RIKEN AICS Software Center*

*Official name TBA

Mission

RIKEN AICS Software Center aims to:

- develop and deploy high quality applications, libraries, programing tools, etc. (called “AICS software”) for many platforms including the K computer and “Post-K”.
- support AICS software users to promote High Performance Computing for various fields of science and engineering.

RIKEN AICS Software Line-up

http://www.aics.riken.jp/en/k-computer/aics-software

More than 30 softwares have already been developed and ported by RIKEN AICS and are available for HPC users.

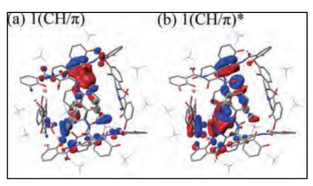
OACIS

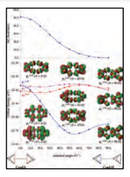
Organizing Assistant for Comprehensive and Interactive Simulations

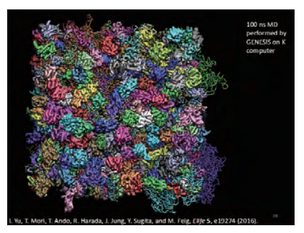
OACIS is a job management software for large scale simulations. With a user-friendly interface of OACIS, you can easily submit various jobs to appropriate remote hosts. After these jobs are finished, all the result files are automatically downloaded from the remote hosts and stored in a traceable way together with logs of the date, host, and elapsed time of the jobs. It also provides APIs.

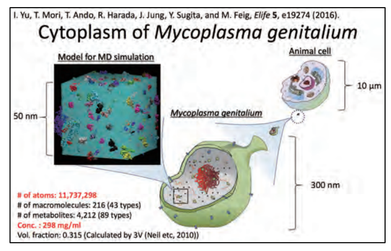
NTChem

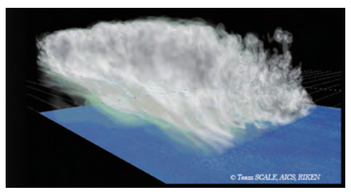
NTChem is a high-performance software package for the molecular electronic structure calculation for general purpose on the K computer. It is a comprehensive new software of ab initio quantum chemistry made in AICS from scratch. NTChem contains not only standard quantum chemistry approaches but our own original approaches.

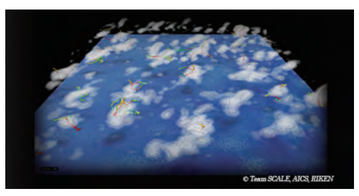






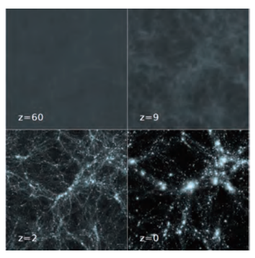


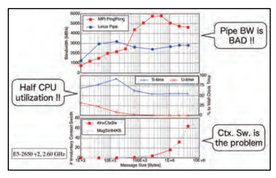


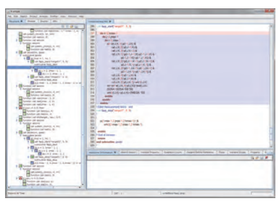


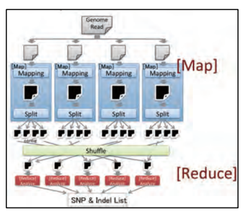
GENESIS

GENESIS (GENeralized-Ensemble SImulation System) is molecular dynamics and modeling software for bimolecular systems such as proteins, lipids, glycans, and their complexes. GENESIS is open source software distributed under the GPLv2 license.







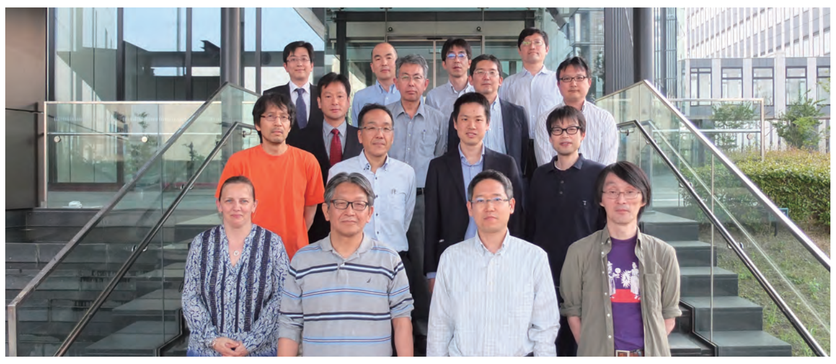


SCALE

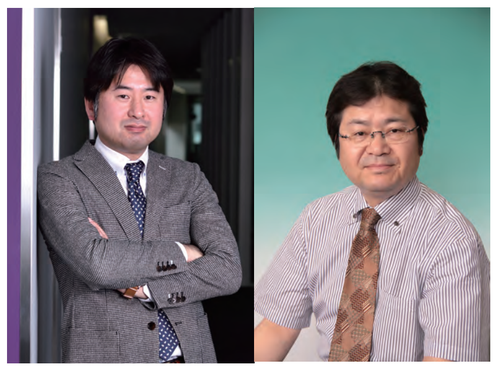
SCALE (Scalable Computing for Advanced Library and Environment), which stands for Scalable Computing for Advanced Library and Environment, is a basic library for weather and climate model of the earth and planets aimed to be widely used in various models.

