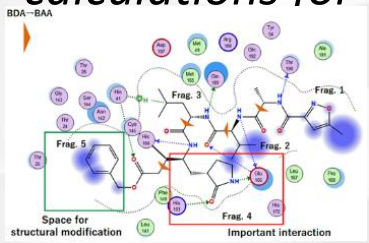
Prediction of conformational dynamics of proteins on the surface of SARS-Cov-2



GENESIS MD to interpolate unknown experimentally undetectable dynamic behavior of spike proteins, whose static behavior has been identified via Cryo-EM

((Yuji Sugita, RIKEN)

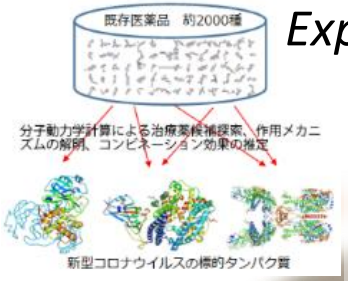
Fragment molecular orbital calculations for COVID-19 proteins



Large-scale, detailed interaction analysis of COVID-19 using Fragment Molecular Orbital (FMO) calculations using ABINIT-MP

(Yuji Mochizuki, Rikkyo University)

Exploring new drug candidates for COVID-19

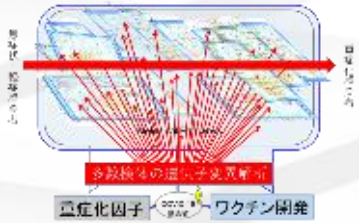


Large-scale MD to search & identify therapeutic drug candidates showing high affinity for COVID-19 target proteins from 2000 existing drugs

(Yasushi Okuno, RIKEN / Kyoto University)

Host genetic analysis for severe COVID-19

Whole-genome sequencing of severe cases of COVID-19 and mild or asymptomatic infections, and identify risk-associated genetic variants for severe disease

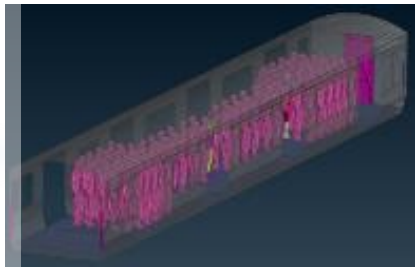


(Satoru Miyano, Tokyo Medical and Dental University)

Societal-Epidemiology

Prediction and Countermeasure for Virus Droplet Infection under the Indoor Environment

Massive parallel simulation of droplet scattering with airflow and hat transfer under indoor environment such as commuter trains, offices, classrooms, and hospital rooms



(Makoto Tsubokura, RIKEN / Kobe University)

Simulation analysis of pandemic phenomena

Combining simulations & analytics of disease propagation w/contact tracing apps, economic effects of lockdown, and reflections social media, for effective mitigation policies

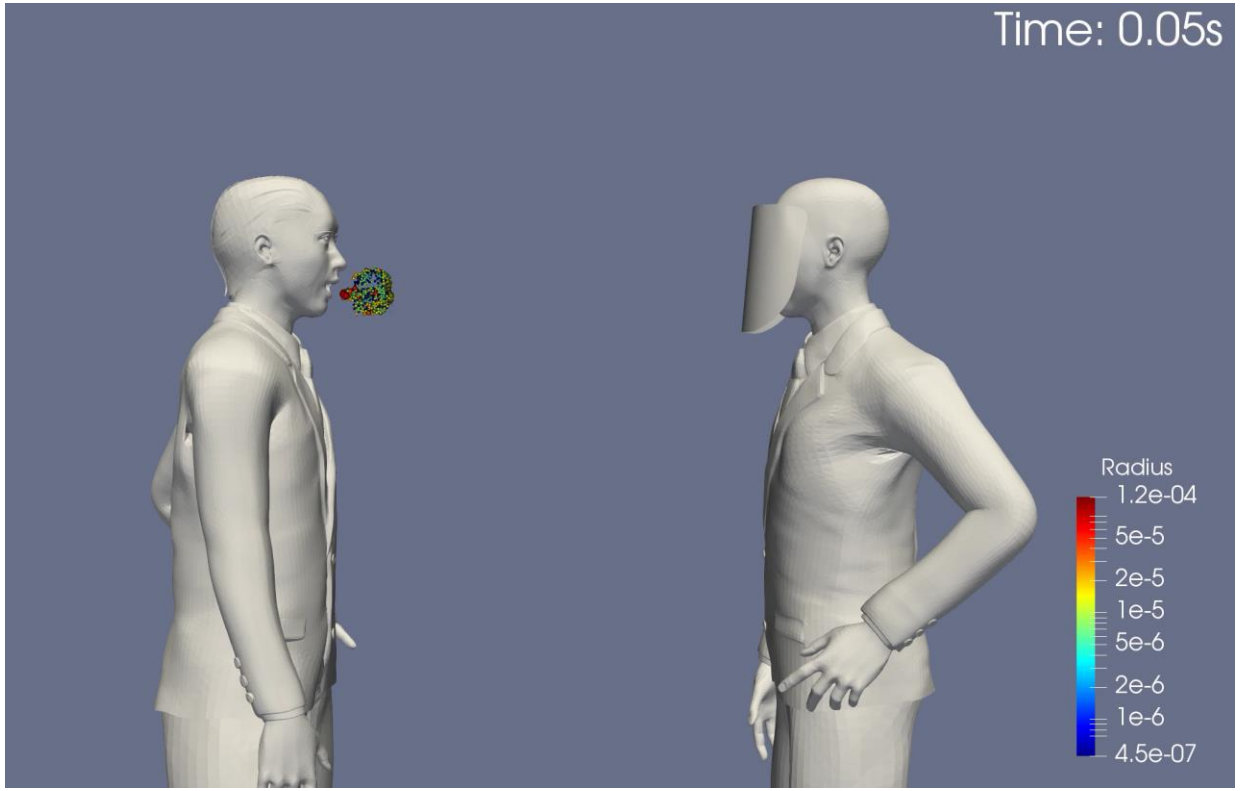
(Nobuyasu Ito, RIKEN)

