

### The New Era of Hybrid-Computing: Vector-Scalar to Vector-Digital Annealing, to Vector-Quantum Annealing

#### Hiroaki Kobayashi

#### **Tohoku University**

Professor of Graduate School of Information Sciences Division Director of the NEC-Joint Research Lab on HPC Division Director of Tough Cyberphysical AI Research Center Special Advisor to President of Tohoku University for Digital Innovation Special Adviser to Director of Cyberscience Center for the HPC Strategy

> koba@tohoku.ac.jp Russian Supercomputing Days September 27, 2021



### Today's Agenda

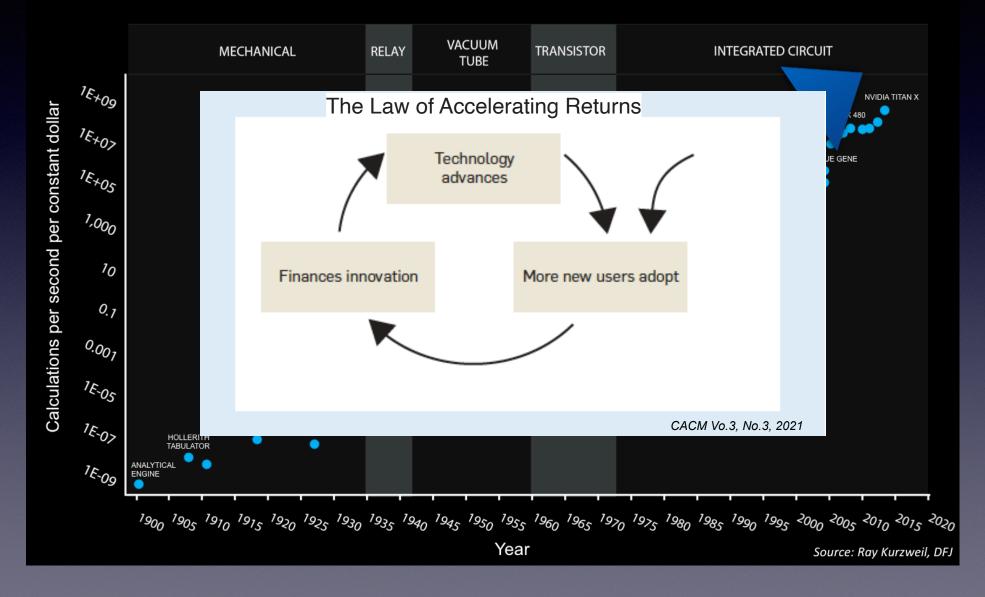
- Background: Shifting from Pursuing General Pursue Monolithic Systems to Heterogeneous Computing Systems with a Wide-Variety of Domain-Specific Architectures (DSAs)
  - ✓ Facing the End of Moore's Law?
  - Maximize Sustained Performance and Power efficiency to select right architectures for different kernels/applications
  - ✓ Orchestrating different DSAs to realize general-purpose functionality, not depending only on a single architecture!

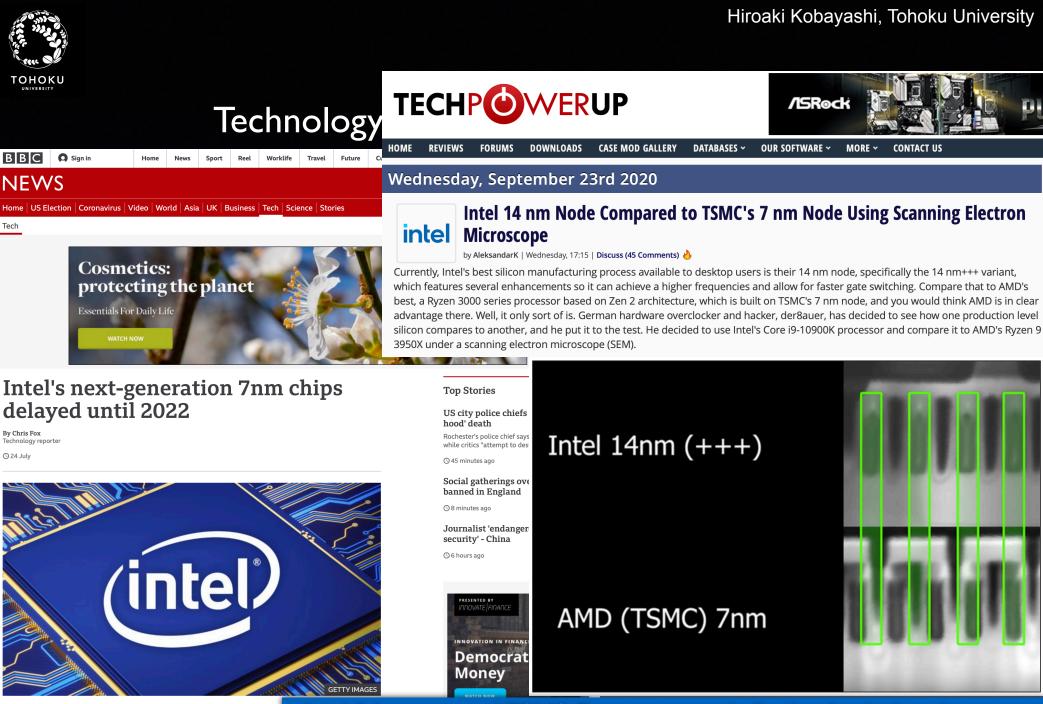
#### R&D of a Quantum Annealing-Classical HPC Hybrid Computing

- Realize a Vector-Scalar and Quantum-Annealing Hybrid Simulation and Data Analysis Environment as a mix of DSAs
- Provide a transparent interface to deductive and inductive computing platforms over the vector-scalar and quantum-annealing hybrid
- ✓ Showcase application design and implementation of Data Analysis assisted by Digital and Quantum Annealing



# 120 Years of Moore's Law?





★Intel 14nm+++ and TSMC 7nm are very similar in physical scale ★Intel 10 nm and TSMC 7nm processes both produce dies with approx 90 million transistors per sq millimeter.

Tech



 $\hat{\mathbf{x}}$ 

and home appliances.

## No More Moore's Law, No More One-Fits-All?!

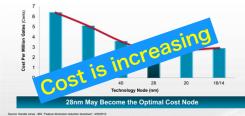
- We are facing the end of Moore's law due to the physical limitations for miniaturization of transistors, and at the same time, the manufacturing cost is hard to reduce gradually,
  - / Tech. is slowing, cost is increasing, and efficiency is lowering!

