

ACIS Annual Report FY2015

Contents

Preface	5
Mission and Overview	7
Organization	9
I Research Division	11
1 System Software Research Team	15
2 Programming Environment Research Team	29
3 Large-Scale Parallel Numerical Computing Technology Research Team	39
4 HPC Usability Research Team	49
5 Field Theory Research Team	57
6 Discrete-Event Simulation Research Team	63
7 Computational Molecular Science Research Team	69
8 Computational Materials Science Research Team	77
9 Computational Biophysics Research Team	85
10 Particle Simulator Research Team	89
11 Computational Climate Science Research Team	95
12 Complex Phenomena Unified Simulation Research Team	101
13 HPC Programming Framework Research Team	111
14 Advanced Visualization Research Team	119
15 Data Assimilation Research Team	129
16 Computational Chemistry Research Unit	143
17 Computational Disaster Mitigation and Reduction Research Unit	147
18 Computational Structural Biology Research Unit	153
II Operations and Computer Technologies Division	159
19 Facility Operations and Development Team	163

20 System Operations and Development Team	167
21 Software Development Team	173
22 HPCI System Development Team	181
III Flagship 2020 Project	189
23 Flagship 2020 Project	191

Preface

Welcome to the RIKEN AICS Annual Report for 2015.

We are very excited about this annual report that illustrates the significant accomplishments that our AICS teams have achieved this year using the supercomputer K.

RIKEN AICS is the Japanese flagship research institution in computer science and computational science. It is an organization charged with operating the K computer. AICS aims to establish international research center which produces breakthroughs through collaboration between computational and computer sciences. Plotting and developing Japan's strategy for HPC, which includes the development of the post K computer, is also an important mission of AICS.

The K computer demonstrates an extraordinary level of stability and provided exceptional service to science, engineering, and research communities. This is due to the dedicated efforts of the teams of the operations and computer technologies division.

In many science and engineering areas, we see many great results since the K computer started full service in September 2012. Many projects that use the K computer would be difficult or impossible to do elsewhere. AICS research teams have developed common computing techniques for utilization of the K computer, and also coded the very efficient parallelized application program packages. Now the K computer is being used in many different fronts and extending the boundaries of computational science. As you see in this report, our AICS research teams made a great contribution to produce many of these exciting results.

The K computer significantly enhanced the resolution both in precision and the system size through massively parallel computation. In addition, capacity computing opened the possibility of prediction in complex and non-linear/chaotic phenomena. We are now tackling highly complex scientific and social problems with inherent uncertainties and unpredictability. The way to deal with such uncertainty is to simultaneously run multiple ensembles, using many different combinations of parameters, and thus be able to explore the solution. The K computer significantly enhanced the capacity computing in addition to the capability computing.

AICS is also charged with the development of the post K computer. The flagship 2020 project division is carrying out the development of the post K computer. The post K computer will be used to work on innovative solutions to current scientific and social issues. The development of application codes to run on the post K computer is also a mission of the project.

We hope you enjoy reading all the wonderful advances made by RIKEN AICS. We look forward to continuing our activities for more fundamental and bigger science and engineering for years to come.

Kimihiko Hirao
Director, RIKEN AICS

Mission and Overview

Organization

Part I

Research Division

This part presents the research activity of the Research Division of RIKEN Advanced Institute for Computational Science (AICS) for the period of 1 April 2015 to 31 March 2016.

This is the sixth year for AICS since its foundation in 2010. The Research Division continued to operate with 16 teams and 3 units as established in 2012. As of 1 April 2014, the number of researchers in the Division was 126 in total.

The research activity in this fiscal year resulted in ??? journal papers, ??? conference reports, ??? invited talks, ??? posters and presentations, ??? patents and deliverables.

Memorable awards bestowed upon AICS researchers included No. 1 ranking on Graph500 in June and November 2015, back from No. 2 in November 2014. Combined with the No. 4 position on Top500 Ranking and No. 2 on HPCG ranking maintained through 2015, they continued to illustrate the well-balanced and versatile performance of the K computer in its 4th year of operation.

Many research-related events took place in this fiscal year. The 6th AICS International Symposium, which became an annual event at Kobe, was held on 22 – 23 February 2016 under the title “Plans and future for international collaborations on extreme scale computing”. AICS also hosted “Lattice 2015 – the 33rd International Symposium on Lattice Field Theory” on 14 – 18 July, 2015, a major conference in the field of computational particle physics.

We hope that this volume conveys the excitement of research and development at the forefront of computational and computer science being conducted at Research Division of RIKEN AICS to readers.

July 2016

Akira Ukawa
Research Division Director
and Deputy Director
RIKEN AICS

Chapter 1

System Software Research Team

1.1 Members

Yutaka Ishikawa (Team Leader)
Atsushi Hori (Senior Scientist)
Yuichi Tsujita (Research Scientist)
Kazumi Yoshinaga (Postdoctoral Researcher)
Akio Shimada (Research Associate)
Masayuki Hatanaka (Research Associate)
Norio Yamaguchi (Research Associate)
Toyohisa Kameyama (Technical Staff)

1.2 Research Activities

The system software team focuses on the research and development of an advanced system software stack not only for the "K" computer but also for towards exascale computing. There are several issues in carrying out future computing. Two research categories are taken into account: i) scalable high performance libraries/middleware, such as file I/O and low-latency communication, and ii) a scalable cache-aware, and fault-aware operating system for next-generation supercomputers based on many core architectures.

1.3 Research Results and Achievements

1.3.1 PRDMA (Persistent Remote Direct Memory Access)

The PRDMA (Persistent RDMA)[[dist-prdma](#)] is an enhancement of MPI persistent communication primitives to reduce the communication latency and to improve the overlap between computation and communication over an RDMA-enabled interconnect. The RDMA-base transfers can progress the non-blocking communication without CPU intervention, and reduce extra copy overheads and memory consumption for data transfers due to the Zero-Copy feature. The MPI persistent communication is defined in MPI standard since MPI version 1.1 specification. For example, when calling with the same communication parameters from an iterative stencil loop, the MPI persistent communication can avoid the redundant setup cost on every call, including the RDMA buffer address exchanges. Also, the initial costs to schedule the communication requests are amortized over a number of stencil iterations.

We implemented the prototype of the PRDMA protocol over the Open MPI provided on the K computer in FY2012. In FY2013, We improved the performance in the ghost cell exchange pattern, such as derived datatype handling and special handling upon non-periodic boundary condition. Furthermore, we applied the PRDMA to an optimized prototype implementation of MPI-3

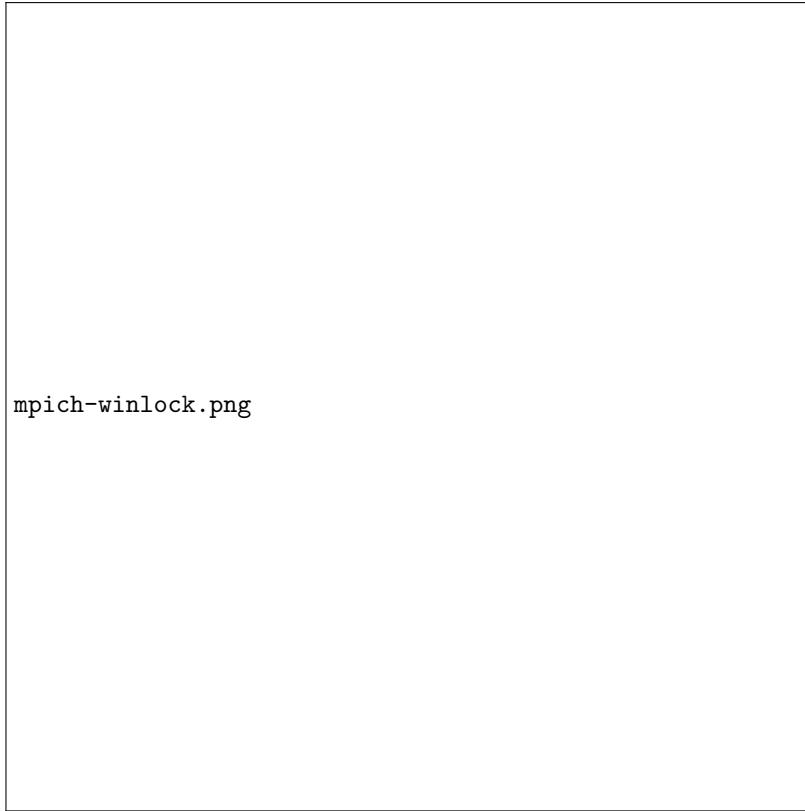


Figure 1.1: Benchmark Result of MPI_Win_lock / unlock with empty critical section

Neighborhood Collectives, as known as MPI_Neighbor_alltoallw, over MPICH on the K computer in FY2014.

In FY2015, we improved the quality of the MPICH on K computer, and made it available for public use on the K computer. In addition, we have been implementing the prototype of the PRDMA-based MPI-RMA implementation on the Tofu2 interconnect of FX100 to compare with the MPI-3 neighborhood collectives. The MPI-RMA passive synchronization such as MPI_Win_lock / unlock does not require the involvement of target process. To implement a truly passive locking on the FX100, we designed a distributed lock queue using RDMA Atomic operations of Tofu2 interconnect. In Figure 1.1, the vertical axis shows the elapsed time in second for 1000 calls of MPI_Win_lock and MPI_Win_unlock with MPI_LOCK_SHARED, and the horizontal axis shows the number of MPI processes which acquires the same lock. The FX10 indicates the result of the Open MPI based generic implementation without RDMA Atomics. The FX100 indicates the result of the MCS-based Readers-Writer lock (Readers Preference) implementation using Tofu2 RDMA Atomics. The FX100 achieves 130 [us] at 256 processes (114 [s] in FX10).

1.3.2 OFI/LLC and RMPI

Two communication libraries have been developed. The first one is called Low-Level Communication Library, which will adopt Open Fabric Interface (OFI) and is called OFI/LLC. The second one is called RIKEN-MPI (RMPI) which is based on MPICH. The relationships between applications, RMPI, OFI/LLC and network drivers are explained by using Figure 1.2. A network driver provides communication functions to OFI/LLC. OFI/LLC provides communication functions to both parallel language runtimes (e.g. MPI library) and applications (e.g. visualization) via a low-level interface. RMPI provides communication functions to applications via a high-level interface.

Two optimizations are performed in FY2015[takagi2015]. The first optimization finds the proper numbers for different kinds of hardware contexts at run-time. The purpose is to maximize the performance while limiting its memory consumption to the amount at which it is possible to run parallel applications with millions of compute nodes.

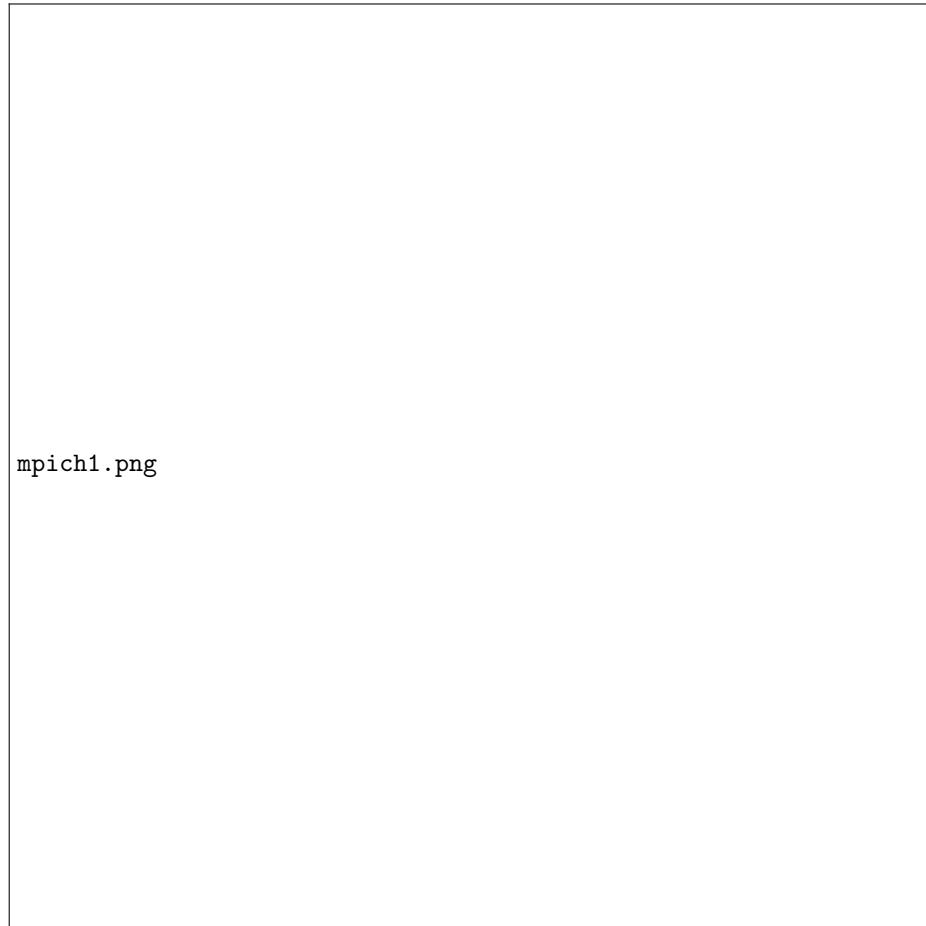


Figure 1.2: Relationships between applications, communication libraries and network driver

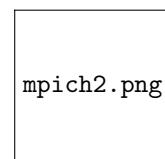


Figure 1.3: Communication latency in all-to-one and one-to-all communication patterns

Table 1.1: Memory consumption per compute node of MPICH and LLC

Software object type	Existing Technique (MiB)	Proposed (MiB)
Communication participants	3,322	204
Software-level communication contexts	15,905	34.5
Hardware-level communication contexts	2,152	106
Buffers for eager protocol	150	151
Buffers for shared-memory	521	118
Table for process management	0.215	0.216
Dynamic buffer and HW context management	-	407
Others	53.3	53.3
Total	22,103	1,074

The conventional network hardware provides a communication model in which the memory-area for communication information for an end-point pair (called context) is never released at run-time. The next generation network hardware adds a new communication model in which the memory-area can be allocated and released at run-time to save memory. However, you lose performance just to make all end-point pairs use the new model because it has a certain amount of performance overhead when allocating. Therefore, it is needed to find how many end-point pairs should use the conventional model and how many the new model. We propose a method to find the proper combination of the number-pair at run-time. This is done by calculating the benefits of different combinations of the number-pair at run-time by using the communication statistics and an analytic model. It was implemented in OFI/LLC and evaluated using two micro-benchmarks performing all-to-one and one-to-all types of communications. The latency is reduced by up to 19% and 13%, respectively, as shown in Figure 1.3.

The second optimization is for the both OFI/LLC and RMPI libraries. It tries to keep only active software contexts in memory. The purpose is to limit the per-node memory consumption with the same target as the first optimization.

The conventional communication libraries prepare software contexts of the number proportional to the number of MPI processes. The proposed method only keeps a constant number of software contexts in memory by releasing inactive contexts when necessary. The mechanism was evaluated by using an analytic model of the memory consumption which is created by analyzing source code of OFI/LLC and MPICH. Table 1.1 shows the memory consumption with 4,194,304 MPI processes on 1,048,576 compute nodes. The proposed method reduces the memory consumption per compute node from 22 GiB to 1 GiB when compared to the existing technique.

1.3.3 Scalable MPI-IO Using Affinity-Aware Aggregation

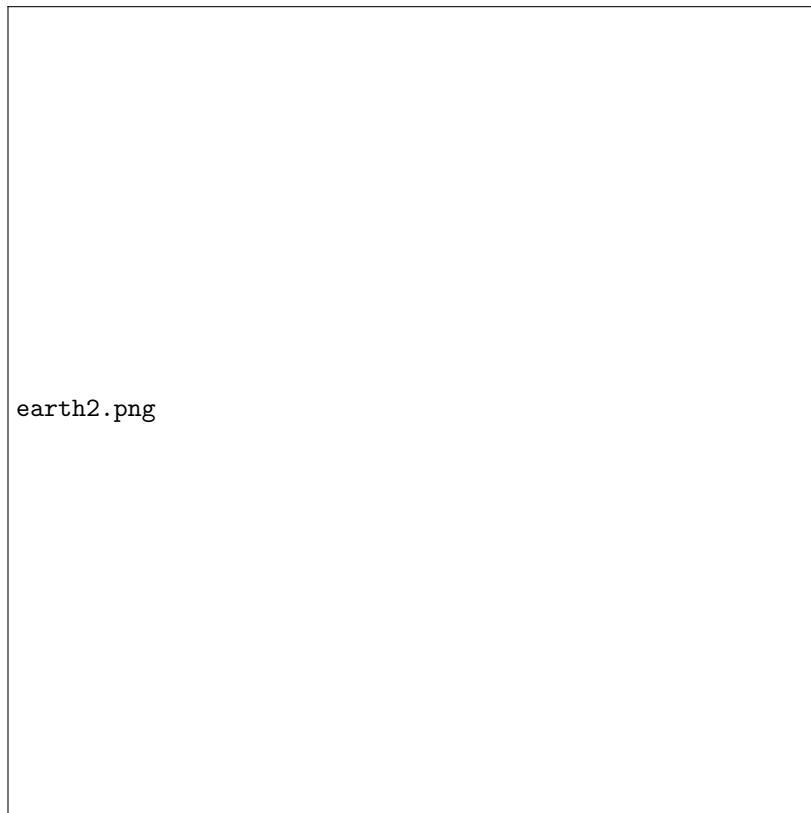
A commonly used MPI-IO library named ROMIO has the two-phase I/O (TP-IO) scheme to improve collective I/O performance for non-contiguous accesses. This research is addressing to optimize TP-IO implementation for further I/O performance improvements than the original one. In the FY2015, we have proposed enhanced TP-IO approach named EARTH (Effective Aggregation Rounds with THrottling) in the MPI library at the K computer[[tsujita2015](#)]. It has been arranged to have cooperative stepwise data exchanges based on the optimized aggregator layout and I/O throttling approach done in the FY2014. Its I/O throttling scheme and stepwise data exchanges are illustrated in Figure 1.4.

Figure 1.4(a) illustrates I/O throttling approach of the EARTH using token-relay. EARTH divided processes into groups which are associated with target Object Storage Target (OST) of the FEFS file system on the K computer. Then the EARTH throttles I/O request generation of each process, where process that receives a token issues I/O request. As a result, network and I/O request contention can be minimized. Furthermore, stepwise data exchanges in Figure 1.4(b) improve data exchange times by splitting all-to-all manner data exchanges into sub-groups which are associated with I/O throttling. This stepwise data exchange scheme has two advantages compared with the original all-to-all manner data exchanges. One is minimization in waiting time to complete data exchanges. Another is mitigation of network contention.

Performance evaluation was carried out using computing nodes ranged from 192 to 3,072 nodes.



(a) I/O throttling approach



(b) associated stepwise data exchanges

Figure 1.4: Optimization Techniques in EARTH

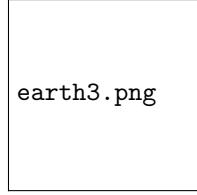


Figure 1.5: I/O throughput of collective write

I/O performance evaluation was done by using the HPIO benchmark with non-contiguous access patterns on a local file system of the FEFS on the K computer. The number of nodes was arranged not to have any interference from other users' applications. In the K computer case, we specified the number of nodes in a 3D manner node allocation, where we chose the following five patterns; 2x3x32, 4x3x32, 8x3x32, 8x6x32, and 8x12x32. We deployed one MPI process per one computing node, thus the number of MPI processes was the same with that of used computing nodes. Figure 1.5 shows I/O throughput values relative to the number of MPI processes.

In this evaluation, we examined the number of I/O requests for 1, 2, 4, and 8 in the I/O throttling scheme indicated by *EARTH(req=1)*, *EARTH(req=2)*, *EARTH(req=4)*, and *EARTH(req=8)*, respectively in addition to the original implementation indicated by *original* and aggregator layout optimization only version (*agg_aw*). The EARTH optimization outperformed the original one and aggregator layout optimization only version. From this evaluation, one I/O request case in the EARTH optimization was the best when we had 3,072 processes. Our future work is the way for tuning the number of I/O requests to gain the best I/O performance.

1.3.4 New Process / Thread Model

From FY2012, we have been developing a new process/thread model that is suitable for the many-core architectures. The many-core architectures are gathering attention towards the next generation supercomputing. Many-core architectures have a large number of low performance cores, and then the number of parallel processes within a single node becomes larger on many-core environments. Therefore, the performance of inter-process communication between the parallel processes within the same node can be an important issue for parallel applications.

