Basic Computational Biology

High Performance Computing Technology(1)

Introduction to HPC systems and Computational Science

M. Sato

Contents

- What is HPC?
- Basics of Parallel Computing and HPC
- Trends of High Performance Computing
  - K computer and “Fugaku”
- Next
  - Basics of Parallel computing
  - Parallel Programming

What is HPC ?

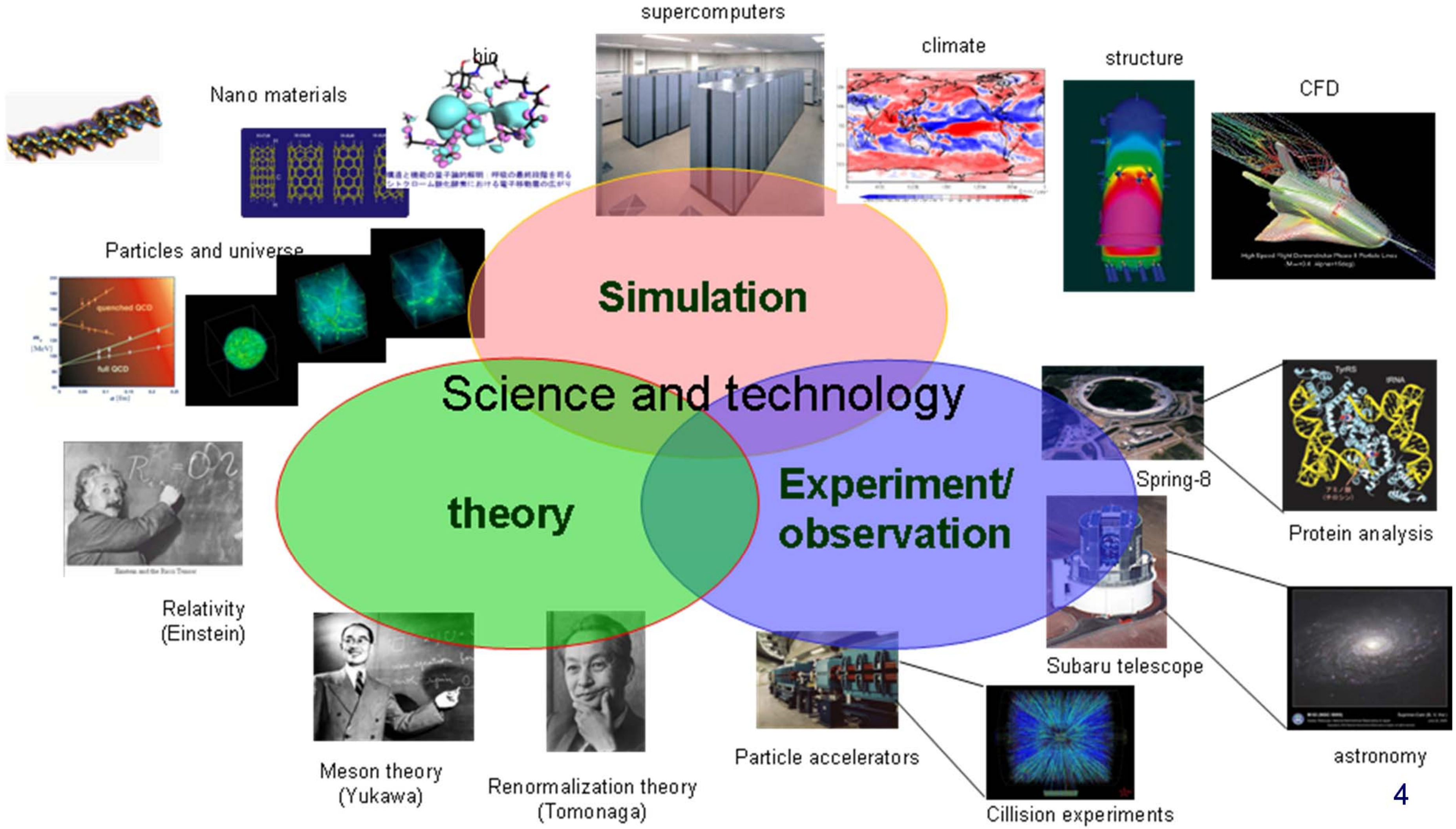
- Today's science (domain science) is driven by three elements
  - – Experiment
  - – Theory
  - – Computation (Simulation)

- In many of these problems, computation performance and capacity are required to be larger and larger
  - – Floating point operation speed
  - – Memory capacity (amount)
  - – Memory bandwidth (memory speed)
  - – Network bandwidth (network speed)
  - – Disk (2nd storage) capacity

- "High Performance" does not mean only the speed but also capacity and bandwidth

Computational science

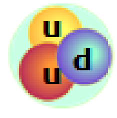
- Large-scale simulations using supercomputers
- Critical and cutting-edge methodology in all of science and engineering disciplines
- The third “pillar” in modern science and technology



What computational science can do ...

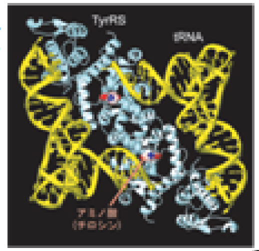
- To explore complex phenomenon which cannot be solved by "paper and pencil"
  - Particle physics to explore origin of materials
  - Phenomenon caused by aggregation of DNAs and protein
- To explore phenomenon which cannot be solve by experiment
  - Origin of universe
  - Global Warming of the earth
- To analyze a large set of data "big data"
  - Genome informatics
- To reduce the cost by replacing expensive experiments
  - car crash Simulation
  - CFD to design air craft

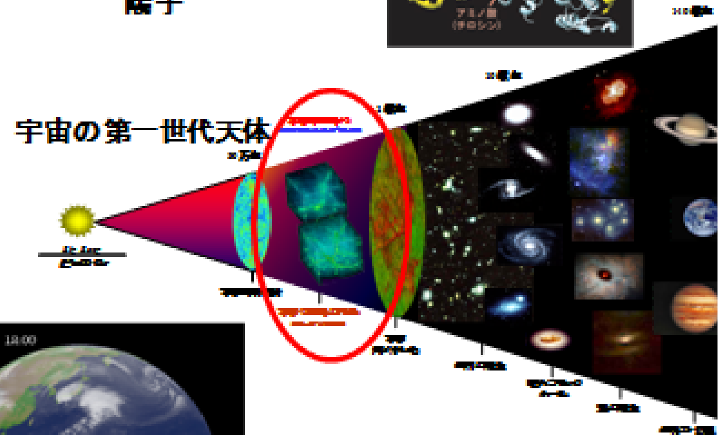
First principal method: computer simulation based on only computation without "experimental parameters. But it may require a huge computations

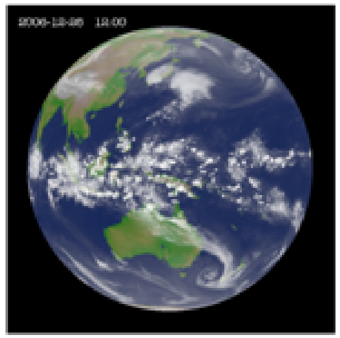


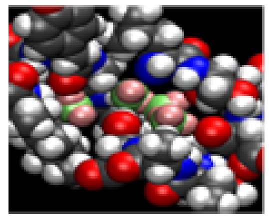
陽子

たんぱく質

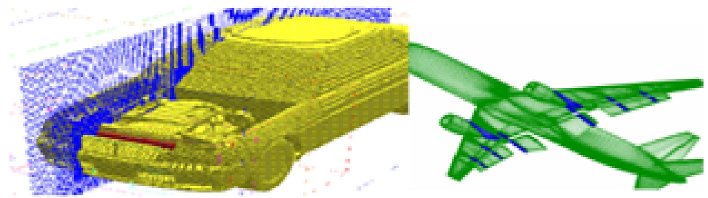








DNA



"First principal computation"...

- Schrödinger equation

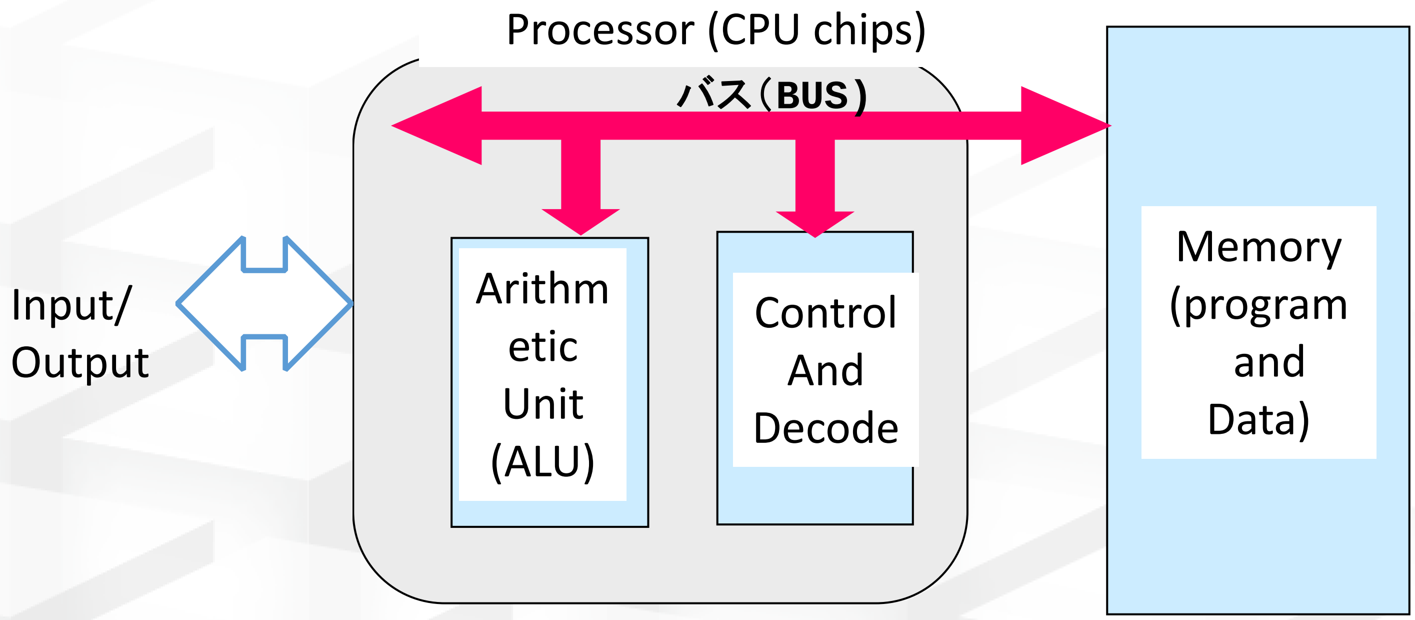
$$\mathbf{H}\psi = E\psi$$

$$i\hbar\frac{\partial \psi}{\partial t} = \left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V(x)\right]\psi$$

