JSS3/TOKI 概要と大規模チャレンジ速報

藤田 直行（宇宙航空研究開発機構）

JSS3/TOKI Overview and Large-Scale Challenge Breaking Report

FUJITA Naoyuki (JAXA)

ABSTRACT

All systems of the third generation HPC system *JSS3* of Japan Aerospace Exploration Agency (JAXA) started operation on December 1st, 2020. This is an outline of *TOKI*, which is the computer infrastructure part of *JSS3,* which also has *J-SPACE* as an archiver infrastructure. In addition, the purpose and the selected challenges of Large-scale Challenge performed at a time when it is easy to secure large-scale computational resources in the early days of *TOKI* operation will be introduced.

1. Introduction

JAXA has been researching and developing numerical simulation technology and its indispensable supercomputer since the days of its predecessor organization (Figure 1). In 1977, *FACOM 230 -75APU* was introduced, and in 1993, the world's fastest *Numerical Wind Tunnel* (NWT) was installed to raise Computational Fluid Dynamics (CFD) to the top level in the world. In the 2000s, numerical simulation began to be used in earnest in design, and entered the era of practical use. In 2009, the first JAXA supercomputer *JSS1* started operation, and after *JSS2*, the third generation JAXA supercomputer *JSS3* started operation in December 2020.

The outline of *TOKI*, which is computer infrastructure part of *JSS3*, is described. *JSS3* also has *J-SPACE* as an archiver infrastructure. In addition, we will introduce the purpose and selected three challenges of the Large-scale Challenge that was implemented at the timing when it was easy to secure a relatively large-scale computing resource at the beginning of *TOKI* operation.