Other softwares:

Profiling/Analisys tool	7 softwares
Job management tool	2 softwares
Language	1 software
I/O support tool	3 sotware
Numerical library	4 softwares
Application/Platform	2 softwares
Visualization	4 softwares
Misc.	9 softwares



Software development and enhancement
Research Div. (16 teams & 2 units)



Promotion and user support
Operations and Computer Technologies Div. (2 teams)

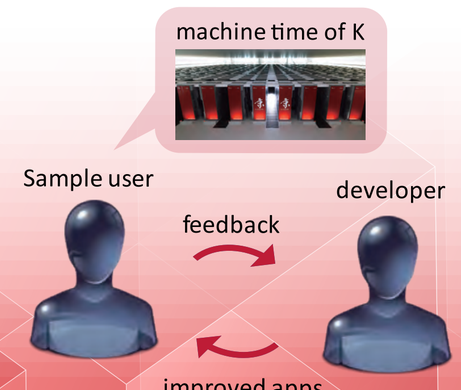
Sample User Program (in preparation)

To accelerate the development of the AICS software and encourage new HPC users, we start “**Sample User Program**”:

1. You apply to join the program as a sample user of AICS software (OACIS, NTChem, GENESIS or SCALE).
2. We provide sufficient machine time of the K computer and special user support for 1 year.

→

- You give us feedback for improvement and enhancement of the software.
- We improve and enhance the software and You experience a world class HPC environment and Apps.



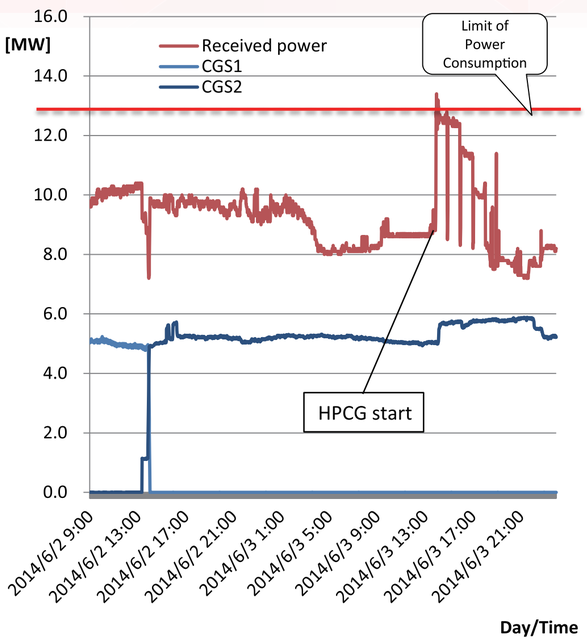
Please contact : **aics-ungi-hud@riken.jp**

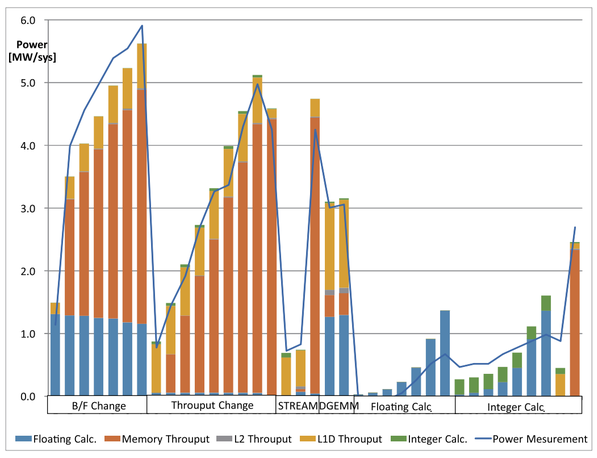
Application Tuning Development Team

Collaborate between system and application to improve system and usability
Team Leader : Kazuo MINAMI contact : minami-kaz@riken.jp

Improvement of Operation Efficiency and Effort for Power Consumption

Recently jobs with large power consumption have increased. And total power consumption exceeded the power consumption limit regulated by contract with power provider. If the total power consumption exceed the power limit, we have to pay additional charge as a penalty.