Partitioned Virtual Address Space (PVAS) is a new execution model to achieve high-performance inter-process communication on the many-core environments. With PVAS, multiple processes run in the same virtual address space as shown in Figure 1.6 to eliminate the communication overhead due to the process boundaries that the current modern OSes introduce for inter-process protection. In PVAS, the data owned by the other process can be accessed by the normal load and store machine instructions, just like the same way accessing the data owned by itself. Thus, high-performance inter-process communication is achieved.

We implemented the prototype of the PVAS execution model in the Linux kernel in FY2012. We improved its quality and published it as open source software in FY2013. To demonstrate the potential of PVAS, Open MPI has been modified to utilize PVAS since then[shimada2015]. Especially in FY2015, proposed and developed PVAS was ported to McKernel.

It is already known that process oversubscription, which binds multiple parallel processes to one CPU core, can hide the communication latency and reduce CPU idle time. However, the lightweight OS kernels for Exascale systems may no longer support OS task scheduling to reduce OS noise. Without OS task scheduling, only one parallel process per CPU core is allowed to run, and the process oversubscription is impossible. Even if the OS task scheduling is supported, the overhead of the context switch hinders the application performance and ruins the advantage of process oversubscription.

To tackle this issue, we proposed user-level process (ULP) in FY2015. The user-level process is a process which can be scheduled in the user-space. The ULP was implemented as an extension of PVAS. ULP has the beneficial features of the user-level thread. Meanwhile, it has its own program code and data like a traditional process. By assigning a role of a parallel process to a user-level process, high-performance process oversubscription can be achieved without OS task scheduling. Moreover, the process oversubscription utilizing ULP does not change the programming model of the parallel application.

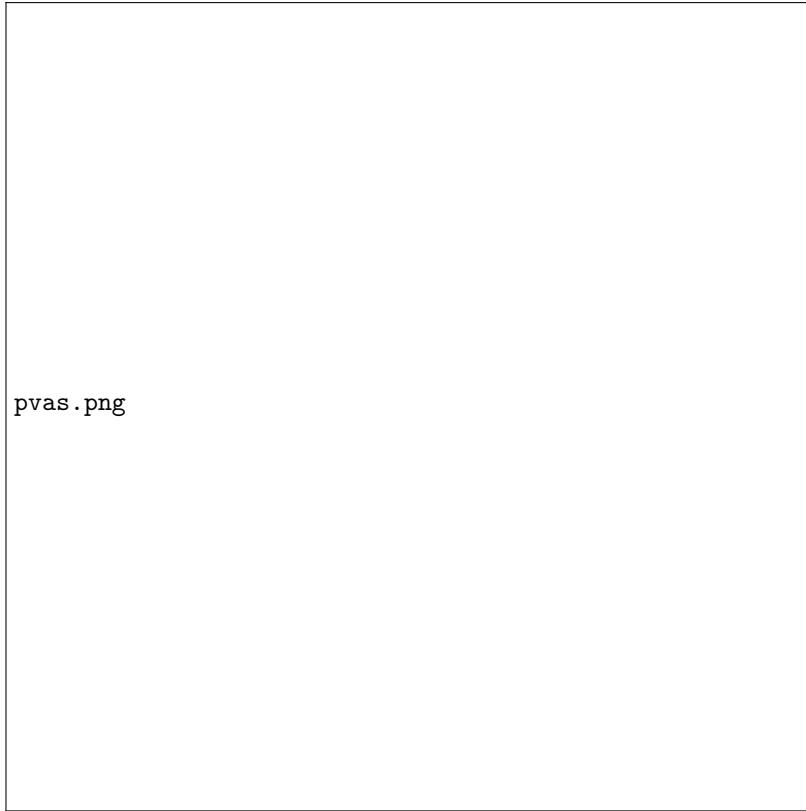


Figure 1.6: Partitioned Virtual Address Space

The context switching times of conventional Linux process, Linux thread and ULP over the number of execution entities are compared in Figure 1.7. Theoretically, there is no need of calling any systemcall to switch user-level processes, however, the privileged FS segment register is used to point Thread Local Storage (TLS) must be switched at the time of context switch on the x86 CPU architecture. As shown in Figure 1.7, the fastest one is ULP without switching the FS register, and send fastest one is ULP with the FS switching. Anyway the context switching times of conventional processes (KLP) and threads (KLP) are much higher than those of ULPs.

1.3.5 Fault Resilience

With the increasing fault rate on high-end supercomputers, the topic of fault tolerance has been gathering attention. To cope with this situation, various fault-tolerance techniques are under investigation; these include user-level, algorithm-based fault-tolerance techniques and parallel execution environments that enable jobs to continue following node failure. Even with these techniques, some programs, such as stencil computation, having no dynamic load balancing function may underperform after a failure recovery. Even when spare nodes are present, they are not always substituted for failed nodes in an effective way.

There are some questions of how spare nodes should be allocated, how to substitute them for faulty nodes, and how much the communication performance is affected by such a substitution. The third question stems from the modification of the rank mapping by node substitutions, which can incur additional message collisions. In a stencil computation, rank mapping is done in a straightforward way on a Cartesian network without incurring any message collisions. However, once a substitution has occurred, the node-rank mapping may be destroyed. Therefore, these questions must be answered in a way that minimizes the degradation of communication performance.

Several spare-node allocation and node-substitution methods had been proposed, compared and analyzed in terms of communication performance(Figure 1.8)[**hori2015**].

In FY2015, the proposed methods were evaluated using BlueGene Q (JUQUUEEN at Jülich Supercomputing Center) and the K computer. Figure 1.9 shows the communication performance



Figure 1.7: Context Switching Time (KLP, KLT and ULP)



Figure 1.8: Message collisions by substituting a failed node (5P-stencil)



Figure 1.9: Proposed Three Failed Node Substitution Methods



Figure 1.10: Communication Performance Degradation by Using Spare Nodes the K computer (12x12x12)

degradation on the K computer[yoshinaga2015]. The black lines represent the simulation results with 3D mesh network having the same size with the K computer evaluation (12x12x12) and the red lines represent the evaluation results of the K computer. In all cases, the hybrid sliding method is used. Left graph shows the cases with 7P stencil communication pattern (4 MiB), upper right graph shows the cases with the barrier collective communication, and the lower right graph shows the allreduce collective communication (16 KiB). Figure 1.10 shows the communication performance on BG/Q with the same way as in Figure citefig:resilience2, unless otherwise noted.

As shown in Figures 1.10 and 1.11, the actual patterns of the stencil communication performance degradation can vary in the K computer and BG/Q. The collective communication performance degradation, however, is relatively constant over the number of failed nodes.

1.3.6 Big data processing on the K computer

This research was conducted by collaboration between the Data Acquisition team of RIKEN SPring-8 Center and the System Software Research team of RIKEN AICS. The goal of this project is to establish the path to discover the 3D structure of a molecule from a number of XFEL (X-ray Free Electron Laser) snapshots. The K computer will be used to analyze the huge data transmitted from RIKEN Harima where SACLÀ XFEL facility is located.

In order to reduce quantum noise, each representative image must be averaged out more than hundred images. In addition to this, the sampled X-ray images must cover all possible orientations of the molecule. The number of images, although it depends on desired accuracy and the size of the molecule, can be one million in typical cases. Thus, the massive power of the K computer is needed.

We had developed a parallel software running on the K computer to analyze images obtained by a light source, SACLÀ. The developed software consists of two components as shown in 1.12.

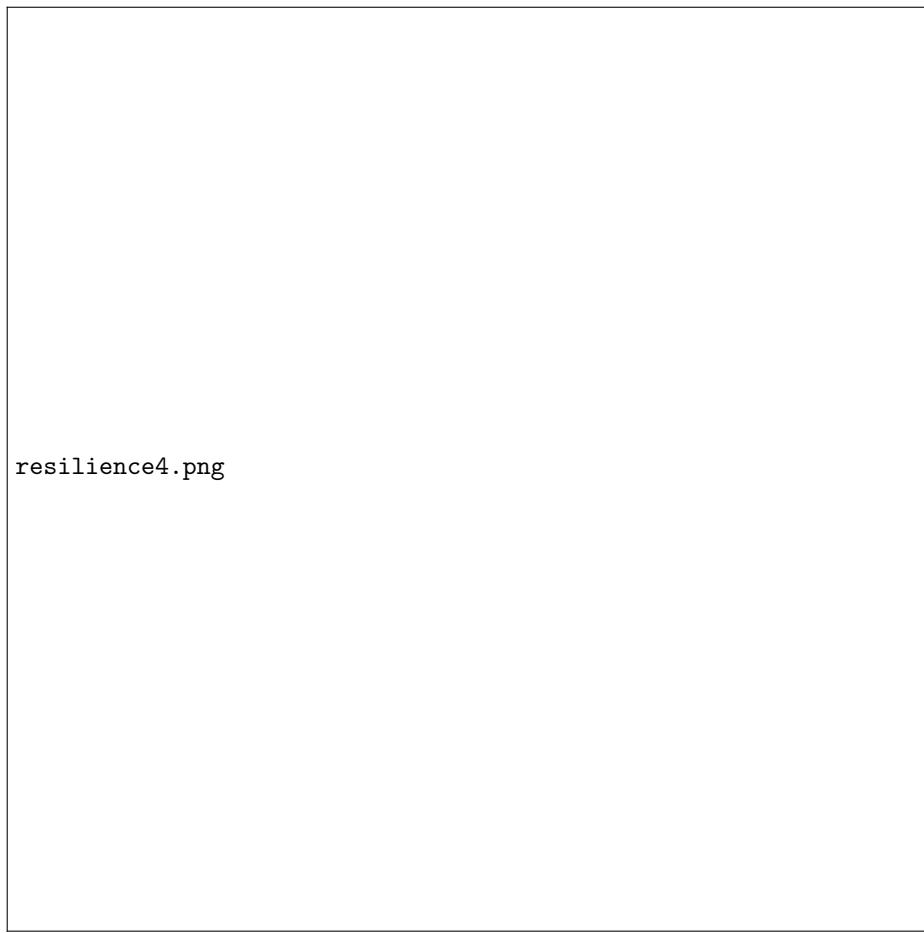


Figure 1.11: Communication Performance Degradation by Using Spare Nodes the K computer (12x12x12)

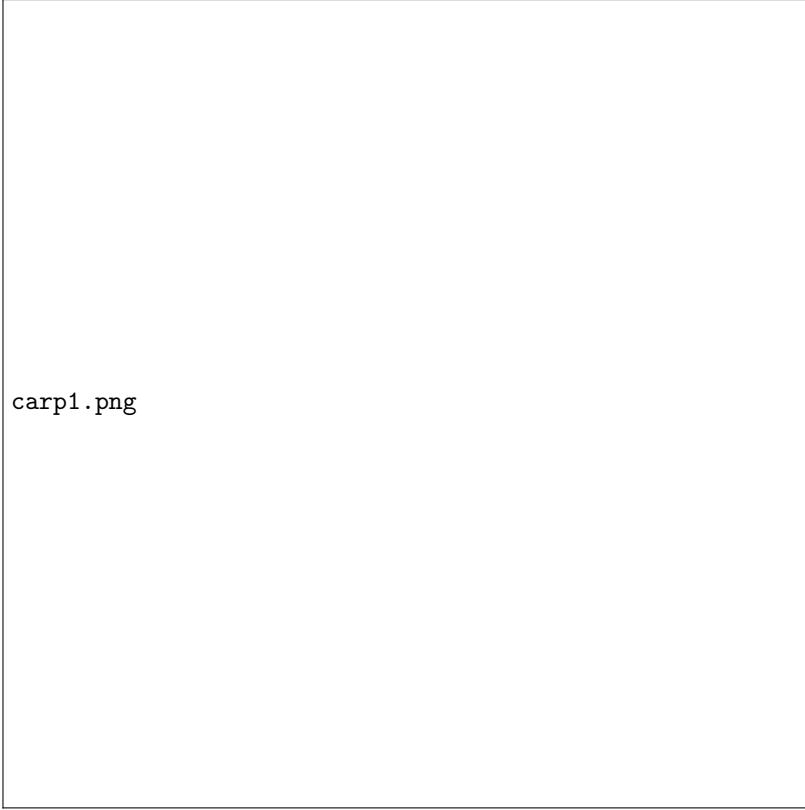


Figure 1.12: Block diagram of the procedure running on the K computer

The first phase is to select the representative images by a classic clustering computation. Thus, all images must be compared with all others. The second phase is to classify the rest of images into the representative images. In this stage, we need to calculate for all possible combinations of representative images and rest of the images.

After developing the first version of the program, we started to develop a new framework, named pCarp[**dist-carp**], to analyze any possible combination of two records in a dataset processed by all participating processes. This parallel processing can be used not only our target application (XFEL), but also can be used to analyze gene sequencing data, images obtained by electron microscopes, and so on. The pCarp framework takes care of parallelization and I/O minimization, while sequential input program to read data from a file and sequential output program to do the computation of two image data and to output the result (Figure 1.13). The most benefit of this framework is that the users do not need to write any parallel programs, but write just two sequential programs; the input and output programs.

In FY2015, we found that the performance of the first version of pCarp was no good because there was a large overhead to transfer data between the pCarp framework and user sequential programs. This overhead, however, was successfully reduced by introducing new data transfer mechanism. New version of pCarp can not only on the K computer, but also on the conventional clusters.

1.4 Schedule and Future Plan

Results of the System Software Research Team are being taken over by the System Software Development Team of Flagship 2020 project. The members mainly will work for the Flagship 2020 project from the next year. The MPI-IO library with the EARTH optimization will be available at the K computer in the next year. The team will mainly maintain the published software.

1.5 Publications

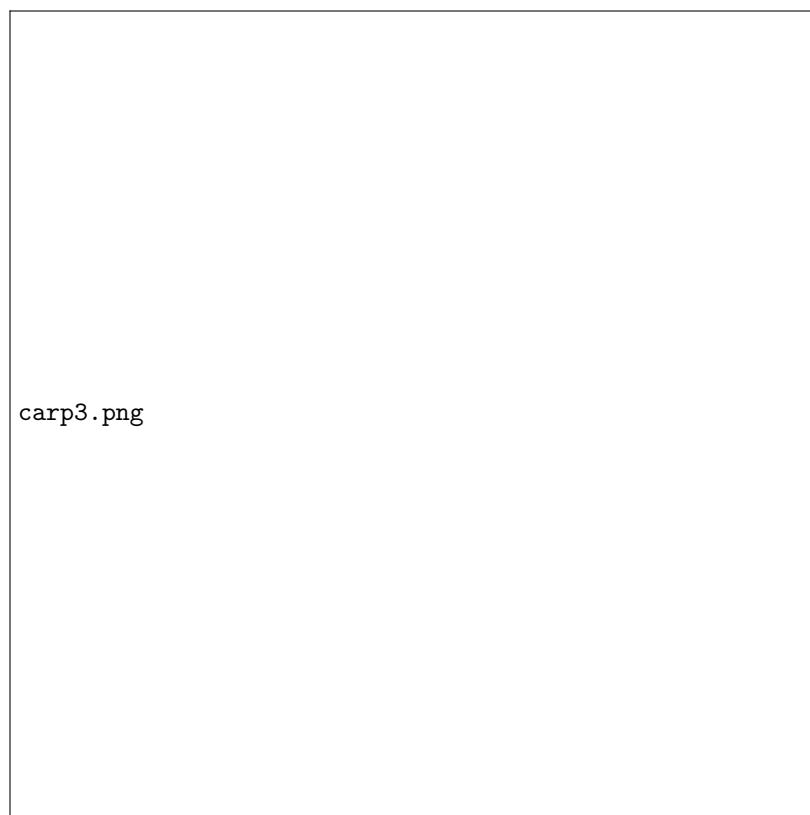


Figure 1.13: I/O minimization and load balancing

Chapter 2

Programming Environment Research Team

2.1 Members

Mitsuhisa Sato (Team Leader)
Hitoshi Murai (Research Scientist)
Miwako Tsuji (Research Scientist)
Masahiro Nakao (Research Scientist)
Jinpil Lee (Postdoc Researcher)
Yuetsu Kodama (Senior Research Scientist)
Hidetoshi Iwashita (Research Associate)
Shinichi Ito (Research Associate)
Makoto Ishihara (Agency Staff)
Masahiro Yasugi (Senior Visiting Researcher)
Hitoshi Sakagami (Senior Visiting Researcher)
Brian Wylie (Visiting Researcher)
Christian Feld (Visiting Researcher)
Kengo Nakajima (Senior Visiting Researcher)
Tomoko Nakashima (Assistant)

2.2 Research Activities

The K computer system is a massively parallel system which has a huge number of processors connected by the high-speed network. In order to exploit full potential computing power to carry out advanced computational science, efficient parallel programming is required to coordinate these processors to perform scientific computing. We conducts researches and developments on parallel programming models and language to exploit full potentials of large-scale parallelism in the K computer and increase productivity of parallel programming.

In 2015FY, in order to archive these objectives above, we carried out the following researches:

- (1) We are working on the development and improvement of XcalableMP (XMP) programming languages. XcalableMP is a directive-based language extension, designed by XcalableMP Specification Working Group (XMP Spec WG) including some members from our team as a community effort in Japan. It allows users to develop parallel programs for distributed memory systems easily and to tune the performance by having minimal and simple notations. In this year, we have improved Coarray functions in Fortran. The feature of coarray of the Omni XcalableMP compiler is implemented for the K compiler. In addition, some benchmarks and an application are parallelized with XcalableMP and their performance is evaluated on the K computer.
- (2) As an extension of XcalableMP to exascale computing, we are proposing a new programming model, XcalableACC, for emerging accelerator clusters, by integrating XcalableMP and OpenACC. We continue working on the language design and the compiler development of XcalableACC. This research is funded by JST CREST project on “post-petascale computing”.
- (3) Co-design for HPC is a bidirectional approach, where a system would be designed on demand from applications, and applications must be optimized to the system. We started the design of tools for co-design, including the SCAMP profiler for the network of large scale systems.
- (4) As the post-K computer will be a large-scale multicore-based system, we are investigating programming models for manycore-based parallel systems as a next version of XcalableMP, including dynamic tasking and load balancing as well as advanced PGAS models for distributed memory systems.
- (5) We conducted several collaborations on the performance evaluation with JSC, University of Tsukuba, Kyusyu Institute of Technology and other groups. In the collaborations with Kyusyu Institute of Technology, a task parallel language Tascell was evaluated on the K computer. We are developing tools for performance analysis of large-scale parallel programs, by enhancing a tuning tool Scalasca, which is being developed by JSC, for the K computer. This tool is used for performance analysis of real applications, in collaboration with their developers.

In addition to the research activities, we conduct promotion activities to disseminate our software. To promote XcalableMP as a means for parallelization of programs, we made the XcalableMP workshop, seminars or lectures as follows.

- XcalableMP workshop and LENS workshop (Oct. 29, 30)
- Tutorial of XMP at Osaka University (Oct 23)
- Tutorial of XMP at University of Tsukuba (Dec 9)
- FOCUS seminar on XMP (Jan 8)

The seminar or tutorials consist of both classroom and hands-on learning

2.3 Research Results and Achievements

We are developing Omni XcalableMP that is an open-source XcalableMP compiler, in cooperation with the university of Tsukuba. The latest version 0.9.2 has been released in November, 2015

2.3.1 Improvement of the coarray feature of XcalableMP

Coarray Fortran (CAF) is a parallel language that is a language extension of Fortran. To support the local view, XMP contains coarray features, which were adopted from Coarray Fortran (CAF) defined as part of Fortran 2008 standard. Based on experience of the implementation of Omni XMP C compiler, we have implemented and improved main part of CAF specification into XMP Fortran compiler.

We performed two improvements for memory allocation / registration and Communication.

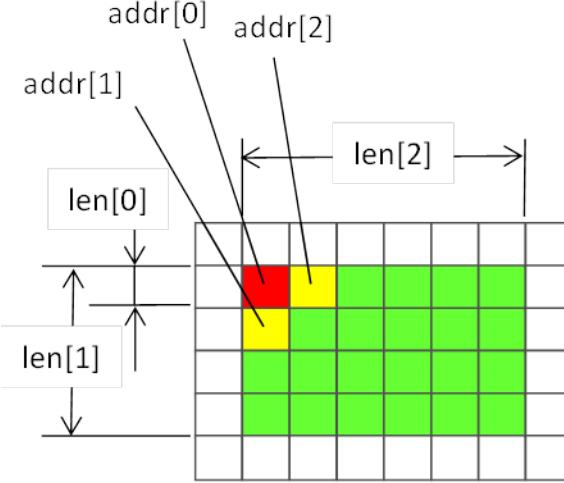


Figure 2.1: Parameters that determines data contiguity

Improvement of Memory Allocation / Registration

Coarrays are variables that can be accessed from other nodes. To allow remote nodes to access the local data, the address of the data must be registered at runtime with the low-level communication library, e.g., Tofu library in case of the K computer. To reduce runtime overhead of this operation for all coarrays, we made a mechanism that registers all static coarrays just before the execution of the program. The compiler generates the initializer for all static coarrays appearing in the program file at compile time and generates the caller calling the initializer at linkage time.