Face Shield Protection – Defensive and Suppressions

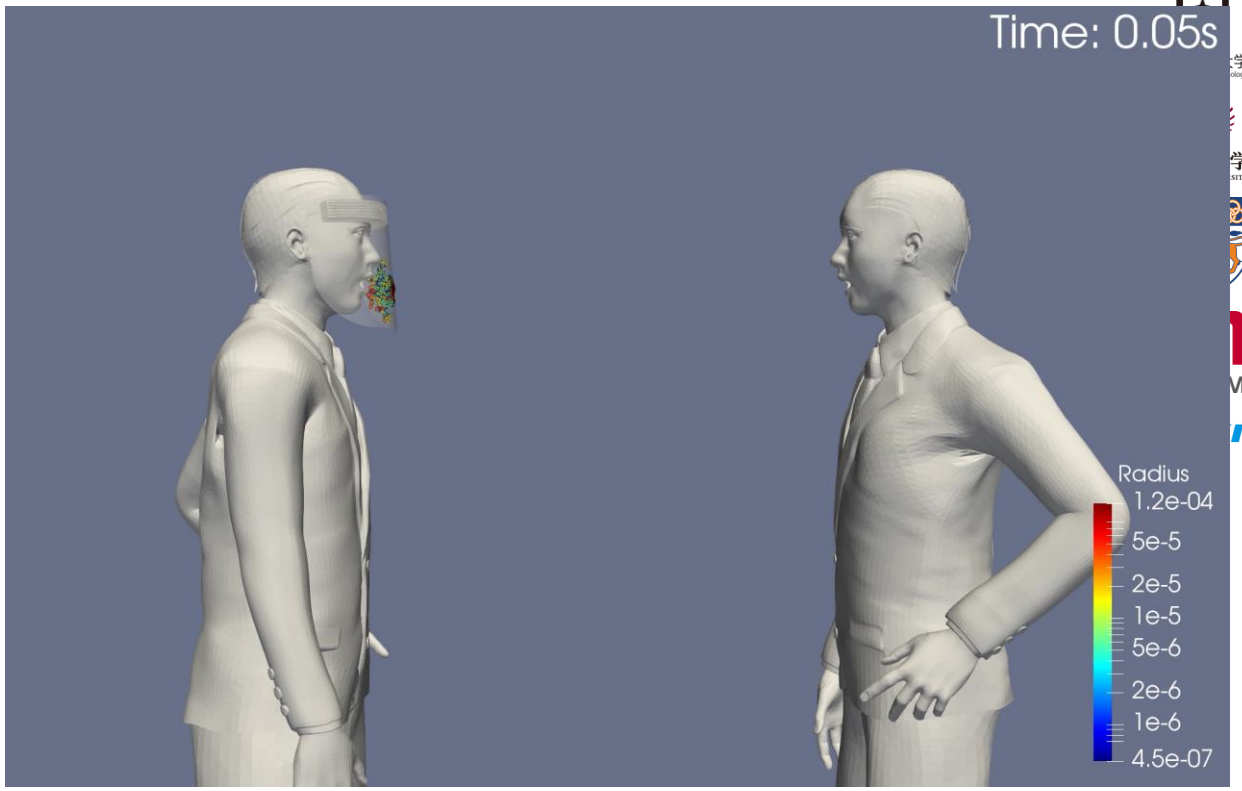
Face shield for defending and suppressing transmissions

- The infected person coughs
- Assumes breathing through both mouth and nose.
- Color indicates the size of droplets (Red: 100 microns Blue: below 2 microns)

Defending from droplets from unprotected infected person



Suppressing spread of droplets from the infected person



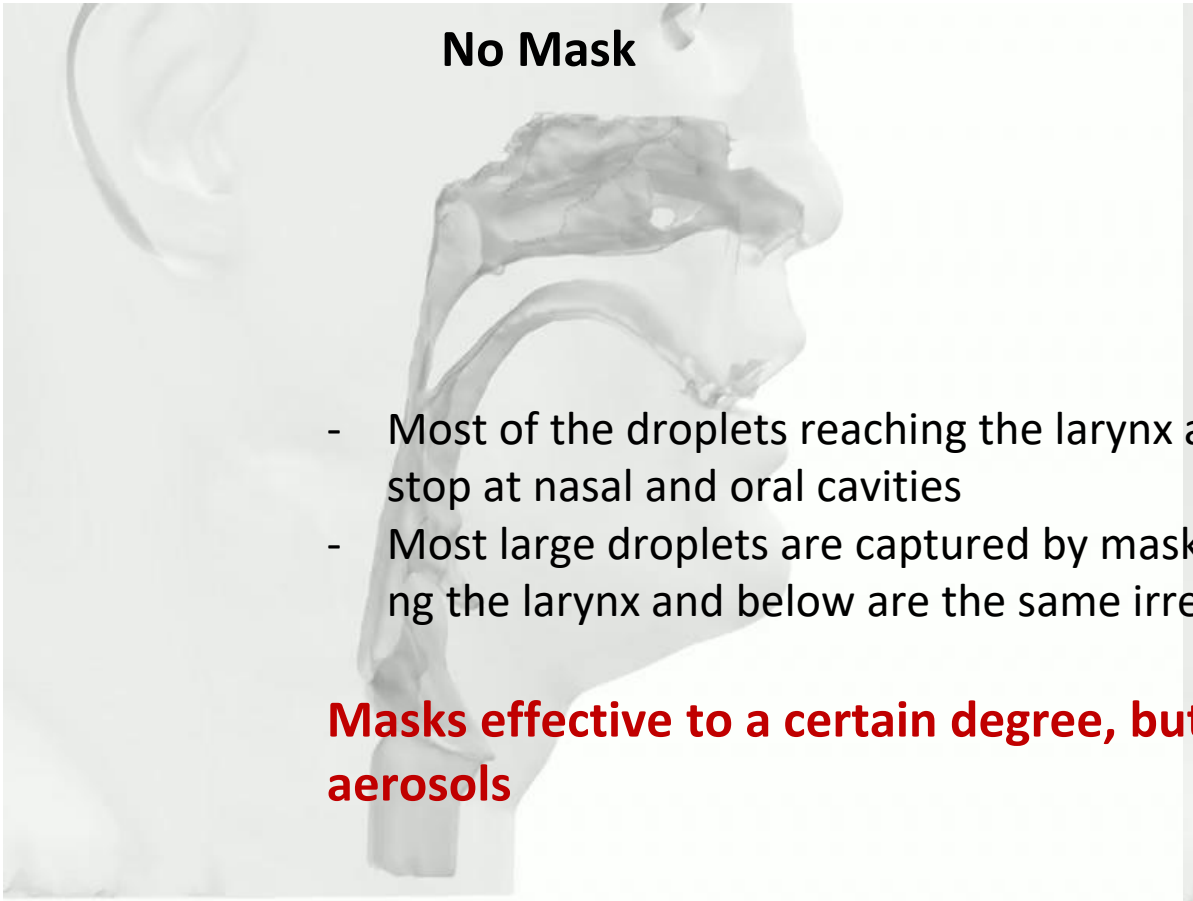
提供：理研・豊橋技科大・神戸大，協力：京工繊大・阪大

Can Masks Defensively Prevent Infection?