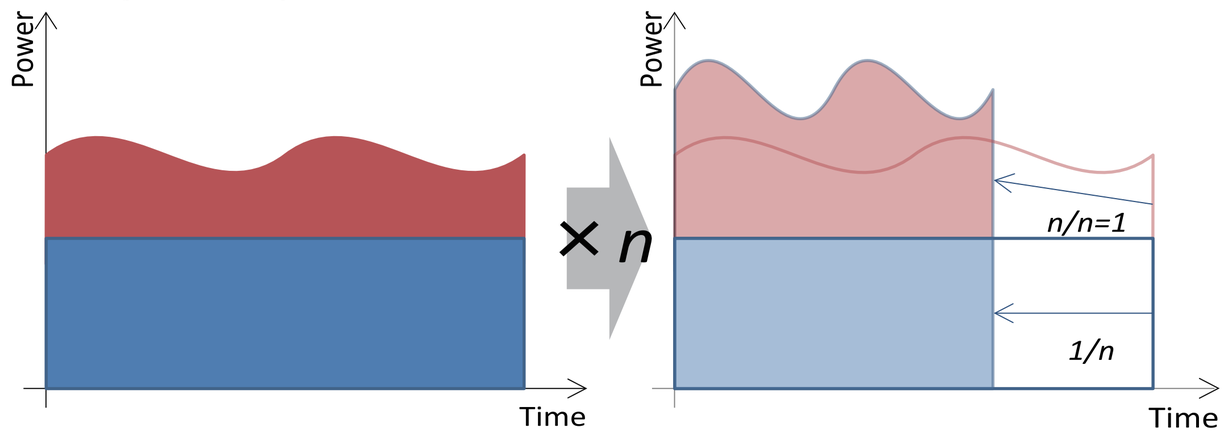


Simple kernel loops

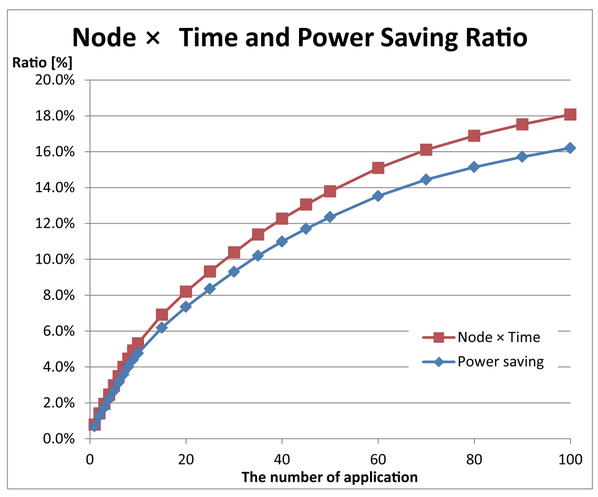
Analysis of correlation with power consumption and application performance was performed. As a result, standby power and memory throughput has a greater impact on the power consumption of the K computer.

$$\delta\text{Power}[\text{MW/sys}] = 8.8128 + 1.3659 \times \text{Floating Calc.}[\%] + 4.3906 \times \text{Memory Throughput}[\%] + 0.0857 \times \text{L2 Throughput}[\%] + 2.3299 \times \text{L1D Throughput}[\%] + 0.2429 \times \text{Integer Calc.}[\%]$$

We performed the trial of improving of power consumption and system operation.



If both the CPU performance and the memory performance improved to field average, NODExTIME can be reducted as below. The performance of each big user's applications has estimated how much differs compared to the field average.

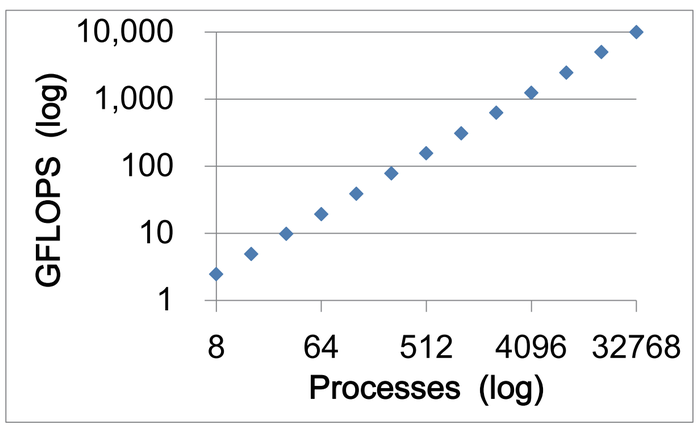


When we improve performance about the 10 (or 50) major user application, we can use 25M (or 66M) NODExTIME.

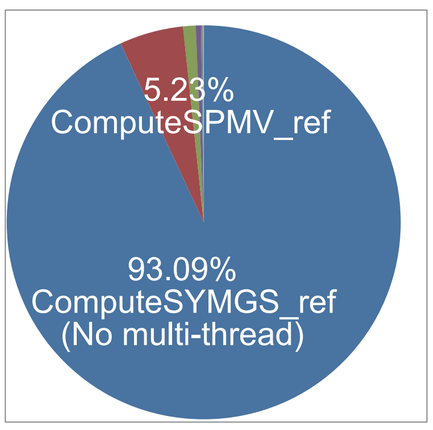
When we improve performance about 10 (or 50) major user application, we can achieve energy savings of 2.7GWh (or 7.0GWh).