Silicon is still fundamental constructing material for computing

platforms just like plastic, steel and concrete for automobiles, buildings





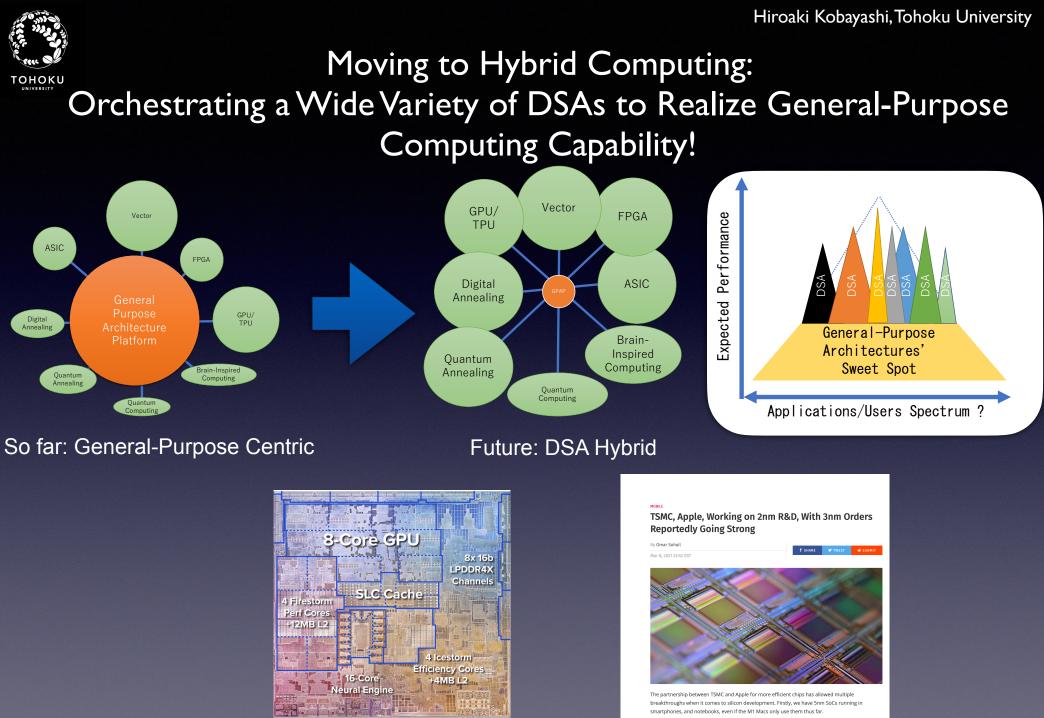


EE Times: http://www.eetimes.com/discussion/other/4238315/Feature-dimension-reduction-slow 975 feature-layerine. All spin served.

Use precious silicon budget (+ advanced device technologies) to effectively design mechanisms that can maximize the sustained performance of individual applications.

Daula	N	0	Rmax	Rpeak	Rmax/
Rank	Name	Cores	(Tflop/s)	(Tflop/s)	Peak
1	Fugaku	7,299,072	415,530.00	513,854.70	80.87
2	Summit	2,414,592	148,600.00	200,794.90	74.01
3	Sierra	1,572,480	94,640.00	125,712.00	75.28
	Sunway	10,649,600	93,014.60	125,435.90	80.87
4	TaihuLi				
	ght				
		1 0 0 1 7 0 0	01 111 50	100 070 70	61.02

It's time to focus on Domain-Specific Architectures for computation-intensive, memory-intensive, I/ O intensive, mixed-precision computing... etc applications to improve silicon/power efficiency, and their orchestration to satisfy the requirements from a wide variety of applications is required! RSD2021



66

Apple M1 SoC Source: Apple

RSD2021

September 27, 2021

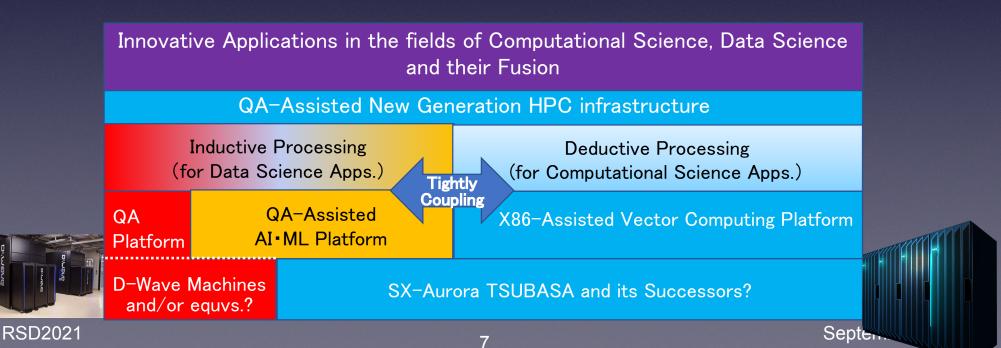
However, phone makers using 3nm chipsets will become a commonplace and shortly after, 2nm parts. To make this transition possible, a new report claims that both TSMC and Apple have teamed up to drive chip development, but we doubt we will see any form of mass production in a few

years



### Toward Realization of Quantum Classical-HPC Hybrid Infrastructure

- ★ Tohoku University has established an interdisciplinary priority research institute, named Q-HPC, for Exploring Quantum Computing-Classical HPC Hybrid, in 2018
- ★ We start a new 5-year research program named "R&D of Quantum Annealing-Assisted HPC Infrastructure", supported by MEXT, in collaboration with NEC and D-wave sys.
  - ✓ provides transparent accesses to not only classical HPC resources but also Quantum Computing one in a unified fashion.
  - Becomes an innovative infrastructure to develop next-generation applications in the fields of computational science, data sciences and their fusions



тоноки

**RSD2021** 

### Quantum Computer: Emerging Domain Specific Architecture

★ Quantum computing is drawing much attention recently as an emerging technology in the era of post-Moore

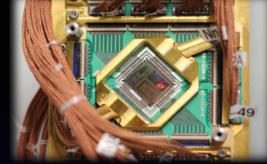
✓ In particular, quantum annealing machines are commercialized by the D-wave systems, and their applications are developed world-widely.

✓ Google, NASA, Volkswagen, Lockheed, Denso...