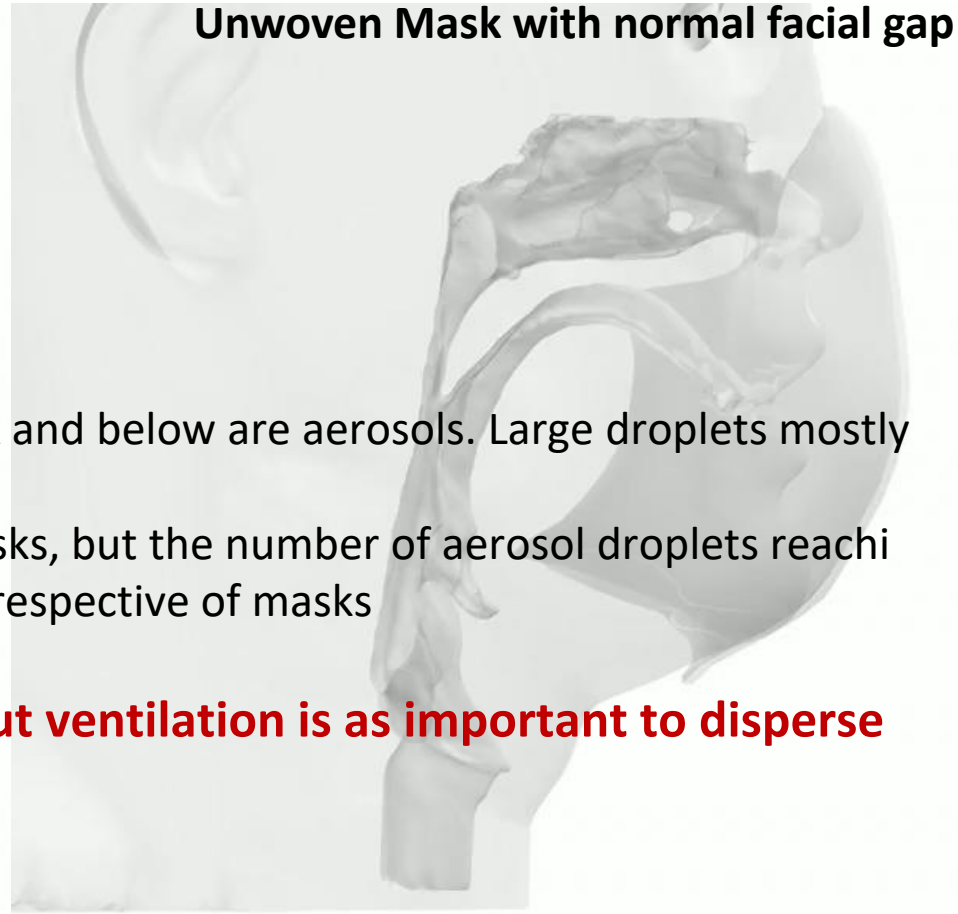
Can masks prevent virus droplets / aerosols from entering the body?

- 3-D Model the upper respiratory tract, simulate how many droplets enter the
- Deep breathing (6 seconds inhale/exhale), simultaneous nose / mouth breathing
- Uniform distribution of droplets / aerosols of varying sizes
- Colors show the size of droplets 色は飛沫のサイズを表す(Red: 100 microns Blue: 1 micron)

No Mask



Unwoven Mask with normal facial gap



- Most of the droplets reaching the larynx and below are aerosols. Large droplets mostly stop at nasal and oral cavities
- Most large droplets are captured by masks, but the number of aerosol droplets reaching the larynx and below are the same irrespective of masks

Masks effective to a certain degree, but ventilation is as important to disperse aerosols

Virus Droplet Simulation & Mitigation - Research, Government, Industry Collaboration

Core Steering Member, Academia & Industry

理化学研究所 R-CCS 神戸大学 Kobe University 東京工業大学 Tokyo Institute of Technology

国立大学法人 豊橋技術科学大学 Toyohashi Univ. of Technology

京都工芸繊維大学 KYOTO INSTITUTE OF TECHNOLOGY

大阪大学 OSAKA UNIVERSITY

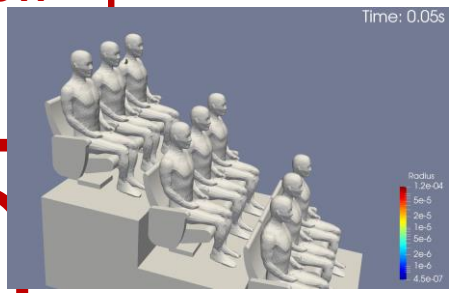
九州大学 Kyushu University

鹿島 KAJIMA CORPORATION

DAIKIN

National U Singapore

Scientific Modeling, Simulation and Physical Experiments of Droplets as well as Environmental Airflow



+

Colaboollating Industry Partners

大王製紙株式会社 DAIO PAPER CORPORATION

ZEN-ON ZEN-ON Music Company LTD Since 1931

TOYOTA

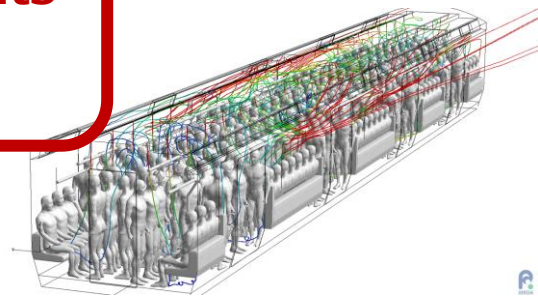
TOYOTA CUSTOMIZING & DEVELOPMENT

FUSO

JAL JAPAN AIRLINES

SUNTORY TOPPAN

Apply to Social Scenarios Transportation, Restaurants & Bars, Theaters, …



Government Collaboration and Funding

文部科学省 MEXT

国土交通省 Ministry of Land, Infrastructure, Transport and Tourism MLIT

内閣官房 Cabinet Secretariat

ひと、くらし、みらいのために 厚生労働省 Ministry of Health, Labour and Welfare MHLW

Kobe City

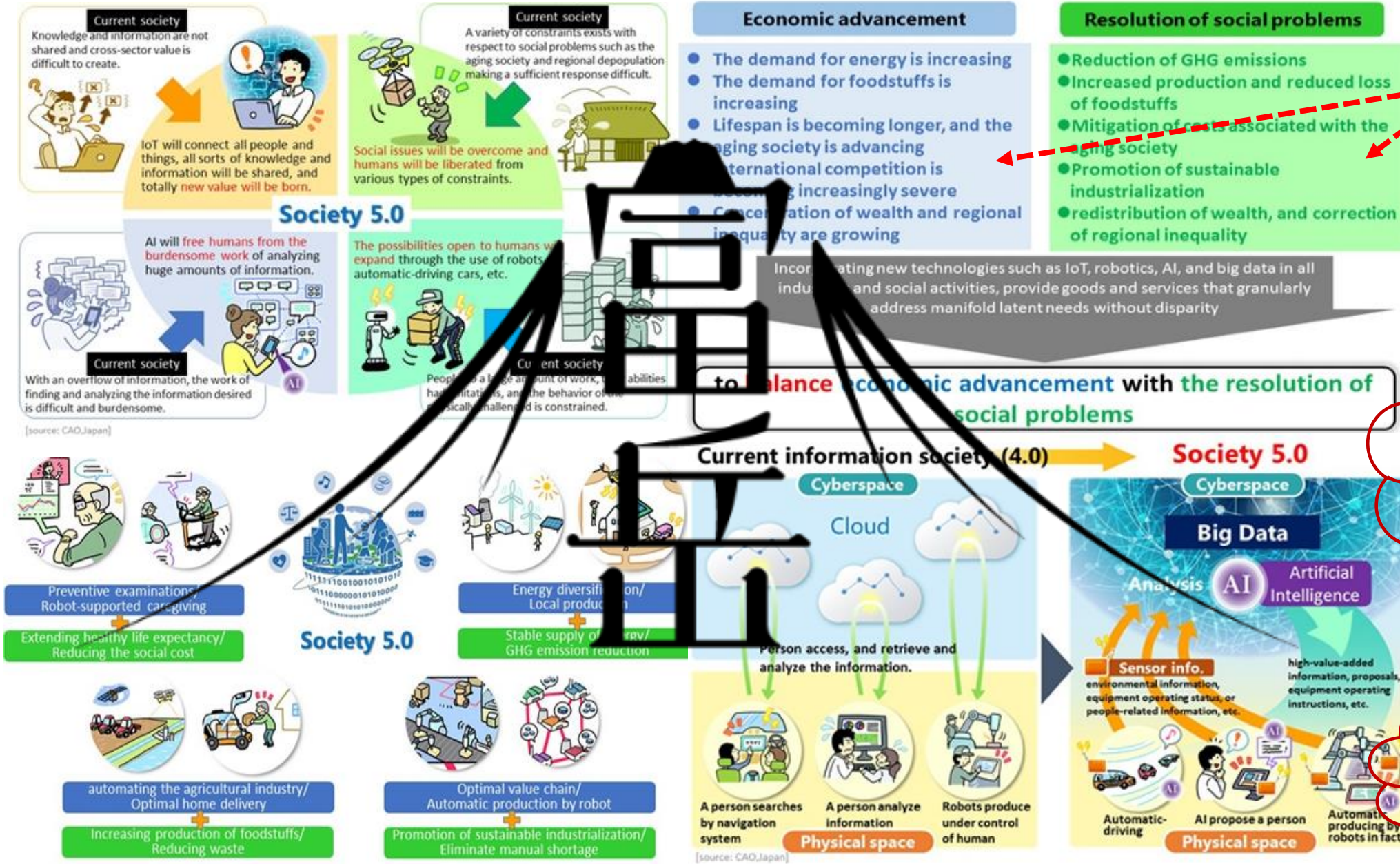
Work with various COVID19 Govt. committees, funding => immediate application to government COVID19 mitigation advisories and policies