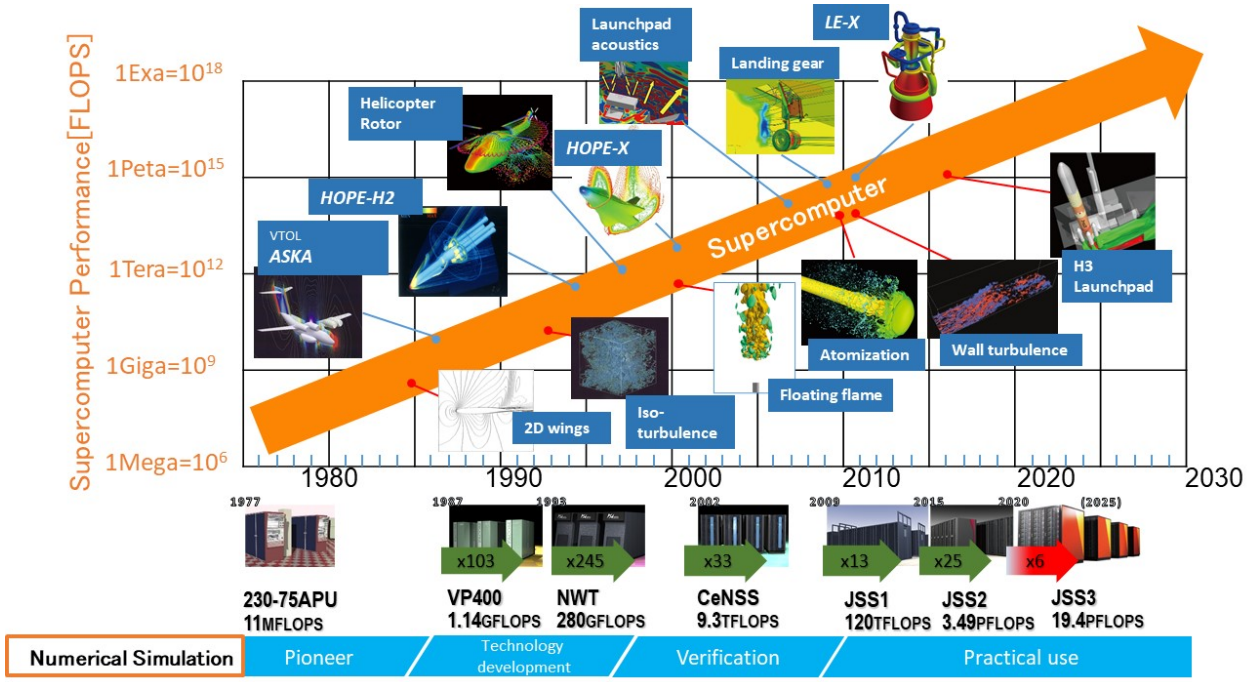

Figure 1 Numerical Simulation and Supercomputer

2. Abstract of JSS3/TOKI

JSS3 consists of *TOKI*: TOkyo and ibaraKI, which is a computer infrastructure, and *J-SPACE*: Jaxa's Storage Platform for Archiving, Computing and Exploring, which is an archiver infrastructure. *J-SPACE* is not the subject of this paper, so details are not shown, but it is a large-scale storage system that manages the disk cache 3PB and tape cartridge 70PB in a hierarchical manner using the software *HPSS*: High Performance Storage System [(1)]. It is interconnected with *TOKI* via supercomputer network named *JSSnet* with a total bandwidth of 80 Gbps.

2.1. Abstract of TOKI

As mentioned above, CFD, which has been developed together with supercomputers, has entered the era of practical use, and as usage fields of supercomputers have expanded not only to CFD but also to data processing such as earth observation data processing and machine learning. The introduction of JAXA's third generation supercomputer has been promoted as a system that plays the following three roles.

(1) Numerical simulation implementation platform to strengthen international competitiveness in the aerospace field

(2) Data center function as a large-scale data analysis platform

(3) Research and development platform to meet new needs

Figure 2 shows the system configuration diagram of *TOKI* and *J-SPACE*. The meaning of each system name is as shown in Table 1.

Next, the role and features of each system will be described.

(1) TOKI-SORA

It is a system for ultra-high parallelism, long-time computation and data processing. It is a Fujitsu FX1000 with 5,760 nodes, peak computational performance 19.4PFLOPS, and total main memory capacity 180TiB. The CPU is an Arm architecture A64FX.

(2) TOKI-RURI

It is a system to calculate large-scale problems that are parallelized in a short time, run many small-scale jobs per unit time, perform processing that requires a large amount of main memory in a single job, run commercial applications, etc. There are four types of nodes (2a) to (2d) shown below, which consists of Fujitsu's PRIMERGY with a total of 416 nodes. The CPU has an intel architecture. In addition, the system configuration is designed with consideration for cooperation with external activities and convenience of use, such as virtualization support for Singularity and VMware, login with remote desktop, etc.

(2a) TOKI-ST

A node with a standard configuration of TOKI-RURI. It has two CPUs in one node and has 192GiB DDR4 memory and one GPU board for visualization processing. The interconnect has one InfiniBand HDR100 port.

(2b) TOKI-GP

Equipped with 4 GPGPUs (NVIDIA V100[(2)]) per node, the main memory is 384GiB DDR4, and the interconnect has 2 InfiniBand EDR ports. It is a node responsible for computations that require medium-scale main memory and machine learning using GPGPU.

(2c) TOKI-XM

A node that uses intel's Optane DC Persistent Memory (DCPMM[(3)]) to achieve a 1-node main memory capacity of 6TiB. DDR4 used for the main memory cache implements 768GiB, which is 1/8 of the DCPMM capacity. This node is used for processing that requires a larger main memory capacity than computing power and processing that requires very huge amount of main memory capacity.

(2d) TOKI-LM

The role and features are the same as TOKI-XM, but it is a node with 1/4 of the main memory capacity. It is equipped with DCPMM 1.5TiB and DDR4 192GiB.

(3) TOKI-LI

It is a *TOKI* login node and there are a total of fourteen nodes. Users login with ssh to compile programs and operate the file system.