Improvement of Communication

To reduce communication latency overhead, contiguous data should be transferred simultaneously. For partially contiguous multidimensional array data, the length of the contiguous portion and the periodic pattern should be detected at compile time or at runtime. We made an algorithm and implemented on the compiler and the runtime library. Figure 2.1 shows an example of partially-contiguous communication data (colored elements) and major parameters in the algorithm. The parameters, lengths and addresses of data elements, are analyzed by the compiler and forwarded to the runtime library to find the contiguity.

Experimental Results

(1) Himeno benchmark

We ported Himeno benchmark program written with MPI to four different CAF programs. They used the same Fujitsu Fortran compiler with the same options including automatic thread parallelization. While the MPI version has 610 lines excluding comment and empty lines, the CAF versions have 402 to 415 lines, 32% to 34% shorter.

The result on grid size XL (1024 512 512) is shown in Figure 2.2. Two CAF programs are respectively 5% and 2% faster than the MPI version in average.

(2) NAS Parallel benchmark

We ported NAS Parallel benchmarks CG, EP, FT and MG written in MPI for CAF respectively. Figure 2.3 shows the result of CG Class-C as an instance and summarizes the history of the CAF program tuning. Finally CAF program V49 exceeds the original MPI version in performance. On EP, the CAF version is only 2% less performance in average than the original without tuning. On FT, the first version of CAF program extract more than 93% of the performance of the original in all evaluation ranges of Class B, C and D. Besides the CAF program has still room for performance tuning. On MG, the final version of CAF provides almost the same performance as the original in average.

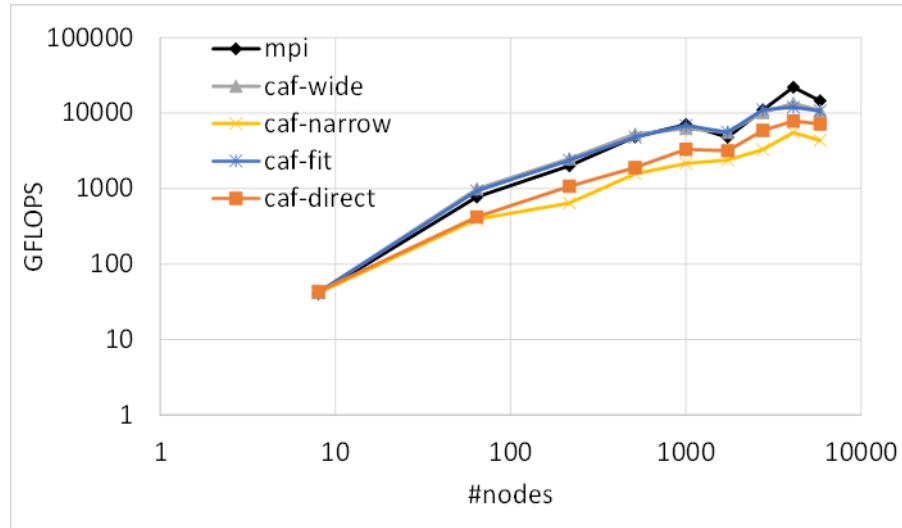


Figure 2.2: Four different CAF programs vs. the original MPI on Himeno benchmark

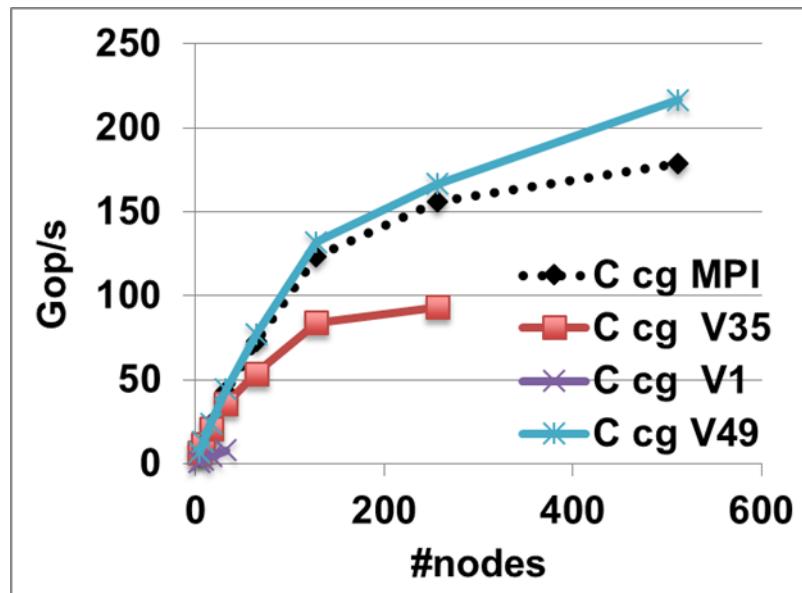


Figure 2.3: Porting and performance tuning of CAF program on NPB CG Class-C

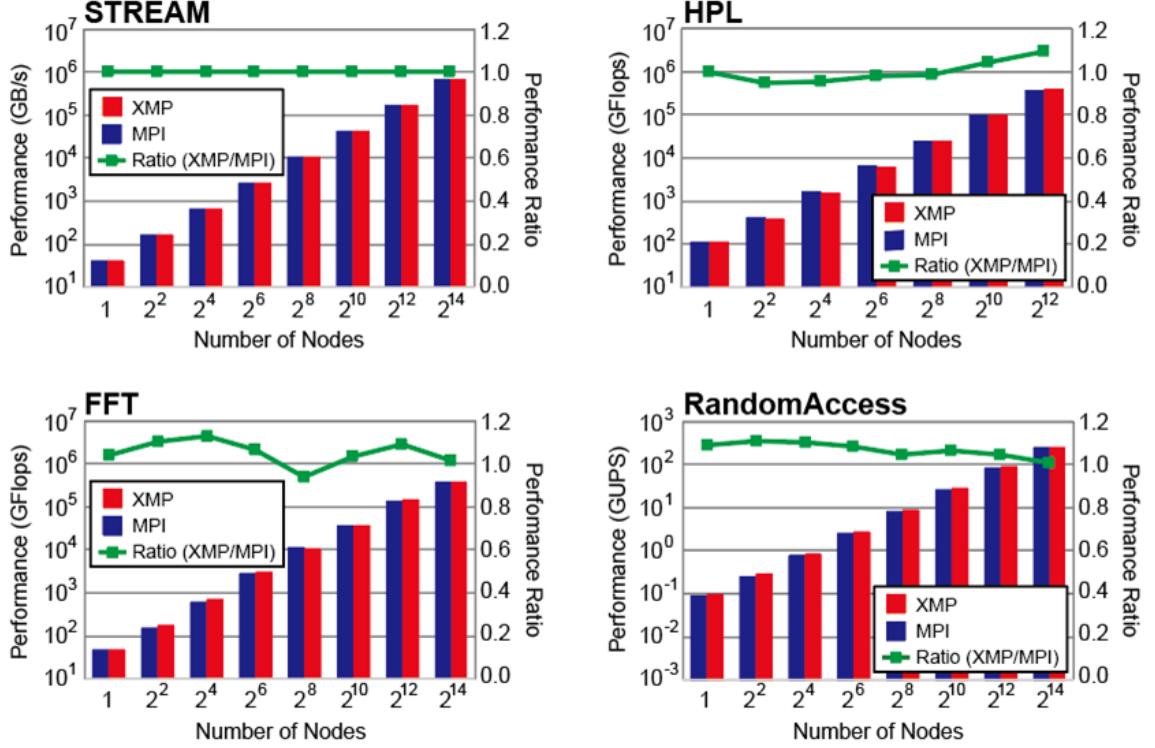


Figure 2.4: Performance of HPC Challenge Benchmark with XcalableMP

	STREAM	HPL	FFT	RandomAccess
XcalableMP	62	517	201	226
MPI	329	8,800	787	938

Table 2.1: SLOC of HPC Challenge Benchmark with XcalableMP

2.3.2 Performance Evaluation of the HPC Challenge Benchmark with XcalableMP

To evaluate productivity and performance of XcalableMP, we have implemented four benchmarks, namely STREAM, HPL, FFT, RandomAccess, in the HPC Challenge Benchmark Suite by using XcalableMP.

The figure 2.4 shows that the performance results of the XMP implementations. For a comparison purpose, we have also evaluated the performances of the MPI implementations which are reference implementations. The horizontal axis means that the number of compute nodes, the left vertical axis means that the performance corresponding to the bar, the right vertical axis means that the ratio of the performance of the XMP implementation to that of the MPI implementation corresponding to the line. When the performance ratio is greater than 1, the performance of the XMP implementation is better than that of the MPI implementation. The figure shows that the performances of XMP are almost the same as those of MPI.

The table 2.1 shows that source lines of code (SLOC) of the benchmarks in XMP and MPI. The table shows that the SLOCs of the XMP implementations are much less than those of the MPI implementations.

2.3.3 Performance of Three-dimensional Fluid Simulation with XcalableMP

The three-dimensional Eulerian fluid code written in Fortran, IMPACT-3D, which performs compressible and inviscid fluid computation to simulate converging asymmetric flows related to laser

#core	$lx=ly=lz$	only Z		both Y and Z		all of X, Y and Z		
		nz	ny	nz	nx	ny	nz	
256	1024	32	8	4	4	4	2	
2048	2048	256	16	16	8	8	4	
16384	4096		64	32	16	16	8	
131072	8192		128	128	32	32	16	

Table 2.2: Simulation parameters

fusion, is parallelized by three different domain decomposition methods, namely the domain is divided in (1) only Z direction, (2) both Y and Z directions and (3) all of X, Y and Z directions using by using directives only for the “global-view” programming model of XcalableMP (XMP). The program is also hand-coded with MPI using the same domain decomposition methods, and the performance difference between XMP and MPI codes is evaluated on the K computer.

As one node consists of 8 cores in the K computer, one process is dispatched onto each node and each process performs parallel computations with 8 threads, which are explicitly described by OpenMP in both XMP and MPI programs. We run both XMP and MPI codes with three different decomposition methods and evaluate the weak scaling on the K computer using Omni XcalableMP/Fortran compiler 0.7.0 and Fujitsu Fortran K-1.2.0.15. A number of cores for execution and corresponding simulation parameters are summarized in Table 2.2. lx , ly , lz are Fortran array size of first, second, third dimension, and nx , ny , nz are a number of division in X, Y, Z direction, respectively.

Performance is measured by a hardware monitor installed on the K computer, and three indexes are obtained. The total number of floating point operations is counted by the hardware monitor and is interpreted to MFLOPS using elapsed time. Finally it is divided by theoretical peak MFLOPS and output as MFLOPS/PEAK value. The average amount of transfer data per second between memory and CPU is also monitored. It is divided by theoretical peak memory access throughput and output as Memory throughput/PEAK value. The hardware monitor counts the number of instructions, and the number of SIMD instructions is divided by the total number of instructions to obtain SIMD execution usage.

MFLOPS/PEAK values for all 6 cases, namely (MPI, XMP) x (only Z, both Y and Z, all of X, Y and Z) are shown in Fig. 2.5 (a). Performance of XMP codes is as same as that of MPI codes, and small differences among three decomposition methods are found. But we can get only 8–9% of peak performance of the K computer. From the hardware monitor, we found that SIMD execution usage was less than 5% in all cases, and this could degrade the performance. Most cost intensive DO loops in IMPACT-3D include IF statements, which are needed to correctly treat extremely low velocity and flow direction change regardless of XMP and MPI codes, and the IF statement prevents the native Fortran compiler from generating SIMD instructions inside the DO loop. Thus relatively low performance is obtained.

As the true rate of the IF statement is nearly 100% in IMPACT-3D, speculative execution of SIMD instruction causes almost no overhead. So forcing the compiler to generate the SIMD instructions could be useful to enhance the performance, and it can be done with “simd=2” compiler option. All codes are recompiled with that option and rerun. SIMD execution usage increases up to around 50% in all cases, and we can expect performance improvement. MFLOPS/PEAK values for all cases are shown in Fig. 2.5 (b). MPI code performance is improved and we can get up to 20% of the peak performance. XMP code performance is also improved, but these are below 15% even XMP code performance is almost same as MPI code performance without “simd=2” option. Although Memory throughput/PEAK values of MPI codes are 55%, those of XMP codes are only 37% and this low memory throughput is one of candidates for low sustained performance.

In the converted code by the XMP/F compiler, all Fortran arrays are treated as allocatable arrays even the original code uses static arrays. The allocatable array prevents the native Fortran compiler from optimizing the DO loop with prefetch instructions because the array size cannot be determined at compilation time, and it could cause low memory throughput. All Fortran static arrays in the hand-coded MPI code for the decomposition method of all of X, Y and Z directions

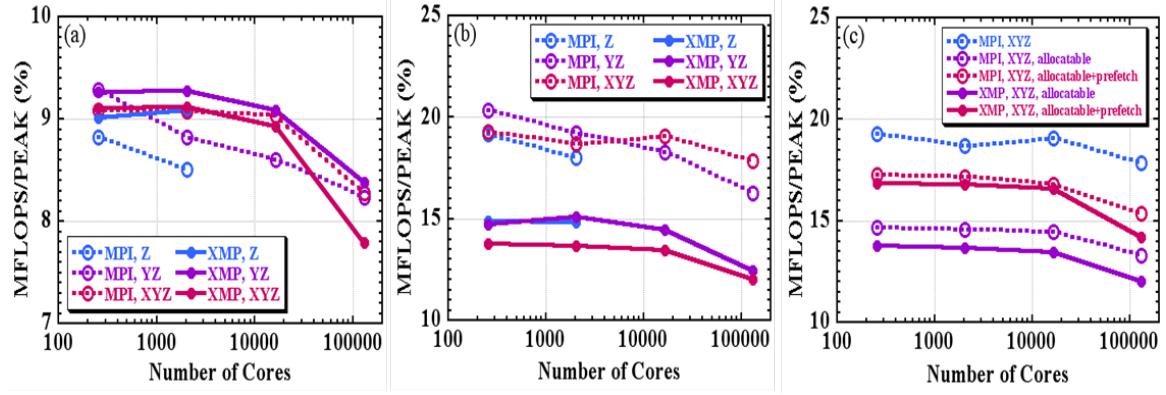


Figure 2.5: Performance comparison between MPI and XMP on the K computer with (a) no optimization for three decomposition methods, (b) SIMD optimization for three decomposition methods and (c) allocatable array optimization for the decomposition method of all of X, Y and Z directions.

are just replaced by allocatable arrays and we check a performance difference. Performance of the MPI code is shown in Fig. 2.5 (c) for static arrays (blue dash) and allocatable arrays (purple dash). MFLOPS/PEAK values are dropped from 20% to 15%, and this performance degradation without the prefetch instructions is confirmed. To force the native Fortran compiler to perform the prefetch optimization, we can use additional “prefetch_stride” compiler option. All codes are recompiled with “simd=2” and “prefetch_stride” options and rerun. Performance improvements by this compiler option are shown in Fig. 2.5 (c) for both MPI (purple dash to red dash) and XMP (purple solid to red solid) codes. MFLOPS/PEAK values are improved by 2–3% with the prefetch optimization.

2.3.4 Design of SCAMP (SCAlable Mpi Profiler) as a co-design tool for large-scale network

Co-design for HPC is a bidirectional approach, where a system would be designed on demand from applications, and applications must be optimized to the system. In order to co-design the network of large scale systems, it is important to evaluate the communication performance of applications. The trace driven simulator estimates the network performance based on trace files. Firstly, user should run their application on a real system in parallel to obtain the trace files from all processes. These trace files should contain MPI function calls, and their arguments and time stamps, etc. Then, the performance of a virtual system is estimated by using the trace files. While the trace driven simulator is straightforward, sometimes it is not appropriate for the simulation of large parallel systems since it is difficult to obtain the number of trace files for the future system if the current system is smaller than the future one. In order to tackle this scaling-problem in the trace driven simulator, we propose a method called SCAMP (SCAlable Mpi Profiler), which creates a large number of pseudo trace files based on the small number of trace files obtained from a small system and drives the network simulator using the pseudo trace files to estimate the performance of the large systems.

According to the experiments using SCAMP and using K-computer, as shown in Figure 2.6, SCAMP overestimates the performance of benchmarks, i.e. the runtime estimated by SCAMP is shorter than the real runtime on K-computer. The reason is that while we have focused only on the network performance, the computation time would change as the number of nodes increases.

2.3.5 Performance evaluation of Tascell on the K computer

Tascell is a task parallel language that supports distributed memory environments. A Tascell worker spawns a real task only when requested by another idle worker. The worker spawns a task after restoring its oldest task-spawnable state by temporarily backtracking. This mechanism eliminates the cost of spawning/managing logical threads. It also promotes the reuse of workspaces and improves the locality of reference since it does not need to prepare a workspace for each concurrently runnable logical thread. Furthermore, a single Tascell program can run efficiently on shared and distributed memory environments.

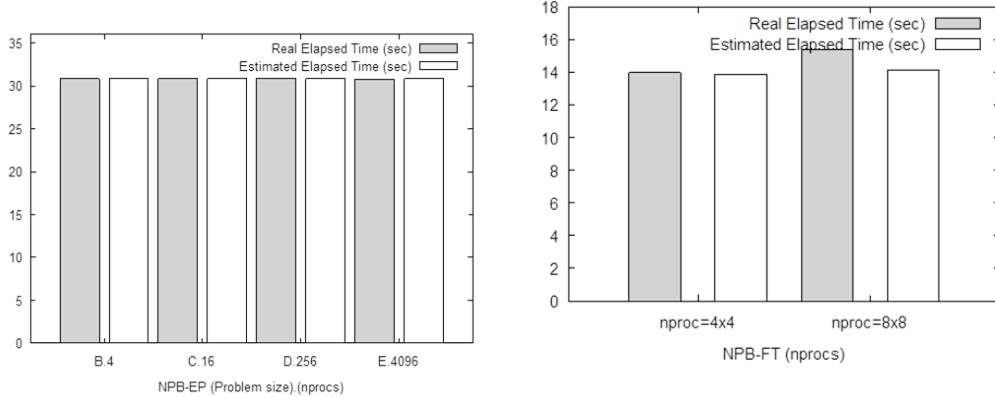


Figure 2.6: Comparison between real time and estimated time by SCAMP

This study aims to evaluate Tascell on massively parallel systems; in particular, we employed 1024 nodes of the K Computer with 8192 cores in total. In addition, we revised the implementation of Tascell to get it working on such systems.

In the conventional implementation of Tascell, inter-node communication is realized by TCP/IP communication via message routing servers called Tascell servers. This implementation is suitable for dynamic addition of computation nodes and wide-area distributed environments. On the other hand, Tascell servers often become communication bottlenecks. Furthermore, in recent supercomputer environments, there may be no appropriate places for deploying Tascell servers, and TCP/IP may not be available for inter-node communication; it is hard or impossible to run the conventional implementation in such an environments.

Therefore, we implemented inter-node communication in Tascell using MPI, which is supported by most practical supercomputer systems. At the same time, we adopted a server-less implementation in order to overcome the deployment and bottleneck problems, excluding the support of wide-area distributed environments. Note that programmers can write Tascell programs without concern about the underlying communication layer.

We evaluate the performance of our MPI-based implementation on the K computer using 7168 workers (7 workers x 1024 nodes). The result is shown in Figure 2.7.

In order to enable our implementation to work with the MPI implementation on the K computer and many other MPI implementations, it only requires the MPI_THREAD_FUNNELED support level, in which only the main thread can make MPI calls, and the two-sided communication paradigm. With such minimum requirements, our MPI-based implementation successfully realized both high performance and deadlock freedom.

2.4 Schedule and Future Plan

From this year, we started the study of the programming models for post-petascale, including programming models and runtime techniques to support manycore. We already propose XcableACC as a solution for accelerator-based system, which is to be explored in the JST CREST project. As the post-K computer will be a large-scale multicore-based system, we will investigate programming models for manycore-based parallel systems including dynamic tasking and load balancing as well as advanced PGAS models for distributed memory systems.

As in recent years, an important action for XcalableMP project is to disseminate our XcalableMP to applications users. As in last years, we organized several schools and hands-on, workshop with potential users also in this year. We will continue these promotion activities while we will study more optimization technique of XcalableMP compiler to improve the performance. As a research agenda especially for the K computer, we will contribute the scalability of large-scale applications for the K computer.