- "first principle calculation(computation)" in computational material

Basic Structure of Computer

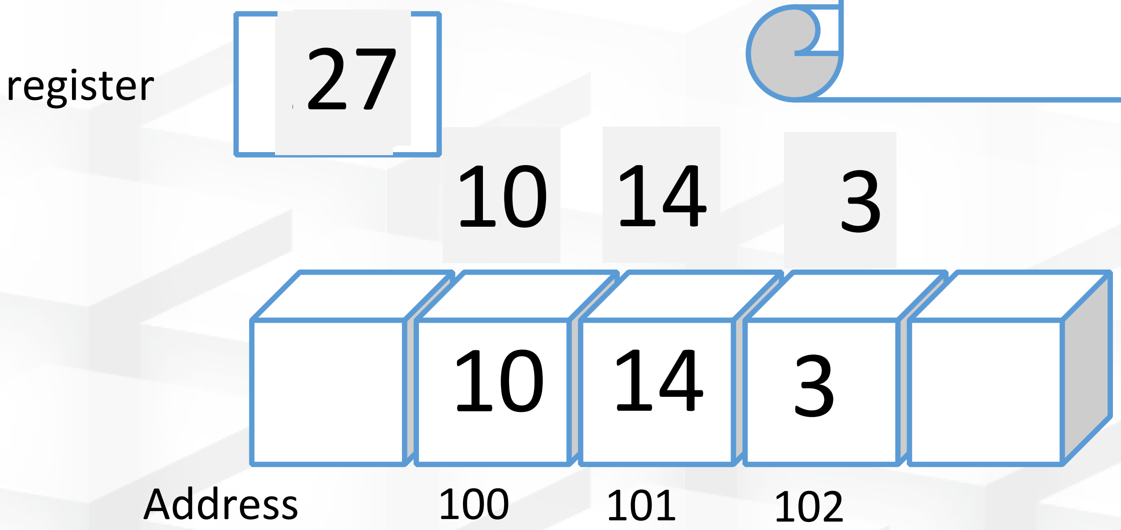
- **The current computer (Neumann type) consists of a processor (CPU, core) and memory.**
  - Memory is a part to store programs and data
  - The processor reads the program and data from its memory and executes the program.
    - Control part: Interpret (?) the program and command
    - ALU (Arithmetic Unit) : Add and Multiply, etc …



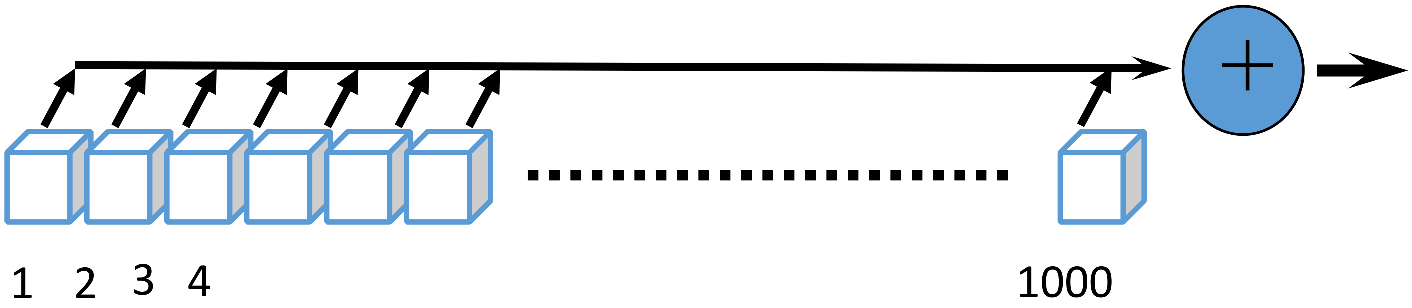
What is “program”?

- **Instructions for computer**
  - Stored in memory, and the CPU reads from the memory one by one and executes it.
  - Memory is a box that store data. There is a number called an address.
  - Register: Temporary memory for arithmetic unit.

1. **Set the register to 0.**
2. **Read data from address 100**
3. **Add to register**
4. **Read data from address 101**
5. **Add to register**
6. **Read data from address 102**
7. **Add to register….**



For calculating the sum of 1000 numbers…



- The program simply reads the numbers in the 1000 memory and adds them.
- It takes 1000 seconds to fetch numbers and add them, assuming that one addition takes 1 second and calculates 1000 numbers.
- By the real computer, one addition can be executed in tens of nanoseconds (one billionth of a second), so it can be calculated in a few microseconds (one millionth of a second)!

How to make Computer fast ...

- ① By making electric circuit work fast
  - Increasing clock speed

  (Frequency of processors used in PC: 2～3GHz)
  - Using fast transistor

- Microprocessor

A computer with all blocks of CPUs in one chip.
Used in PCs, the current microprocessor is far faster than an old supercomputer!

A progress of "computer" speed!

- Metric of speed of computation: arithmetic operations (floating point) per second
  - MFLOPS: Millions of FLoating Point OPerationS.
  - GFLOPS：10^9 ops，TFLOPS：10^{12} ops，PFLOPS：10^{15} pos、Exa

- Rapid progress of microprocessor (all components in a chip) used for PC --- "killer micro"
  - Moore's Low: integration (density) of transistor increase double per 1.5 year
  - 4004(first microproessor、1971、750KHz) 8008(1972、500KHz、Intel) 8080(1974、2MHz、Intel)
  - Pentium 4 (2000、～3.2GHz)

  - Clock speed increased from 1MHz to 1GHz in the last three decades

To make computer fast …

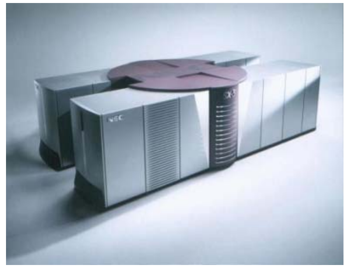
- ② By good mechanisms (architecture) in computer
  - mechanism to execute many instruction at a time (in one clock ...)
  - Vector supercomputer: a computer with computing unit to execute vector computation frequently used in scientific computation

    (1980's)

Fujitsu VPP500

Fujitsu VPP5000

NEC SX-4

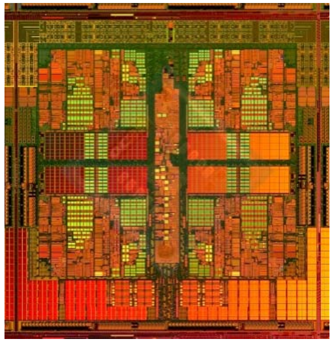


NEC SX-5

To make computer fast …

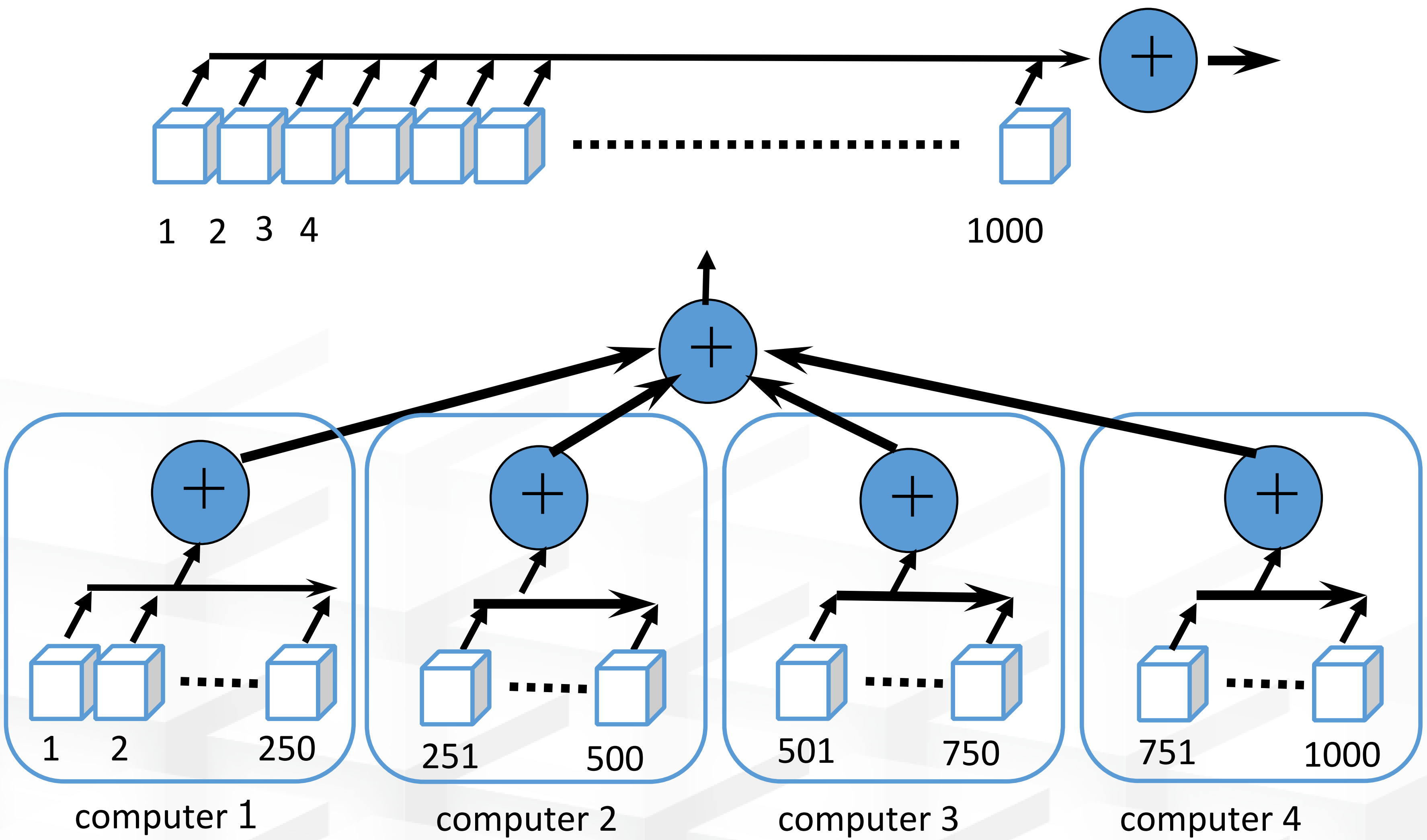
- ③ by using many computer at a time
  - Parallel computers, parallel processing ...
  - This is a main stream in supercomputer !

  - You can find 2 or 3 processors in a PC or "smart phone"!



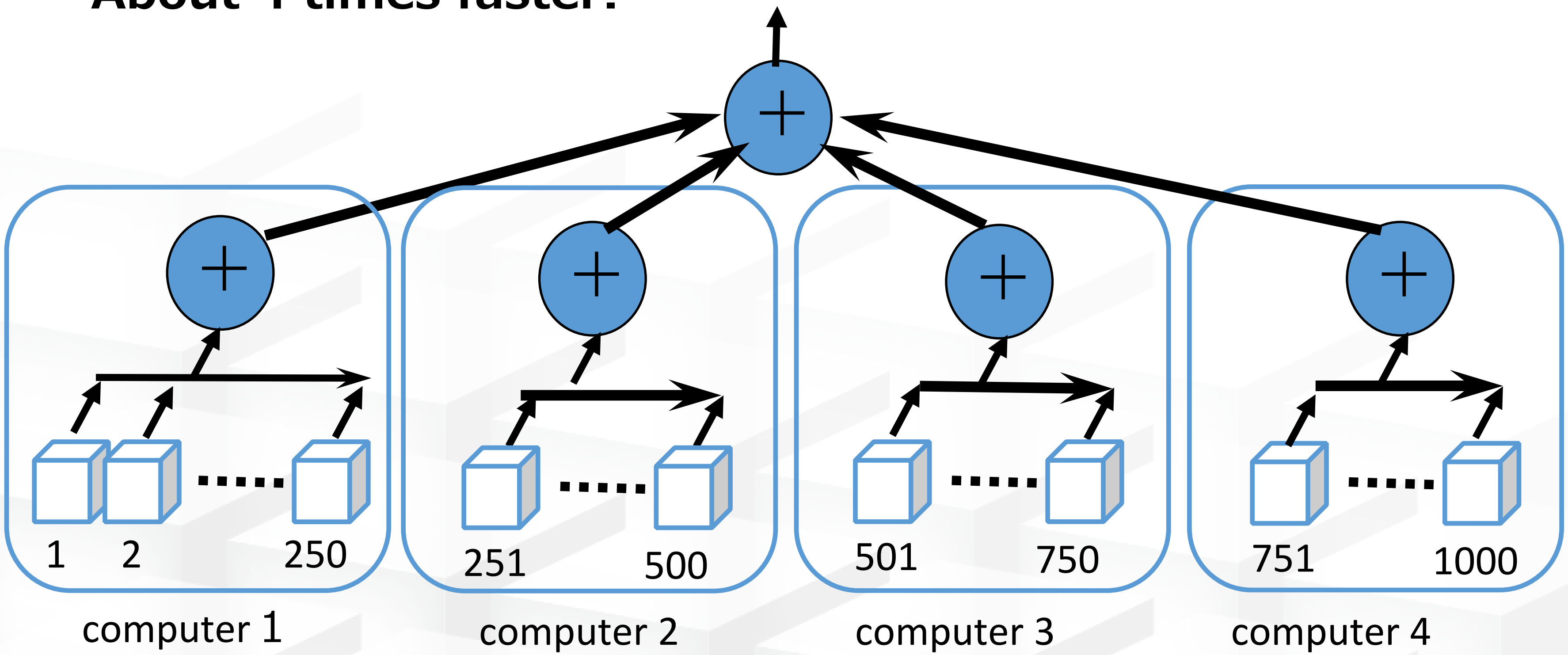
AMDのquad coreのプロセッサ

Adding 1000 numbers using 4 computers ...