HPCG Performance Tuning on the K computer

Evaluate Original HPCG on the K



Weak-scaling performance



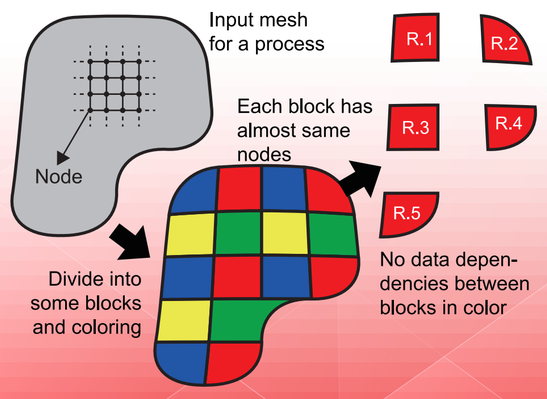
Cost components in CG

The original HPCG code evaluation on the K computer gave the following information.

- Linear scalability was obtained, so parallelization tuning is not necessary
- Single CPU performance was low since SYMGS is not multi-thread

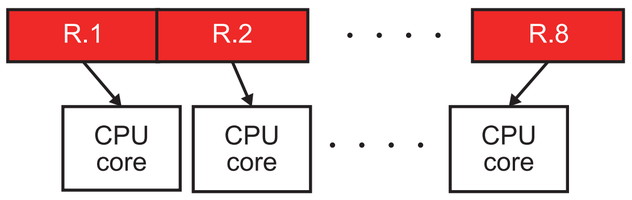
Therefore we have aimed the single CPU tuning

Tune: Coloring for SYMGS

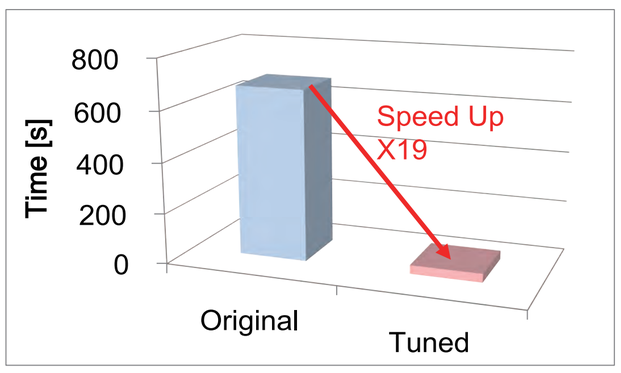


In the original code, SYMGS is not able to be multi-thread since there are data dependencies between rows.
To avoid data dependencies, we employed the colored blocking that divide the mesh into some blocks and do the coloring to blocks.

There are no data dependencies between same colored blocks, so the blocks are able to be processed with multi thread.



Significant Improving Obtained



Comparing CG run time for 5 CG calls

We tried these additional ways.

- Memory serialize for matrix
- Data access ordering improvement for SYMGS
- Loop optimization for SPMV and SYMGS
- Parameter adjustment
- Improvement miscellaneous routines

Then, **19 times speed up** was obtained finally.

Result of ISC2017

Rank	Computer	Country	HPL PFLOPS	HPCG PFLOPS	Ratio to HPL %
1	K computer	Japan	10.510	0.6027	5.7%
2	Tianhe-2	China	33.863	0.5801	1.7%
3	Sunway TaihuLight	China	93.015	0.4808	0.5%
4	Piz Daint	Swiss	19.590	0.4767	2.4%
5	Oakforest-PACS	Japan	13.555	0.3855	2.8%

http://www.hpcg-benchmark.org/custom/index.html?lid=155&slid=291

K Computer Hardware and Operations

System Configuration

The K computer
Compute nodes

Number of CPUs : 82,944
Total memory capacity : 1.27PB

Tofu interconnect
(6D mesh/torus network)