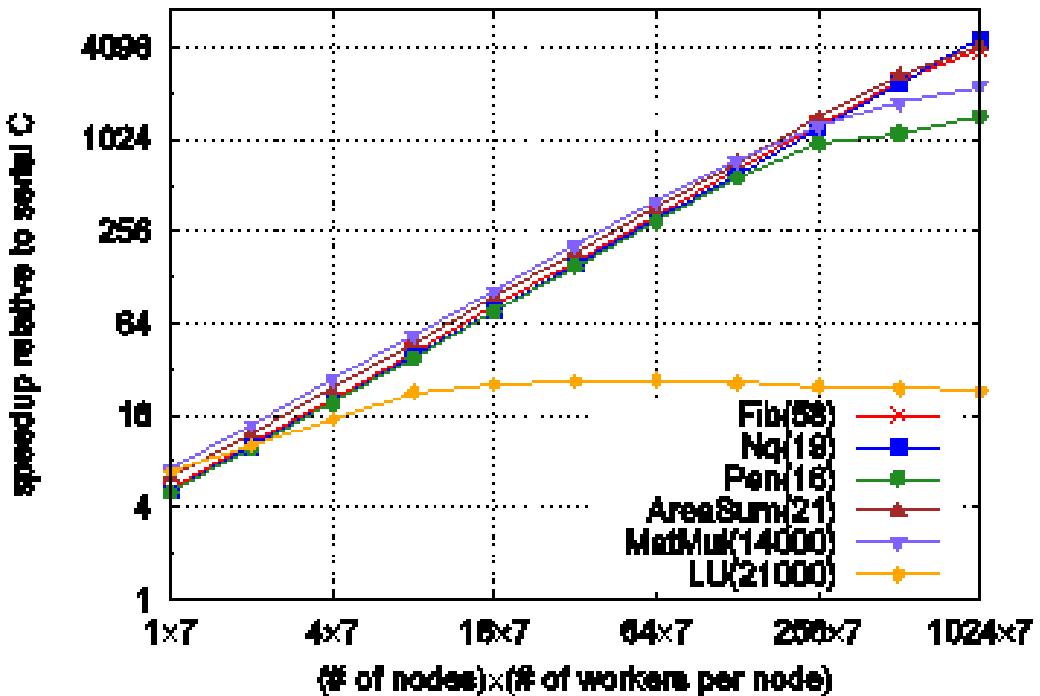


Figure 2.7: Evaluation results (Speedup) of Tasccell programs on the K computer

2.5 Publications

Chapter 3

Large-Scale Parallel Numerical Computing Technology Research Team

3.1 Members

Toshiyuki Imamura (Team Leader)
Yoshiharu Ohi (PostDoctoral Researcher)
Yusuke Hirota (PostDoctoral Researcher)
Daichi Mukunoki (PostDoctoral Researcher)
Daisuke Takahashi (Senior Visiting Researcher)
Franz Franchetti (Visiting Researcher)
Yoshio Okamoto (Visiting Researcher)
Takeshi Fukaya (Visiting Researcher)
Cong Li (Student Trainee)
Doru Thom Popovich (Student Trainee)
Yukiko Akinaga (Assistant)

3.2 Research Activities

The Large-scale Parallel Numerical Computing Technology Research Team conducts research and development of large-scale, highly parallel and high-performance numerical software for K computer. Simulation programs require various numerical techniques to solve systems of linear equations, to solve eigenvalue problems, to compute and solve non-linear equations, and to do fast Fourier transforms. In order to take advantage of the full potential of K computer, we must select pertinent algorithms and develop a software package by assembling numerical libraries based on the significant concepts of high parallelism, high performance, high precision, resiliency, and scalability. Our primary mission is to develop and deploy highly parallelized and scalable numerical software on K computer, namely KMATHLIB. It comprises several components such as for solving

- systems of linear equations,
- eigenvalue problems,
- singular value decomposition,
- fast Fourier transforms, and

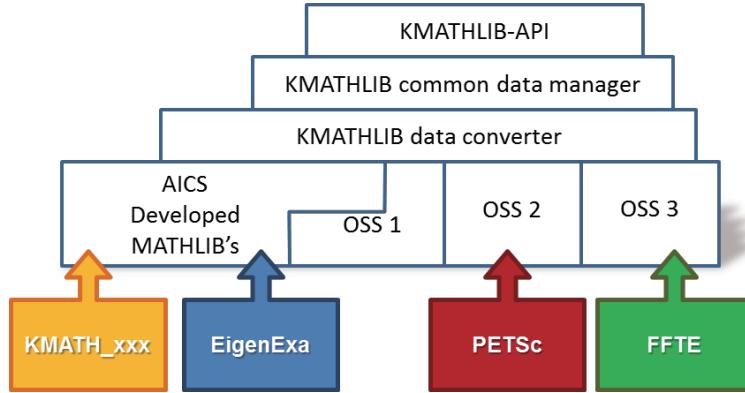


Figure 3.1: Software layer of KMATHLIB

- nonlinear equations.

The K-specific topics and technical matters for emerging supercomputer systems are also our challenging works such as

- Tofu interconnect,
- parallel I/O,
- fault detection (soft-error), and
- higher accuracy computing.

We are going to complete this project through a tight collaboration among computational science (simulation), computer science (hardware and software), and numerical mathematics. Our final goal is to establish fundamental techniques to develop numerical software libraries for next generation supercomputer systems based on strong cooperation within AICS.

3.3 Research Results and Achievements

Following series of the annual reports from 2012-13 to 2014-15, we summarize the latest results of our running projects, mainly focused on 1) development of KMATHLIB, 2) development of EigenExa, 3) investigation of FDTD related methods, and 4) other fundamental studies to optimize the BLAS kernels through automatic parameter tuning. The plans and the publication list are also presented in the last section.

3.3.1 KMATHLIB Project

Development of KMATHLIB for the integration of OSS packages

Since FY2012-2013, we have developed an integration framework named KMATHLIB, which supports a broad range of numerical libraries, and its API covers the computation resources from hundreds of nodes to the whole system of K computer.

Fig. 3.1 draws the schematic of KMATHLIB. KMATHLIB API is on the top layer and is accessed by users directly. Since we designed a flexible plugin mechanism and APIs, favorite OSS can be plugged in like the bottom highlighted part of Fig. 3.1. We intended to develop KMATHLIB API so that it encapsulates any components related to the libraries and conceals the differences of APIs and data structures. KMATHLIB API adopts a modern API style of the standard numerical libraries, like PETSc and FFTE. Thus, we only have to use a unique procedure to use the numerical solver plugged in the KMATHLIB package. In this FY2015-2016, we have updated the plugin mechanism to enable users to enhance the KMATHLIB library according to their computational environment [EigenExa, KMATH-EIGEN-GEV, KMATH-RANDOM].

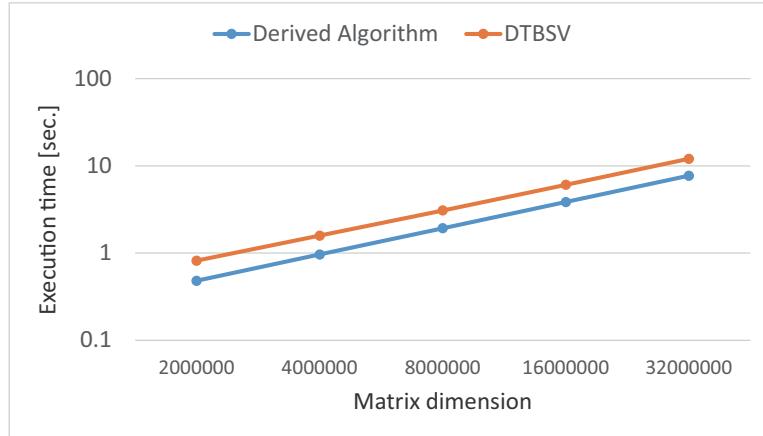


Figure 3.2: The execution time of DTBSV in Intel Math Kernel Library and the implementation of the derived algorithm

Maintenance and modification of KMATHLIB API

In FY2015, as a part of KMATHLIB project, we developed KMATHLIB API. KMATHLIB API is an application programming interface for development of computational science software. It provides a common interface and related functions for using various numerical libraries to reduce the cost of the development and maintenance of computational science software. Based on application and investigation on real simulation codes[**JKIS-JCPC2016**, **MSII-MC+SNA+MC2015**], the mechanism of software plugin and kernel functions are designed and implemented. The current KMATHLIB API contains the user interface and functions to use several basic (built-in) numerical libraries.

In March 2016, we organized a tutorial program for the KMATHLIB project, and the KMATHLIB API was available on a Fujitsu FX10 computer at that time. In the end, the tarball and user's manual of KMATHLIB API has been released (23 May, 2016) [**KMATHLIB-API**].

Feasibility study of asynchronous algorithms for applied mathematics

The asynchronous algorithms and their representation methods have been investigated as a part of research on the development of algorithms for manycore processors. In FY2015, we have investigated some conventional asynchronous algorithms, including the incomplete LU factorization algorithm proposed by E. Chow et al., a Jacobi/Gauss-Seidel like algorithm, and a logic of parallel adders. We classified the algorithms into two groups: (1) algorithms based on the approximation of operators and (2) algorithms which represent the original solution with the ones of other problems. Based on the classification, we derived a new back substitution algorithm of a narrow banded matrix for manycore processors. We carried out preliminary experiments on an Intel Xeon Phi 3120P, which show that the derived algorithm achieved 1.8 times speedup over a sequential implementation of the DTBSV of Intel Math Kernel Library in the case of a 32×10^6 dimensional band matrix with bandwidth 3 (shown in Fig. 3.2).

A solver for generalized eigenvalue problems of banded matrices

We have researched on the solver for generalized eigenvalue problems of banded matrices (GEPBs) since FY2013-2014. In FY2015-2016, we studied techniques to implement the algorithm proposed in FY2014-2015 for manycore systems. We implemented communication hiding technique on an Intel Xeon Phi system and evaluated its performance. Fig. 3.3 shows that the implementation run on 'a CPU(16threads) + a Xeon Phi' outperforms the routine DSYGV, a de facto standard numerical library, run on 'a CPU(16threads)' about 7.1 times by elapsed time. Also, the simultaneous use of a CPU and a Xeon Phi accelerates the performance 2.2 times over the single use of a CPU. The related results were presented in the SIAM LA [**HI-SIAMLA2015**] and EPASA [**HI-EPASA2015**] as oral and poster presentations in 2015, respectively.

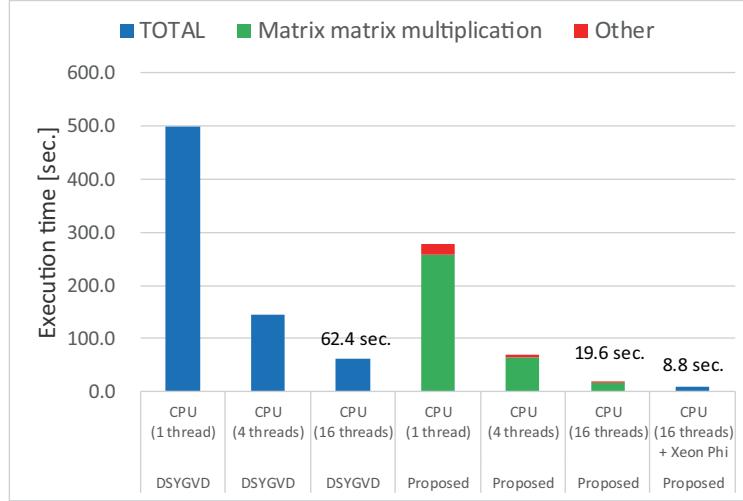


Figure 3.3: The execution time of DSYGVGD and the implementations of the proposed algorithm for a CPU, and the implementation for an Intel Xeon Phi system. The test matrix is a 10,000-dimensional random matrix.

3.3.2 EigenExa Project

We conducted the EigenExa Project with the grant support of ‘Development of System Software Technologies for post-Peta Scale High Performance Computing’ by JST CREST (the project code name was ‘Development of an Eigen-Supercomputing Engine using a Post-Petascale Hierarchical Model’ and the leader was Prof. Tetsuya Sakurai, University of Tsukuba) during FY2010-FY2015. In FY2015-16, we concluded our EigenExa project, on which we developed a parallel dense eigenvalue solver. The EigenExa library was already released in August 2013, and the current release version is 2.3d (31 August 2015).

Following FY2014-2015, we continued to promote the library and evaluated the performance on some available supercomputer systems, such as a Fujitsu FX10, an NEC SX-ACE, an IBM Blue-Gene/Q, and Intel Cluster systems. In the performance evaluation, we reported the preprocessing part of tri-diagonalization and pent-diagonalization at PDSEC2015 [**FI-PDSEC2015**], and the divide and conquer part at EPASA2015 [**FI-EPASA2015**].

Communication avoiding algorithm for the Householder tri-diagonalization

Communication avoidance (CA) is considered to be a promising technology to overcome the drawbacks resulting from the communication latency. Well-known examples of the CA algorithm reported in the literature are the tall skinny QR decomposition (TSQR) algorithm and the matrix powers kernel (MPK) algorithm. We investigated the related algorithm CholQR2, and showed a policy or a sort of new performance metric of the Chebyshev basis conjugate gradient (CBCG) method on K computer [**KFTHFIS-PPAM2015**].

Even though the CA algorithms induce more flops counts, the CA algorithms tend to reduce the number of communication, which is often dominant part of parallel computing on modern systems. To derive a new communication avoiding Householder scheme, we applied two simple principles (or simple rule) for transformation; i) distributive property of linear operators, ii) combining a couple of communication into one. In Fig 3.4, the underlined statement requires two `MPI_Allreduce`'s per iteration. This is the optimal version because matrix-vector multiplication needs at least two collective communications when we take advantage of the symmetric property of the matrix. Since the naive Householder tridiagonalization has to call five `MPI_Allreduce`'s per iteration, the proposed version drastically reduces the number of communications and leads to better parallel scalability. The proposed algorithm CAHTR(3) was presented in ParCo2015 conference [**IFHYM-ParCo2015**], and Communication Hiding (CH) technique was also presented at SIAM LA 2015 [**I-SIAMLA2015**]. At the moment, the parallel implementation of the present version 2.3 or later yields good performance acceleration on K computer compared with the non-CA/CH optimized version (see Fig. 3.5).

$$\left[\begin{array}{c|c} v & d \\ \hline s & t \\ C_U & \gamma_U \\ C_V & \gamma_V \\ \hline g & v_1 \\ v_1 & a_{11} \end{array} \right] = \left[\begin{array}{c|ccc} I & 0 & 0 & 0 \\ \hline 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \\ \hline u^t & 0 & 0 & 0 \\ e^t & 0 & 0 & 0 \end{array} \right] \left[\begin{array}{c|cccc} A & u & U & V & \\ \hline u^t & 0 & 0 & 0 & \\ U^t & 0 & 0 & 0 & \\ V^t & 0 & 0 & 0 & \\ \hline u & e & & & \\ 0 & 0 & & & \\ 0 & 0 & & & \\ 0 & 0 & & & \end{array} \right]$$

$s = \text{sign}(\sqrt{s}, -t)$
 $[u, v] := [u, v] - s[e, d]$
 $[C_U; C_V] = [C_U; C_V] - s[\gamma_U; \gamma_V]$
 $v := v - (U C_V + V C_U)$
 $f = g - 2C_U^T C_V - s(2v_1 - sa_{11})$
 $v := v - afu$

Figure 3.4: Communication avoiding Householder tridiagonal transformation

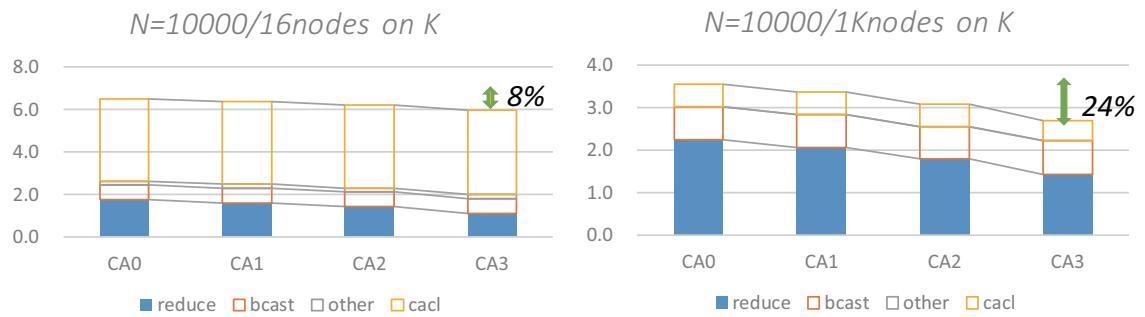


Figure 3.5: Big impact of performance improvement by CAHTR (Communication Avoiding Householder TRidiagonalization). The blue bars correspond to the elapsed time of MPI_Allreduce operations.

3.3.3 Investigation of the FDTD Related Methods

FDTDM (Finite-Difference Time-Domain Method)

Since electronic apparatuses downsize in a short period and the cost-cut of development is strongly demanded, numerical simulation is thought to be useful to a industrial design process. High definition and large-scale simulations must be indispensable for reliable evaluation. The finite-difference time-domain method (FDTDM) is applied for numerical simulations of electromagnetic wave propagation phenomena, while most of the preexistent software of FDTDM are commercial. Modification to the software or change of a simulation scenario sometimes are limited due to the software licencing. Therefore, we decided to develop open source software based on FDTDM for K computer. Since FY2014-2015, we have surveyed the computational electromagnetics and social contribution of numerical simulation by using FDTDM and related methods.

MTDM (Meshless Time-Domain Method)

In the simulation using FDTDM, the node arrangement of the electric and magnetic fields based on a staggered grid often becomes a significant difficulty when we treat a complex shaped domain including a curved surface. The hybrid idea of FDTDM and a meshless method yields a novel spatial discretization scheme of the meshless time-domain method (MTDM). The meshless method is a mathematical approach to find an approximate solution of the boundary value problem of the partial differential equation without using the grid which is used in the finite element method. Therefore, it is expected that analysis of electromagnetic wave propagation phenomena in a complex shaped domain can be easily executed by using MTDM. In FY2015-2016, we developed a test version of the three-dimensional MTDM simulator [OI-PFR2015].

3.3.4 A study for development of high-performance linear algebra libraries on future architectures

Traditional linear algebra libraries such as Basic Linear Algebra Subprograms (BLAS) are still important building blocks for computational science. As processor architectures become more and more complex, more challengings on development and code optimization are met. Also, they are required not only to achieve high performance, but also to support accurate, fault-tolerant, and energy-efficient computations toward the Exascale computing. Therefore, we are conducting a study for developing such linear algebra kernels on modern many-core architectures such as GPUs.

High performance memory-bound BLAS routines with automatic thread-block size adjustment on CUDA

In the previous FYs, we proposed a sophisticated implementation of general and symmetric matrix-vector multiplication (GEMV and SYMV) routines on CUDA [ASPEN.K2, MUBLAS-GEMV]. In this FY2015-2016, we have extended the study to other memory-bound linear algebra kernels [MIT-HPC150, IMYM-HPC151]. The performance of CUDA kernels often depends on the number of threads per thread-block (thread-block size), and the optimal thread-block size often differs according to the GPU hardware running the kernel and the given data size to the kernel. We proposed a method to determine the nearly optimal thread-block size for the DGEMV kernel. Our proposed method automatically and theoretically determines the thread-block size using an occupancy model for thread-blocks on a grid along with warp-occupancy and some rules (Fig. 3.6). Also, we improved and extended our GEMV and SYMV implementations to support multi-GPU environments. Our implementations, especially GEMV kernels, achieved better throughput and performance stability with respect to the matrix size on up to 4 Kepler GPUs when compared to the existing implementation

Short length floating-point formats (SLFP) for fast and energy efficient computation (work-in-progress)

The required precision depends on the purpose of the computation. However, most numerical libraries only support IEEE754 32- and 64-bit floating point formats. To optimize the performance of numerical software prominently with respect to the precision, we proposed new floating-point formats that have shorter bit-length than IEEE standards on both CPUs and GPUs [MI-HPC152].

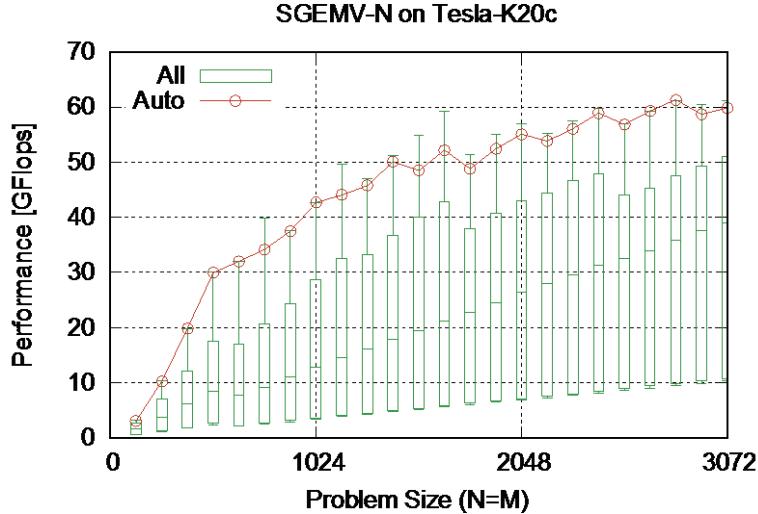


Figure 3.6: Performance of SGEMV-N (single-precision, non-transposed) on Tesla K20c Kepler GPU. The green line (All) shows the performance distribution obtained by all possible configurations of thread-block size. The red line (Auto) shows the performance obtained by our method.