(4) TOKI-FS

Until now, JAXA's supercomputer system has been configured to provide performance with sequential access, which is often used in numerical simulations, but *TOKI*'s file system TOKI-FS has also designed as a file system with high random-access performance. SSD: Solid State Drive device has a capacity of 10PB and HDD: Hard Disk Drive device

has a capacity of 40.5PB, for a total capacity of 50.5PB. Lustre-based software FEFS: Fujitsu Exabyte File System[4] manages these devices.

(5) TOKI-Txx

(1) to (4) are installed at the JAXA Chofu Aerospace Center (CAC), but similar system with a reduced scale except for (1) TOKI-SORA and (2c) TOKI-XM are installed at JAXA Tsukuba Space Center (TKSC). By installing it in TKSC, the processing capacity will be reduced compared with TOKI-RURI, non-stop supercomputer operation can be realized, even if CAC is power outage or Chofu system is under maintenance. The system installed at TKSC is called TOKI-TRURI /-TLI/-TFS.

Table 2 shows the introduction of each system introduced in Section 2.1.

Table 1 Name of each system

Name	Meaning
JSS3	Jaxa Supercomputer System generation 3
TOKI	TOkyo and ibaraKI, and Time (TOKI-SORA means time and space), Solve, Crested ibis, in Japanese
SORA	Supercomputer for earth Observation, Rockets, and Aeronautics
RURI	all-RoUnd Role Infrastructure
ST	STandard
GP	GPgpu
XM	eXtra large Memory
LM	Large Memory
FS	File System
LI	LogIn
-Txx	-Tsukuba xx
J-SPACE	Jaxa's Storage Platform for Archiving, Computing and Exploring