Timey Simulations and Media Dissemination

- We have been staging multiple press conferences on the latest research results
- Extremely high interest from the media, with immediate national news coverage
- Most people in Japan have seen the Fugaku COVID19 news, esp. droplet simulation, with high trust in being scientifically grounded
- Visualization extremely effective in raising public understanding & awareness of COVID19 & its mitigation
- Prime Minister Suga holds a press conference 22 Nov., urging everyone to wear masks even during group dining, as “it’s effectiveness has been proven by a supercomputer (Fugaku)”.

Fugaku as the Centerpiece of Society 5.0

(background slides: https://www8.cao.go.jp/cstp/english/society5_0/index.html)



Same area as Fugaku Apps Co-Design Priority Areas

"Digital Twin in Cyberspace" is what traditional simulation is all about

Fugaku is No.1 in Simulation, Big Data & AI

Where do we go from here? Five pillars of research at R-CCS for future system & Society5.0

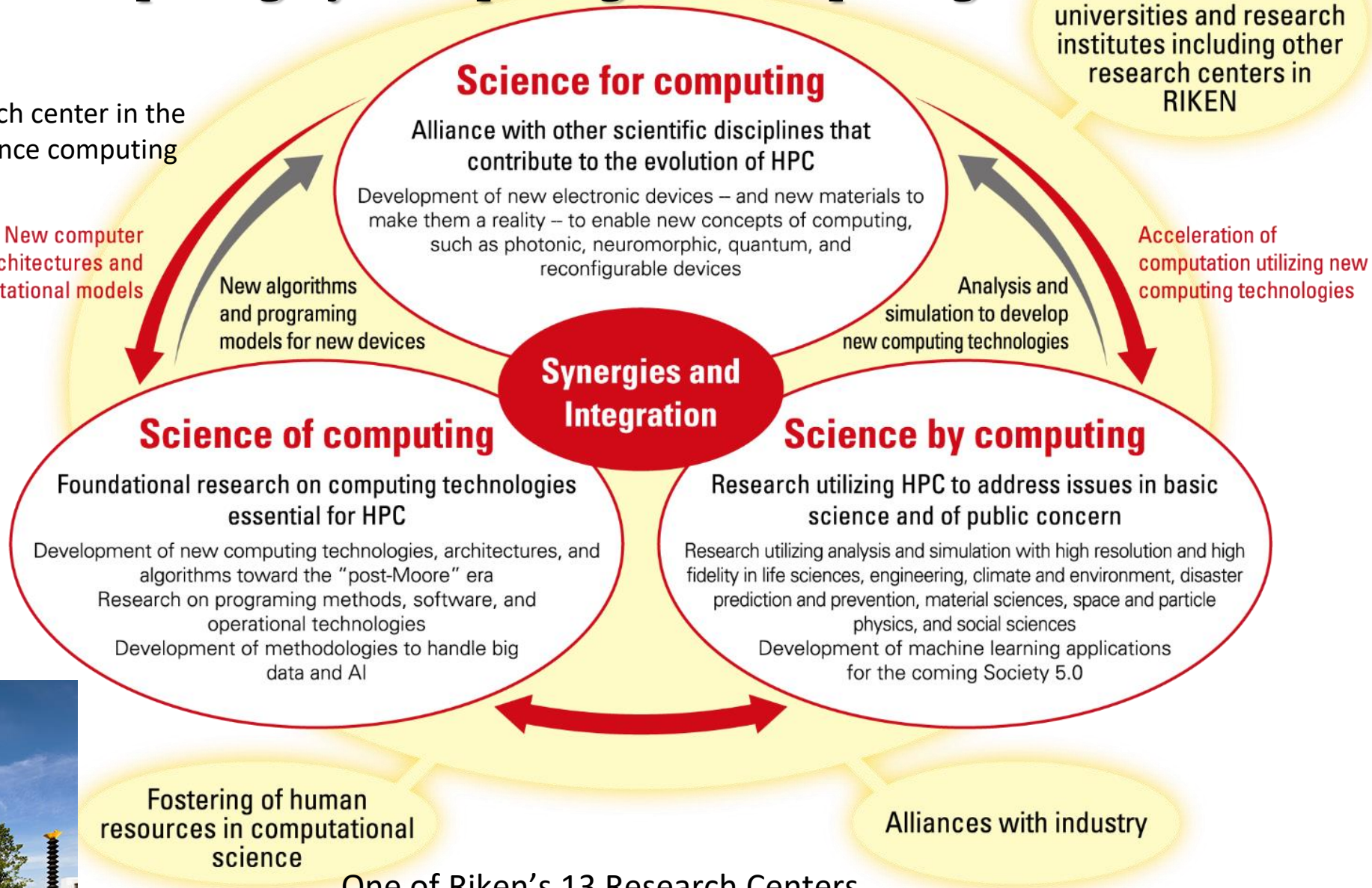
1. **Further S&T grand challenges R&D + future leadership IT**
2. **Driving "Simulations First" Society 5.0**
3. **The science of convergence of 'first-principle simulations', 'empirical AI methods', and 'big data instrumentation' on large scale HPC systems**
    1. 'Ad-hoc' integration => foundational 'science of computing'
4. **Broadening of workload analysis and increased generality of HPC to broad IT ecosystem**
    1. HPC technology fundamental to future IT, from IoT to Clouds
    2. Benchmarking & analyzing such workloads for acceleration
5. **Platform to investigate new computing paradigm**
    1. Large scale simulation of quantum, neuromorphic, …