By eliminating waste data movement in the computation using the shorter floating-point formats, we expect to improve the computation speed and energy efficiency. In our preliminary evaluation on a GPU, the proposed method achieved better performance and energy efficiency. A preliminary version is available from our webpage [[SLFP](#)].

3.3.5 Seminar

We hosted researchers from foreign research institutes, and organized a part of AICS HPC seminars in this FY2015-2016 (see also the webpage <https://sites.google.com/site/aicshpcseminar/>).

- 4-th AICS HPC Seminar, Friday, September 11, 2015,
 1. Cong Li (the Department of Computer Science at the University of Tokyo (Ph.D. student)), ‘Evaluation of Communication-Avoiding Block Chebyshev Basis Conjugate Gradient Method Based On Large Scale Experiment’
 2. Prof. Bruno Lang (Computer Science (Algorithms) at University of Wuppertal), ‘High-performance large-scale eigenvalue computations’
- 7-th AICS HPC Seminar, Tuesday, March 29, 2016,
 1. Dr. Osni Marques (Lawrence Berkeley National Laboratory, US), ‘Tuning the Coarse Space Construction in a Spectral AMG Solver’
 2. Prof. Yusaku Yamamoto (The University of Electro-Communications, Japan), ‘Roundoff Error Analysis of the CholeskyQR2 and Related Algorithms’
 3. Dr. Susumu Yamada (Japan Atomic Energy Agency, Japan), ‘Quadrature precision basic linear algebra subprograms with FMA instruction and its applications’
 4. Dr. Toshiyuki Imamura (RIKEN AICS, Japan), ‘EigenExa: Dense Symmetric eigenvalue solver for distributed parallel systems’

3.3.6 International Collaborations

Our team has joined a new international project which runs on the Joint Laboratory for Extreme Scale Computing (JLESC) framework. The project is a collaboration with Inge Gutheil from Juelich Supercomputer Center (JSC) (<https://jlesc.github.io/about/>). The title of the proposed project is ‘HPC libraries for solving dense symmetric eigenvalue problems’, and we plan to evaluate the several aspects of the routines on the existing production and prototype supercomputers

available to give the users recommendations which routine to use for their specific task. In the 4th JLESC meeting in Bonn, December 2015, we had a pre-meeting of this project concerning a research update and planning in 2016 of each member.

At the 6-th AICS International Symposium in February 2016 at AICS, we also presented key research topics that would be potential themes for collaboration such as

1. application of EigenExa to application codes,
2. numerical algorithms, and
3. higher/reduced precision numerical kernels.

In order to specify more detailed plans and roadmap for the collaborations, we discussed with a couple of institutes that signed up MoU.

3.4 Schedule and Future Plan

3.4.1 KMATHLIB project

To promote KMATHLIB, the eternal maintenance of useful plugged-in OSS is of significance. From FY2014-2015, we extended plugin solvers such as for sparse linear equations, GEBPs, and SVD for tensors. This FY2015-2016, we only did the investigation of the GEBPs solver on distributed highly parallel computers such as K computer due to the lack of human resources, whereas we will be able to release the solver package for GEBP as a part of KMATHLIB in FY2016-FY2017. In addition, we already completed the preliminary studies for the solver on many-core accelerators such as an Intel Xeon Phi, Knight Corner, aka KNC, thus, we are going to continue to study on both KNC and extend it to the emerging processor, an Intel Xeon Phi Knight Landing, aka KNL, in FY2016-2017.

3.4.2 EigenExa project

After the first release of the EigenExa library, several application codes adopt to use EigenExa. Continual maintenance of the EigenExa library becomes our important mission. In addition, the performance improvement and scalability towards the future systems such as a post-K computer. Even though, we introduced the Communication Avoiding (CA) and Communication Hiding (CH) techniques to the present EigenExa implementation, to apply these techniques to block algorithms must be established because other eigenvalue solver projects adopt this approach naturally.

Also, a lightweight or flexible implementation of BLAS kernels for small linear algebra is also an important issue. Through reviews to the application users on K computer system, major groups demand a very wide variety of spectrum and problem sizes. Since the present version of EigenExa was intended to accelerate and scale up for the ultra-scale problem, we recognized to increase the performance on a diagonalization of a small dimension matrix, such as a couple of thousand dimensions. Also, we will modify the EigenExa library to calculate not only full spectrum but a part of the spectrum, as other modern eigenvalue solvers do.

3.4.3 FDTD related method

One of the future works related to the FDTD project is to investigate the behavior of the three-dimensional MTDM scheme. In fact, we confirmed numerical errors in three-dimensional MTDM simulations when particular collocation point was selected. Since the goal of MTDM is to apply for practical and industrial simulation codes, we recognize that it is significant to free numerical instability as well as to obtain performance improvement. Also, we need to promote MTDM, which has been originally studied in our project.

3.4.4 Other issues

In this annual report, we have untouched topics, fault tolerance (or resilience), and high precision computing such as double-double format computing. Some of these topics have been already investigated as one of the keywords for the petascale computing. For example, following issues were researched in previous FY's;

1. algorithmic-based fault tolerance,
2. numerical reproducibility,
3. quad precision numerical tools, and
4. dynamical process mapping.

We are going to present them at international conferences soon.

In particular, we recognized that the research related to numerical precision, higher precision and reproducibility become important in near future. The higher-precision numerical framework or software toolkit must be organized on modern or future supercomputer systems as well as the post-K computer. For the feasibility study, we already have started research on a higher precision computational kernel of a spectral method [**SYMMI-JSIAM2015**]. We will soon merge a higher-precision numerical library into KMATHLIB, and investigate the effect of numerical precision on real application codes.

3.5 Publications

Chapter 4

HPC Usability Research Team

4.1 Members

Toshiyuki Maeda (Team Leader)

Masatomo Hashimoto (Research Scientist)

Peter Bryzgalov (Research Scientist)

Itaru Kitayama (Technical Staff I)

Yoshiki Nishikawa (Visiting Scientist, University of Tokyo)

Yves Caniou (Visiting Scientist, Université Claude Bernard Lyon 1)

Judit Gimenez (Visiting Scientist, Barcelona Supercomputing Center)

Sameer Shende (Visiting Scientist, University of Oregon)

Fabien Delalondre (Visiting Scientist, École polytechnique fédérale de Lausanne)

Pramod Kumbhar (Visiting Scientist, École polytechnique fédérale de Lausanne)

Tatsuya Abe (Visiting Scientist, Chiba Institute of Technology)

Sachiko Kikumoto (Assistant)

Yumeno Kusuvara (Assistant)

4.2 Research Activities

The mission of the HPC Usability Research Team is to research and develop a framework and its theories/technologies for liberating large-scale HPC (high-performance computing) to end-users and developers. In order to achieve the goal, we conduct research in the following three fields:

4.2.1 Computing portal

In a conventional HPC usage scenario, users live in a closed world. In other words, users have to play roles of software developers, service providers, data suppliers, and end users. Therefore, a very limited number of skilled HPC elites can enjoy the power of HPC, while the general public sometimes gives a suspicious look to the benefit of HPC. In order to address the problem, we are designing and implementing a computing portal framework that lowers the threshold for using, providing, and aggregating computing/data services on HPC systems, and liberates the power of HPC to the public.

4.2.2 Virtualization

Virtualization is a technology for realizing virtual computers on real (physical) computers. One big problem of the above mentioned computer portal that can be used by wide range of users simultaneously is how to ensure safety, security, and fairness among multiple users and computing/data service providers. In order to solve the problem, we plan to utilize the virtualization technology because virtual computers are isolated from each other, thus it is easier to ensure safety and security. Moreover, resource allocation can be more flexible than the conventional job scheduling because resource can be allocated in a fine-grained and dynamic way. We also study lightweight virtualization techniques for realizing virtual large-scale HPC for test, debug, and verification of computing/data services.

4.2.3 Program analysis/verification

Program analysis/verification is a technology that tries to prove certain properties of programs by analyzing them. By utilizing software verification techniques, we can prove that a program does not contain a certain kind of bug. For example, the byte-code verification of Java VM ensures memory safety of programs. That is, programs that pass the verification never perform illegal memory operations at runtime. Another big problem of the above mentioned computing portal framework is that one computing service can be consists of multiple computing services that are provided by different providers. Therefore, if a bug or malicious attack code is contained in one of the computing services, it may affect the whole computing service (or the entire portal system). In order to address the problem, we plan to research and develop software verification technologies for large-scale parallel programs. In addition, we also plan to research and develop a performance analysis and tuning technology based on source code modification history.

4.3 Research Results and Achievements

4.3.1 Design and Implementation of a Computing Portal Framework for HPC

In FY2015, we enhanced the computing portal framework developed in FY2012-FY2014 so that the users of the K computer are able to build and publish their own computing portal with their own authority and computing resources on the K computer. More specifically, we modified the computing portal framework so that it is able to directly communicate with the login nodes of the K computer. In addition, the framework is able to generate SSH public keys (and hide their corresponding SSH private keys inside) so that the users of the K computer are able to associate the generate public keys to their accounts. Thus the users are able to launch their own computing portal by installing/deploying the framework on their servers and registering the SSH public keys generated by the framework to their K accounts.

In FY2016, we will continue to develop the computing portal framework. Especially, we will create a VM image in which the framework is installed and ready-to-run so that the users of the K computer is able to launch their own computing portal with the framework by simply deploying the VM image to, e.g., AWS. In addition, we will also investigate to develop a Docker (<https://www.docker.com>) container image.

4.3.2 Virtualization Techniques

Lightweight Virtualization for Testing/Debugging Parallel Programs

Writing a program which makes full use of massively parallel HPC environments (e.g., the K computer) is extremely difficult because debugging parallel programs is extremely difficult because of inherent non-determinacy of parallel programs and hard-to-reproduce bugs. Moreover, writing massively parallel programs also tend to suffer from a performance problem. For example, even if a program scales well on a PC cluster system whose size is small-to-moderate, the program may not scale on massively parallel HPC systems. Even worse, the performance may severely degrade and will be worse than on a small PC cluster system or even a single PC. Actually, this is not uncommon and the reason is that communication costs between computing nodes may largely vary and sometimes incurs unacceptable heavy overheads.

In order to address the abovementioned problem, we have been developing a lightweight network virtualization system for testing/debugging programs for massively parallel programs without actually using real massively parallel HPC environments. With our system, users can run several hundreds of virtual computing nodes on a single physical computing node.

There are two key ideas in our system: library-hooking and decentralized management of routing information. Library-hooking is a kind of virtualization technology which intercepts function calls for system operations, and modify their parameters and/or return values in order to trick the programs as if they run on in isolated multiple computing nodes, even though they run on a single physical computing node. More specifically, in our lightweight virtualization system, we mainly hook network related operations (and some file I/O) from user programs. One benefit of the library-hooking approach is that it introduces little overheads to program execution (compared to other virtualization techniques, e.g., CPU level virtualization, OS level virtualization, and so on) because it can be achieved by user-level operations and requires no interaction with OS.

When implementing a lightweight network virtualization system, decentralized management of routing information is necessary in order to avoid maintaining routing information in a single or a few physical nodes. Our lightweight virtualization system has to manage routing information by itself because it virtualizes network environments. If the routing information is managed in a single physical node, all the other physical nodes have to ask the single node in order to correctly route network packets from one virtual node to another. Therefore, when the numbers of virtual nodes and physical nodes are huge, the single node will become a performance bottleneck and severely degrade the overall performance of our lightweight virtualization system.

More specifically, our lightweight virtualization system statically distributes the information which virtual node runs on which physical node before executing user programs on virtual computing nodes. In ordinary HPC systems, it is uncommon that computing nodes are directly allocated during job execution. In addition, in order to virtualize port numbers, our lightweight virtualization system let physical nodes exchange the information about virtualized port numbers when one virtual node on one physical node communicates with another virtual node on another physical node.

Based on the abovementioned approach, in FY2012-FY2014, we implemented the prototype of our lightweight virtualization system and it successfully ran on the K computer under a restricted environment. More specifically, it successfully ran 20000 virtual computing nodes on 1000 physical computing nodes. In theory, however, it must be able to run more virtual computing nodes on a single physical computing node and run on more physical computing nodes, but this was impossible because the K computer restricts the number of user processes on a physical computing node and the operating system kernel of the computing nodes of the K computer has a serious fault which is related to memory management.

In order to avoid the abovementioned restriction of the K computer and run more number of virtual computing nodes, in FY2015, we implemented workarounds for the prototype of our lightweight virtualization system. More specifically, we refactored our prototype implementation and the amount of memory usage were reduced to about 20 % of the original one. We also modified the prototype implementation in order to keep it from accessing the /etc/hosts file excessively when establishing connections, by directly passing IP addresses to mpirun. This modification addressed the problem of memory imbalance among physical computing nodes. In addition, we upgraded the version of OpenMPI in order to reduce the amount of memory usage.

In FY2016, we will continue the development of our system and evaluate it with more numbers of virtual computing nodes and physical computing nodes. In addition, we plan to study an approach of tricking performance profiling tools so that they feel as if they run on real computing nodes and emit profiling data which represents characteristics of real massively-parallel computing environments.

Container Technologies for HPC

Container technologies are a kind of lightweight virtualization technology. Although they tend to be less efficient than the library-hooking approach described in the previous section (Sec. 4.3.2), they provide more complete image of virtual execution environments. For example, Docker (<https://www.docker.com>) provides multiple isolated virtual Linux execution environments on a host Linux system. Because Docker is built and depends on several functionalities provided by the Linux kernel, it is not able to host non-Linux virtual execution environments unlike full-virtualization technologies (e.g., KVM, QEMU, and so on), but far more efficient than them.

One big problem of the current typical HPC systems compared to today's so-called cloud services

from viewpoint of software developers/publishers is that the HPC systems are less flexible and/or responsive. For example, they are not allowed to install and/or modify system/middleware programs in the HPC systems, while the cloud services provide fully-virtualized environments to them and they can freely modify the environments. In addition, the typical HPC systems are operated with conventional batch schedulers and it sometimes takes time to launch jobs, while the cloud services launch virtual execution environments instantly when requested by them.

The reason why the conventional HPC systems are less flexible and/or responsive is that their primary purpose is to compute scientific applications efficiently as much as possible, thus the overheads that may be introduced by utilizing full virtualization technologies are unacceptable.

On the other hand, as described above, the recent advance in the container technologies achieves very small overheads yet provides sufficiently flexible virtual execution environments, thus we predict that the container technologies will play important role in forthcoming HPC usage.

In FY2015, we utilized Docker to improve the usability of K-scope, which is a Fortran source code analysis tool developed by Software Development Team of AICS. More specifically, we created a Docker container image in which K-scope is installed so that users are able to use K-scope without manually installing it. In addition, we also extended K-scope so that users are able to analyze their programs seamlessly on the remote server without modifying their source code and/or build scripts (please note that, in FY2014, we implemented the extension of K-scope so that users are able to analyze their programs on the remote server without installing the analysis engine (more specifically, XcalableMP, which is developed of Programming Environment Research Team of AICS), but users may have to modify their source code and/or build scripts because the paths on the local machine of users and the remote server may not match).

4.3.3 Program Verification and Analysis

Memory Consistency Model-Aware Program Verification

A memory consistency model is a formal model that specifies the behavior of the memory that is shared and simultaneously accessed by multiple threads and/or processes. Under the recent multicore CPU architectures and shared memory multithread/distributed programming languages (e.g., Java, C++, UPC, Coarray Fortran, and so on), the shared memory sometimes behaves in an unexpected way because they adopt relaxed memory consistency models. For example, under some relaxed memory consistency models, the effects of the memory operations performed sequentially by one thread (e.g., A → B) may be observed in a different order by the other threads (e.g., B → A). Moreover, the threads may not agree on the orders of the effects of the memory operations (e.g., one thread observes A → B, while the other observes B → A, and so on) they observe. The reason why the recent CPUs and shared memory languages adopt relaxed memory consistency models is that enforcing sequential (non-relaxed) memory consistency incurs huge synchronization overheads among a large number of the threads/nodes that share a single address memory space.

From the viewpoint of program verification, there are two problems in handling relaxed memory consistency models. First problem is that the conventional program verification does not consider relaxed memory consistency models. Thus, they cannot be applied directly to relaxed memory consistency models because they may yield false results. Second problem is that there exist various kinds of relaxed memory consistency models and each CPU architecture/each programming language adopts different memory consistency models. Thus, it is very tedious to define and implement program verification for each CPU and programming languages of relaxed memory consistency models.

To address these problems, we have been studying three approaches. First approach is to define a new formal system which is able to represent various relaxed memory consistency models. More specifically, we define a very relaxed memory consistency model as a base model. Then, we define various memory consistency models as additional axioms on the base model. In fact, we are able to define a broad range of memory consistency models from CPUs to shared-memory programming languages (e.g, Intel64, Itanium, UPC, Coarray Fortran, and so on), in the single formal system.

Second approach is to design and implement a model checker that supports various relaxed memory consistency models based on the formal model of the abovementioned first approach. More specifically, we define a non-deterministic state transition system with execution traces where each execution trace represents a possible permutation of instruction executions. Roughly speaking, given a target program, our model checker explores all the reachable states in the non-deterministic transition system of the target problem for all the possible execution traces (that is, permutations of instructions). In our model checker, memory consistency models can be defined as constraint rules

on execution traces. For example, the sequential consistency model can be defined as a constraint which allows no permutation on the execution traces. With our model checker, we were able to verify the examples programs of the specification manuals of the memory consistency models of Itanium and UPC.

Third approach is to define a program logic for a shared-memory parallel process calculus under a relaxed memory consistency model. More specifically, we define an operational semantics for the process calculus, and then define a sound (and relatively-complete) logic to the semantics. There are two key ideas in our program logic. First idea is that a program is translated into a dependence graph among instructions in the program, and the operational semantics and the logic are defined in terms of the dependence graph. One advantage of handling dependence graphs is that while loops, branch statements, and parallel composition of processes can be handled in a uniform way. In addition, another advantage is that multiple memory consistency models can be handled by adopting different translation approaches for each memory consistency model. Second idea is that we introduce auxiliary variables in the operational semantics that temporarily buffer the effects of memory operations.

In FY2015, we further improved the implementation of McSPIN (developed from FY2013), and studied the memory consistency model of the programming language Chapel by request from a research developer of Chapel. Moreover, we also studied several memory management algorithms on various memory consistency models with external researchers.

Evidence-Based Performance Tuning

To get the maximum of HPC systems, it is inevitable to optimize the performance of applications. However, performance tuning for massively parallel HPC systems is very difficult because it is not apparent how to improve programs except for highly skilled programmers. In addition, generally speaking, modifying correctly working programs is a bothering task from the viewpoint of developers. Thus, performance tuning requires experienced craftsmanship, and relies on intuition and experience.

In order to address the problem, we have been working on evidence-based performance tuning. More specifically, we store the results of performance profiling in a database where the results are associated with source code modification history. With the database, developers are able to know, for example, what kinds of optimization were applied in the past, what kinds of optimization are effective for improving a certain performance profiling parameter, and so on.

In FY2015, we tried to increase the number of tuning cases in order to conduct detailed evaluation and improve accuracy of the analysis, but it turned out that it is hard to collect data directly because we could not find any researcher/developers who have such the data in and out of AICS. To work around the problem, we studied an approach of predicting performance of programs only from their source code modification history. In fact, we analyzed several thousands of Fortran projects registered in GitHub (<https://github.com>).

In FY2016, we will continue to try to increase the number of tuning cases in order to conduct detailed evaluation and improve accuracy of the analysis. More specifically, we will manually develop a training set by using the results of analysis of the Fortran projects obtained in FY2015, in cooperation with the two members of Software Development Team of AICS.

Python-Based Aggregation of Multiple Software for HPC

In the world of HPC, programs are usually written in somewhat old-fashioned programming languages such as Fortran/C/C++ for historical reasons, thus writing programs for HPC is painful because we cannot use useful features of modern sophisticated programming languages. On the other hand, it is not realistic so far to write a whole program in modern programming languages because of performance problems.

In order to address the problem and achieve both productivity of program development and performance of program execution, we studied an approach of using Python for writing HPC applications. More specifically, we write non-performance critical large part of a program in Python, and performance critical small part in Fortran/C/C++. The reason why we choose Python is that Python provides a rich set of foreign language interfaces. For example, Fortran programs can be interfaced with f2py (NumPy: <http://www.numpy.org/>), C programs can be interfaced with ctypes and Cython, and C++ programs can be interfaced with Boost.Python and Cython.