Adding 1000 numbers using 4 computers ...

- It takes 1000 seconds to fetch numbers and add them, assuming that it is 1 second and 1000 numbers.
- When using four computers, store 250 numbers on each computer, so you can calculate them in 250 seconds + α. About 4 times faster!

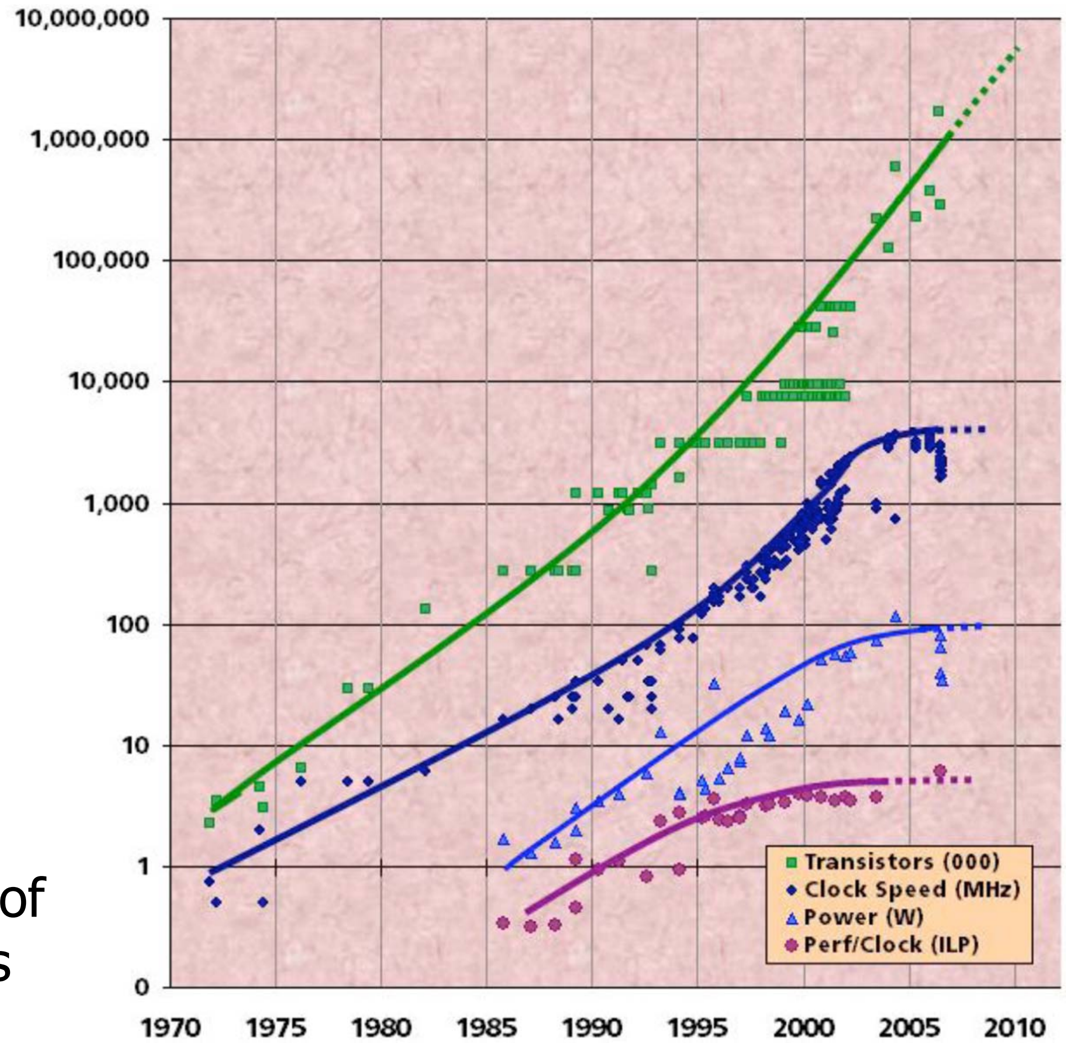


Moore's Law re-interpreted

- Progress of clock speed stops after 2000's
- Still increasing the number of transistors

↓

- Multicore
  - Core (computer) in onechip
  - double in the number of cores every 18 months



TOP 500 List: How to measure (rank) performance of supercomputers

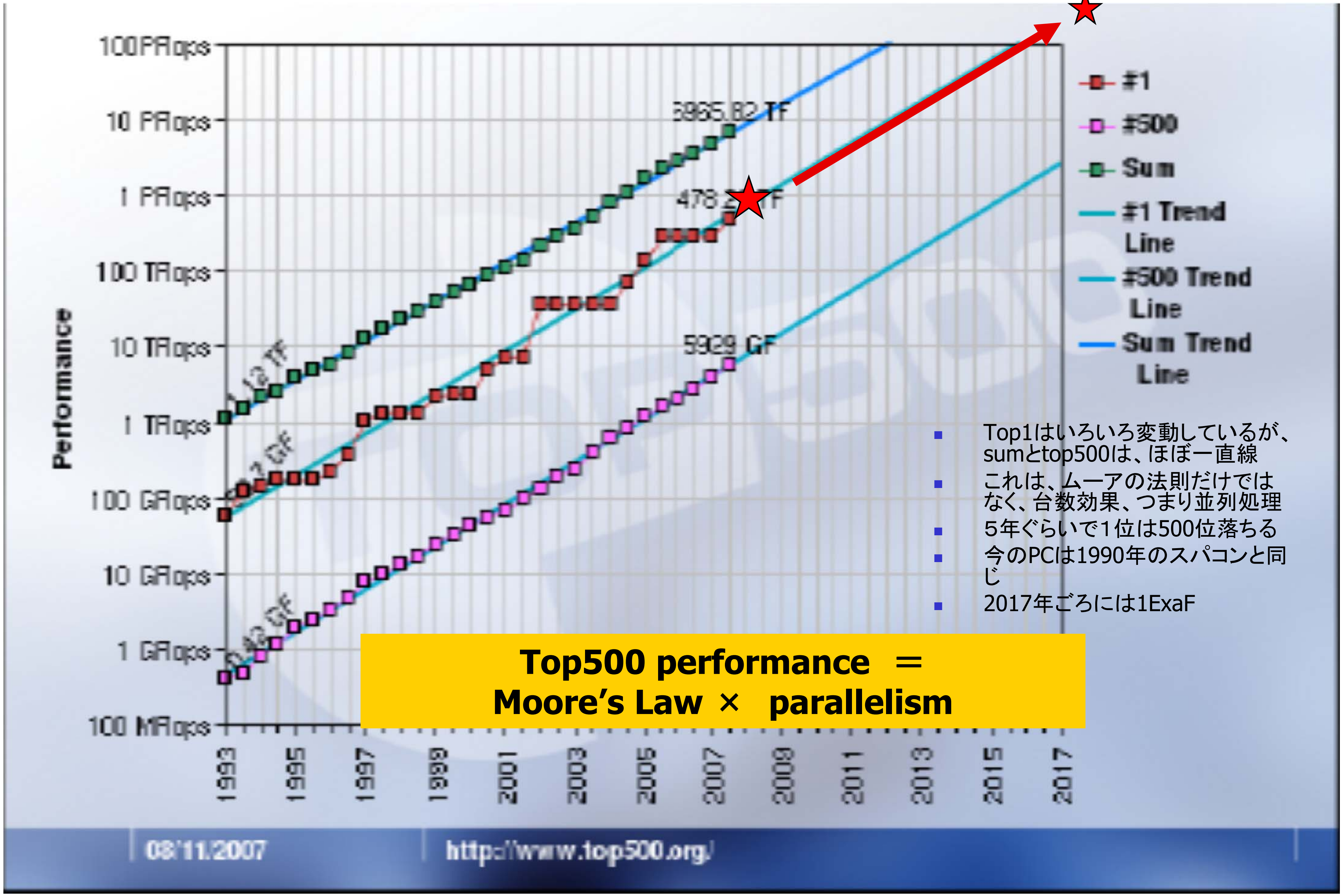
http://www.top500.org/

- Ranked by the performance of benchmark program "LINPACK"
  - LINPACK solves a huge size of linear equations
  - the size is more than 10 millions

- Different from the performance of "real" applications
  - It does not necessarily reflect the performance of "real" applications

- The power consumption is indicated since 2008
  - The power saving is import now !