- The base model named the Ising model to design and implement the D-wave machines has been proposed by Prof. Nishimori et al of Tokyo Inst. Tech. In 1998.
- ★ The quantum annealing is a metaheuristic for finding the global minimum of a given objective function over a given set of candidate solutions (candidate states), by a physical process named quantum fluctuations

### An ideal solver for combinatorial problems!

Hiroaki Kobayashi, Tohoku University





Source by D-Wave Sys.

Transverse magnetic field type quantum annealing Chip and System (D-Wave)

**Optimal solution** 

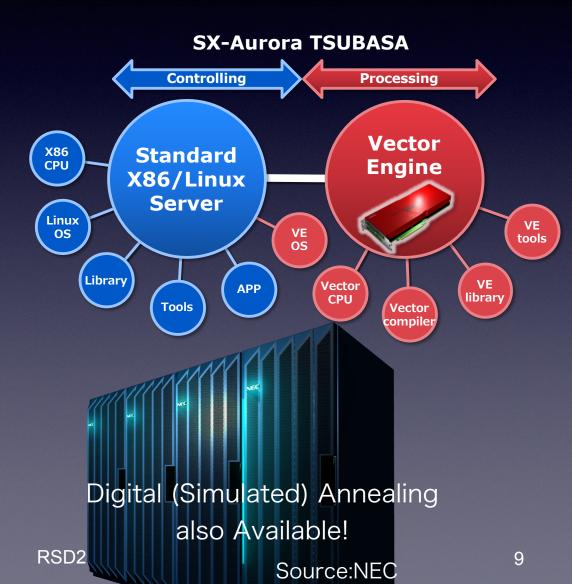


reach optimal one by Quantum Fluctuation September 27, 2021



### Why Vector System: SX-Aurora-TSUBASA? ~Pursuing Balanced Architectures for High Sustained Performance~

Two types of balancing: computing performance and memory performance, and standardization and customization



- Customization for realization of the balanced vector architecture for memoryintensive apps
  - ✓ Highest Mem. BW
    - Largest Single Core Performance
- ★ Standardization for realization of the user-friendly environment and control-intensive apps.
  - 🗸 x86 Linux Environment
    - New execution model centralized on vector computing

September 27, 2021

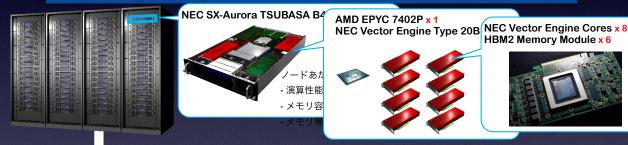


### New Supercomputer System at Tohoku University

#### 🖈 Start Servicing in Oct. 2020

- Peak Performance of I.8Pflop/s
- 20+ x performance-enhanced in 2022-23

#### Vector Supercomputer SX-Aurora TSUBASA (2nd Gen VEs)





## HPCI for Nation-wide service





#### Interconnect Fabric (InfiniBand HDR)



RSD2021 X86 Cluster System(AMD EPYC 7720)

Hiroaki Kobayashi, Tohoku University



# Hardware Specification of SX-Aurora TSUBASA (2nd Gen in 2020)

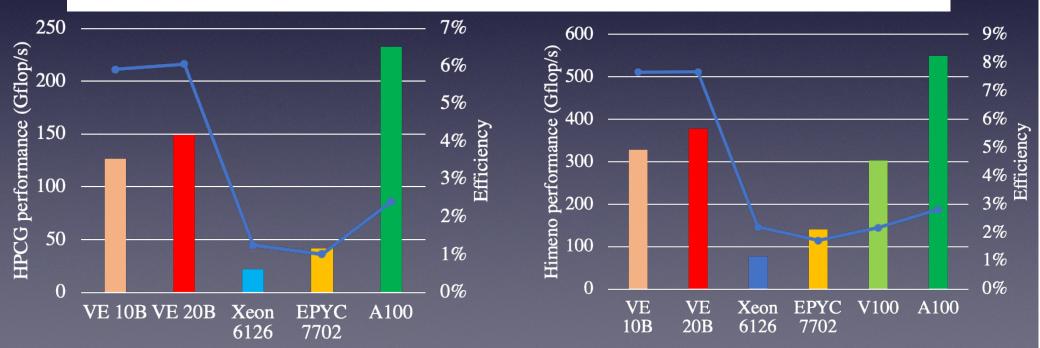
S	SX Vector Processor				
SX-Aurora TSUBASA A300-2 #00001 HBM2 VE processor	BM2 GB Core 307.2 GF Core 307.2 GF Core 307.2 GF SM2 GB CORE 307.2 GF CORE 307.2 GF CORE 307.2 GF SM2 GB CORE 307.2 GF SM2 GB CORE 307.2 GF SM2 GB SM2 GB SM2 GB SM2 GB SM2 GB SM2 GB SM2 GB SM2 SM2 SM2 SM2 SM2 SM2 SM2 SM2 SM2 SM2	12      2x UPI x20      PCIe* x6      PCIe        12      PCIe* x6      PCIe* x6      PCIe* x6      PCIe* x6        12      CHAYSF/LLC      CHAYSF/LLC      CHAYSF/LLC      CHAYSF/LLC        12      DDR4      NS      CHAYSF/LLC      CHAYSF/LLC        12      DDR4      NS      CHAYSF/LLC      CHAYSF/LLC	AMA  Image: CHAYSF/LLC    FILIC  CHAYSF/LLC    SKX Core  SKX Core    SKX Core  SKX Core    CHAYSF/LLC  CHAYSF/LLC		
н	Core Core 307.2 GF 307.2 GF BM2 GB Core Core 307.2 GF 807.2 GF 8 GF	12	F/LIC  CHAYSF/LLC  CHAYSF/LLC    Core  SKX Core  SKX Core    F/LIC  CHAYSF/LLC  SKX Core    F/LIC  CHAYSF/LLC  CHAYSF/LLC		
Vector Engine (VE)	Туре 20В	Vector Engine (VE)	Туре 10В		
Frequency	1.6 GHz	Frequency	1.4GHz		
Performance / Core	614 GF (SP), 307 GF (DP)	Performance / Core	537.6 GF (SP), 268.8 GF (DP)		
# cores	8 14%↑	# cores	8		
Performance / socket	4.91 TF (SP) 2.45 TF (DP)	Performance / socket	4.30 TF (SP) 2.15 TF (DP)		
Memory Subsystem	HBM2 8 GB x6	Memory Subsystem	HBM2 8 GB x6		
Memory Subsystem Memory Bandwidth	HBM2 8 GB x6 1.53 TB/s	Memory Subsystem Memory Bandwidth	HBM2 8 GB x6 1.22 TB/s		