In FY2013, we modified EigenExa (a high performance Eigen-solver developed by the Large-scale Parallel Numerical Computing Technology research team of AICS) so that it can be used as

a shared library and a Python module (these modifications were feedbacked to the upstream). In addition, integration of Lotus (a quantum chemistry library developed by Tomomi Shimazaki, the Computational Molecular Science research team of AICS) and EigenExa were ongoing mainly by Tomomi Shimazaki.

In FY2014 and FY2015, in collaboration with Tomomi Shimazaki, a non-performance critical large part of Lotus were refactored and written in Python. With the refactoring, we were able to utilize various existing libraries (e.g., EigenExa: http://www.aics.riken.jp/labs/lpnctr/EigenExa_e.html, SMASH: <http://smash-qc.sourceforge.net/>, ASE: <https://wiki.fysik.dtu.dk/ase/>, etc.) in Lotus, and demonstrated that the features of Lotus can be easily extended. More specifically, we extended Lotus by request of Yukio Kawashima of the Computational Chemistry research unit of AICS with only several tens of lines of code addition. We further refactored Lotus by using Cython in FY2015.

Porting Performance Analysis Tools to the K computer

Because massively parallel supercomputers, such as the K computer, are very different from single computer systems or small size cluster systems, simply porting existing applications to the K computer typically does not work due to performance problems (many existing conventional applications do not consider massively-parallel systems). Therefore, performance profiling is necessary to understand the behaviors of applications on massively parallel systems and tune the applications.

To address the problem, we are porting/deploying existing performance analysis tools to the K computer, in cooperation with external research institutes. In FY2014, we ported Scalasca (<http://www.scalasca.org/>) and Extrae (<https://www.bsc.es/computer-sciences/extrae>) to the K computer. Scalasca was ported in cooperation with a research team of Jülich Supercomputing Centre, and Extrae was ported in cooperation with a research team of Barcelona Supercomputing Center. Using the ported tools, we actually analyzed the behavior of ABYSS, a parallel genome sequence assembler (<http://www.bcgsc.ca/platform/bioinfo/software/abyss>). We also analyzed the behavior of SIONlib, a parallel I/O library, in cooperation with Jülich Supercomputing Centre, and deployed it on the K computer.

In FY2015, we continued to port/deploy existing performance analysis tools to the K computer. Especially, we ported Eclipse PTP (<https://eclipse.org/ptp/>), which is an extension framework of Eclipse (<https://eclipse.org/>) for parallel program development/execution, to the K computer, in cooperation with a research team of University of Oregon. In addition, we modified and integrated Extrae and SIONlib so that Extrae is able to use SIONlib for its I/O processing (this should be useful for handling very large trace data).

4.4 Schedule and Future Plan

In FY2016, we will continue to develop the computing portal framework. Especially, we will create a VM image in which the framework is installed and ready-to-run so that the users of the K computer is able to launch their own computing portal with the framework by simply deploying the VM image to, e.g., AWS. In addition, we will also investigate to develop a Docker (<https://www.docker.com>) container image.

Regarding the virtualization technologies, we will continue the development of our lightweight network virtualization system and evaluate it with more numbers of virtual computing nodes and physical computing nodes. In addition, we plan to study an approach of tricking performance profiling tools so that they feel as if they run on real computing nodes and emit profiling data which represents characteristics of real massively-parallel computing environments.

Regarding the program verification and analysis, we will pursue our evidence-based performance tuning approach. More specifically, we will continue to try to increase the number of tuning cases in order to conduct detailed evaluation and improve accuracy of the analysis. More specifically, we will manually develop a training set by using the results of analysis of the Fortran projects obtained in FY2015, in cooperation with the two members of Software Development Team of AICS.

In addition to the above mentioned individual research topics, we also design/implement integration of the research results of the virtualization technologies and the software verification with the computing portal.

4.5 Publications

Chapter 5

Field Theory Research Team

5.1 Members

Yoshinobu Kuramashi (Team Leader)

Yoshifumi Nakamura (Research Scientist)

Hiroya Suno (Research Scientist, Joint Position with the Nishina Center for Accelerator-based Research)

Eigo Shintani (Research Scientist)

Yuya Shimizu (Postdoctoral Researcher)

Yusuke Yoshimura (Postdoctoral Researcher)

Ken-Ichi Ishikawa (Visiting Scientist, Hiroshima University)

Takeshi Yamazaki (Visiting Scientist, University of Tsukuba)

Shinji Takeda (Visiting Scientist, Kanazawa University)

5.2 Research Activities

Our research field is physics of elementary particles and nuclei, which tries to answer questions in history of mankind: What is the smallest component of matter and what is the most fundamental interactions? This research subject is related to the early universe and the nucleosynthesis through Big Bang cosmology. Another important aspect is quantum properties, which play an essential role in the world of elementary particles and nuclei as well as in the material physics at the atomic or molecular level. We investigate nonperturbative properties of elementary particles and nuclei through numerical simulations with the use of lattice QCD (Quantum ChromoDynamics). The research is performed in collaboration with applied mathematicians, who are experts in developing and improving algorithms, and computer scientists responsible for research and development of software and hardware systems.

Lattice QCD is one of the most advanced case in quantum sciences: Interactions between quarks, which are elementary particles known to date, are described by QCD formulated with the quantum field theory. We currently focus on two research subjects: (1) QCD at finite temperature and finite density. We try to understand the early universe and the inside of neutron star by investigating the phase structure and the equation of state. (2) First principle calculation of nuclei based on QCD. Nuclei are bound states of protons and neutrons which consist of three quarks. We investigate the hierarchical structure of nuclei through the direct construction of nuclei in terms of quarks.

Successful numerical simulations heavily depend on an increase of computer performance by improving algorithms and computational techniques. However, we now face a tough problem that the trend of computer architecture becomes large-scale hierarchical parallel structures consisting of tens of thousands of nodes which individually have increasing number of cores in CPU and arithmetic accelerators with even higher degree of parallelism: We need to develop a new type of algorithms

and computational techniques, which should be different from the conventional ones, to achieve better computer performance. For optimized use of K computer our research team aims at (1) developing a Monte Carlo algorithm to simulate physical system with negative weight effectively and (2) improving iterative methods to solve large system of linear equations. These technical development and improvement are carried out in the research of physics of elementary particles and nuclei based on lattice QCD.

5.3 Research Results and Achievements

5.3.1 QCD at finite temperature and finite density

Establishing the QCD phase diagram spanned by the temperature T and the quark chemical potential μ in a quantitative way is an important task of lattice QCD. We have been working on tracing the critical end line in the parameter space of temperature, chemical potential and quark masses in 3 and 2+1 flavor QCD using the $O(a)$ -improved Wilson quark action and the Iwasaki gauge action. As a first step we have determined the critical end point at zero chemical potential $\mu = 0$ in 3 flavor case. Our strategy is to identify at which temperature the Kurtosis of physical observable measured at the transition point on several different spatial volumes intersects. This method is based on the property of opposite spatial volume dependences of the Kurtosis at the transition point between the first order phase transition side and the crossover one. We have obtained $T_E=133(2)(1)(3)$ MeV and $m_{PS,E}=306(7)(14)(7)$ MeV for the temperature and the pseudoscalar meson mass at the critical end point. This is the world's first determination of the critical end point in 3 flavor QCD providing a significant step forward in our understanding of the phase diagram. As a next step we have investigated the phase structure in the presence of finite chemical potential $\mu \neq 0$ in 3 flavor QCD[cel'3f]. We have successfully determined the curvature of the critical end line on the $\mu/T-(m_{PS})^2$ plane near the vanishing chemical potential employing the Kurtosis intersection method and the multi-parameter reweighting method. After the investigation with and without the chemical potential in 3 flavor QCD, we embark on the determination of the critical end line of the finite temperature phase transition in 2+1 flavor QCD. We first focus on the behavior of the critical end line around the SU(3) symmetric point ($m_{ud} = m_s$) at the temporal size $N_T = 6$ with the lattice spacing as low as $a \approx 0.19$ fm[cel'2+1f]. Figure 5.1 plots the critical end line on the $m_\pi^2-m_{\eta_s}^2$ plane, where the pink line indicates $m_{ud} = m_s$ ($m_\pi = m_{\eta_s}$). We confirm that the slope of the critical end line takes the value of -2 at the SU(3) symmetric point, and find that the second derivative is positive. At present our investigation is extended to wider range of quark masses away from the SU(3) symmetric point.

5.3.2 Nuclei in lattice QCD

In 2010 we succeeded in a direct construction of the ${}^4\text{He}$ and ${}^3\text{He}$ nuclei from quarks and gluons in lattice QCD for the first time in the world. Calculations were carried out at a rather heavy degenerate up and down quark mass corresponding to $m_\pi=0.8$ GeV in quenched QCD to control statistical errors in the Monte Carlo evaluation of the helium Green functions. As a next step we investigated the dynamical quark effects on the binding energies of the helium nuclei, the deuteron and the dineutron. We performed a 2+1 flavor lattice QCD simulation with the degenerate up and down quark mass corresponding to $m_\pi=0.51$ GeV. To distinguish a bound state from an attractive scattering state, we investigate the spatial volume dependence of the energy difference between the ground state and the free multi-nucleon state by changing the spatial extent of the lattice from 2.9 fm to 5.8 fm. We observed that the measured ground states for all the channels are bound. This result raises an issue concerning the quark mass dependence. At the physical quark mass, namely in experiment, there is no bound state in the dineutron channel. So we expect that the bound state in the dineutron channel observed in our simulation at $m_\pi=0.51$ GeV may disappear at some quark mass toward the physical value. A new simulation at $m_\pi=0.30$ GeV, however, reveals that the ground states in all channels are bound states showing rather weak quark mass dependence in the region from $m_\pi=0.30$ GeV to 0.51 GeV[nuclei'mpi300]. In order to understand the quark mass dependence more systematically, we are currently working on the calculation of the binding energies for the helium nuclei, the deuteron and the dineutron at the physical point on a 96^4 lattice. Figure 5.2 shows a preliminary result for the effective energy difference of ${}^3\text{He}$ nuclei, which should

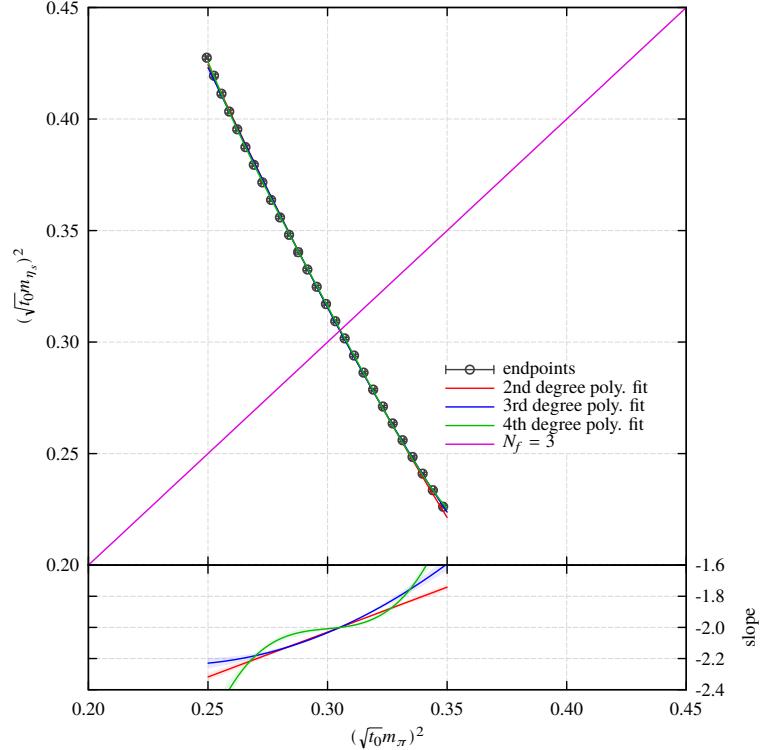


Figure 5.1: Critical end line on the m_π^2 - $m_{\eta_s}^2$ plane (top) and the slope along the critical end line (bottom). $\sqrt{t_0}$ denotes the Wilson flow scale.

represent the binding energy in large time region. Although the statistical precision is not sufficient at this stage, we plan to diminish the error bars with increasing statistics.

5.3.3 Development of algorithms and computational techniques

Application of z-Pares to lattice QCD on K computer

Eigenvalue problem for a given large sparse matrix is common across various computational sciences including lattice QCD. Sakurai group in University of Tsukuba, who has been working on the eigenvalue problem for sparse matrices for a long time, is now developing a software package for massively parallel eigenvalue computation for sparse matrices called z-Pares (short for Complex Moment-based Parallel Eigen-Solvers). We have applied z-Pares to a large scale calculation of lattice QCD on K computer. We need to solve the Wilson-Dirac equation $D\vec{x} = \vec{b}$ in lattice QCD, where D is an $N \times N$ complex sparse non-Hermitian matrix with N the number of four dimensional space-time sites multiplied by 12. In current typical simulations the dimension N is $O(10^9)$. z-Pares implements a complex moment based contour integral eigensolver: It computes eigenvalues inside a user-specified discretized contour path on the complex plane and corresponding eigenvectors. In most cases of lattice QCD calculations our interest is restricted to $O(10)$ eigenvalues around the origin so that lattice QCD should be a good example of application of z-Pares. Figure 5.3 shows a test study comparing the analytic results (black crosses) and the numerical ones (red circles) for the eigenvalues of the free (without interactions with gauge fields) Wilson-Dirac matrix on a 96^4 lattice. We observe both results show good agreement inside the contour. In Fig. 5.4 we plot the numerical results for the eigenvalues of the $O(a)$ -improved Wilson-Dirac matrix on a 96^4 lattice used for a state-of-the-art calculation in lattice QCD, whose configuration properties are explained in Ref. [lat15•075]. We find four eigenvalues (red circles) near the origin. We are now working on an algorithmic improvement to efficiently solve the shifted Wilson-Dirac equation on the discretized contour.

Tensor network scheme in path-integral formalism

The Monte Carlo simulation of lattice gauge theory is quite powerful to study nonperturbative phenomena of particle physics. However, when the action has an imaginary component like the θ

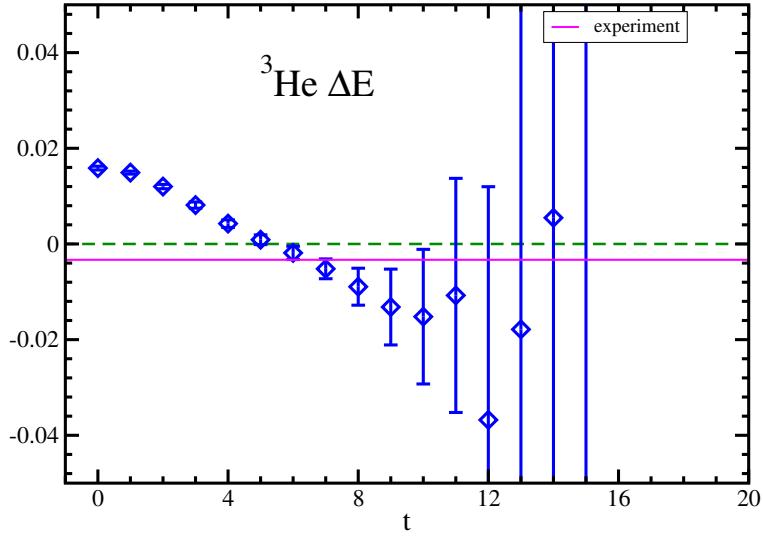


Figure 5.2: Effective energy difference for ${}^3\text{He}$ nuclei as a function of time. Solid line indicates the experimental result for the binding energy.

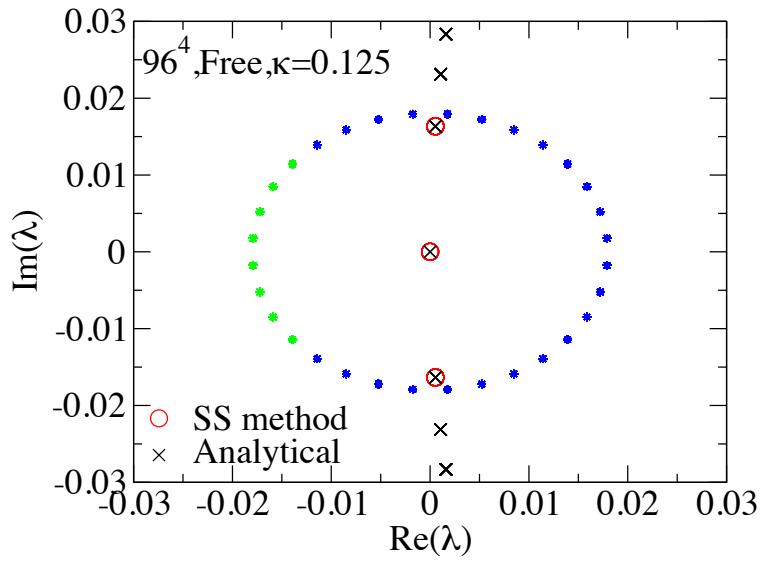


Figure 5.3: Eigenvalue distribution of the free (without interactions with gauge fields) Wilson-Dirac matrix on a 96^4 lattice. κ is a parameter to control the mass of the Wilson quark. Green and Blue star symbols denote the discretized contour around the origin of the complex plane. Black crosses denote the analytic results for the eigenvalues and red circles is for the numerical ones obtained by z-Pares.

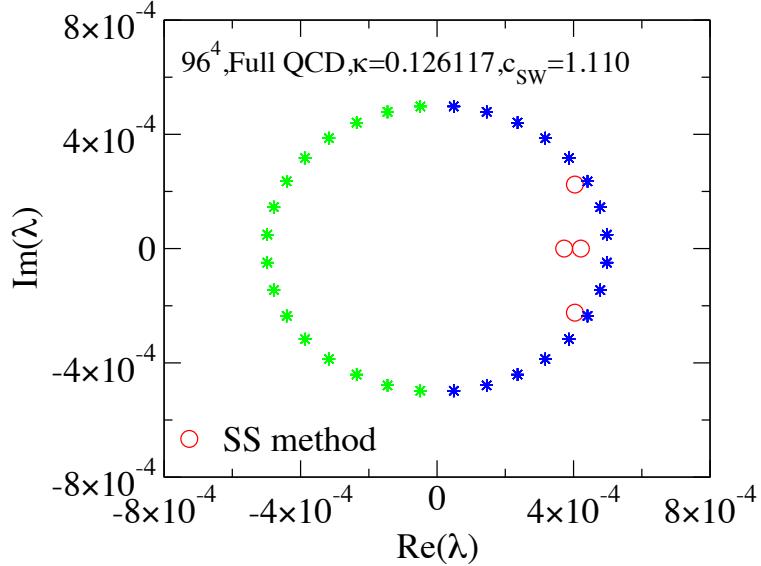


Figure 5.4: Same as Fig. 5.3 for the $O(a)$ -improved Wilson-Dirac matrix in 2+1 flavor QCD.

term, it suffers from the numerical sign problem, failure of importance sampling techniques. The effect of the θ term on non-Abelian gauge theory, especially quantum chromodynamics (QCD) is important, because it is related to a famous unsolved problem, strong CP problem. The difficulty is also shared with finite density lattice QCD. So development of effective techniques to solve or by-pass the sign problem leads to a lot of progress in the study of the QCD phase diagram at finite temperature and density. The tensor network scheme is a promising theoretical and computational framework to overcome these difficulties. So far we have developed the Grassmann version of the tensor renormalization group (GTRG) algorithm in the tensor network scheme, which allows us to deal with the Grassmann variables directly. The GTRG algorithm was successfully applied to the analysis on the phase structure of one-flavor lattice Schwinger model (2D QED) with and without the θ term showing that the algorithm is free from the sign problem and the computational cost is comparable to the bosonic case thanks to the direct manipulation of the Grassmann variables. This was the first successful application of the tensor network scheme to a Euclidean lattice gauge theory including relativistic fermions. Toward the final target of 4D QCD we are currently working on three research subjects in the tensor network scheme: (i) non-Abelian gauge theories, (ii) higher dimensional (3D or 4D) models, and (iii) development of computational techniques for physical observables. Figure 5.5 presents a preliminary result for analysis of the phase transition in the 4D Ising model. The precision of the GTRG algorithm is controlled by the parameter D_{cut} . We observe that the critical (inverse) temperature K_c converges close to the previous Monte Carlo result (blue line) as D_{cut} increases. It should be noted that our result is directly obtained on a 1024^4 lattice, while the Monte Carlo result was obtained by extrapolating the data on 80^4 and smaller lattices to the thermodynamic limit.

5.4 Schedule and Future Plan

5.4.1 QCD at finite temperature and finite density

We are now investigating the phase structure in 2+1 flavor QCD with the Kurtosis intersection method and the multi-parameter reweighting method. The first target is to determine the critical end line on the $m_{\text{ud}}\text{-}m_s$ plane. Especially, we are interested in small m_{ud} region.