TOP500: The list of the fastest computers



- Top1はいろいろ変動しているが、sumとtop500は、ほぼ一直線
- これは、ムーアの法則だけではなく、台数効果、つまり並列処理
- 5年ぐらいで1位は500位落ちる
- 今のPCは1990年のスパコンと同じ
- 2017年ごろには1ExaF

Top500 performance ＝
Moore’s Law × parallelism

京コンピュータ “The K computer"

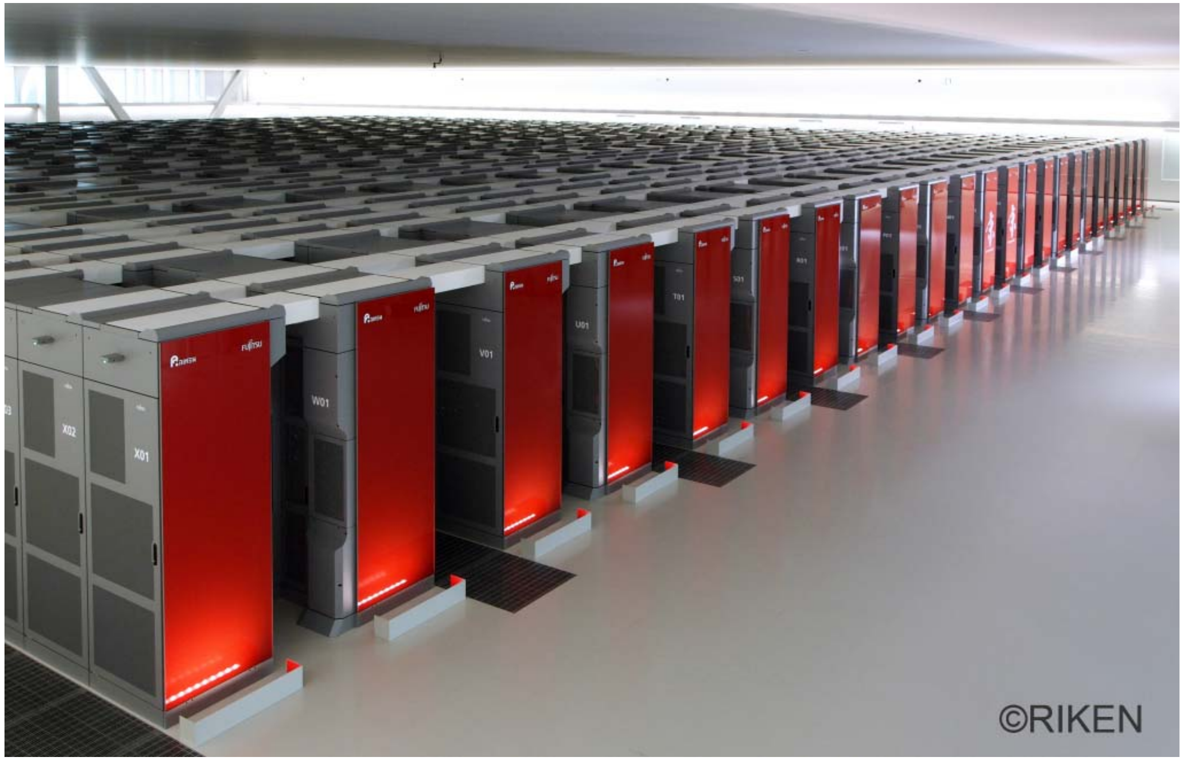
Facts of the K computer

- The number of racks (boxes) 864
- the number of chips 82,944
- The number of cores (computers) 663,552
- Linpack perf

  10.51PF

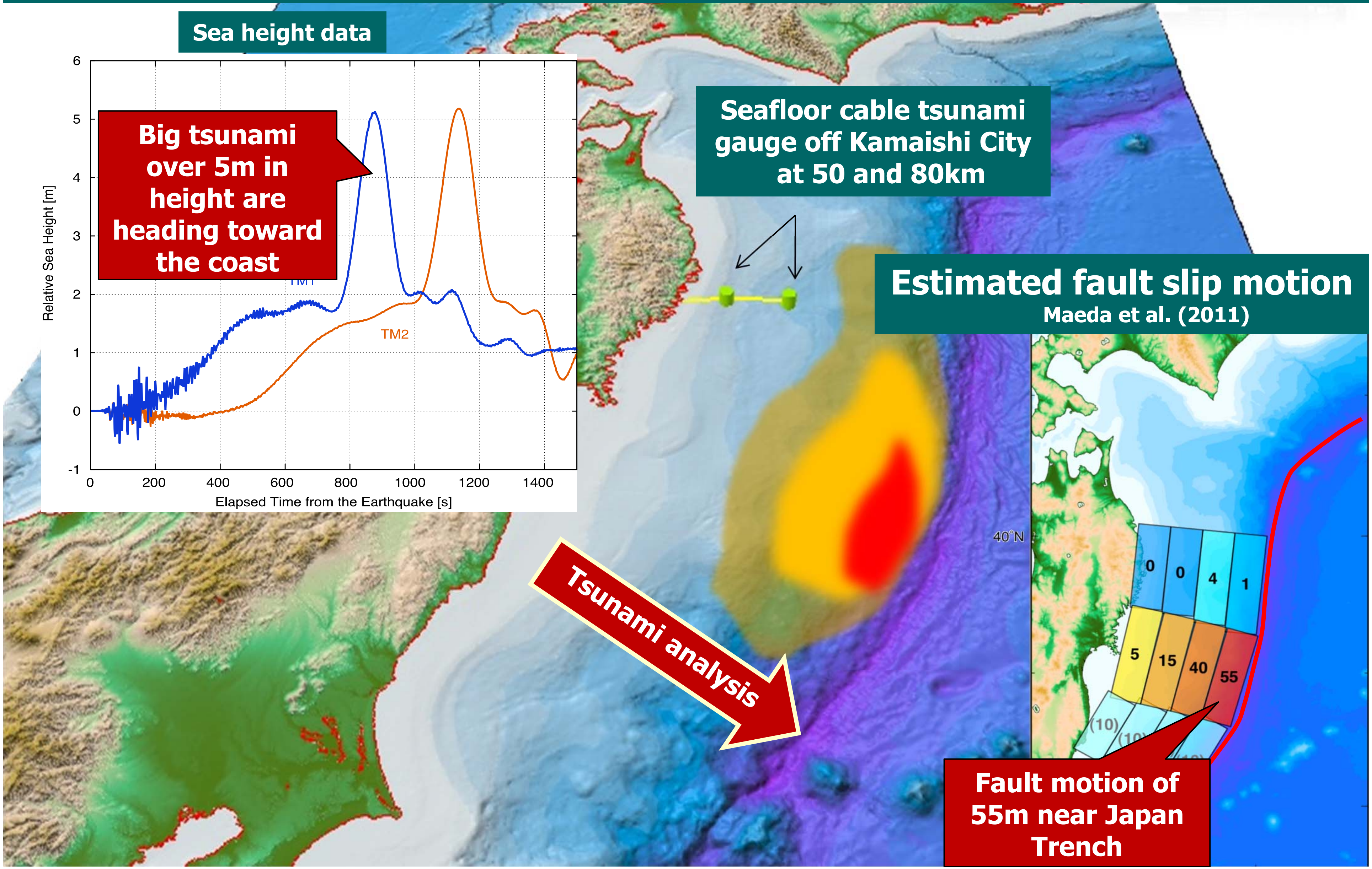
  (Power 12.66MW)

  2011/11月



The tsunamis were captured by seafloor pressure gauge

Courtesy: T. Furumura (U. Tokyo)

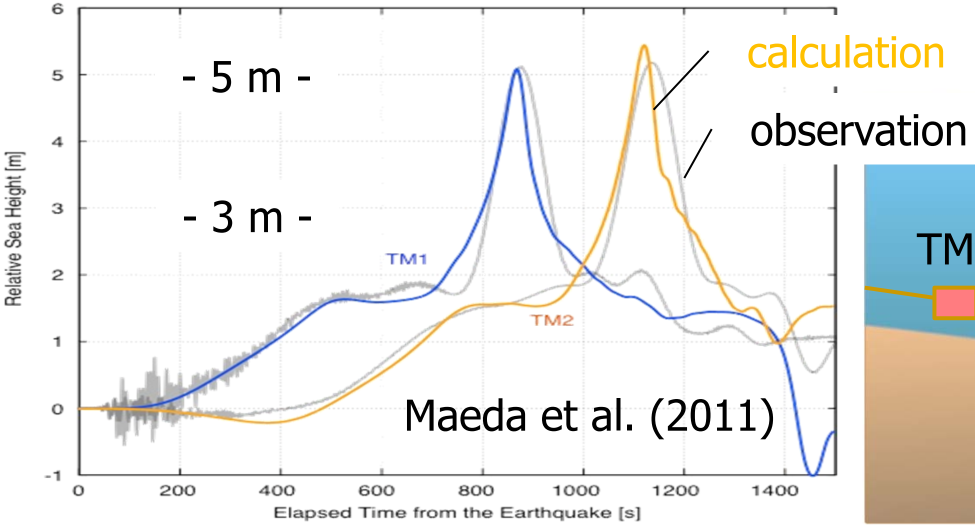


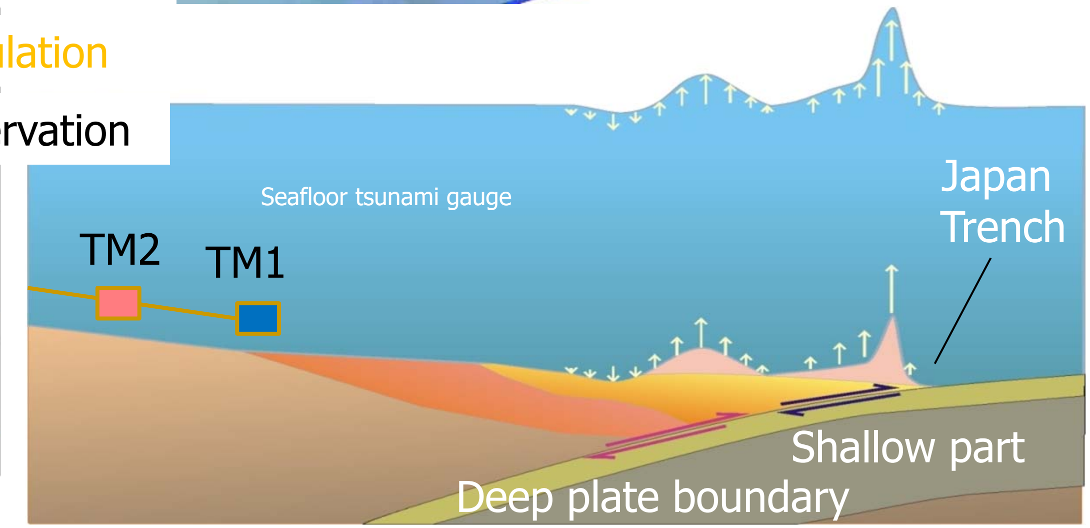
The combination of deep and shallow plate slips generated the big tsunamis

Courtesy:
T. Furumura (U. Tokyo)

(a) Slip of deep plate boundary only

(b) Slip of deep *and* shallow plate boundary





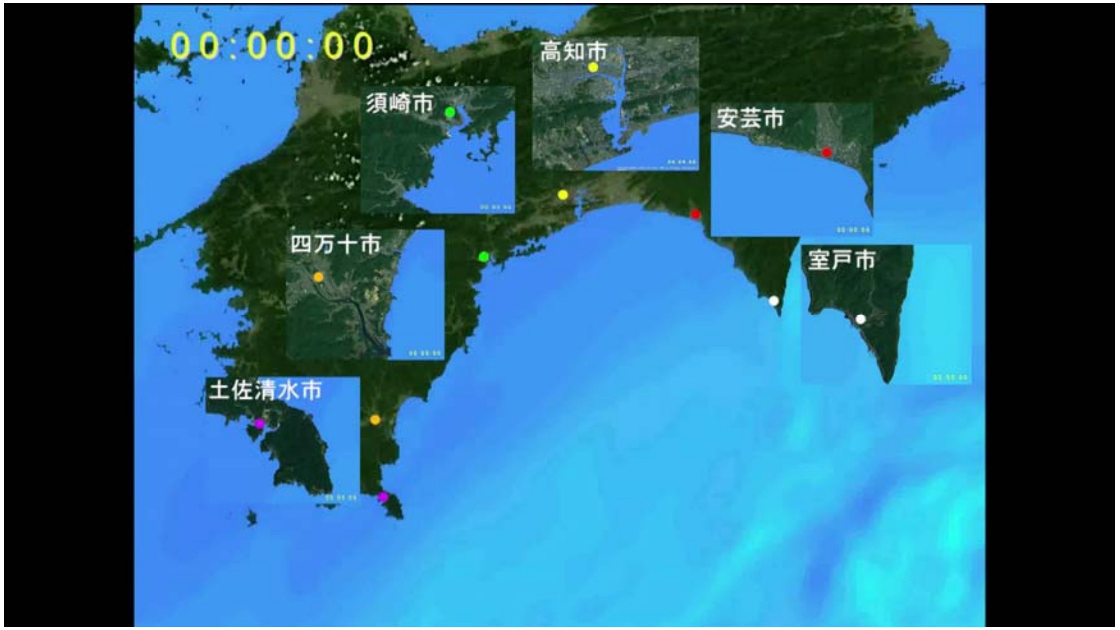
高精細なシミュレーションによる災害に対する防災・減災

巨大地震により引き起こされる①強い揺れ，②地殻変動（海底や海岸の隆起・沈降），そして③津波を，地震発生からの時間を追って詳細に評価して地震防災・避難計画に活用するために「地震 - 津波同時シミュレーション」を開発

