Looking Back on Supercomputer Fugaku Development Project

Yutaka Ishikawa
RIKEN Center for Computational Science
Project Leader, Flagship 2020 Project

For the purpose of leading the resolution of various social and scientific issues surrounding Japan and contributing to the promotion of science and technology, strengthening industry, and building a safe and secure nation, RIKEN (The Institute of Physical and Chemical Research), and Fujitsu have worked on the research and development of the supercomputer Fugaku (hereafter, Fugaku) since 2014. The installation of the hardware was completed in May 2020 and, in June, the world’s top performance was achieved in terms of the results of four benchmarks: Linpack, HPCG, Graph500, and HPL-AI. At present, we are continuing to make adjustments ahead of public use in FY2021. This article looks back on the history of the Fugaku development project from the preparation period and the two turning points that had impacts on the entire project plan.

1. Introduction

The supercomputer Fugaku (hereafter, Fugaku) is one of the outcomes of the Flagship 2020 Project (commonly known as the Post-K development), initiated by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) starting in FY2014. This project consists of two parts: development of a successor to the K computer and application development for social and scientific issues (priority issues) on which to focus by using the developed computer. At the start of the project, the name of the successor had not yet been decided, so Post-K name used (this alias is used in this article). The development of the Post-K was carried out for the purpose of leading the resolution of various social and scientific issues surrounding Japan and contributing to the promotion of science and technology, strengthening industry, and building a safe and secure nation. In order to realize a supercomputer with comprehensive capabilities combining system features including the world’s highest-level performance per power consumption, computing capability, user-friendliness and convenience, and creation of breakthrough results, two performance targets to achieve were set for the system: system power consumption of 30 to 40 MW, and 100 times higher performance than the K computer in some actual applications.

This article looks back on the history of the Post-K project since the preparation period and the two turning points that had impacts on the entire project plan.

2. Conditions before start of project

The development of the Post-K computer can be traced back to the Workshop on Strategic Direction/Development of High Performance Computers (SDHPC), which the author started as a grass-roots activity in 2010, four years before the start of the project, together with the University of Tokyo, University of Tsukuba, Tokyo Institute of Technology, and Kyoto University. At that time, after the budget screening in autumn 2009, the MEXT positioned supercomputers such as those at various university information technology centers as the innovative High-Performance Computing Infrastructure (HPCI) centering on the K computer and started the building of the computing environment to meet the needs of various users. Also, supercomputers based on the T2K Open Supercomputer specification, which was formulated as a result of joint research by the
University of Tsukuba, University of Tokyo, and Kyoto University, were in operation at that time. Meanwhile, the International Exascale Software Project (IESP) \[1\] launched in 2008, and discussions started on the technical issues for exascale machines and the roadmap.

The following is the author's explanation of the aim made at the first SDHPC Workshop:

“We will discuss with application developers and developers of numerical computing libraries, programming languages, middleware, system software, and hardware what kind of system specifications are possible for high-performance parallel computer systems that can be operated at the center in five years and what research and development should be done for them. All participants will have about 10 minutes to express their opinions and discuss using the projector. We look forward to the personal opinions of researchers and engineers independent of the organizations to which they belong. We welcome participation by young researchers and engineers.”

The following two points were ground-breaking in this workshop.

- It provided an opportunity for young researchers and developers of applications, system software, and hardware to gather together for discussions.
- They could discuss not from the standpoint of the organizations to which they belong but from an individual perspective.

Subsequently, the Working Group on the Study of Future HPC Technology R&D was launched under the HPCI Plan Promotion Committee [2] of the MEXT, under which the Application Working Group and the Computer Architecture/Compiler/System Software Working Group were established [3]. SDHPC was integrated with the activities of the latter working group and, with the Workshop held 11 times in total, the White Paper on the HPCI Technology Roadmap was organized as part of the Report on the Future HPC Technology Development, which was submitted to the HPCI Plan Promotion Committee in 2012.

In addition, applications were publicly invited for the MEXT’s Feasibility Study on Future HPC Systems (hereafter, Feasibility Study) and, for two years from July 2012 to March 2014, the Information Technology Center, the University of Tokyo, of which the author served as the Director, acted as the representative organization to carry out the Feasibility Study on Advanced

and Efficient Latency Core-based Architecture for Future HPCI System together with companies such as Fujitsu. The results of this study provided the basis of Fugaku.

3. Project background

While the Feasibility Study made progress, RIKEN proposed in 2013 a system composed of a general-purpose CPU and acceleration unit for the computation node based on the idea of the following three system design concepts as the Post-K.

1) Design in a science-driven manner. That is, design shall be advanced for the purpose of providing computing resources necessary for resolving social and scientific issues based on the Computational Science Roadmap (Ver. 2) in the Report on the Future HPC Technology Development.

2) Sustainable system. The system shall be one that inherits the assets of the K computer as a successor of the K computer in view of trends in development of future computer systems.

3) System with total cost of ownership (TCO) taken into consideration. A system shall be designed with low power consumption, high software portability, and high fault tolerance.

As stated in the materials [4] of the Council for Science and Technology (the present Council for Science, Technology and Innovation), this system was specified to have a 1 exaFLOPS-class theoretical computing performance with a power consumption of 30 to 40 MW. The development goals set were to aim at an effective application performance 100 times higher than that of the K computer and to carry out application and hardware cooperative design (hereafter, co-design) together with application and hardware developers for the purpose of realizing an exascale machine operable starting in 2020 with excellent performance-power ratio and wide-ranging application execution environments.