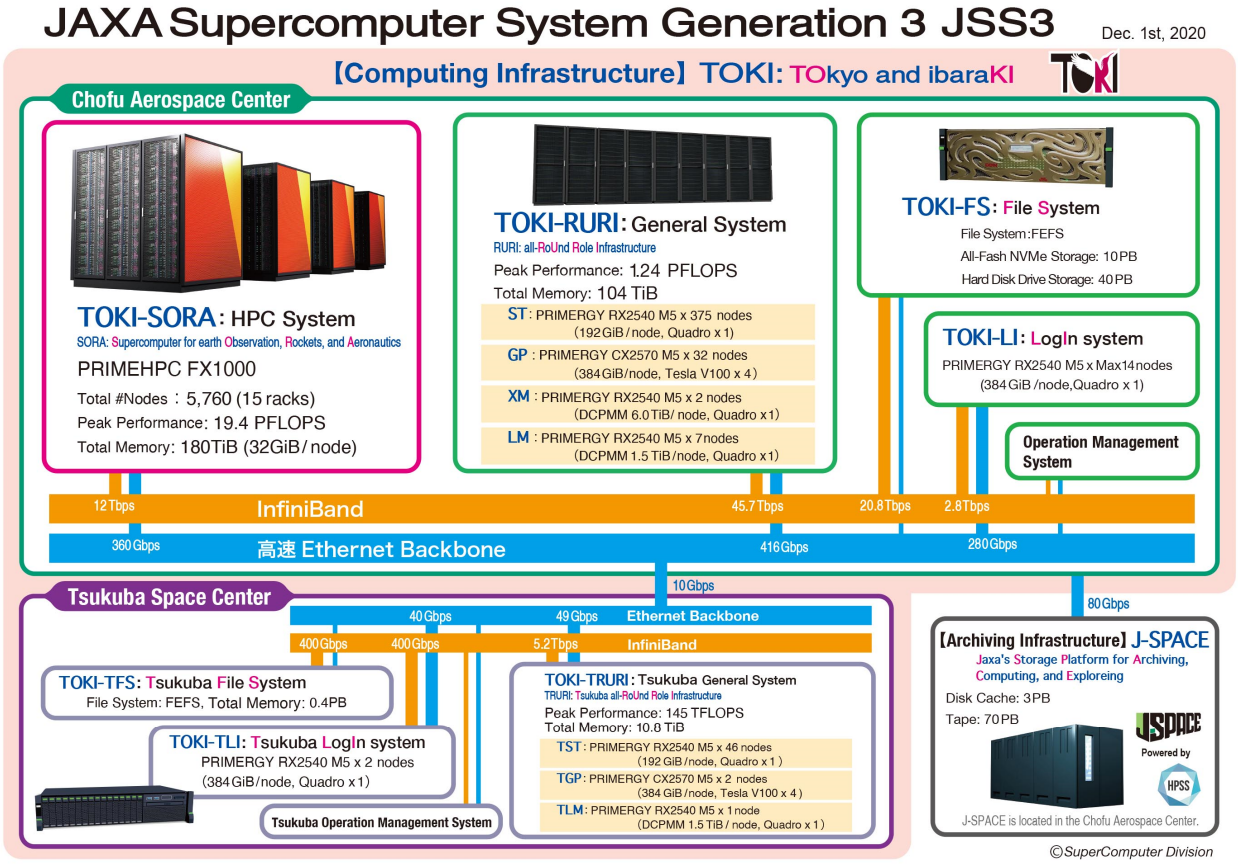

Figure 2 System configuration diagram of TOKI/J-SPACE

2.2. Usage of TOKI

Not to mention *TOKI*, JAXA supercomputers are installed and operated to contribute to the steady implementation of JAXA projects and to maximize research and development results. This contribution includes the followings; joint research and development with outside organizations, space science research activities closely related to inter-university aerospace-craft project (scientific satellites, rockets, balloons) etc. , a system for paid use by private companies, universities, research institutes, etc. in the aerospace-related field. Table 3 summarizes the usage system. The Supercomputer Division is the contact point for inquiring to regard the use of *TOKI* (https://www.jss.jaxa.jp/en/inquiry/).

Table 2 Introduction to each system of TOKI

System name / Introduction	TOKI-SORA	TOKI-ST [TOKI-TST]	TOKI-GP [TOKI-TGP]	TOKI-XM	TOKI-LM [TOKI-TLM]
Number of nodes	5,760	375 [46]	32 [2]	2	7 [1]
Total peak calculation performance	19.4PFLOPS	TOKI-RURI totally 1.24PFLOPS [TOKI-TRURI totally 145TFLOPS]			
Total main memory	180TiB	TOKI-RURI totally 104TiB [TOKI-TRURI totally 10.8TiB]			
Model name	PRIMEHPC FX1000	PRIMERGY RX2540 M5	PRIMERGY CX2570 M5	PRIMERGY RX2540 M5	PRIMERGY RX2540 M5
CPU	A64FX	Intel Xeon Gold 6240	Intel Xeon Gold 6240	Intel Xeon Gold 6240L	Intel Xeon Gold 6240
Number of cores/CPU	48	18			
Number of CPUs/node	1	2			
Main memory/node	32GiB	192GiB	384GiB	6.0TiB	1.5TiB
GPU	0	NVIDIA Quadro P4000 x1	NVIDIA Tesla V100(SXM2) x4	NVIDIA Quadro P4000 x1	NVIDIA Quadro P4000 x1

Table 3 TOKI usage system

Usage system name	Overview	Target users
Specified Project Use	Use on strategic themes	It can be used by companies, universities, research institutes, etc. for joint research and development related to the JAXA business.
General Use	Standard use	
Small scale Use	Experimental and sprouting use	JAXA staff only
JSS Inter-University Research	Use as Inter-University joint research	Researchers belonging to universities and national and public research institutes nationwide, or researchers equivalent thereto. Research members are limited to graduate students and those of equivalent or higher.
Facility Utilization	Paid use by aerospace-related field organizations. We prepare a trial use.	It can be used by companies, universities, research institutes, etc. based on JAXA's facility service system.