東日本大震災の再現

前田、古村（東大）

南海トラフ地震・津波の予測



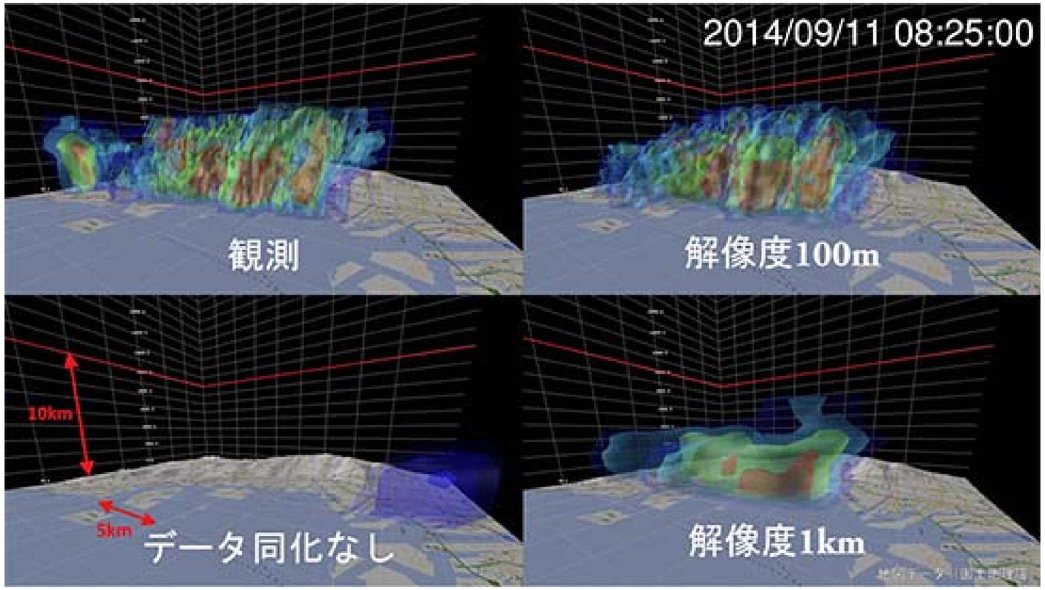
南海トラフ巨大地震　広域詳細な津波計算　　馬場 (JAMSEC)

たくさんのシナリオを用意し、地方自治体と連携し、防災計画、ハザードマップの作成に寄与

「京」と最新鋭気象レーダを組み合わせたゲリラ豪雨予測

- 現在の天気予報は、2kmの解像度でシミュレーションを行い、1時間毎に新しい観測データを取り込んで更新するため、わずか数分の間に局地的にゲリラ豪雨を引き起こす積乱雲を予測することは困難。

- 「京」を使った解像度100mの高精細シミュレーションに30秒毎の観測データを組み合わせた時間的・空間的に桁違いのシミュレーションを世界で初めて実現し、実際のゲリラ豪雨の動きを詳細に再現することに成功。