# Potentials of SX-Aurora TSUBASA

### **Sustained Performance**

Year	2017	2020	2018	2020	2017	2020
	<b>VE 10B</b>	<b>VE 20B</b>	Xeon 6126	EPYC 7702	V100	A100
Number of cores	8	8	12	64	5120	6912
Peak SP (Tflop/s)	4.30	4.92	1.766	4.096	14	19.5
Peak DP (Tflop/s)	2.15	2.46	0.883	2.048	7	9.7
Memory	$6 \times \text{HBM2}$	$6 \times \text{HBM2E}$	$6 \times DDR4$	$8 \times DDR4$	$4 \times \text{HBM2}$	$6 \times \text{HBM2E}$
Mem. BW (GB/s)	1228	1536	128	204.8	900	1555
Mem. Cap. (GB)	48	48	192	256	32	40
LLC BW (TB/s)	2.66	3.00	-	-	2.70	6.88
LLC Cap. (MB)	16	16	19.25	256	6	40

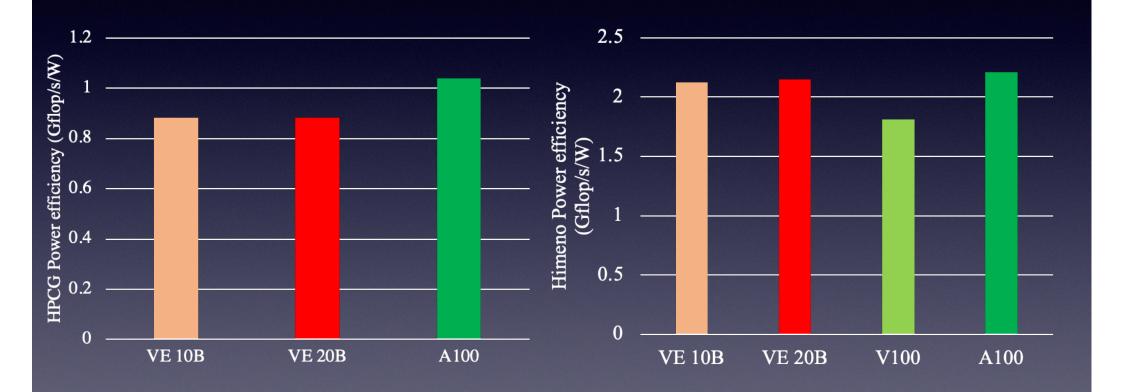


RSD2021

Hiroaki Kobayashi, Tohoku University

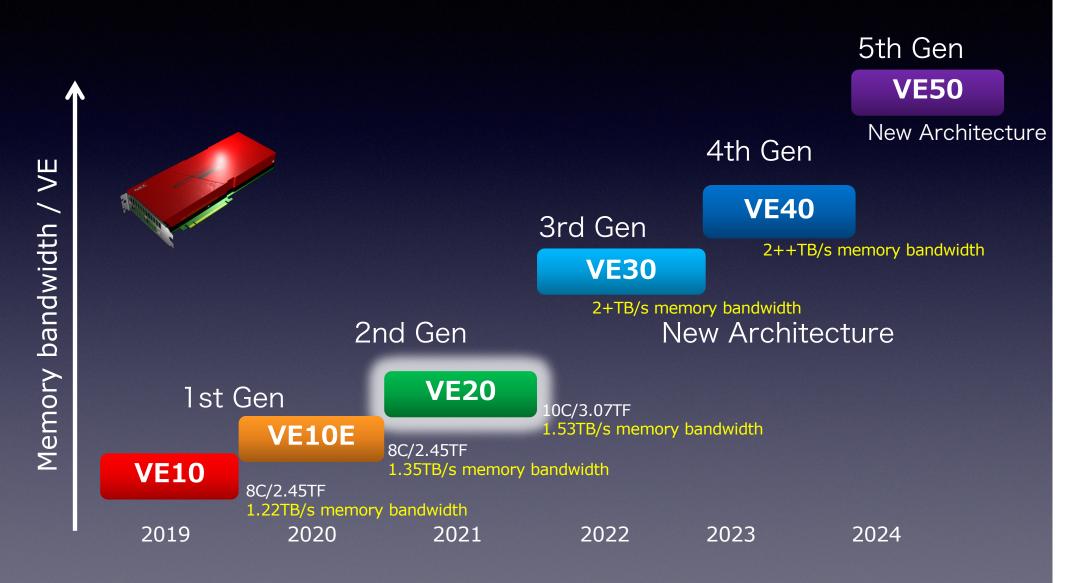


#### Potentials of SX-Aurora TSUBASA Power Efficiency



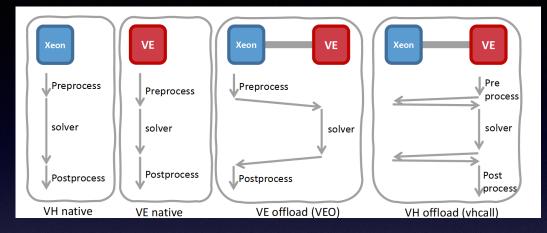


### SX-Aurora TSUBASA Tick-Tock Roadmap

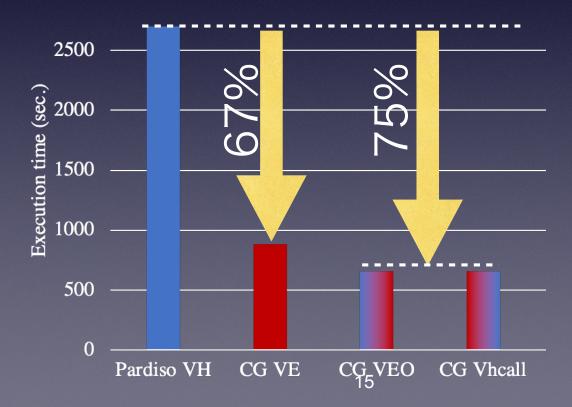




### Potentials of SX-Aurora TSUBASA VE-VH Hybrid (Offloading)



3000



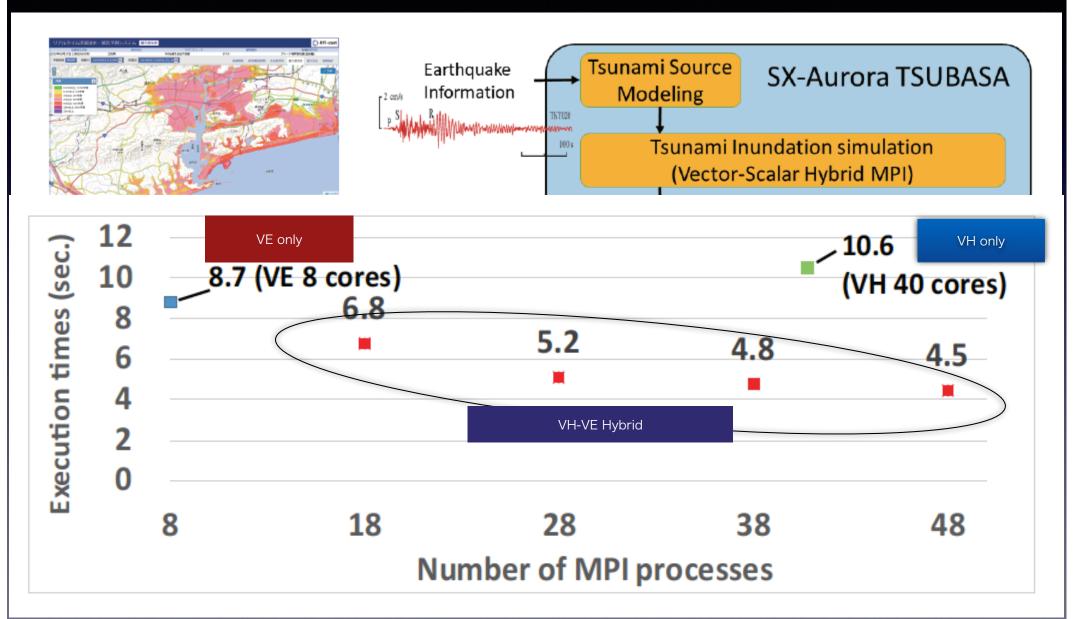
September 27, 2021

**RSD2021** 

Hiroaki Kobayashi, Tohoku University



### Potentials of SX-Aurora TSUBASA VE-VH Hybrid (MPI-Hybrid)





### Special Issue on SX-Aurora TSUBASA in Supercomputing Frontiers and Innovations 2021, Vol.8., No.2.

SUPERCOMPUTING FRONTIERS AND INNOVATIONS	
HOME ABOUT FOR AUTHORS CURRENT ISSUE ARCHIVES Home > Archives > Vol 8, No 2 (2021) VOL 8, NO 2 (2021)	JOURNAL CONTENT Search
DOI: 10.14529/jsfi2102	Search
The special issue on Advance Methods and Technologies on Vector Computing and Data-Processing Using NEC SX-Aurora TSUBASA.	By Issue By Author

#### **Invited Editors:**

- Prof. Hiroaki Kobayashi, Graduate School of Information Sciences, Tohoku University.
- Shintaro Momose, Manager of NEC Corporation, and Visiting Associate Prof. Tohoku University.

Published: 2021-09-14

#### FULL ISSUE

View or download the full issue

#### TABLE OF CONTENTS

ARTICLES	
Accelerating Seismic Redatuming Using Tile Low-Rank Approximations on NEC SX-Aurora TSUBASA	PDF
Yuxi Hong, Hatem Ltaief, Matteo Ravasi, Laurent Gatineau, David Keyes	6-26