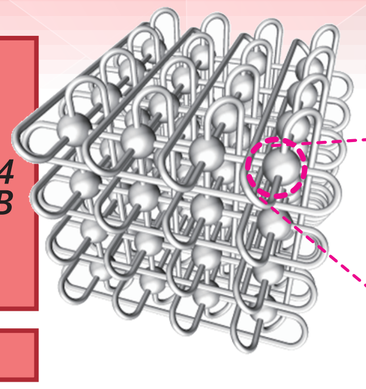
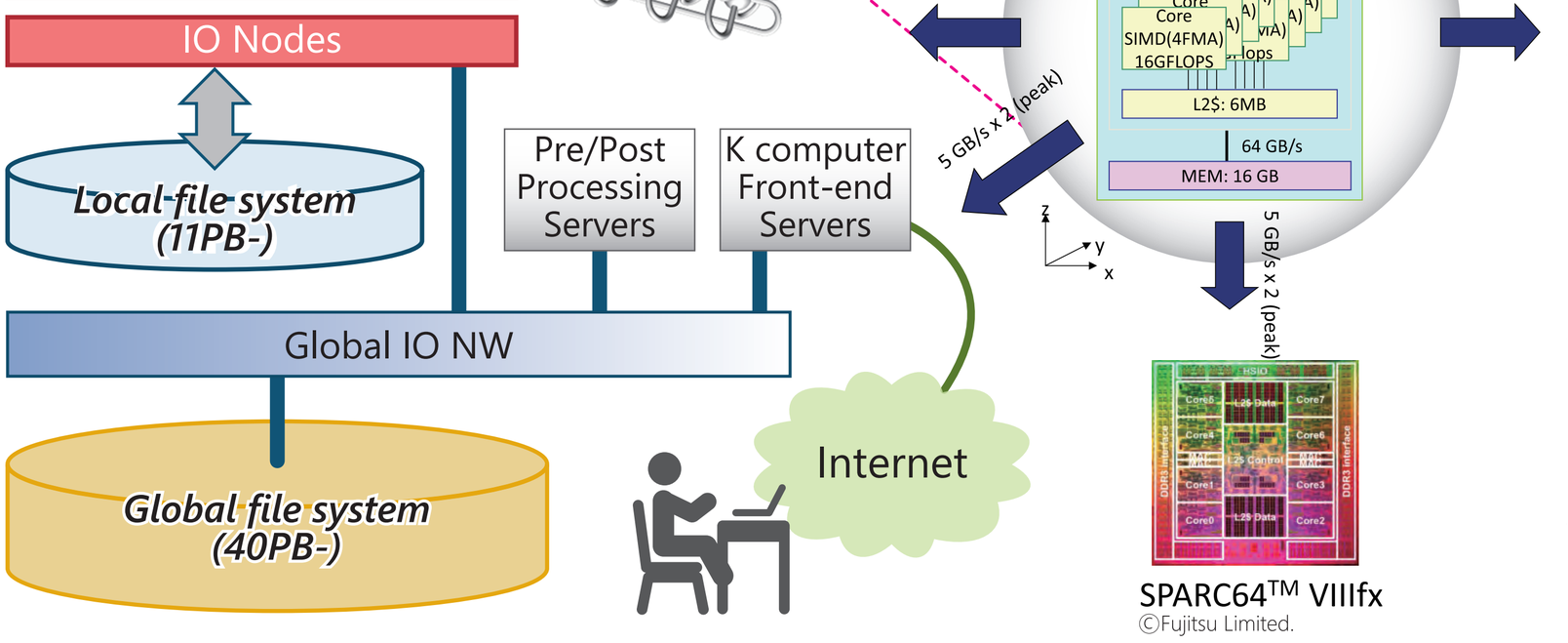


Image of 3D torus
©Fujitsu Limited.



SPARC64™ VIIIfx
©Fujitsu Limited.

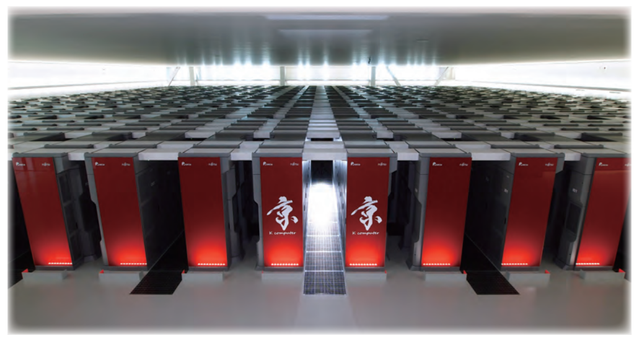
SPARC64™ VIIIfx
Superscalar Multi-core processor

	Specification
Performance (peak)	128 GFLOPS (16 GFLOPS x 8 cores)
Core	8
Clock	2.0 GHz
Floatingpoint Execution units (Core spec)	FMA x 4 (2 SIMD) DIVIDE x 2 COMPARE x 2
	Floating-point register (64bit) : 256 General purpose register (64bit) : 188
Cache	L1I$: 32 KB (2way) L1D$: 32 KB (2way) L2$: Shared 6 MB (12way)
Memory throughput	64 GB/s (0.5B/F)

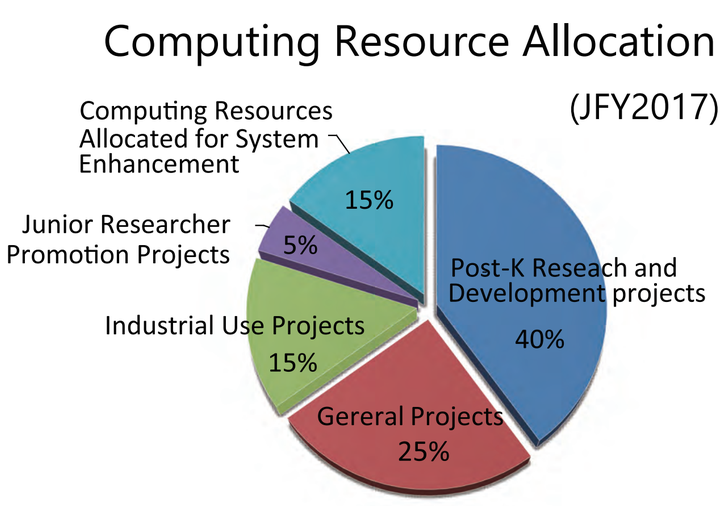
- Peak Performance: 10.62 PFLOPS
- Total Memory Capacity: 1.26 PiB (16 GiB/ node)
- Interconnect: Tofu
  - 6D mesh/torus, Logical-1,2 or 3D torus
  - Peak bandwidth: 5 GiB/s x 2
  - Bisection bandwidth: 30 TiB/s