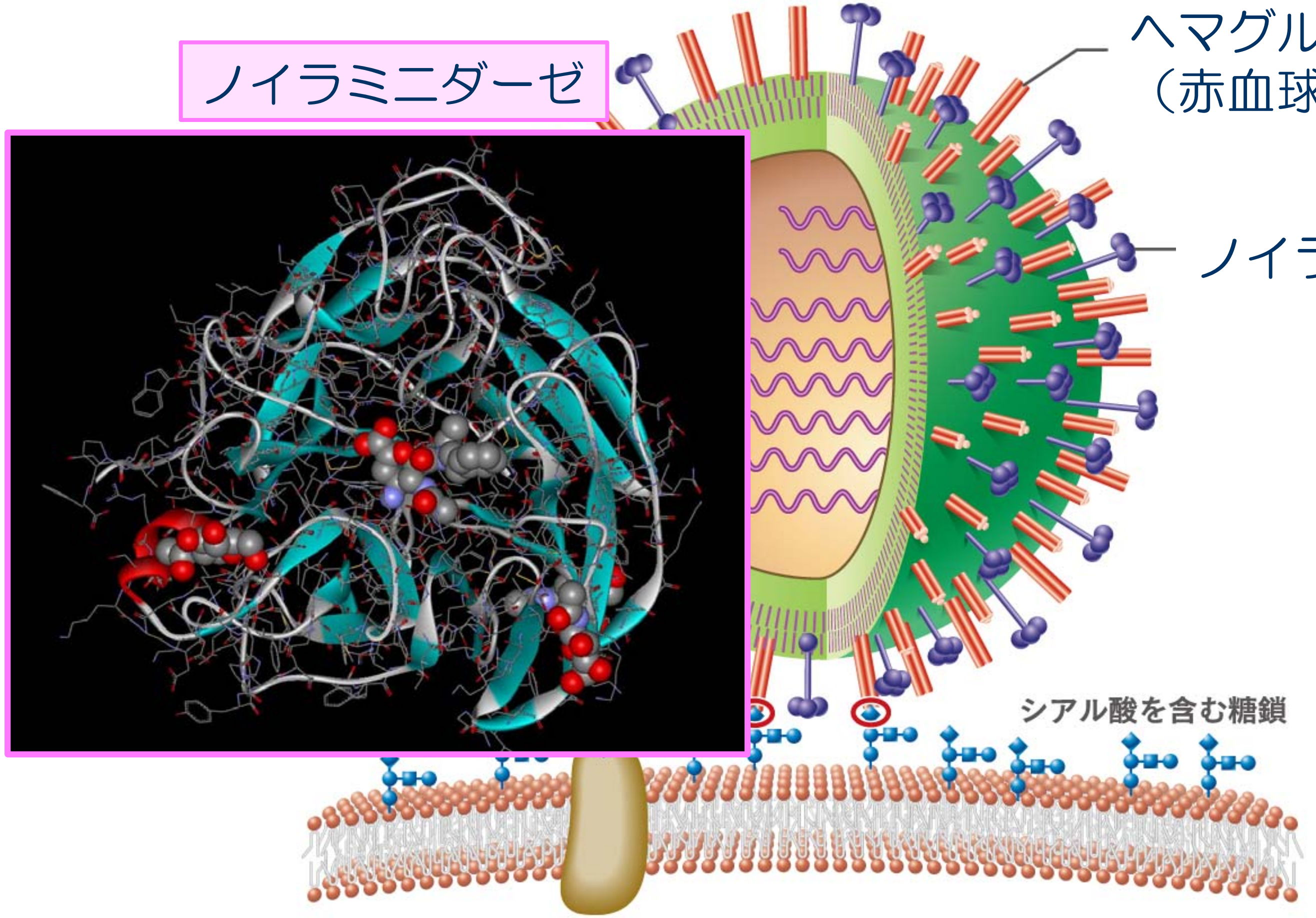


〈2014年9月11日午前8時25分の神戸市付近における雨雲の分布〉

解像度100mのシミュレーション結果は、積乱雲内部の微細構造や降水分布が観測データに非常に近いことが分かる。

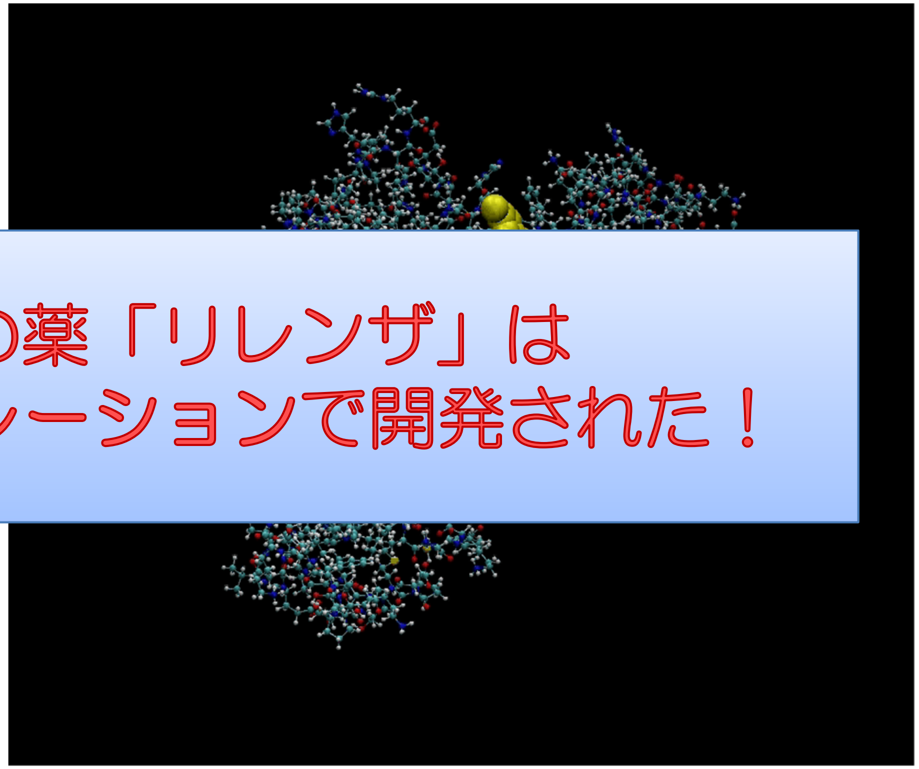
インフルエンザウイルスの働き

ノイラミニダーゼ



出典： 生命科学教育用画像集　http://csls-db.c.u-tokyo.ac.jp
インシリコサイエンス社　http://www.pd-fams.com

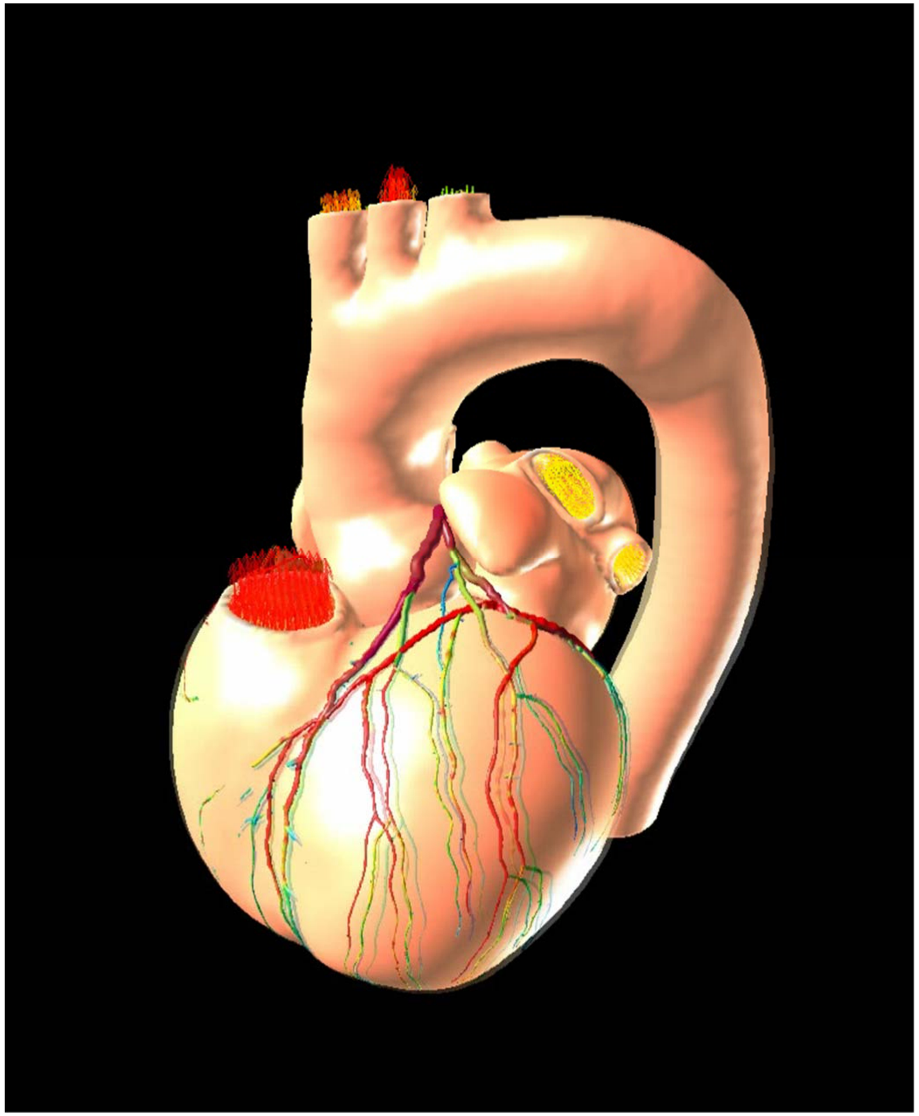
病気の原因となるたんぱく質と薬の
ドッキングシミュレーション



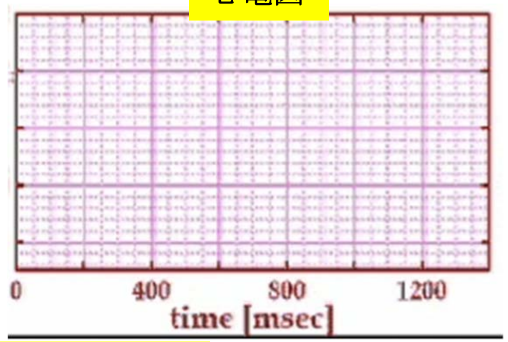
インフルエンザの薬「リレンザ」は
コンピュータ・シミュレーションで開発された！

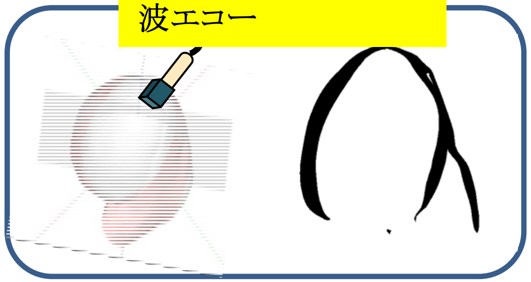
ちょっと前までの計算　　スーパーコンピュータ「京」で計算

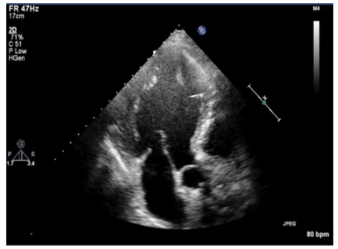
細胞モデルからの心臓シミュレーション



株式会社UT-Heart研究所　協力：富士通株式会社







- 心筋細胞内のたんぱく質の確率的運動から細胞の収縮、心拍動、血液駆出、冠循環までを一貫してシミュレート。
- シミュレーションから超音波エコー、流速ドップラー、心電図、カテーテル検査などの精緻なデータが再現される。そのデータを基に病態の解析が可能に。仮想手術や薬の副作用予測（不整脈予測）にも応用。

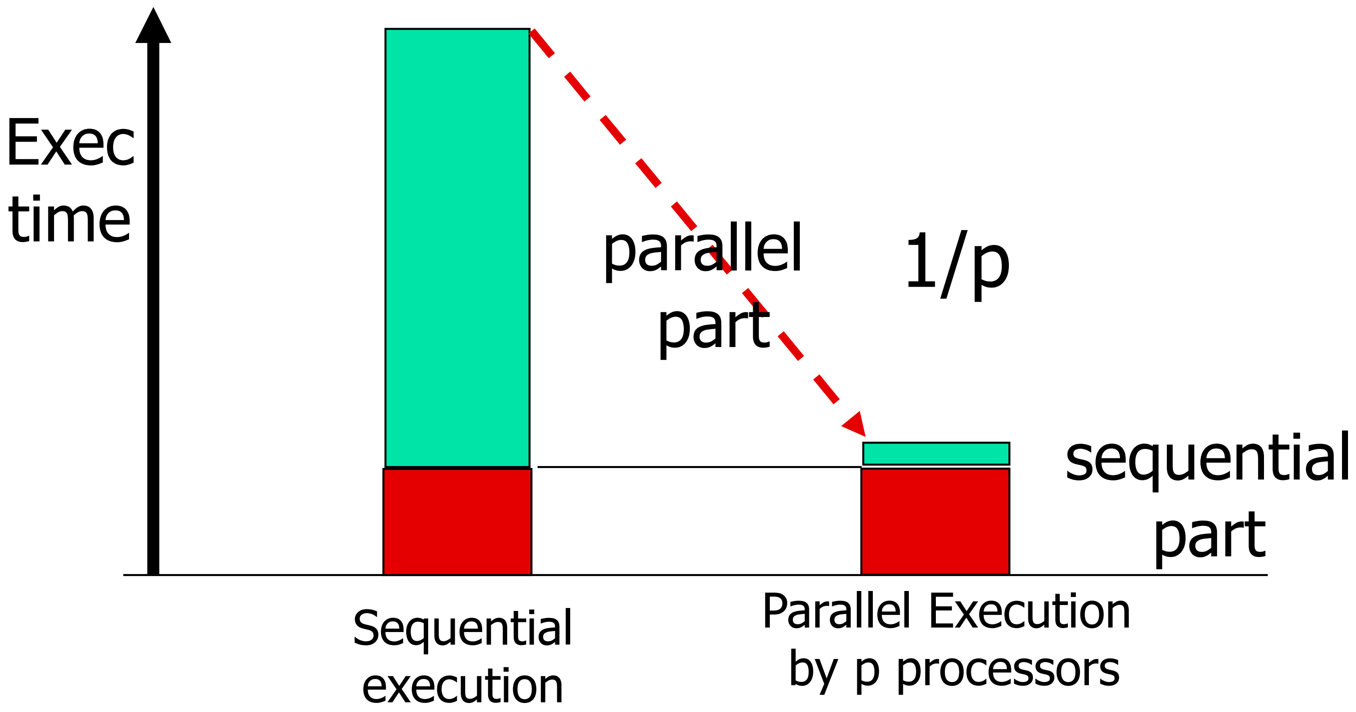
Amdahl's low

- Question: How much do parallel computers became fast by increasing the number of processors???

ジーン・アムダール（Gene Amdahl、1922年11月16日 - ）は、アメリカ人のコンピュータアーキテクトで、企業家ある。彼の業績はIBMおよび彼の創設した会社(特にアムダール社)における、メインフレームの設計である。並列コンピューティングの基本的な理論としてアムダールの法則がよく知られている。(wikipediaより)

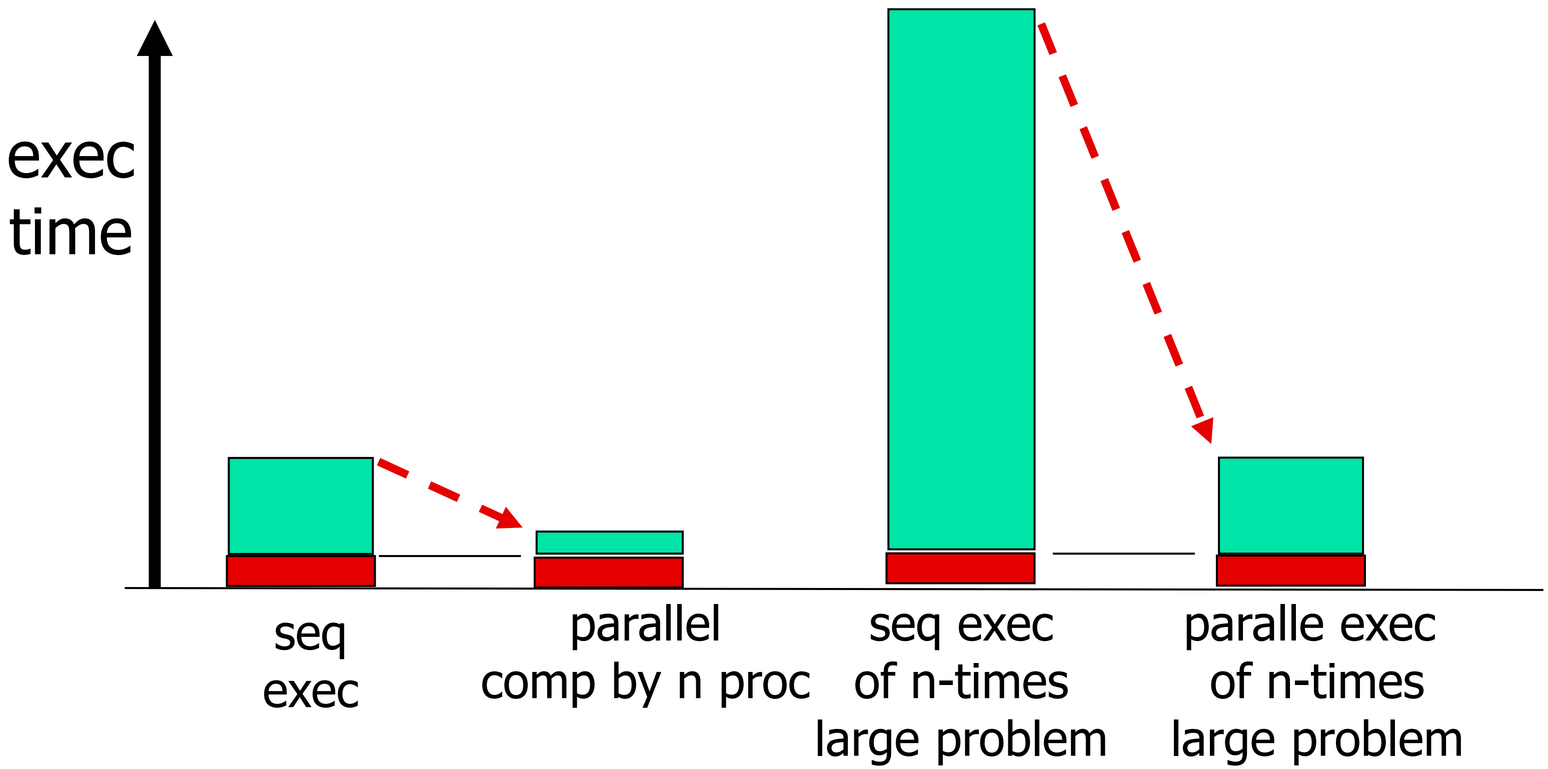
Speedup by parallel computing: "Amdahl's low"

- Amdahl's low
  - Suppose execution time of sequential part T_1, ratio of sequential part α, execution time by parallel computing using p processors T_p is (no more than) $T_p = \alpha * T_1 + (1-\alpha) * T_1/p$
  - Since some part must be executed sequentially, speedup is limited by the sequential part.