5.4.2 Nuclei in lattice QCD

So far our study reveals that the dineutron channel is a bound state at heavier quark masses corresponding to $m_\pi \geq 300$ MeV. We currently make a large scale simulation at the physical quark mass. We are also investigating possible sources of the systematic errors.

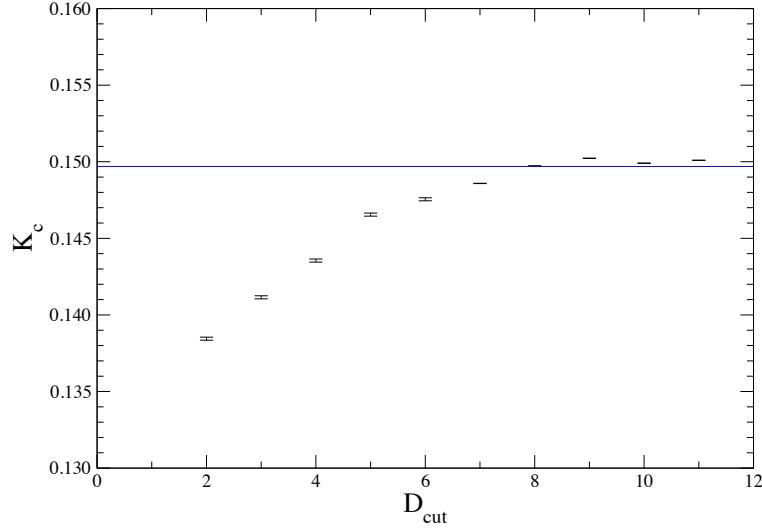


Figure 5.5: D_{cut} dependence of critical (inverse) temperature K_c in 4D Ising model on a 1024^4 lattice. Blue line denotes the previous Monte Carlo result.

5.4.3 Development of algorithms and computational techniques

Application of z-Pares to lattice QCD on K computer

In collaboration with Sakurai group at University of Tsukuba we work on an algorithmic improvement to efficiently solve the shifted Wilson-Dirac equation on the discretized contour.

Tensor network scheme in path-integral formalism

We are now try to apply the tensor network scheme to various spin models and non-Abelian lattice gauge theories on higher dimensions. It is also an interesting subject to apply the GTRG method to the chiral fermion.

5.5 Publications

Chapter 6

Discrete-Event Simulation Research Team

6.1 Members

Nobuyasu Ito (Team Leader)

Hajime Inaoka(Research Scientist)

Yohsuke Murase(Research Scientist)

Naoki Yoshioka (Research Scientist)

Takeshi Uchitane (Postdoctoral Researcher)

Shih-Chieh Wang (Postdosctoral Researcher)

Tomio Kamada (Guest Researcher)

6.2 Research Activities

Discrete-event simulations cover much wider fields than discretized simulations of continuous models. They comprise various kinds of models, for example, particles, agents, automata, games and so on, and their applications are from material and biomedical sciences to ecological and environmental problems. Social designs and controls have becoming the more interesting target since so-called the "big data sciences" became popular.

One characteristic feature of discrete-event simulations is their variety both in model parameters and behaviors. Different parameters of discrete models often result in qualitatively quite different behaviors. For example, two particles just pass through when they do not collide with each other, but they will be scattered to different orbits when they collide. A automaton reacts specifically when their inputs satisfy its activation condition. A system with such discrete elements will behaves unpredictable way. This feature is much different from the case of "continuous" simulations which are often characterized by continuous change in behaviors when input parameters are slightly modified.

Another feature of discrete-event simulations is network structure. Relations between elements are often characterized by graphs and the graphs have usually nonuniform. For example, in a system of hard particles, colliding particles are connected and noncolliding ones are not. The connection changed at every collision. Another example is human relations. Some are friendly connected, and some are competing. Such human relations are known to be characterized by a small-world structure.

Activities of the Discrete-event simulation research team(DESRT) are challenging these two features with the K and the post-K scale massive parallel supercomputers. The target problems are potentially in very vast fields of sciences, technologies and humanities, and the DESRT is now focusing mainly on social systems.

6.2.1 Parameter-space explorer

The DESRT has been developing job management tools named OACIS and CARAVAN to challenge the huge-parameter space, the first feature. The name OACIS is an abbreviation of "Organizing Assistant for Comprehensive and Interactive Simulations"**[MUIA]**.

The OACIS have been released for public use as an AICS software**[OACISA]**, and the CARAVAN is now being tested with its prerelease version. Both of these two tools are used with user's simulation and analysis softwares. A user register one's simulation and analysis softwares together with their input parameters and available host computers to these tools, then the tools control submitting and analyzing jobs.

Each simulation and each analysis for many parameters will be properly processed by computers if they are specified correctly, but error will easily be introduced by human side. In this sense, supercomputers are demanding not only greater programming skill but also more reliable operation and smarter decision of simulation parameters. The OACIS and CARAVAN are designed to solve this demand. A difference of the two tools is number of jobs and/or parameter sets. The OACIS is designed for jobs up to $10^6 \sim 10^7$ different parameters, and the CARAVAN to 10^9 and more.

The OACIS, a job management tool for simulations and analyses are designed and developed in DESRT. It is coded using Ruby, Ruby-on-rail framework and MongoDB. After installation, users register their applications for simulations and analyses, and their computers from PC to supercomputers like K to the OACIS. Then they can design and order executions of simulations and analyses on its web-browser front end. The ssh connection is used to operate the registered remote computers and Job states are supervised by the OACIS. Current prototype transfers output files of simulations and analysis to the local computer operating the OACIS from remote computers. The results and historical data are preserved in local computer using MongoDB. The first version was released last year, and the second version in this year 2015. Now users are extending to universities and research institutes. Lecture courses have been organized on demand not only in AICS but also in these sites. In this year, the OACIS activities are mainly to support user groups and minor version-ups have been continued.

The CARAVAN is coded with a PGAS language X10 implemented to the K computer. It is now under development and it will be released in following years.

6.2.2 Graph simulation

A challenge to network structures, the second feature of the discrete-event simulation, have been cultivated in contexts of phenomenological layers of social systems, especially, traffic, economics and social relations. These three are major basic components of modern society, and agent-based modeling of them has been developed in these decades.

Car traffic simulations with agent-based model of each car on a single linear road will be simple. But real roads form irregular network, and therefore treatment of networks becomes a major ingredient of simulations for real roads.

Agent-based models of car traffic simulations have been well developed since 1980s, and some simulation softwares are available now. But parallelized simulation software is not available yet. So the DESRT has been developing a parallelized car traffic simulator. Each computer process simulates a part of a road map, and simulations are geometrically parallelized.

Although the simulator is still random-walking cars without routing, the simulator had achieved simulations of all over Japan scale(see Table 6.1) and ones of all over the world was achieved in this year. Using quarter nodes of the K computer, realtime global simulation was achieved(see Table 6.2). Preliminary survey of parallelized routing algorithm has been conducted this year, and full routing will be equipped in the simulator in the following years.

To develop analysis of traffic simulation, the DESRT had developed a car traffic simulator of Kobe city**[AIIUUA]** using unparallelized simulator, SUMO. An ensemble of OD sets is assumed and Monte Carlo sampling of OD sets are executed. For each OD set, car traffic was simulated with the Kobe traffic simulator. Then the factor analysis of multivariate statistics had been done for traffic of each road segment**[UIA]**. Estimated factors(Fig. 6.1) are reasonably explain what we are experiencing in our daily lives in Kobe. Another issue of urban network traffic is whether it is stable and how it behaves macroscopically. A network flow model assuming a basic fundamental diagram between car density and flux was analyzed, and it was concluded that a network traffic is not stable in general, and therefore so-called the macroscopic fundamental diagram is fragile**[YSIA]**,

Table 6.1: performance of our parallelized traffic simulator on the K computer is listed for the road network of Japan using the Open street map(<https://openstreetmap.jp/>), which contains totally 1,284,452 Km of road with 8,143,352 road segments and 5,887,609 crossings. Totally 11,775,218 cars are simulated and each car selects next way randomly at every corner. This simulator firstly executes initialization mainly the map preparation and file I/O took 51 to 55 seconds. Then car movements are simulated. Performance data of traffic for one hour are listed for several number of nodes up to the quarter nodes of the K computer. Simulation time step was 0.01 second and therefore totally 360,000 steps were executed after the initialization.

number of node	simulation time(sec)	MPI time (sec)	map generation time(sec)	elapsed time(sec)
81	75,020	4,285	103	75,1
324	19,059	2,082	74.0	19,1
1296	4,865	1,067	68.3	4,96
5184	712	328	63.1	789
20736	386	281	65.2	499

Table 6.2: performance of our parallelized traffic simulator using the quarter nodes of the K computer is listed for the road network of all over the world using the Open street map(<https://openstreetmap.jp/>), which contains totally 30,887,952 Km of road with 104,743,486 road segments and 79,441,144 crossings. Totally 100,000,000 cars are simulated and each car selects next way randomly at every corner. This simulator firstly executes initialization mainly the map preparation and file I/O took 561 seconds. Then car movements are simulated. Performance data of traffic for 100 seconds are listed. Simulation time step was 0.01 second and therefore totally 10,000 steps were executed after the initialization.

number of node	simulation time(sec)	MPI time (sec)	map generation time(sec)	elapsed time(sec)
20736	116	41.4	74.8	857

NIA, YSIB, SYIB, YSIE]. So the real urban traffic is strongly influenced by the traffic rules and controls.

6.2.3 Other activities

Studies on ecosystems[**MSIRA, SMIA**], pedestrian and other self-propulsive and dissipative particles[**HSIA, KYSIA, KYSIB**], models of social-relation networks[**MJTKKA, MTJKKA, MTJKKB, MSIRB, MTJKKC**], molecular dynamics simulation of nonequilibrium phenomena[**ISIWHA, ISIWHA, WMIMTIA**], material breakdown[**YSIC, YSID**] were also be achieved.

During this year, a post-doc(Dr. Shih-Chieh Wang) and a researcher(Dr. Naoki Yoshioka) have arrived for new projects of modelings and simulations of disease propagation and simulations of quantum computers. Some preliminary research of these subjects are initiated.

6.3 Schedule and Future Plan

In the following, we describe our schedule and plan.

Users of the OACIS have been growing, and the more requests are coming. Formation of the OACIS user group will be the next issue, together with international extension. In addition to the user support, an API to control OACIS from a software will be developed to make intelligent and automatic model analysis possible.

The first version of CARAVAN will be released in the following years. For this purpose, fixing some problems in the X10 library is a key process.

Car routine with geometrically parallelized map over nodes will continue to be developed, and parallelized traffic simulator will be released. Simulations of disease propagation and quantum

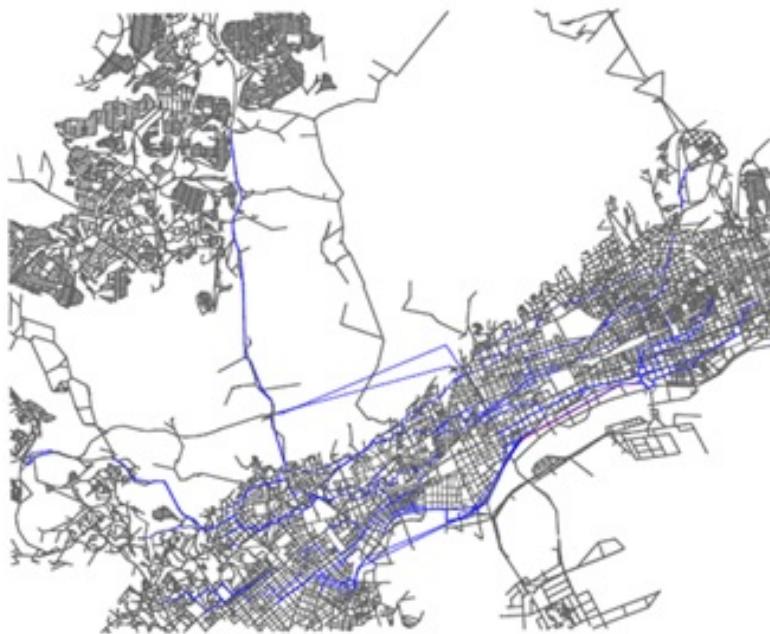


Figure 6.1: Road segments appeared in the 33 factors from 160 Monte Carlo events were shown with blue lines in a road map of the Kobe city. It is observed that the national roads, high way, and other major roads are extracted from the simulations.

computers will be conducted.

Together with these activities, future roadmap and perspective of social studies with discrete-event simulations are summarized in Fig.6.2.

6.4 Publications

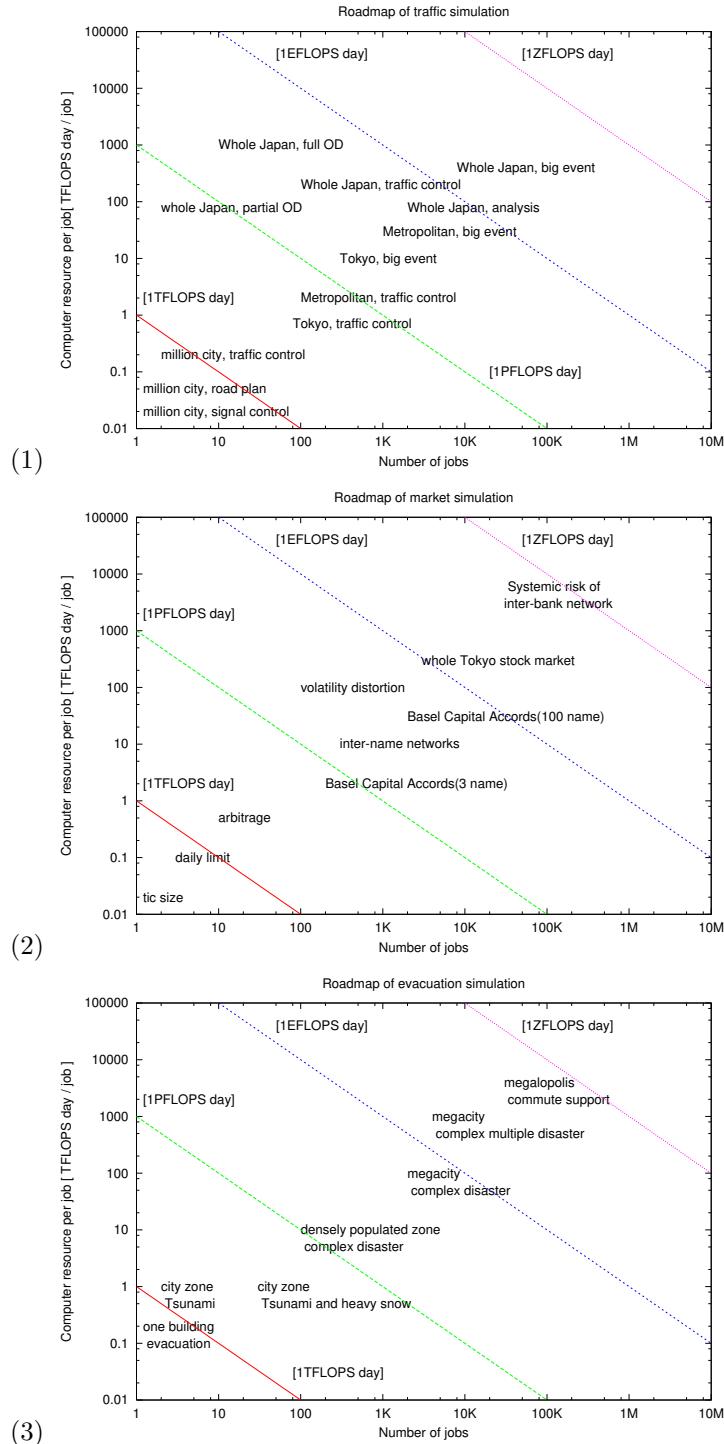


Figure 6.2: Roadmaps of agent-based simulations of (1) car traffic, (2) market and (3) evacuation are plotted. Horizontal axis shows estimated number of simulation jobs to achieve each task. Vertical axis shows computer resources of each job in a unit of TFLOPS·day, Lines of -45 degree correspond to lines of equal computational effort[NIYMKMYHA].

Chapter 7

Computational Molecular Science Research Team

7.1 Members

Takahito Nakajima (Team Leader)

Tomomi Shimazaki (Senior Scientist)

Noriyuki Minezawa (Senior Scientist)

Michio Katouda (Research Scientist)

Taichi Kosugi (Research Scientist)

Tomoki Kobori (Research Scientist)

Keisuke Sawada (Research Scientist)

Muneaki Kamiya (Visiting Scientist)

Yutaka Nakatsuka (Visiting Scientist)

7.2 Research Activities

7.2.1 Development of original molecular theory

An atomic- and molecular-level understanding of drug actions and the mechanisms of a variety of chemical reactions will provide insight for developing new drugs and materials. Although a number of diverse experimental methods have been developed, it still remains difficult to investigate the state of complex molecules and to follow chemical reactions in detail. Therefore, a theoretical molecular science that can predict the properties and functions of matter at the atomic and molecular levels by means of molecular theoretical calculations is keenly awaited as a replacement for experiment. Theoretical molecular science has recently made great strides due to progress in molecular theory and computer development. However, it is still unsatisfactory for practical applications. Consequently, our main goal is to realize an updated theoretical molecular science by developing a molecular theory and calculation methods to handle large complex molecules with high precision under a variety of conditions. To achieve our aim, we have so far developed several methods of calculation. Examples include a way for resolving a significant problem facing conventional methods of calculation, in which the calculation volume increases dramatically when dealing with larger molecules; a way for improving the precision of calculations in molecular simulations; and a way for high-precision calculation of the properties of molecules containing heavy atoms such as metal atoms.

7.2.2 New molecular science software NTChem

Quantum chemistry software comprises immensely useful tools in material and biological science research. Widely diverse programs have been developed in Western countries as Japan has lagged. In fact, only a few programs have been developed in Japan. The mission of our research team is to provide K computer users with a high-performance software for quantum molecular simulation. In the early stage of the K computer project, no quantum chemistry software was available for general purpose and massively parallel computation on the K computer because not every program was designed for use on it. Therefore, we have chosen to develop a new comprehensive ab initio quantum chemistry software locally: NTChem. NTChem is completely new software that implements not only standard quantum chemistry approaches, but also original and improved theoretical methods that we have developed in our research work. The main features of the current version, NTChem2013, are the following:

- 1). Electronic structure calculation of the ground state of atoms and molecules based on Hartree–Fock (HF) and density functional theory (DFT) methods.
- 2). Linear-scaling or low-scaling DFT: Gaussian and finite-element Coulomb (GFC) resolution-of-the-identity (RI) DFT, pseudospectral DFT/HF, and dual-level DFT.
- 3). Low-scaling SCF calculation using diagonalization-free approaches: purification density matrix, pseudo-diagonalization, and quadratic convergence SCF.
- 4). Excited-state DFT calculation: time-dependent DFT (TDDFT) and transition potential (DFT-TP).
- 5). Accurate electron correlation methods for ground and excited states: Møller–Plesset perturbation theory, coupled-cluster (CC) theory, and quantum Monte Carlo (QMC) method.
- 6). Massively parallel computing on the K computer and Intel-based architectures: HF, DFT, resolution-of-the-identity second-order Møller–Plesset (RI-MP2) method, and QMC method.
- 7). Two-component relativistic electronic structure calculation with spin–orbit interactions: Douglas–Kroll (DK) (DK1, DK2, and DK3), regular approximation (RA) (zeroth-order RA (ZORA) and infinite-order RA (IORA)), and Relativistic scheme for Eliminating Small Components (RESC).
- 8). Model calculations for large molecular systems: quantum mechanics/molecular mechanics (QM/MM) and Our own N-layered Integrated molecular Orbital and molecular Mechanics (ONIOM).
- 9). Calculation of solvation effects: COnductor-like Screening MOdel (COSMO) (interfaced with the HONDO program), averaged solvent electrostatic potential/molecular dynamics (ASEP/MD), and QM/MM-MD.
- 10). Efficient calculation for chemical reaction pathway.
- 11). Ab initio molecular dynamics calculation.
- 12). Calculation of magnetic properties: nuclear magnetic resonance (NMR) chemical shifts, magnetizabilities, and electron paramagnetic resonance (EPR) g tensors.
- 13). Population analysis: Mulliken and natural bond orbital (NBO) analysis (interfaced with NBO 6.0).
- 14). Orbital interaction analysis: maximally interacting orbital (MIO) and paired interacting orbital (PIO) methods.