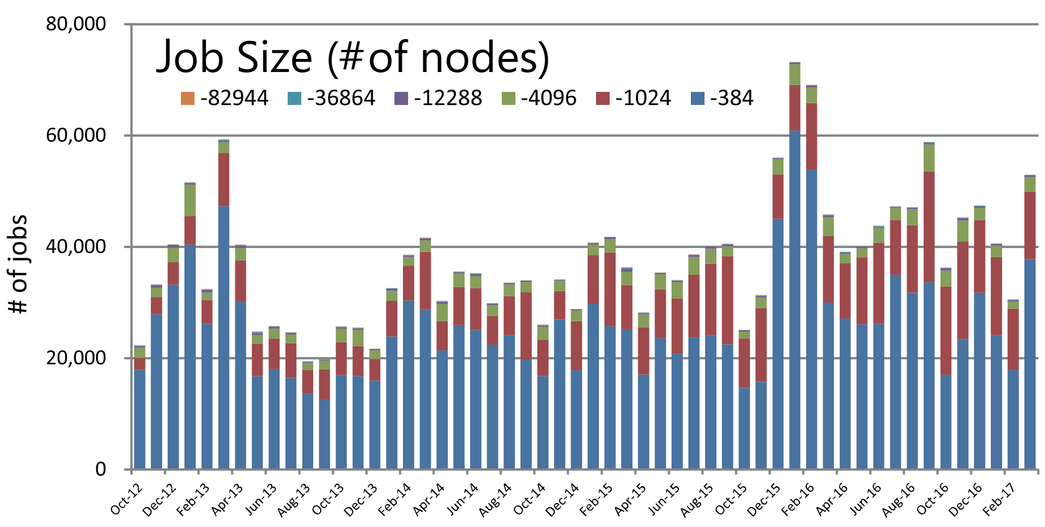
Tofu (Torus Fusion) interconnect

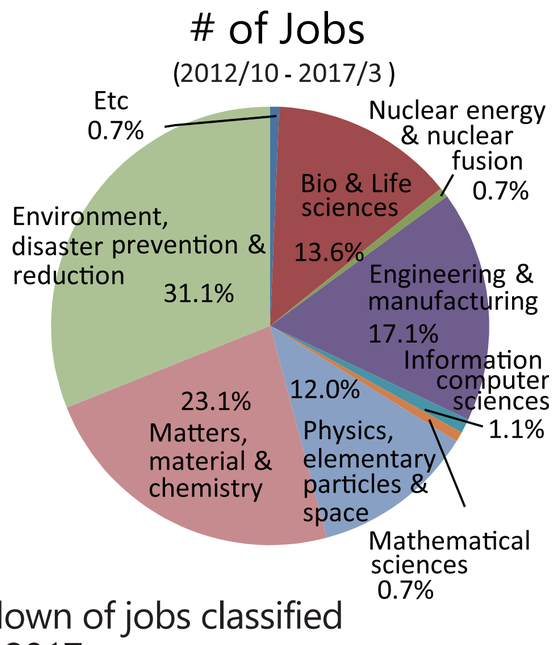
- High communication performance and fault-tolerant network
- Network topology: 6D mesh / torus network
  - 10 links (5 GiB/s x 2 bandwidth / link) on each node
  - Axis: X, Y, Z, a, b, c
    - X,Z,b: torus (Z=0: IO node), Y, a, c: Mesh
  - Network size: (X, Y, Z, a, b, c) = (24, 18, 17, 2, 3, 2)
  - 1, 2 or 3D torus network configurable from user's programming point of view



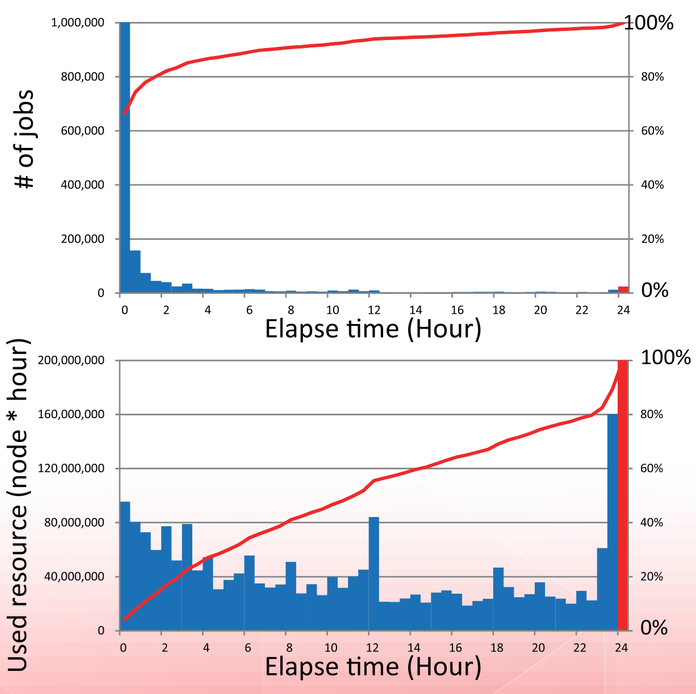
Operation Statistics (Oct. 2012 – Mar. 2017)