Science of Computing by Computing for Computing

Riken R-CCS

International core research center in the science of high performance computing (HPC)

21 research teams +
5 ops units
(more to come)



One of Riken's 13 Research Centers
For details https://www.r-ccs.riken.jp/en/

R-CCS R&D Teams&Units 2021=>2022

Science of Computing = Computer Science

Area Lead
Mitsuhisa Sato
(Deputy Director)

Programming Environment
M. Sato (Area Lead)

High Performance Big Data Systems
K. Sato

Performance Modeling & Instrumentation
=> New PI public call

Advanced Processor Architectures
K. Sano

Next Gen High Performance Architecture
M. Kondo

Digital Twin & Smart Cities
=> New PI public call

Parallel Numerical Technology
T. Imamura

High Performance AI Systems
S. Matsuoka
=> New PI public call

Science by Computing = Application Sciences

Area Lead
Kengo Nakajima
(Deputy Director)

Discrete Event Simulation
N. Ito

Climate Science
H. Tomita

Disaster Mitigation & Reduction
S. Oishi

Molecular Science
T. Nakajima

HPC Engineering Applications
M. Tsubokura

Data Assimilation
T. Miyoshi

Quantum Physics
S. Yunoki

Field Theory
Y. Aoki

Structural Biology
F. Tama

Biophysics
Y. Sugita

New/Merger Apr. 1, 2021
HPC- and AI-driven Drug Development Platform Division

Biomedical Computational Intelligence
Yasushi Okuno
(Division Leader)

Medicinal Chemistry Applied AI
Teruki Honda

Molecular Design Computational Intelligence
Mitsunori Ikeguchi

AI-driven Drug Discovery Collaborative
Yasushi Okuno

Operational and Computer Technologies Division(& research)

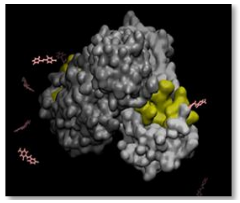
Facility Operations and Development
T. Tsukamoto

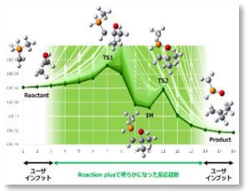
System Operations & Development
A. Uno

(new) Software Development Technology Unit
F. Shoji
(Interim)

HPC Usability Development
F. Shoji
(Division Lead)

Advanced Operations Technologies
K. Yamamoto



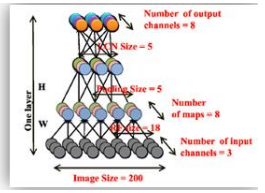


First Principle Based Simulation

Training Data Synthesis

Acceleration via Surrogates, Paretho optimization

Empirical AI-Based Prediction



Trajectory Correction

Inter/Extra polation

Data Assimilation

Big Data Instrumentation

Prediction & Actuation

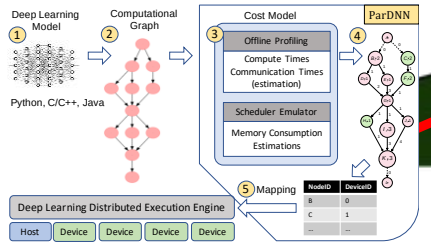
Live Data Training

What are the fundamental theories that define such relationships between compute, intelligence, and data?

What will be the underlying system (HW&SW) that will facilitate such convergence effectively?

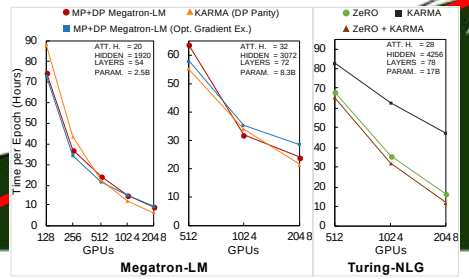
HPC (&Cloud) Infrastructure

Exploring and Merging Different Routes to O(100,000s) Nodes Deep Learning



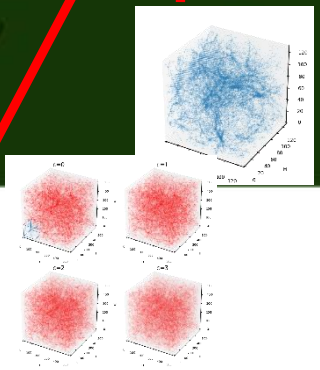
Non-intrusive graph-based partitioning strategy for large DNN models achieving superlinear scaling [1]

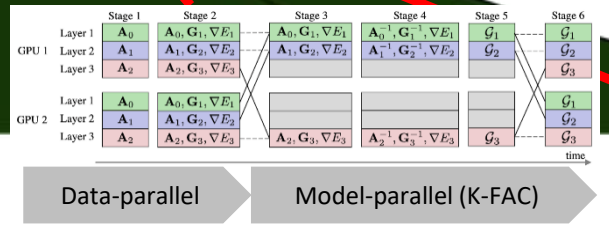
AIST, Koc U.



KARMA: Out-of-core distributed training (pure data-parallel) outperforming SoTA NLP models on 2K GPUs [2]

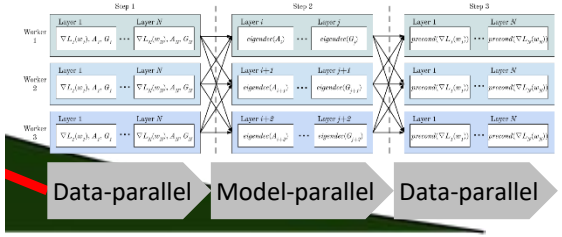
AIST, Matsuoka-lab, RIKEN





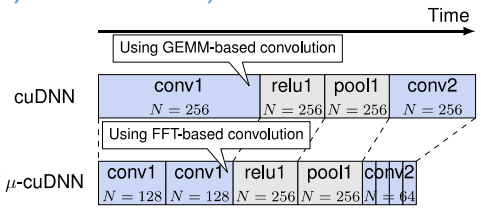
A model-parallel **2nd-order method (K-FAC)** trains ResNet-50 on **1K GPUs** in 10 minutes [4]

TokyoTech, NVIDIA, RIKEN, AIST



Layer-wise distribution and inverse-free design further accelerate K-FAC [5]

UT Austin, UChicago, ANL

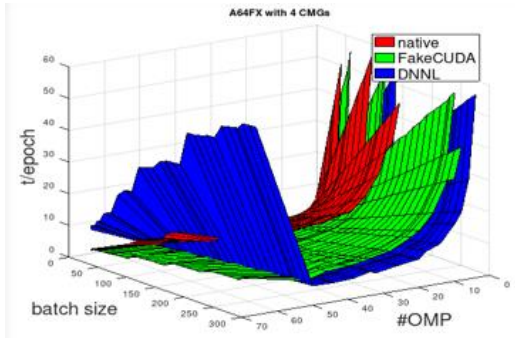


Model-parallelism enables 3D CNN training on **2K GPUs** with 64x larger spatial size and better convergence [3]

Matsuoka-lab, LLNL, LBL, RIKEN

Layer-wise loop splitting accelerates CNNs [6]

Matsuoka-lab, ETH Zurich



MocCUDA: Porting CUDA-based Deep Neural Network Library to A64FX and (other CPU arch.)