Porting and Optimizing Molecular Docking onto the SX-Aurora TSUBASA Vector Computer	PDF
Leonardo Solis-Vasquez, Erich Focht, Andreas Koch	27-42
First Experience of Accelerating a Field-Induced Chiral Transition Simulation Using the SX-Aurora TSUBASA	PDF
Shinji Yoshida, Arata Endo, Hirono Kaneyasu, Susumu Date	43-58
Evaluating the Performance of OpenMP Offloading on the NEC SX-Aurora TSUBASA Vector Engine	PDF
Tim Cramer, Boris Kosmynin, Simon Moll, Manoel Römmer, Erich Focht, Matthias S. Müller	59-74
Performance and Power Analysis of a Vector Computing System	PDF
Kazuhiko Komatsu, Akito Onodera, Erich Focht, Soya Fujimoto, Yoko Isobe, Shintaro Momose, Masayuki Sato, Hiroaki Kobayashi	75-94
Distributed Graph Algorithms for Multiple Vector Engines of NEC SX-Aurora TSUBASA Systems	PDF
Ilya V. Afanasyev, Vadim V. Voevodin, Kazuhiko Komatsu, Hiroaki Kobayashi	95-113
Optimizing Load Balance in a Parallel CFD Code for a Large-scale Turbine Simulation on a Vector Supercomputer	PDF
Osamu Watanabe, Kazuhiko Komatsu, Masayuki Sato, Hiroaki Kobayashi	114-130



PDF



#### Target Applications Design and Implementation for QA/DA Hybrid on and with SX-Aurora TSUBASA



### Numerical Turbine: High Performance Turbine Simulator on SX Systems as a Digital Twin of a Real Turbine

Numerical Turbine developed by Prof. Yamamoto of Tohoku University

- is a simulation code realizing High-performance and High-reliable Future Turbines and
- has been accelerated on the SX series of Cyberscience Center at Tohoku University.



Gas turbine for plants



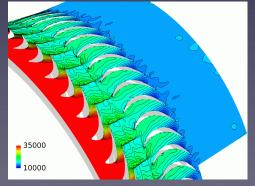
Gas turbine for airplanes



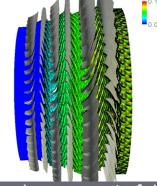
Steam turbine

#### Only Numerical Turbine has achieved the following simulations in the world.

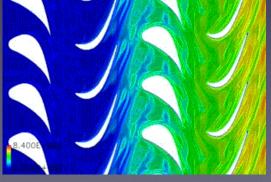
- Unsteady flows with wetness and shocks n actual gas turbines and steam turbines → Resolving such complex flows is crucial for developing high-performance and high-reliable turbines
- Full annulus (maru-goto) simulation for resolving unsteady wet-steam and moist-air flows in actual turbines and compressors