3. TOKI Large-scale Challenge

When introducing *TOKI*, we carried out Large-scale Challenge computations that cannot be performed in normal operation. The purpose is to produce impactful results for society through applications in the aerospace field. To *JSS* users, including inside and outside JAXA, we recruited the theme. As a result, the following three research have been selected and *TOKI* was being used in a truly challenging manner. The allocated computational resource was less than 967,680 [node*hours] per theme (2,880 nodes (9.7 PFLOPS), memory 90Tib, 14 days). In addition to the analysis execution period for Large-scale Challenges, a program preparation period was set about one month, and the minimum number of nodes required for debugging was allocated.

In this part, we will briefly quote from the *JSS* usage result report[(5)] and some application materials, and introduce the challenges that each research aims for.

3.1. LES of Full-scale Liquid Rocket Engine Combustor

Focusing on the spatial distribution and time fluctuation of the wall heat flux assumed in the actual machine, by performing a complete 3D analysis of the actual shape, acquire knowledge that cannot be obtained by conventional simplified analysis (one to several injectors, cake-cut shape assuming symmetry in the circumferential direction). The analysis target is the full-scale combustor (528 injectors) of the LE-X engine. A large-scale combustion LES analysis was performed using an efficient table-referenced combustion model. There has never been an example of such analysis with more than 100 injectors. This will be possible for the first time by combining the computing performance of JSS3 with the next-generation large-scale high-speed solver LS-FLOW-HO that has been developed so far.

3.2. Large-scale analysis of multi droplet evaporation by Interface-resolved DNS

Reduction of NOx emissions is required in aircraft engine development, and combustion with a homogenized fuel distribution is the key. However, in the spray model used for the analysis of fuel "atomization" ⇒ "dispersion" ⇒ "evaporation", it is necessary to adjust parameters by combining with experiments to obtain an analytical solution of the fuel distribution that can be compared with the current measured values. This is one of the major reasons why the use of numerical analysis is not progressing in combustor development. In this challenge, "evaporation", which is particularly difficult to measure and is not sufficiently modeled, will be modeled from a detailed numerical analysis approach, and a spray evaporation model that can be used in the upstream stage of combustor design will be established.

3.3. 3D visualization of large-scale full-color tomography data: Break through the resolution of the human retina

Visualization technology to convert meteorites and various rocks of the earth into large-scale data has been developing, but the amount of information contained in this high-resolution data is enormous and cannot be processed by laboratory workstations. In addition, it is a size that could not be accommodated by the computing power normally available at supercomputers. Therefore, we aim to overcome this bottleneck by this Large-scale Challenge and achieve ultra-high resolution 3D visualization.

4. Conclusion

JSS3 / *TOKI*, which plays three roles, was prepared and started full-system operation in December 2020. At the beginning of operation, three challenges aiming for the world's first results have been adopted as Large-scale Challenges.

One of the roles of *JSS3* / *TOKI* is to strengthen the international competitiveness of the aerospace field. Not only will it contribute to the steady implementation of the JAXA project, but also it will continue to work with JAXA software to make it a powerful tool for the aerospace industry. We will meet demand through various system operations such as joint research, JSS Inter-University research, paid use. If you have any questions or concerns, please feel free to contact us [(6)].

Reference

(1) High Performance Storage System, https://www.hpss-collaboration.org/, (Referred on 2021-09-13)
(2) NVIDIA V100 Data Sheet, https://images.nvidia.com/content/technologies/volta/pdf/volta-v100-datasheet-update-us-1165301-r5.pdf, (Referred on 2021-09-13)
(3) Intel Optane Persistent Memory, https://www.intel.com/content/www/us/en/architecture-and-technology/optane-dc-persistent-memory.html, (Referred on 2021-09-13)
(4) White paper Advanced Software for the FUJITSU Supercomputer PRIMEHPC FX1000, P.15, https://www.fujitsu.com/downloads/SUPER/primehpc-fx1000-soft-en.pdf, (Referred on 2021-09-13)
(5) JAXA Supercomputer System Annual Report (April 2020 – March 2021), https://www.jss.jaxa.jp/en/ar/e2020/15595/, (Referred on 2021-09-13)
(6) JSS3 public relations page, https://www.jss.jaxa.jp/en/inquiry/, (Referred on 2021-09-13)