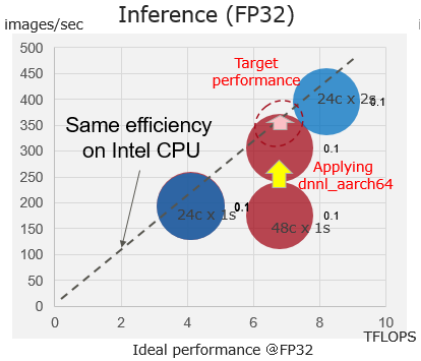
RIKEN, Matsuoka-lab, AIST

Engineering for Performance Foundation

Merging Theory and Practice

Porting High Performance CPU-based Deep Neural Network Library (DNNL) to A64FX chip

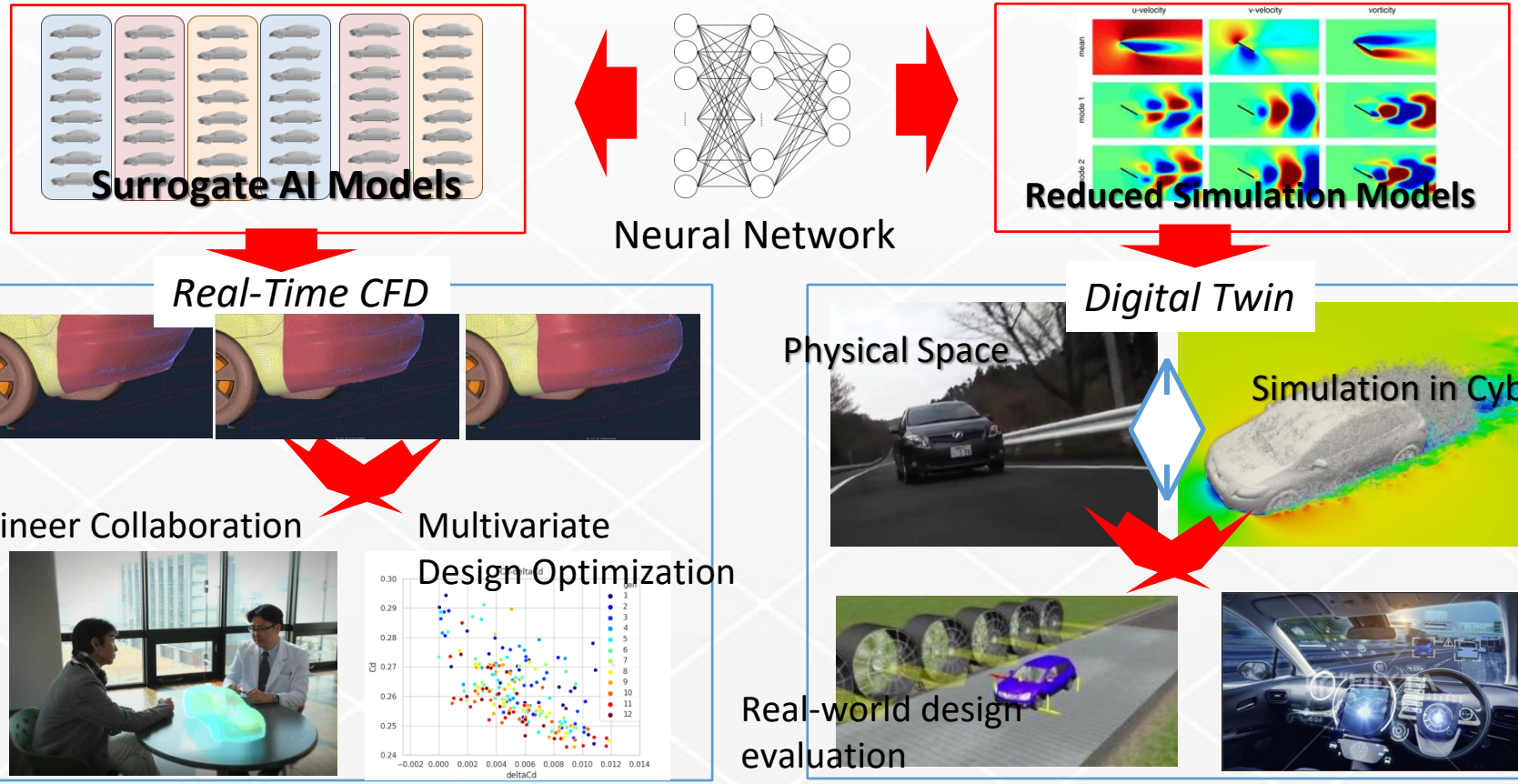
Fujitsu, RIKEN, ARM



[1] M. Fareed et al., "A Computational-Graph Partitioning Method for Training Memory-Constrained DNNs", Submitted to PPoPP21
[2] M. Wahib et al., "Scaling Distributed Deep Learning Workloads beyond the Memory Capacity with KARMA", ACM/IEEE SC20 (Supercomputing 2020)
[3] Y. Oyama et al., "The Case for Strong Scaling in Deep Learning: Training Large 3D CNNs with Hybrid Parallelism," arXiv e-prints, pp. 1–12, 2020.
[4] K. Osawa, et al., "Large-scale distributed second-order optimization using kronecker-factored approximate curvature for deep convolutional neural networks," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2019-June, pp. 12351–12359, 2019.
[5] J. G. Pauloski, Z. Zhang, L. Huang, W. Xu, and I. T. Foster, "Convolutional Neural Network Training with Distributed K-FAC," arXiv e-prints, pp. 1-11, 2020.
[6] Y. Oyama et al., "Accelerating Deep Learning Frameworks with Micro-Batches," Proc. IEEE Int. Conf. Clust. Comput. ICCC, vol. 2018-September, pp. 402–412, 2018.

HPC and AI Convergence for Society 5.0 Manufacturing [Tsubokura et. al., R-CCS]

- **Combining ML/Deep Learning, Data Assimilation, Multivariate Optimization with Simulation for new generation manufacturing**
- **Use output of high-resolution simulation data to train AI**
  - Construct AI surrogate model training on simulation data, allowing real-time CFD to facilitate designer-engineer collaboration, multivariate design optimization, etc.
  - Use NN to derive reduced simulation model, allowing digital twin in cyberspace corresponding to entities in physical space for real-time interactions