Breaking "Amdahl's low"

- "Gustafson's low": what about performance of real apps?
  - The fraction of parallel part often depends on the size of problem
  - For example, n-times larger problem to be solve by n-times larger parallel computers.
  - Weak scaling – Scaling with constant size per processor ← in the case of large scale scientific applications
  - Strong scaling —Scaling with constant size problem ← We need fast one-processor.



How different between the K computer and your PC?

- The processors (computer) used are almost the same!
  - Even slow clock for the K computer, but some enhancement in computing unit.
- The K computer consists of many "processors"
  - 80,000 chip、0.64 M cores
  - Fast network between processors is required!
- The programmer is forced to make parallel program to make use of many processors
  - The program running on the PC (sequential program) does not run fast !

FLAGSHIP2020 Project

Missions

- Building the Japanese national flagship supercomputer, "Fugaku" (a.k.a Post-K), and
- Developing wide range of HPC applications, running on Fugaku, in order to solve social and science issues in Japan

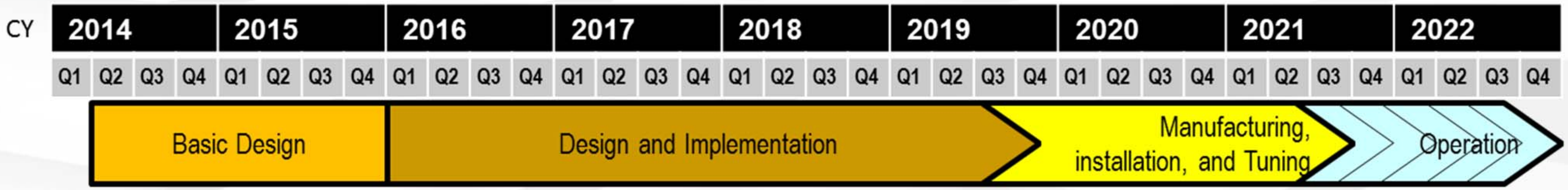
Planned Budget (from 2014FY to 2020FY)

- 110 billion JPY (about 1 billion US$ if 1US$=110JPY, total) includes:
  - Research and development, and manufacturing of the Fugaku system
  - Development of applications

Project organization

- System development
  - RIKEN is in charge of development
  - Fujitsu is vendor partner.
  - International collaborations: DOE, JLESC, CEA ..

- Applications
  - The government selected 9 social & scientific priority issues and their R&D organizations.
  - Additional projects for Exploratory Issues were selected in June 2016



Target science: 9 Priority Issues

R-CCS

①Innovative Drug Discovery

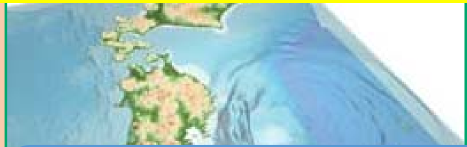
Society with health and longevity

RIKEN Quant. Biology Center

②Personalized and Preventive Medicine

Inst. Medical Science, U. Tokyo

③Hazard and Disaster induced by Earthquake and Tsunami

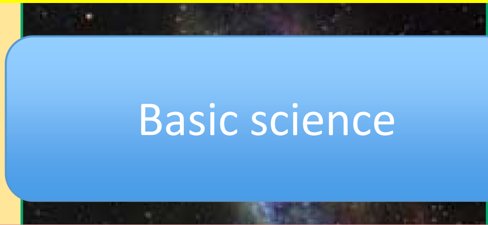


Disaster prevention and global climate

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

Industrial competitiveness

⑨Fundamental Laws and Evolution of the Universe



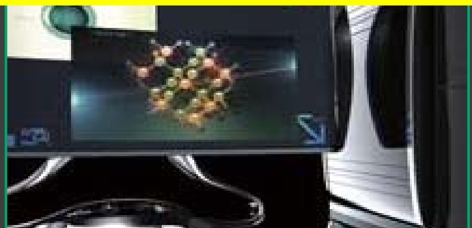
Basic science

Cent. for Comp. Science, U. Tsukuba

④Environmental Predictions with Observational Big Data

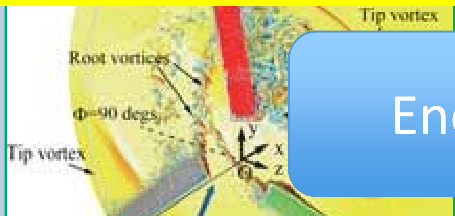
Center for Earth Info., JAMSTEC

⑦New Functional Devices and High-Performance



Inst. For Solid State Phys., U. Tokyo

⑥Innovative Clean Energy Systems



Grad. Sch. Engineering, U. Tokyo

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

Energy issues

Inst. Molecular Science, NINS

Target science: Exploratory Issues

Projects (more than 10 teams) were selected in Jun 2016

Interactive Models of Socio-Economic Phenomena and their Applications

Frontiers of Basic Science - challenge to extremes -

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

Mechanisms of Neural Circuits for Human Thoughts and Artificial Intelligence

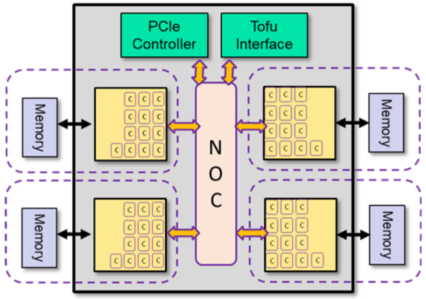
The name of our system (a.k.a post-K) was announced as “Fugaku” (May 23, 2019)

富岳 (Fugaku)
=
Mt. Fuji

- The highest mountain in Japan
- Wide foot area around the mountain

http://ww

FLAGSHIP2020 Project: Status



Fujitsu A64FX processor

☐ Overview of Fugaku architecture

Node: Manycore architecture

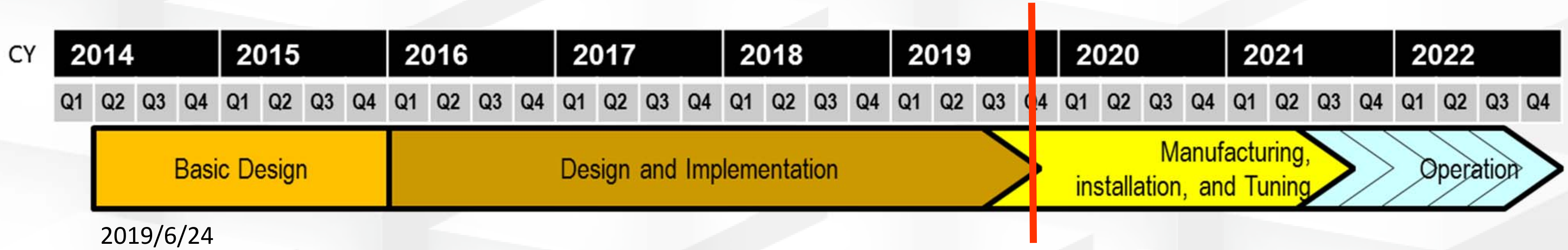
- Armv8-A + SVE (Scalable Vector Extension)
- SIMD Length: 512 bits
- # of Cores: 48 + (2/4 for OS) (> 2.7 TF / 48 core)
- Co-design with application developers and high memory bandwidth utilizing on-package stacked memory (HBM2) 1 TB/s B/W
- Low power : 15GF/W (dgemm)

Network: TofuD

- Chip-Integrated NIC, 6D mesh/torus Interconnect

☐ Status and Update

- March 2019: The Name of the system was decided as "Fugaku"
- Aug. 2019: The K computer decommissioned, stopped the services and shutdown (removed from the computer room)
- Oct 2019: access to the test chips was started.
- Nov. 2019: Fujitsu announce FX1000 and FX700, and business with Cray.
- Nov 2019: Fugaku clock frequency will be 2.0GHz and boost to 2.2 GHz.
- Mov 2019: Green 500 1st position!
- Oct-Nov 2019: MEXT announced the Fugaku "early access program" to begin around Q2/CY2020
- Around Jan 2020: Installation of "Fugaku" will be started.



2019/6/24

No.1 in Green500 at SC19!

R-CCS

Announce from
Fujitsu at SC19

Green500, Nov. 2019

FUJITSU

A64FX prototype –
Fujitsu A64FX 48C 2GHz
ranked **#1** on the list

768x general purpose A64FX
CPU w/o accelerators

- 1.9995 PFLOPS @ HPL, 84.75%
- 16.876 GF/W
- Power quality level 2

The GREEN 500

HOME GREEN500 LISTS RESOURCES ABOUT MEDIA KIT

Home / Lists / November 2019

NOVEMBER 2019

- The most energy-efficient system and No. 1 on the Green500 is a new Fujitsu A64FX prototype installed at Fujitsu, Japan. It achieved 16.9 GFlops/Watt power-efficiency during its 2.0 Pflop/s Linpack performance run. It is listed on position 160 in the TOP500.
- In second position is the NA-1 system, a PEZY Computing / Exascaler Inc. system which is currently being readied at PEZY Computing, Japan for a future installation at NA Simulation in Japan. It achieve 16.3 GFlops/Watt power efficiency. It is on position 421 in the TOP500.
- The No 3 on the Green500 is AiMOS, a new IBM Power systems at the Rensselaer Polytechnic Institute Center for Computational Innovations (CCI), New York, USA. It achieved 15.8 GFlops/Watt and is listed at position 25 in the TOP500.

Green500 List for November 2019

Listed below are the November 2019 The Green500's energy-efficient supercomputers ranked from 1 to 10.

Note: Shaded entries in the table below mean the power data is derived and not measured.

Rank	TOP500 Rank	System	Cores	Rmax (TFlop/s)	Power (kW)	Power Efficiency (GFlops/watts)
1	159	**A64FX prototype** - Fujitsu A64FX, Fujitsu A64FX 48C 2GHz, Tofu interconnect D , Fujitsu Fujitsu Numazu Plant Japan	36,864	1,999.5	118	16.876
2	420	**NA-1** - ZettaScaler-2.2, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 700Mhz , PEZY Computing / Exascaler Inc. PEZY Computing K.K. Japan	1,271,040	1,303.2	80	16.256
3	24	**AiMOS** - IBM Power System AC922, IBM POWER9 20C 3.45GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Volta GV100 , IBM Rensselaer Polytechnic Institute Center for Computational Innovations (CCI) United States	130,000	8,045.0	510	15.771
4	373	**Satori** - IBM Power System AC922, IBM POWER9 20C 2.4GHz, Infiniband EDR, NVIDIA Tesla V100 SXM2 , IBM MIT/MGHPCC Holyoke, MA United States	23,040	1,464.0	94	15.574
5	1	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148,600.0	10,096	14.719

FUJITSU CONFIDENTIAL 3

Advances from the K computer

	K computer	Fugaku	ratio
# core	8	48	
Si tech. (nm)	45	7	
Core perf. (GFLOPS)	16	> 64	4
Chip(node) perf. (TFLOPS)	0.128	>3.0	24
Memory BW (GB/s)	64	1024	
B/F (Bytes/FLOP)	0.5	0.4	
#node / rack	96	384	4
Rack perf. (TFLOPS)	12.3	>1179.6	96
#node/system	82,944	> 150,000	
System perf.(DP PFLOPS)	10.6	> 460.8	43