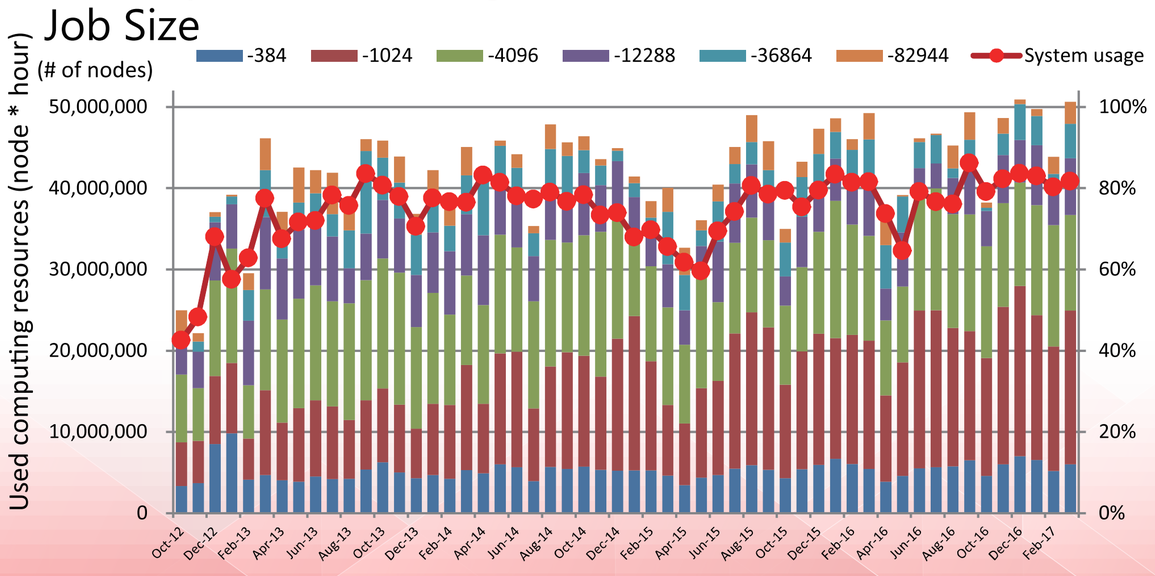


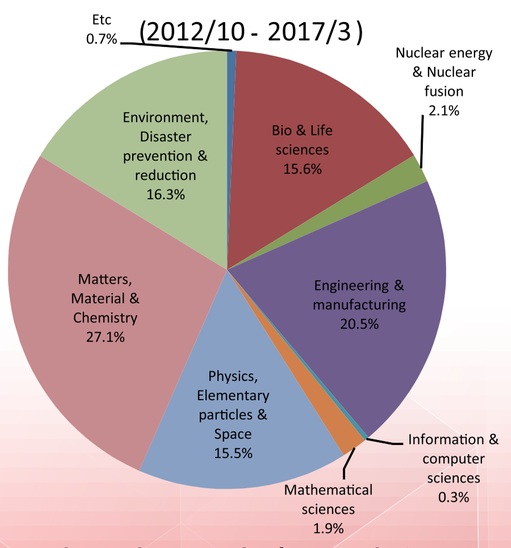




Number of executed jobs of each month (left) and a breakdown of jobs classified by research areas (right) between October 2012 and March 2017.







Number of executed jobs (upper) and amount of resources used (lower) by elapsed job time (rightmost red rectangle denotes jobs over 24 hours elapsed time; solid lines are cumulative values)

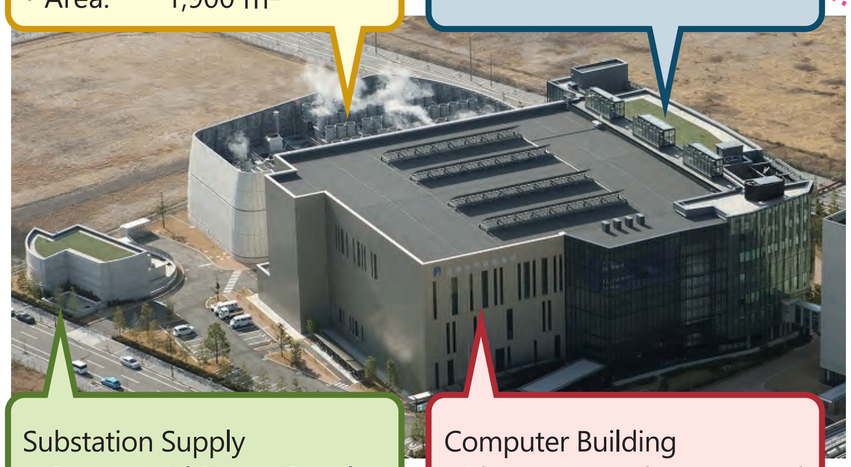
Amount of used computing resources (product of number of used nodes and elapsed time) and system usage of each month (left) and a breakdown of used resources classified by research areas (right) between October 2012 and March 2017.