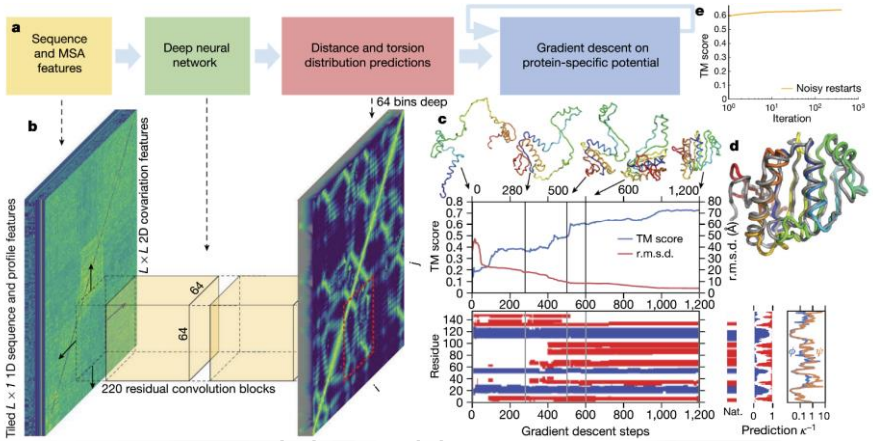
Example: Intelligent Genomics/Proteomics Drug Design based on Convergence

[New HPC/AI pharma division @ R-CCS]

Genomic Sequencers & Federated DB

Genomic Sequence

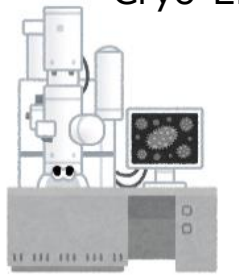
CNN (AI) × Simulation



Alpha-Fold

Feedback & Train

Cryo-EM



3d Structure

Structural Variations

Dynamics

DEFMap

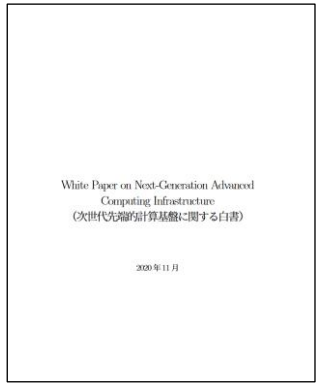
HPC (&Cloud) Infrastructure

Performance projection of many-core CPU systems based on IRDS roadmap

Predictions based on the IRDS Roadmap(2020 ed.), extrapolation of traditional many core architectures relying merely on advances of semiconductor technologies will achieve only 1.8EFLOPS Peak (3.37x c.f. Fugaku), if a machine with broad applicability will be built

- Methodoogies(CPU part): Assumptions from IRDS Roadmap Systems and Architectures
  - Cores/socket=70 cores
  - SIMD width=2048-bit x 2
  - Clock frequency=3.9GHz
  - Socket TDP = 351W
- System assumptions
  - System Power=30, 40, 50MW
  - PUE=1.1
  - CPU power occupy=60,70,80%

https://sites.google.com/view/ngaci/home



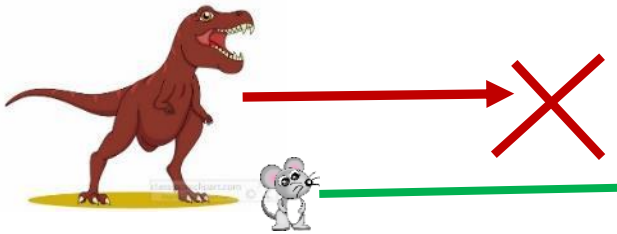
NGACI white paper

From NGACI white paper

	30MW			40MW			50MW		
	60%	70%	80%	60%	70%	80%	60%	70%	80%
Socket Cores	46620	54390	62160	62160	72520	82880	77700	90650	103600
	3.3×10^6	3.8×10^6	4.4×10^6	4.4×10^6	5.1×10^6	5.8×10^6	5.4×10^6	6.3×10^6	7.3×10^6
	815	950	1086	1086	1267	1448	1358	1584	1810
DDR 総 BW (PB/s)	102	120	137	137	160	182	171	200	228
HBM 総 BW (PB/s)	307	358	410	410	478	547	512	598	683
DDR 総容量 (PB)	17	20	23	23	27	31	29	34	39
HBM 総容量 (PB)	4	5	5	5	6	7	7	8	9
Injection BW(Tb/s)	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6
総I/O 性能 (TB/s)	34	34	34	34	34	34	34	34	34
Storage (EBytes)	3.45	3.45	3.45	3.45	3.45	3.45	3.45	3.45	3.45

最もアグレッシブなシステム構成（50MW電力バジェット、CPUで80%電力消費）においても1.8EF程度の性能と予測

Many Core Era



Post Moore
Cambrian Era

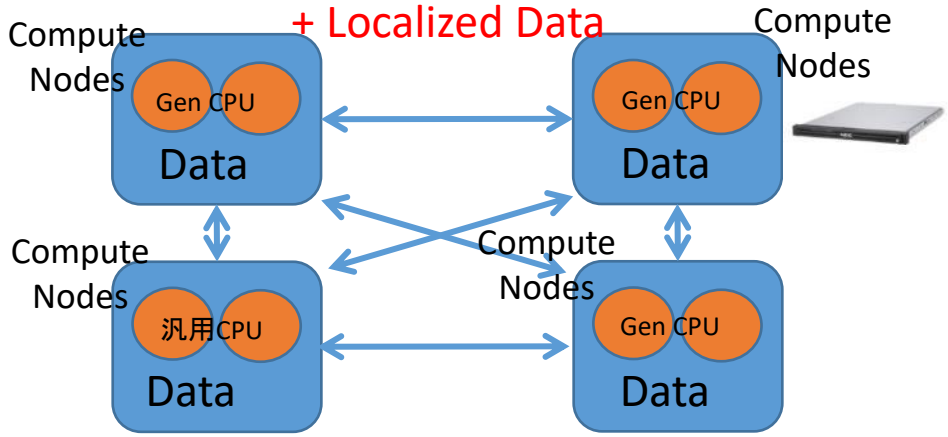
Flops-Centric Monolithic Algorithms and Apps

Flops-Centric Monolithic System Software

Hardware/Software System APIs
Flops-Centric Massively Parallel Architecture

~2025
M-P Extinction
Event

Homogeneous General Purpose Nodes
+ Localized Data



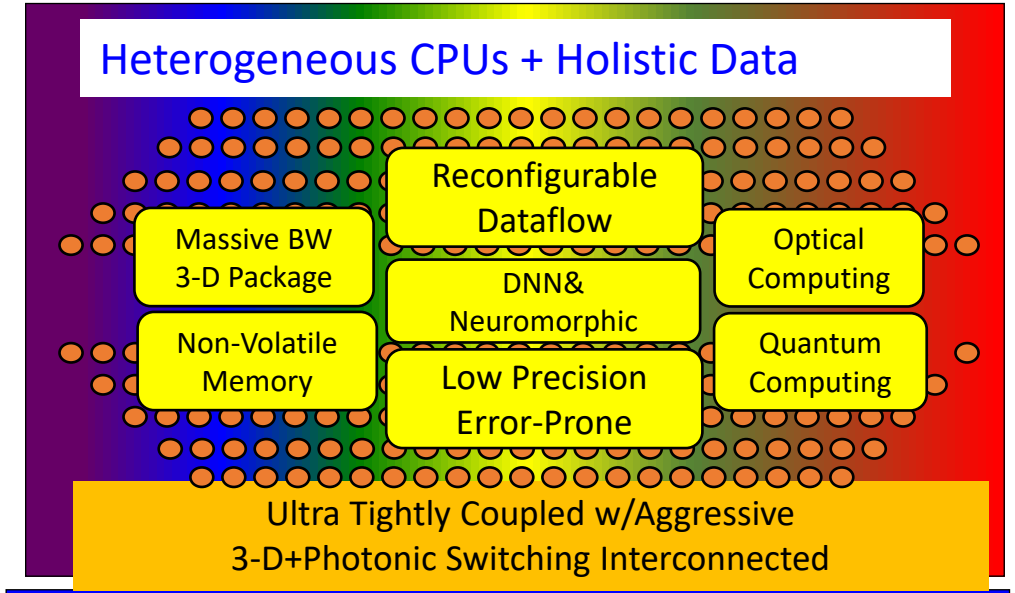
Loosely Coupled with Electronic Interconnect