Si Tech

SVE

CMG&Si Tech

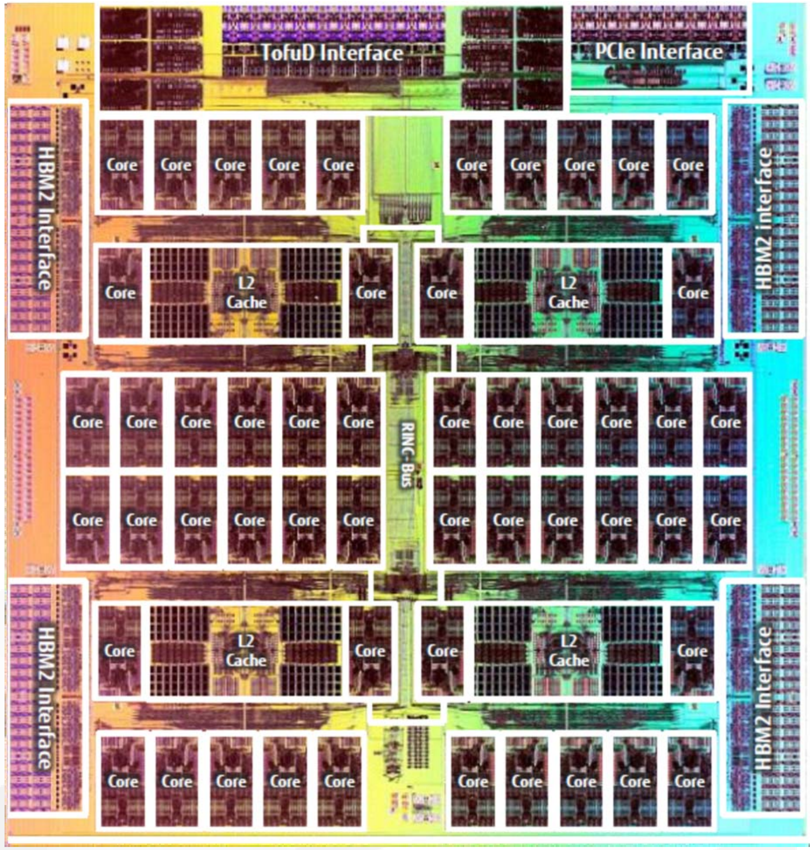
HBM

More than 7.5 M General-purpose cores!

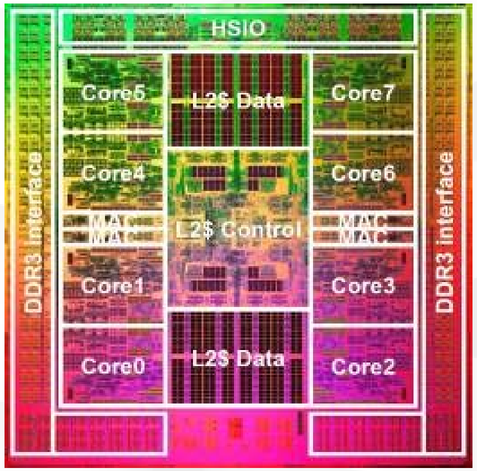
- SVE increases core performance
- Silicon tech. and scalable architecture (CMG) to increase node performance
- HBM enables high bandwidth

Comparison of Chips

Memory mounted in Silicon substrate

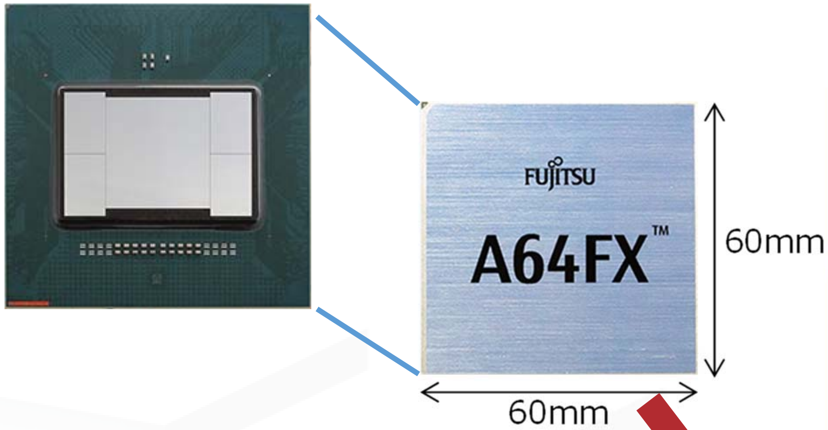


Fugaku A64FX chips
48 core（+ 4 core）、
NIC and IO (PCIe) integrated

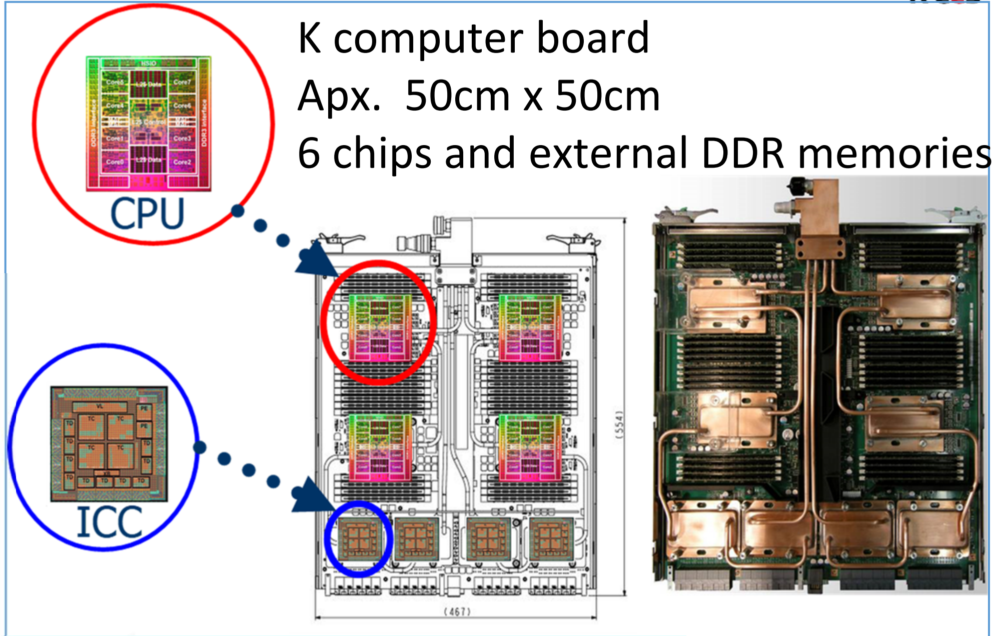


K computer chip、 8 cores

Comparison of Boards

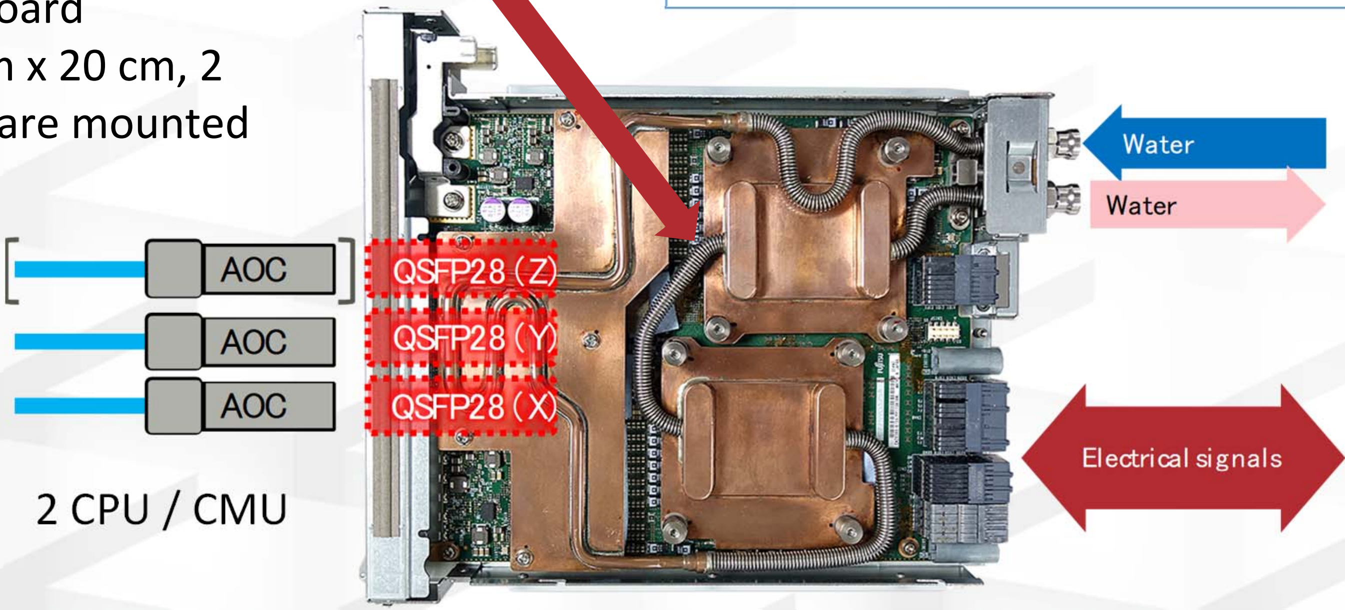


K computer board
Apx. 50cm x 50cm
6 chips and external DDR memories



Fugaku's board
Apx. ..20cm x 20 cm, 2
CPU Chips are mounted

Optical
Cables

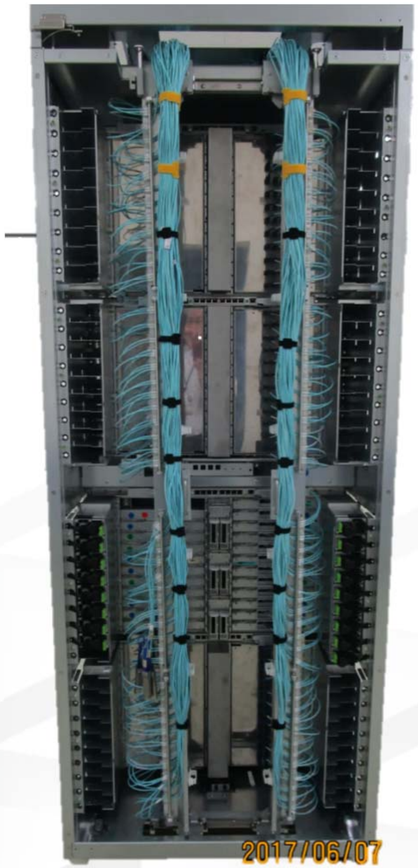


2 CPU / CMU

Water
In/Out

Comparison of Rack

Fugaku’s Rack
384 CPUs



Installed on both
Front and Back
side

K computer rack
96ノード（CPU)

System Performance

- **Peak performance, more than 38 times faster.**
- **#node is more than 150K**
- **10 racks of Fugaku will provide the same peak performance to the K system**
- **Power consumption 、12MW ⇒ 30～40 MW (about 3 times larger)**

X 10 =

Challenges for the future supercomputer

- **More power-performance**
  - It is necessary to increase the power to achieve higher performance, but the power is limited.
  - Supercomputers in the US (and China) use accelerator mechanism (GPU, etc.) to improve power efficiency, but they cannot be applied to all apps or have to rewrite the program.
- **The slow-down of the progress of silicon technology**
  - The end of “Moore’s” low, post-Moore tech, a new device …
  - A new computing paradigm, such as quantum computing
- **Integration of Big data and AI technologies**
  - Since it will be possible to execute many cases of relatively large simulations, data processing and integration with data processing are required
  - A new market called AI or a new computing technology called AI