Unsteady shocks generated in turbine stages RSD2021



Unsteady wetness in full annulus turbine stages



Unsteady wet-steam flow in turbine stages

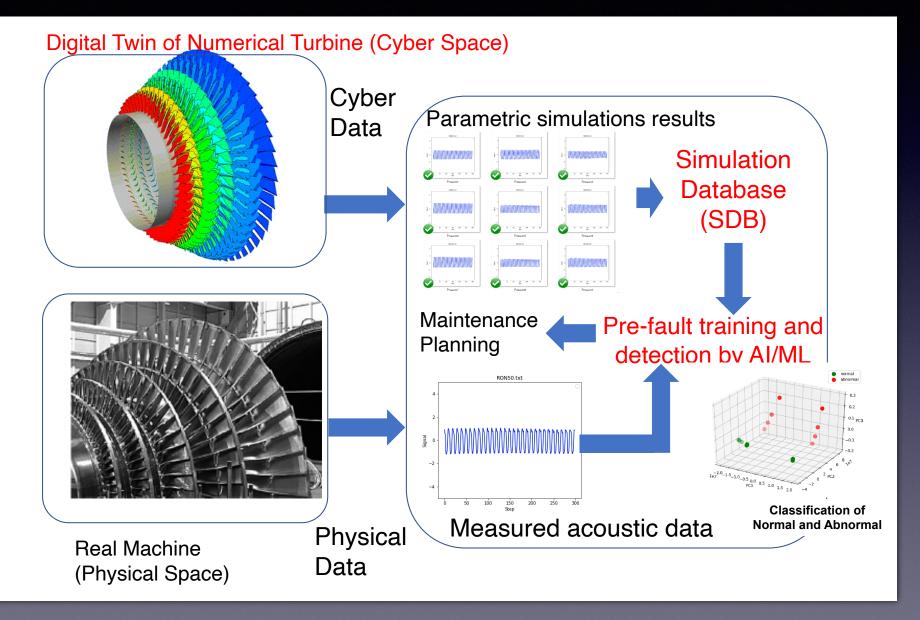
September 27, 2021

September 27, 2021



Simulation-Driven AI (Simulation-Data for AI)

## Target App I: Realization of a Digital Twin of a Real Turbine



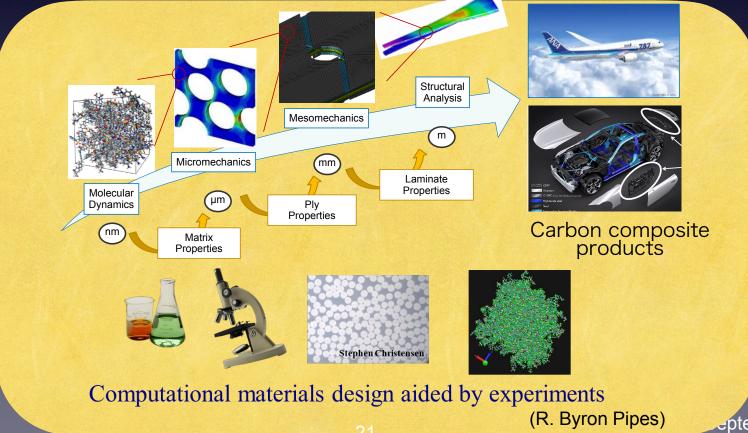


**RSD2021** 

### Target App. 2: Materials Integration System R&D

Network Polymer (thermoset polymer) is a key material for industrial products with carbon composite

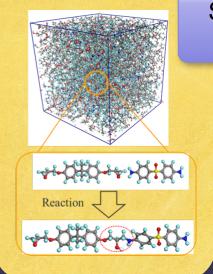
high deformation resistance and high durability to extreme environmental conditions
 Its design needs high performance simulation, starting from molecular level simulation, up to system level one, such as aircrafts, by using multiscale analysis combined with experiments
 Its also needs efficient identification of candidate materials that satisfy the required properties from a plenty of simulation results by using data science approaches



eptember 27, 2021



### Target App II : QA-Assisted Materials Integration System

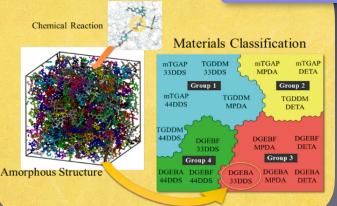


Simulation assisted by next-generation vector-type supercomputing



- More accurate and faster reaction model incorporated into MD simulation for cross-linked network formation in thermosetting resins
- Faster multi-scale simulation for predicting various thermo-mechanical properties

#### Quantum Annealing-assisted ML frameworks



- Hierarchical screening involving clustering approach
  Highly accurate machine learning model based on polymer physics
- ✓Inverse problem-based optimum design for screening of polymeric materials



eptember 27, 2021



### Combinatorial DA Clustering on SX-Aurora TSUBASA Hybrid Computing of QUBO Generation on VH and Digital Annealing on VE

#### ★ Data Clustering

- Well-know method to analyze data for data mining such as pattern recognition, image analysis, information retrieval and machine learning as well
- ✓ K-means is popular, but realizes a approximate clustering

#### ★ Combinatorial Clustering

- $\checkmark$  precise clustering by considering all the data distances
- Clustering is defined as a combinatorial optimization problem that can be accelerated by quantum and simulated(digital) annealing.
- ★ A newly-developed combinatorial clustering for SX-Aurora TSUBASA with digital annealing on VE
  - The one-hot constraint is given separately from the objective function to be minimized, but controlled independently to still keep the one-hot constraint situation.
    - A QUBO is used for definition of the objective function only to have enough precision to represent, resulting in higher clustering quality
  - An Efficient Hybrid Computing of QUBO generation and post-processing of clustering on VH and Digital Annealing on VE