The first turning point came in early 2014, when the project was officially started. RIKEN closely examined the development and production costs, including the acceleration unit, and consulted the MEXT’s Feasibility Study on Future HPCI System. The result of the assessment was that, regarding the development of the acceleration unit, there was sufficient feasibility for the technology itself but the development and production costs were estimated to be high and applicability to wide-range applications was limited. This led to the
fmitusc enquiry aed dpa of the k development project. As an alternative, we
considered the use of a GPU as the acceleration unit, but we decided that the adoption of a GPU was not ap-
propriate to carry out the Post-K development without delay because of the uncertainty as to when a GPU sat-
sifying the required performance would appear.

Given these circumstances, it was necessary to
make the system capable of utilizing a wide range of
applications with high effective performance within the
range of the total project cost. Accordingly, we chose
to use the relevant resources for the extension of the
general-purpose CPU without adopting the computing acceleration unit. This is where the value of the theo-
retical computing performance of 1 exaflop class, the
original system target, posed an issue.

In the Feasibility Study together with Fujitsu,
eight types of architectures were assessed. The use of
one as the basis provided a possibility to achieve the
original theoretical performance target. However, al-
though this architecture provided the expectations for
performance of 1 exaflop in Linpack, it was not ac-
cepted in terms of providing a wide range of application execution environments, the project goal, and we did not adopt this architecture. We proposed to the MEXT's System Study Working Group on the Next Flagship System a system that does not hinder the provision of computing resources required for resolving social and scientific issues, aims for an effective performance at
the application level up to 100 times higher than the
K computer by co-design, and is composed only of a
general-purpose CPU. This proposal was discussed and
organized into the Report of the System Study Working
Group on the Next Flagship System [5].

The research and development contract between
RIKEN and Fujitsu started in October 2014 through pub-
ic invitation. The co-design of hardware and software
was carried out together with the nine priority issue
implementation organizations selected by the MEXT.
Applications were selected from each priority issue
implementation organization, and hardware and soft-
ware were designed to improve the execution efficiency
of those target applications. We established 13 working
groups (WGs) for the development of hardware and
system software and nine WGs for applications. At the
beginning, multiple WG sessions were held every day
to vigorously carry out the design. Thanks to the close
collaboration with application developers on the design
and development of hardware and system software
from the initial period of the development, the system
developers could understand the features of the applic-
ations and the application developers could recognize
the structure and performance limits of the hardware
and system software [6]. It also enabled the applica-
tion developers to implement software tuning at an
early date.

The other turning point was the delay in semi-
conductor manufacturing technology. The research and
development plan for this project was decided on the
basis of future trends in semiconductor manufactur-
ing technology investigated around 2012 during the
Feasibility Study period. At that time, we expected that
the Post-K CPU could be implemented using a 10 nm
process technology. In 2016, however, this 10 nm pro-
cess technology had a degree of uncertainty, and it was
found out that the technology would fail to achieve the
performance we needed. There were two alternatives—
to proceed by lowering the target performance without
changing the development deadline, or to delay the
development with the next 7 nm process technology as
the target. After various discussions, we decided to ex-
tend the development period for one to two years and
include new added values.

Around the end of May 2016, RIKEN approached
Fujitsu on the possibility of the implementation of
half-precision floating point arithmetic as one of the
added values. Fujitsu was apprehensive about further
delays in the development process and the idea was
initially denied. However, as half-precision floating
point arithmetic instructions were defined in the final
specification of Arm V8 scalable vector extension (SVE),
Fujitsu implemented the instructions in the A64FX.
Half-precision floating point arithmetic is used in the
AI application field, and Fugaku is now posed to be
used in the AI application field as well. I am grateful
to Fujitsu for making such a bold decision at that time.

In summer 2018, the Post-K prototype started
operation at the Fujitsu Numazu Plant. A performance
as designed was achieved, and the Linux kernel and
tools such as the emacs editor also worked. This made
me recognize anew Fujitsu’s high technological capa-
bilities. Outside Japan, even the OS kernel often fails
to function properly with the first version of a CPU and
most CPUs can only be put to practical use after three
modifications. Fujitsu subsequently brought the final version to tape-out (design completion).

Based on the evaluation results of the Post-K prototype, production was approved subject to deliberation by the MEXT HPCI Plan Promotion Committee and the Council for Science, Technology and Innovation, and the production of the real machine was started. Transportation of the real machine to RIKEN started on December 3, 2019 and ended on May 13, 2020. Installation adjustments by Fujitsu will then continue until the end of December 2020.

4. Conclusion

This article looked back on the history of the Fugaku development project since the preparation period, including two turning points that had impacts on the entire project plan.

At the time of writing in July 2020, part of the machine is offered while in the period of installation adjustment to the respective users of the Gordon Bell Prize Challenge, the MEXT’s Program for Promoting Researches on the Supercomputer Fugaku, and the development of applications for measures against COVID-19. I would like to express my gratitude to the people of Fujitsu who are working together with RIKEN on the management of computing resources. I also appreciate the assistance given in all-night benchmarking for the TOP500, HPGC, Graph500, and HPL-AI benchmarks beginning in late May 2020. Fugaku ranked number one in these four performance benchmarks in June 2020. While the goal of this development project is not to attain the first place in benchmarks, I am very pleased to hear that the people have felt cheered and encouraged.

During the development of Fugaku, some problems occurred that were not mentioned here. These were overcome every time together with Fujitsu. We must continue to work on stabilization and performance improvements ahead of the official start of operation in the spring of 2021.

References and Notes


This article first appeared in Fujitsu Technical Review, one of Fujitsu’s technical information media. Please check out the other articles.

Fujitsu Technical Review
https://www.fujitsu.com/global/technicalreview/