K Computer
Hybrid Cooling System

- **High-efficiency**
  - Efficient Power generation and eco-system (CGS)
  - PUE～1.34
- **Safety**
  - Earthquake preparedness
  - Countermeasures to power interruptions and voltage dips without using a UPS

Chillers & Power Generator
- Chillers : 10,500 USRt (～37MW)
- CGS: 6MW (peak)×2
- Area: 1,900 m^2

Research Building
- Six stories above ground and one below
- Area: 1,800 m^2



Substation Supply
- Commercial Power Supply : 11.4MW / 77,000V
- Area: 200 m^2

Computer Building
- Three stories above ground and one below
- Area: 4,300 m^2

Computer Building

Computer room
(50m×60m = 3,000m^2 pillar-free)
K computer system

Cold Air / Water

Air/Water-Handling Unit Room
Air-Handling Unit &
Water-Handling Unit

Data Storage Room
Global File System

Air-Handling Unit Room
Air-Handling Unit

Research Building
Room
Room
Room
Room
Room
Room

Earthquake preparedness
- Laminated-rubber seismic isolation device
- Lead damper
- Steel damper

Protected area from power interruptions and voltage dips

Chillers & Power Generator

CGS: Co-Generation System

Energy efficiency 75%

Energy 100% → Gas → CGS

Electric power 30%

Steam

Absorption Refrigerating Chillers

Water-cooling 45%

Exhaust 25%

- High-efficiency air/water hybrid cooling system
  - CPU temperature: 17-25℃
  - Room temperature: 24℃ (peak)
- High reliability
  - MTBF (All SBs): 2-3 days

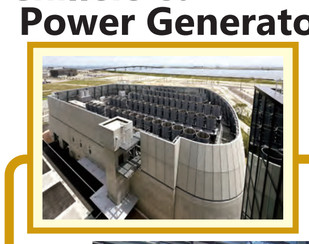
Substation Supply

11.4MW (average)
(77,000V⇒6,600V)

System-separation Circuit Breaker
Protect from
- Power interruptions
- voltage dips

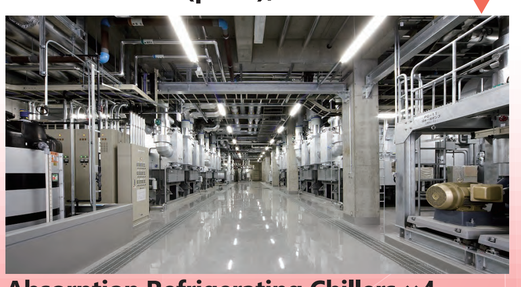
Protected Area
- Data Storage room
- Research Building

Chillers & Power Generator



CGS ×2: 6MW (peak), 6600 V

Steam



Absorption Refrigerating Chillers ×4
Centrifugal Water Chillers ×3

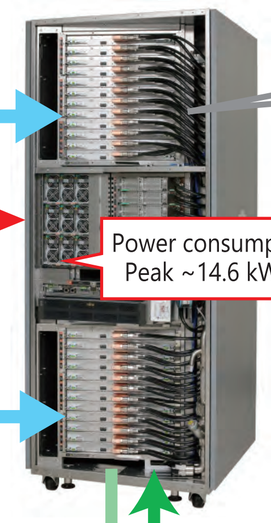
Computer Building

Voltage Converter

400V

Air Outlet ~23℃

Air Inlet ~17℃



Power consumption / rack
Peak ~14.6 kW (210 V)

44ℓ over / min / rack

2nd Coolant Loop

Water outlet ~17℃

Water inlet ~15℃

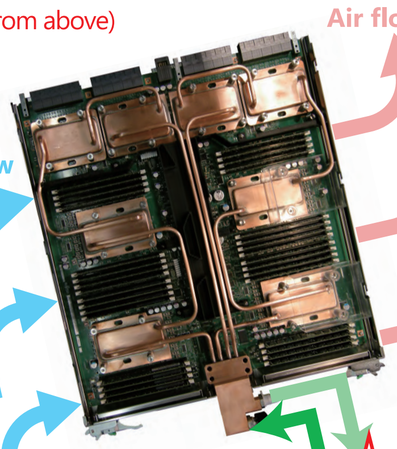
Air-Handling Unit ×50

400V

Water-Handling Unit ×5

1st Coolant Loop

System Board
(View from above)



- Power consumption / CPU 58W (ICC: 28W)
- CPU temperature keep range: 17-25℃
- Cold water circulate about 20 times per minute