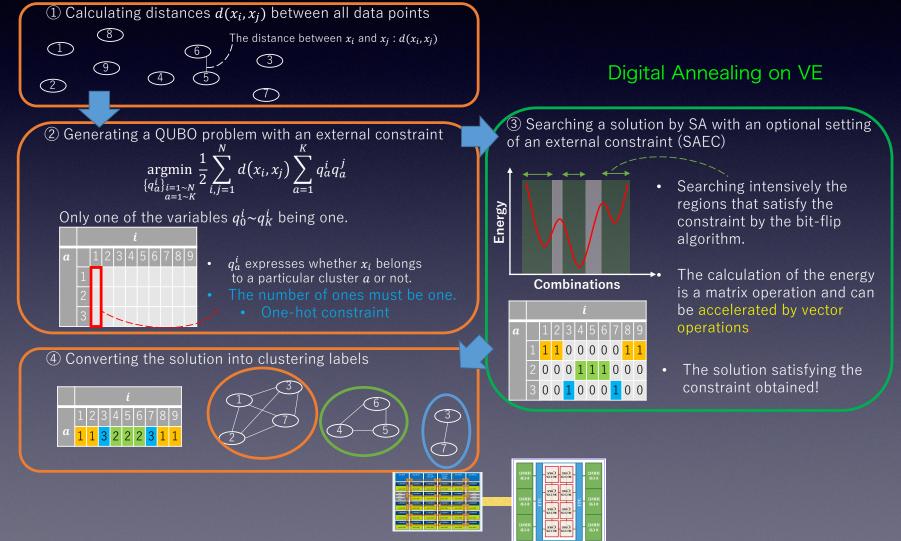


September 27, 2021



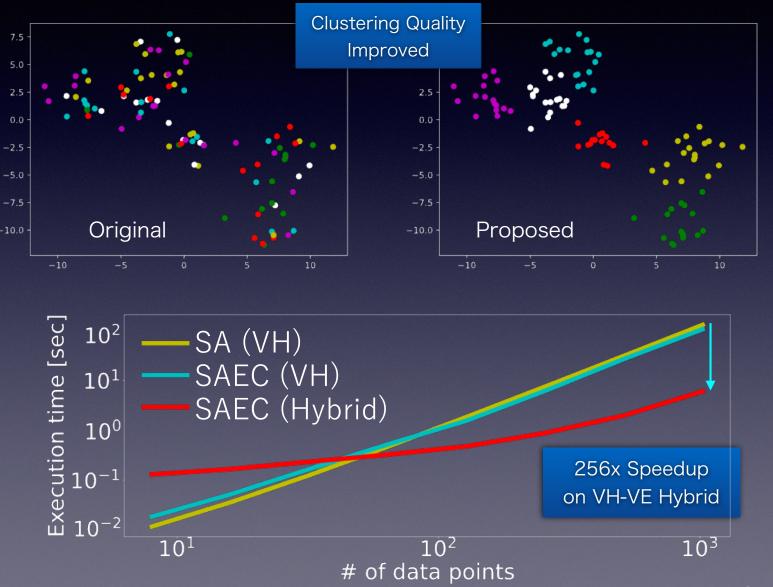
### Combinatorial DA Clustering on SX-Aurora TSUBASA Hybrid Computing of QUBO Generation on VH and Digital Annealing on VE

Pre&Post Processing on VH





### Combinatorial DA Clustering on SX-Aurora TSUBASA Hybrid Computing of QUBO Generation on VH and Digital Annealing on VE



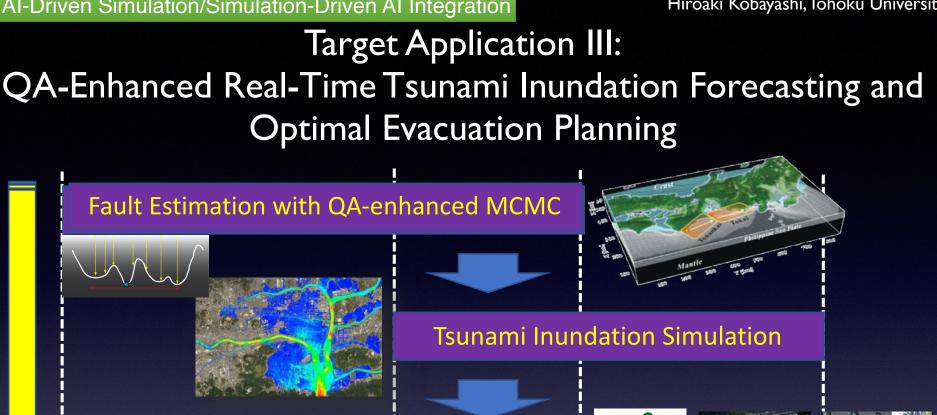
RSD2021

September 27, 2021



Time

#### AI-Driven Simulation/Simulation-Driven AI Integration



**Optimal Evacuation Planning** with Quantum Annealing





#### **Integrated Programming Framework**





dation

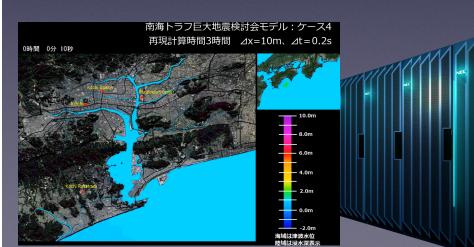
Execution TIme for

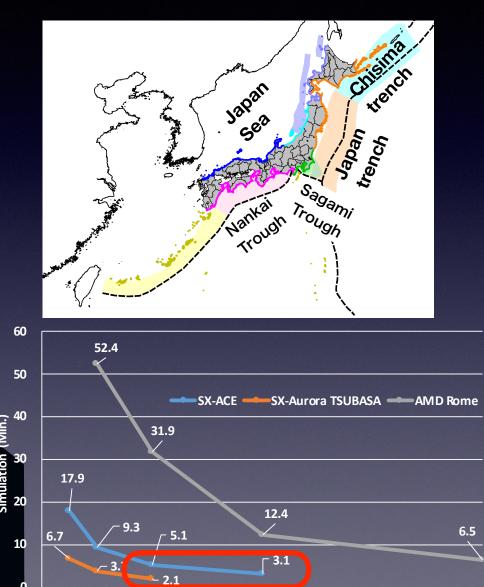
#### ★Area Covered

✓ From two points (Kochi and Shizuoka areas) to Japan coastal areas along coastlines of 8,000km from Kagoshima up to Ibaraki at a 10mx10m mesh resolution.

#### Expanding the covered area nationwide

Up to Hokkaido and thenDown to the side of Japan Sea





256

128

384

512

Number of Cores September 27, 2021

768

1024



### Real-Time Evacuation Guidance based on Multi-Agent Simulation Powered by Quantum Annealing

# After the inundation damage estimation is obtained, evaluation guidance to the safety zones will be estimated

- three candidates for safety evacuation routes obtained by using I) shortest path algorithm, 2) enforced learning based multi-agent simulation and 3)their combination under the consideration of the inundation damage simulation results,
- Best evacuation paths are selected based on the locations by using quantum annealing to solve an optimal combinatorial problem to maximize survivers from the inundation