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

Cambrian Heterogeneous Algorithms and Apps

Cambrian Heterogeneous System Software

Hardware/Software System APIs
"Cambrian" Heterogeneous Architecture

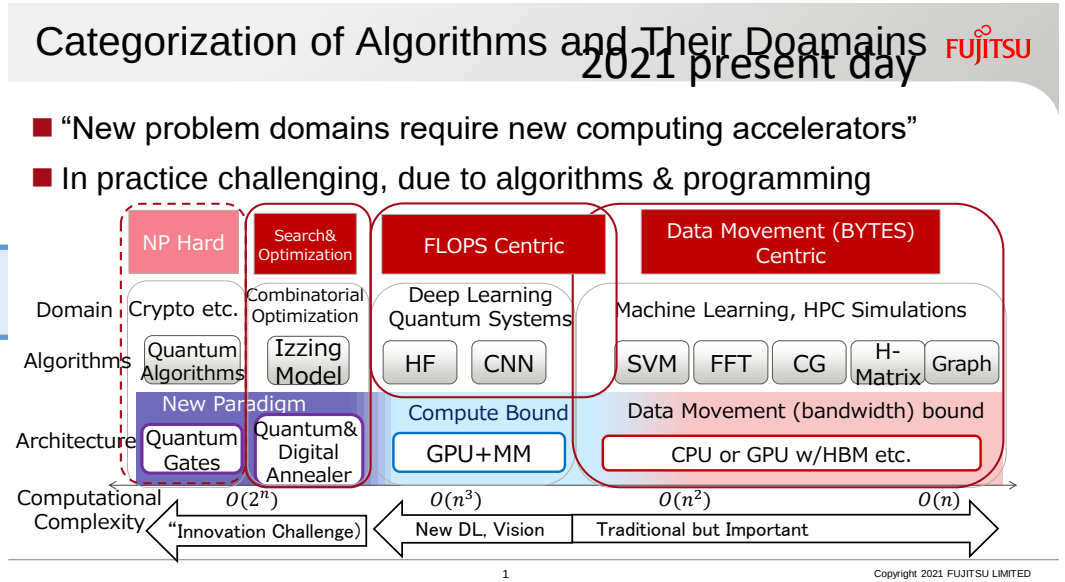


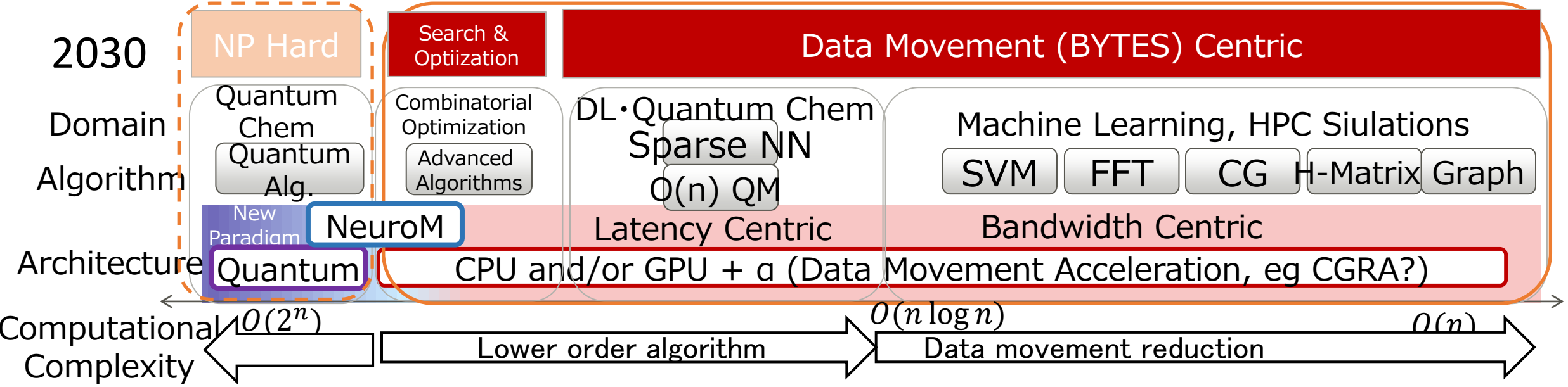
Novel Devices + CMOS (Dark Silicon)
(Nanophotonics, Non-Volatile Devices etc.)

Post-Moore Algorithmic Development

- **Towards 2030 Post-Moore era**

- End of ALU compute (FLOPS) advance
- Disrupritve reduction in data movement cost with new devices, packaging
- Algorithm advances to reduce the computational order (+ more reliance on data movement)
- Unification of BD/AI/Simulation towards data-centric view





Our Project: Exploring versatile HPC architecture and system software technologies to achieve 100x performance by 2028

Problems to be solved and goals to be achieved

- General-purpose computer architectures that will accelerate a wide range of applications in the post-Moore era have not yet been established.
- What is a feasible approach for versatile HPC systems based on bandwidth improvement?
- **Goal:** to explore architectures that can achieve 100x performance in a wide range of applications around 2028

Approaches and subtasks

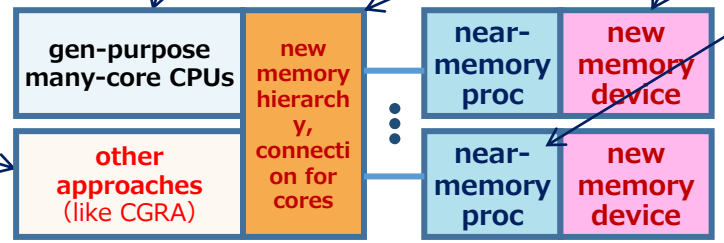
- Exploration of future CPU node architectures and necessary technologies

Subtask1.1 Performance characterization and modeling with benchmarks to identify directions for exploration and improvement (Riken R-CCS)

Subtask2 Exploring innovative memory architectures with ultra-deep and ultra-wide bandwidth (Tokyo Tech.)

Subtask1.2 Exploring a reconfigurable vector data-flow architecture (CGRA) that can exploit increased data transfer capability (Rijen R-CCS)

Subtask3 Exploring near-memory computing for highly effective bandwidth and cooling efficiency for general purpose computing (U-Tokyo)



目標とするノードアーキテクチャの例

Planned

Subtask4 Exploration of node architectures as extension of existing many-core CPUs with non von-Neumann methods (unnamed company)

Plan