2nd

BEST

1 st

BEST







D-wave Machine Integrated Programming Framework

3rd

BEST

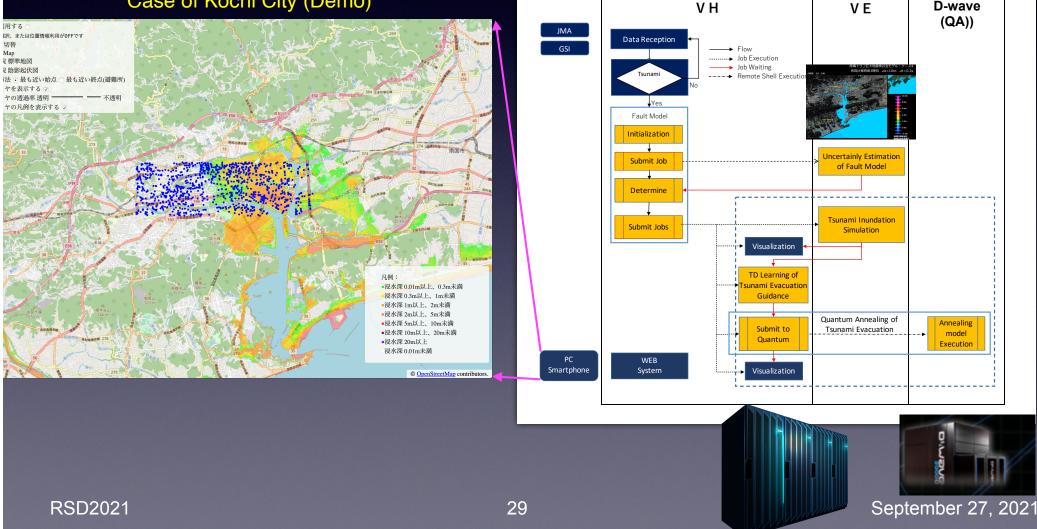
Aurora TSUBASA Vector Host (Xeon) Aurora TSUBASA Vector Engine

September 27, 2021



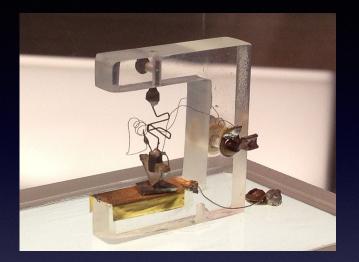
#### A Workflow of QA-Classical HPC Hybrid Computing for TSUNAMI Inundation Simulation and Optimal Evacuation Path Planning

#### Obtained Optimal Evacuation Path Results in the Case of Kochi City (Demo)

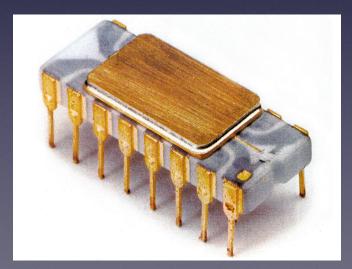




#### We are seeing The Dawn of Quantum Computing!?



The first transistor ever made, built by John Bardeen, William Shockley and Walter H. Brattain of Bell Labs in 1947.



The Intel 4004 was the world's first microprocessor a complete general-purpose CPU on a single chip. Released in March 1971,



## Summary

- Realization of general-purpose computing by ensemble of domain specific architectures as the next generation computing infrastructure toward post Moore's era
  - ✓ Maximize computing performance per cost and/or power best suited for a specific domain
  - Best mix of domain specific architectures that satisfies the demands of a wide variety of applications
- ★ R&D of a next generation HPC infrastructure: Fusion of Quantum-Annealing and classical HPC in a unified way
  - SX-Aurora TSUBASA, combination of vector engine and X86 engine, has a great potential to achieve a high sustained performance because of its best mix of vector architecture for memory-intensive apps. and x86 architecture for complicated control-intensive apps.
  - ✓ D-wave machine, A Quantum annealing machine, is the best domain specific architecture for combinatorial problems in the post-Moore era

#### **\*** R&D of three innovative killer apps:

- $\checkmark$  digital twin of a power generating Turbine for its effective operation and maintenance, and
- $\checkmark$  material informatics for efficient carbon composite products design
- ✓ real-time optimal Tsunami inundation evaluation planning,
- 🖈 Quantum annealing has a potential as a game changer toward the post-Moore era, but still is in its infancy
  - ✓ We are seeing The Dawn of Quantum Computing!?
  - $\checkmark$  Yes it needs more efforts and breakthrough to make it happen!
  - Digital Annealing on SX-Aurora TSUBASA is reasonable choice until the quantum annealing becomes practical!

## Acknowledgments

#### TOHOKUMembers of Association for Real-time Tsunami

#### Science (ARTS)

- ★ Tohoku University
  - IRIDES Shun-ichi Koshimura Takashi Abe
  - Graduate School of Science
  - Cyberscience Center



Ryota Hino Yusaku Ota Kenji Oizumi

#### ★ KOKUSAI KOGYO Co, LTD.

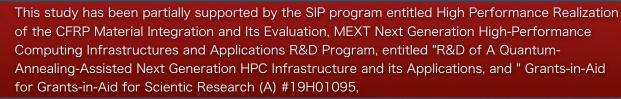
- Yoichi Murashima (Visiting Prof. of Tohoku Univ.)
- Muneyuki Suzuki
- Takuya Inoue

#### ★ NEC

- Akihiro Musa (Visiting Prof. of Tohoku Univ.)
- Osamu Watanabe (Visiting Researcher of Tohoku Univ.)
- ★ NEC Solution Innovator LTD.
  - Yoshihiko Sato
- ★ Osaka University
  - Cybermedia Center
    - Shinji Simojyo
    - Susumu Date

#### 🗙 A2 Corp.

Masaaki Kachi



Members of Research Division of High-Performance Computing (jointly organized with NEC), Tohoku University

- Hiroyuki Takizawa (Cyberscience Center)
- Akihiro Musa (Visiting Prof., NEC)
- Mitsuo Yokokawa (Visiting Prof., Kobe Univ.)
- Ryusuke Egawa (Visiting Prof, Tokyo Denki Univ)
  - Shintaro Momose (Visiting Assoc. Prof., NEC)
  - Kazuhiko Komatsu (Cyberscience Center)
  - Masayuki Sato (GSIS)
  - Technical Staff members (all from Cyberscience Center)
    - Kenji Oizumi
    - Satoshi Ono
    - Tsuyoshi Yamashita
    - Atsuko Saito
    - Tomoaki Moriya
    - Daisuke Sasaki
  - Visiting Researchers (all from NEC)
    - Shigeyuki Aino
    - Kazuto Nakada (Research Prof, Tohoku Univ)
    - Noritaka Hoshi
    - Takashi Hagiwara
    - Osamu Watanabe
    - Yoko Isobe
    - Yasuhisa Masaoka
    - Takashi Soga
    - Yoichi Shimomura
    - Soya Fujimoto





September 27, 2021



🥌 KOKUSAI KOGYO CO., LTD.

大阪大学 サイバーメディアセンター

### ▶ 株式会社エイツー