7.3 Research Results and Achievements

7.3.1 Development of original molecular theory and software

Fast estimations of two-electron repulsive integrals using Pseudospectral method

A fast estimation of two-electron repulsive integrals (ERIs) is an important and imperative subject in any ab initio quantum chemical calculationsa. Since the computational cost of the ERIs formally increases as N^4 , where N is the number of basis functions, we often suffer from much time-consuming estimations in large molecular systems. In order to address the tough problem, several methodologies

have been developed to date. Among them, the pseudospectral (PS) method is a strong candidate for a quick and efficient evaluation of the ERIs. In the PS method, one analytical integral is replaced by a numerical summation consisting of discrete grid points so that the computational cost is reduced from $O(N^4)$ to $O(MN^2)$, where M is the number of grid points. Because of the discretization of a continuous integral space, the PS method is not only fast in estimations of the ERIs but also suitable for recent massively parallel computations using numerous CPU cores. Nevertheless, ab initio quantum chemical calculations with the PS method have never demonstrated in large molecular systems which contains more than 1,000 atoms. To this end, we implement the PS and PS-GAP methods into our NTChem program and investigate the performances of these methods using the MPI-parallelized code. The PS-GAP method is a further accelerated method that the PS and Gaussian and plane-wave (GAPW) methods are combined. In the following all quantum chemical calculations, the pure density functional theory was adopted, and the Kohn–Sham orbital was expanded by Def2-SVP Gaussian basis set. We employed one-dimensionally polymerized glucose-alanine chain and three-dimensionally distributed H₂O molecular cluster as test systems. The largest size system is constituted by 1,226 atoms and 10,128 basis functions. The LIBINT library was utilized in computations of analytical integrals, and the parallel FFTW library was exploited in the GAPW and PS-GAP methods. The performance check of PS and PS-GAP methods was conducted on the Research Center for Computational Science (RCCS) and the K computer system. First, we investigated computational scalings of the PS and PS-GAP methods with respect to the number of basis functions. We performed MPI computations using 16 CPU cores on the RCCS and compare the computational times of the ERIs among usual analytical, GAPW, PS, and PS-GAP methods. Figure 7.1 exhibits the computational times as a function of the number of basis functions. We find that the PS-GAP method is much faster than the other methods whereas the PS method is the slowest among all methods. Moreover, the PS-GAP method is found to achieve the low-dimensional scaling by less than square. This suggests that the PS-GAP method allows us fast calculations even for much larger systems than present maximum size one. Next, we examined parallel efficiencies of PS and PS-GAP methods in terms of large-scale parallel computations. The glucose-alanine chain with 8,202 basis functions and H₂O molecular cluster with 10,128 basis functions were adopted for test systems. We carried out MPI/OpenMP hybrid computations using eight threads per one MPI process. One MPI process was assigned to one node on the K Computer. Figure 7.2 shows the computational time of the ERIs as a function of used nodes. As well as the test of computational scalings on the RCCS, we realize the fast evaluation of the ERIs using the PS-GAP method. For the efficacy of large-scale parallelization, the PS method shows a good parallel efficiency whereas the PS-GAP method seems to quickly saturate with nodes. On the other hand, we find that the conventional analytical method also shows a good efficiency and is faster than the PS method in the case of the H₂O molecular cluster.

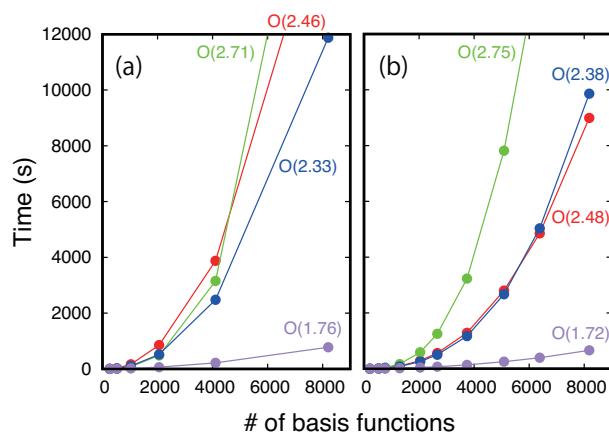


Figure 7.1: Computational time of two-electron repulsive integrals in glucose-alanine chain (a) and H₂O molecular cluster (b) as a function of the number of basis functions. The red, blue, green, and purple circles denote the analytical, GAPW, PS, and PS-GAP methods, respectively.

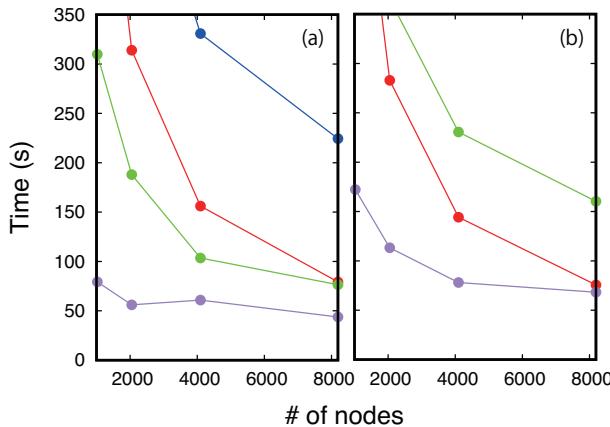


Figure 7.2: Computational time of two-electron repulsive integrals in glucose-alanine chain with 8,202 basis functions (a) and H₂O molecular cluster with 10,128 basis functions (b) as a function of used nodes. The red, blue, green, and purple circles denote the analytical, GAPW, PS, and PS-GAP methods, respectively.

Development of massively parallel algorithm of linear response time-dependent density functional theory for excited-state calculations on peta-flops supercomputers

Linear response time-dependent density functional theory (LR-TDDFT) is an efficient approach for the calculation for excited-states of large molecules. In the previous study, we developed the MPI/OpenMP hybrid parallel code of LR-TDDFT calculations with random phase approximation (RPA) and Tamm–Dancoff approximation (TDA) [Int. J. Quant. Chem. 115,349 (2015).] in NTChem software. However, the parallel scalability of this code was limited up to about 1,000 MPI processes on the K computer due to the problems of the parallelization scheme and the I/O overhead of the Davidson diagonalization in the LR-TDDFT calculations. In this FY, we have developed a new MPI/OpenMP hybrid parallel algorithm for LR-TDDFT/RPA and LR-TDDFT/TDA calculations. In this algorithm, we have devised an efficient parallelization scheme of the Davidson diagonalization utilizing more than 1,000 MPI processes on the peta-scale supercomputers such as the K computer. In this scheme, the Davidson diagonalization is parallelized utilizing the distributed memory avoiding the I/O operation reading and writing data to disks. Vectors needed for Davidson diagonalization (trial vectors, sigma vectors, and residual vectors) are divided to the two-dimensional blocks, and these blocks are distributed to each process. Matrix-matrix and matrix-vector multiplications evaluating the residual vector and orthogonalization of trial vectors are the computation demanding tasks in the Davidson diagonalization. These multiplications are performed efficiently applying the OpenMP parallelized basic linear algebra subprograms (BLAS) library. The scalability of parallel algorithm has been evaluated performing some test calculations of large nano-size molecules. The largest calculation is the calculation of crambin protein (PDB ID: 1CRN) at the LR-TDDFT/RPA/def2-SVP level (642 atoms 6,177 atomic orbitals (AOs), and 50 excited states) using from 768 to 7,680 nodes of the K computer. The parallel scalability is good up to 4,608 nodes. Parallel speedups are 80% and 57% using 1,536 nodes and 4,608 nodes, respectively. The elapsed time is 251 minutes using 4,608 nodes, which demonstrates that the LR-TDDFT excited state calculations of molecular materials and biological molecules containing up to 600 atoms and 6,000 AOs become routinely possible utilizing the K computer. The new parallel LR-TDDFT code has been released with NTChem 2013 Version 7.0 on Dec. 2015.

Development of analytical energy gradient for two-component relativistic time-dependent density functional theory with spin-orbit interaction

Precise information on excited state potential energy surfaces is the most important prerequisite for a deeper understanding of photochemical reaction or the shape of absorption and luminescence spectra. In particular, the inclusion of spin-orbit coupling and other relativistic effects is crucial for a proper description of excited-state characters, relaxation dynamics, radiative and nonradiative decay pathways, as well as lifetimes and reactivity for systems containing heavy elements. The development and efficient implementation of an analytic gradient theory for reliable theoretical mod-

els incorporating both electron correlation and relativistic effects has been awaited. Recently, we have developed two-component relativistic time-dependent density functional theory with spin-orbit interactions (SO-TDDFT), which is becoming popular methodologies for computing excited states containing heavy elements because of its reasonable cost and relatively high accuracy. In this work, we have implemented the analytical gradient for two-component relativistic TDDFT with spin-orbit interactions in the NTChem program. Our implementation is based on the derivation of the geometrical derivatives for nonrelativistic TDDFT presented by Furche and Ahlrichs. The noncollinear exchange-correlation potential presented by Wang *et al.* has been applied. The higher-order matrix elements of the noncollinear exchange-correlation kernel for the relaxed one-particle and two-particle density matrices have been derived and implemented into efficient computer codes with the aid of a newly-developed computerized symbolic algebra system. In addition, various DFT functionals including the recently proposed range-separated hybrid functionals are applicable to the calculations of excitation energy for spin-orbit coupled states.

Development of a method for electronic-structure calculations based on self-energy functional theory

Electronic-structure calculations based on the density functional theory (DFT) have been used in quantum chemistry and providing reliable explanations and predictions of many molecular systems. While the elaboration of techniques of DFT-based calculations is continued in the community, the limitation of the framework of DFT is known to exist and should be overcome, which is problematic particularly for strongly correlated systems. For such systems, model Hamiltonians have been used traditionally, in which only few adjustable parameters looking important are often introduced by hand. Among many approaches proposed so far, the approach based on the self-energy functional theory (SFT) [Potthoff, Eur. Phys. J. B 32 (2003) 429.] is promising. It has been applied to systems with model Hamiltonians such as a Hubbard chain [Potthoff, Eur. Phys. J. B 36 (2003) 335.] and NiMnSb [Allmaier *et al.*, Phys. Rev. B 81 (2010) 054422.], however, no calculation based purely on SFT has not been reported to our knowledge. Since a SFT-based calculation uses the result of exact diagonalization for the subspace appropriately chosen from the entire Hilbert space, it is expected that the many-body effects are taken into account accurately for describing the strongly correlated system. We started to develop a method for SFT-based quantum chemistry calculations. Specifically, we first construct the second-quantized Hamiltonian by calculating one- and two-electron integrals for a molecule and divide the entire Hilbert space into the subspaces so that as many the molecular orbitals in the vicinity of the Fermi level are contained in one of the subspaces as computationally possible. We then perform exact diagonalization for the subspaces and calculate the thermodynamic potential of the molecule [Potthoff, Eur. Phys. J. B 32 (2003) 429.] from the self-energies Σ of the subspaces as $\Omega = F[\Sigma] - \text{Tr} \ln(-G_0^{-1} + \Sigma)$, where $F[\Sigma]$ is the universal functional of the self-energies and G_0 is the non-interacting Green's function.

Theoretical study on the cooperative exciton dissociation process based on dimensional and hot charge-transfer state effects in an organic photocell

In recent years, organic electronics devices such as electroluminescent displays, transistors, and photocells, have actively been studied because of various properties of organic materials, such as lightweight, thin film structures, flexibility, manufacturing costs, design, and so on. We have been studying organic photocell devices. It is critical to understand the energy conversion mechanism from the solar photon flux in organic semiconductors, especially the dynamics of strongly bound electron-hole pairs (excitons). In organic photocells, excitons are first created by absorbing photons. They then diffuse to the donor-acceptor interface of the organic semiconductor. Finally, they are dissociated to create free electrons and holes after the electron (charge) transfer process. Here, we studied the cooperative exciton dissociation process based on dimensional and hot charge transfer state effects. In last FY, we introduced a local temperature to handle with the hot charge-transfer (CT) state, and calculated the exciton dissociation probability based on the one-dimensional organic semiconductor model. Although the hot CT state plays essential roles, the probabilities calculated are not enough high to efficiently separate bounded electron-hole pairs. In this FY, therefore, we focused on the dimensional effect together with the hot CT state effect. We showed that cooperative behaviors between both effects can significantly improve the exciton dissociation process.

Development of trajectory surface hopping algorithm for the time- dependent density functional theory: Toward the understanding of working mechanisms of organic photovoltaic solar cells

The solar cell can convert solar energy into electric energy and is a valuable resource of clean energy that is free from environment pollution. While people have used silicon-based solar cells widely in a real life, organic photovoltaic devices are considered to be promising in a next generation due to their versatility, easy processing, low cost, and so on. The photo-energy conversion efficiency is a critical factor in the development of new solar cells. So it is necessary to derive a valuable relationship between the conversion efficiency and the properties of constituent organic molecules. Nowadays, quantum chemistry calculations routinely afford (1) the energy gap between the highest occupied molecular orbital and lowest unoccupied molecular orbital, (2) excitation energies, and (3) oscillator strengths. Although these data give guidelines to develop new photo-functional materials, they are insufficient to reveal the details of photochemical processes after light absorption; how solar cells work relies on the mechanisms of the exciton (electron-hole pair) generation and its separation/migration. Therefore, it is desirable to simulate the relaxation processes on the excited-state potential energy surfaces (PESs) directly and to observe the real-time behavior of electron–hole pair. A large size of organic dye molecules hampers the on-the-fly molecular dynamics simulation by sophisticated wavefunction approaches. Furthermore, photochemical processes involve a large number of PESs, and one must take accounts of non-adiabatic transition between these electronic states. To satisfy these requirements, we have implemented a combined method of time-dependent density functional theory (TD-DFT) and trajectory surface hopping (TSH) algorithm. For now, the TSH input data was taken from the TDDFT output given by the program package GAMESS. The TSH program needs the non-adiabatic coupling vector (NACV), which is not trivial at the TDDFT level of theory. Instead of deriving analytic NACV, we introduced isotropic scalar NAC as the overlap of wavefunctions at consecutive time steps and computed it on the fly. Some test calculations in progress show the efficiency and robustness of the proposed method.

7.3.2 Applications of original molecular theory to new materials and drugs

Theoretical study on spin-forbidden transitions of metal complexes by two-component relativistic time-dependent density functional theory

Spin-forbidden transitions of metal polypyridyl sensitizers are studied by the two-component relativistic time-dependent density functional theory with spin-orbit interaction based on Tamm–Dancoff approximation. The spin-forbidden transitions for a phosphine-coordinated Ru(II), DX1, as well as the modified DX1 complexes whose Ru is replaced with Fe and Os, are calculated. The role of the central metals in spin-forbidden transitions is discussed toward the exploration for new efficient sensitizers. In addition, we study spin-forbidden transitions of Os polypyridyl sensitizers. The absorption spectra, including spin-forbidden-transition peaks, for the Os complexes are reasonably reproduced in comparison with the experimental ones. The extension of the conjugated lengths in the Os complexes is investigated and found to be effective to enhance photo absorption for spin-allowed transitions as well as spin-forbidden ones. This study provides fruitful information for a design of new dyes in terms of conjugation lengths.

Large-Scale QM/MM calculations of hydrogen bonding networks for proton transfer and water inlet channels for water oxidation - Theoretical system models of the oxygen-evolving complex of photosystem II

In order to confirm theoretical system models of photosystem II (PSII), quantum mechanics (QM)/molecular mechanics (MM) calculations using a large-scale QM model (QM Model V) have been performed to elucidate hydrogen bonding networks and proton wires for proton release pathways (PRPs) of water oxidation reaction in the oxygen-evolving complex (OEC) of PSII. Full geometry optimizations of PRP by the QM/MM model have been carried out starting from the geometry of heavy atoms determined by the recent high-resolution X-ray diffraction (XRD) experiment on PSII refined to 1.9 Å resolution. The optimized MnMn and CaMn distances by large-scale QM/MM are consistent with the EXAFS results, removing out the discrepancy between the refined XRD and EXAFS. Computational results from QM/MM calculations have demonstrated the labile nature of the Mn₄O₅Mn_d

bond of the CaMn₄O₅ cluster in the OEC of PSII which allows left (L)-opened, quasi-central (CQ)-, and right (R)-opened structures. This confirms the feasibility of the left- and right-hand scenarios for water oxidation in the OEC of PSII that are dependent on the hydrogen bonding networks. The QM/MM computations have elucidated the networks structures: hydrogen bonding O...O(N) and O...H distances and O(N)H...O angles in PRP, together with the ClO(N) and Cl...H distances and O(N)H...Cl angles for chloride anions. The obtained hydrogen bonding networks are fully consistent with the results from XRD and available experiments such as EXAFS, showing the reliability of our theoretical system model that is crucial for investigations of functions of PSII such as water oxidation. The QM/MM computations have elucidated possible roles of chloride anions in OEC of PSII for proton transfers. The QM/MM computational results have provided useful information for the understanding and explanation of several experimental results obtained with mutants of the OEC of PSII. The possible implications of the present results are discussed in relation to our theoretical system models of PSII, strong or weak perturbations of the system structures by mutations, damage-free X-ray free-electron laser structure of PSII, and bioinspired working hypotheses for the development of artificial water oxidation systems which use 3d transition metal complexes.

Full geometry optimizations of the CaMn₄O₄ model cluster for the oxygen evolving complex of photosystem II

Full geometry optimizations of ([CaMn₄O₄(CH₃COO)₈(py)(CH₃COOH)₂], (py: pyridine) (1)) were performed at the UB3LYP theoretical level. This model 1 is a theoretical model for the synthetic model ([CaMn₄O₄(ButCOO)₈(py)(ButCOOH)₂], (But: t-butyl) (2)) which closely mimics the native oxygen evolving complex (OEC) in photosystem II. It was shown that the X-ray structure of 2 was well reproduced by 1 in the (Mn1(III), Mn2(IV), Mn3(IV), Mn4(III)) valence state with the unprotonated O₅ (O₅ = O₂⁻), and two different valence states were obtained in the one-electron oxidized state. Importance of the Jahn-Teller effect of the Mn(III) site for the structural deformations was presented.

Theoretical study on selective recognition of biomolecule in supramolecule

The purpose of this study is to analyze and understand of molecular recognition using quantum chemical calculations. In 2009, Sawada *et al.* observed only single nucleotide duplex could form stably in water inside artificial cage-like supramolecule [Sawada *et al.* Nature Chemistry 1, 53 (2009)]. Since it was said that at least four complementary nucleotide base pairs had been necessary in order to form stable structure in water, this result can be called great progress. In our study, we approached this issue in terms of the molecular orbital (MO) method and tried to clarify structural stability and selectivity mechanism of the target. The selectivity in this target is that single base pair is spontaneously formed as anti-Hoogsteen (AH) type, not Watson–Crick (WC) type. In order to elucidate this selectivity, we prepared both AH-type and WC-type nucleotide duplex and performed structure optimizations with/without surrounding cage-like supramolecule. According to the result, AH-type structure is more stable than WC-type one by 53.1 kcal/mol inside cage-like supramolecule and the same energy gain becomes only 0.2 kcal/mol without the supramolecule. Moreover, our calculation indicates the importance of non-covalent bonding between the cage and nucleotide duplex which originates from π-π/CH-π interaction.

7.4 Schedule and Future Plan

In the next financial year, we will continue to develop new algorithms and improve the parallel efficiency of the NTChem2013 suit of program. In the present implementation of LR-TDDFT, the replicated memory algorithm is adopted for the evaluations of Fock like matrix. This situation limits the application of the present code to the systems having less than 7,000 AOs on the K computer. To overcome this problem, we are developing the distributed-memory parallel code of Fock-like matrix calculations. Furthermore, the spin-orbit coupling often plays an important role for the excited states containing heavy elements. We are developing the massively parallel algorithm and code of two-component relativistic LR-TDDFT to account for the spin-orbit interactions explicitly in the excited state calculations. We will also develop transition properties and nonadiabatic coupling constant matrix elements, which are the key quantity in the description of excited-state dynamics. In addition, for further acceleration of the PS and PS-GAP methods, we are planning to do coding and

tuning of related programs in NTChem so that we will apply the improved PS and PS-GAP methods to several large molecular systems. For self-energy functional theory (SFT), we further will continue to parallelize the calculations of one- and two-electron integrals and of exact diagonalization and of Green's functions. The implementation of SFT-based electronic-structure calculations for periodic systems are in progress. It is interesting to develop a unified approach to the understanding of the internal conversion (IC) and intersystem crossing (ISC) processes. The suppression of unfavorable quenching pathways may help the design of new organic materials with high efficiencies. One of the attractive points in NTChem program is the availability of relativistic methods. The spin-orbit TDDFT (SO-TDDFT), in particular, can deal with both singlet and triplet states on an equal footing. The SO-TDDFT/TSH simulation involving both spin states can be performed in a similar way as the singlet-state-only TSH because a natural extension of the conventional TDDFT coupling yields the SO-TDDFT counterpart. Another direction is to develop a hybrid method of quantum mechanics/molecular mechanics (QM/MM). Bulk heterojunction solar cell, for example, is a composite system of p- and n-type layers. It is natural to split the whole system to reduce the computational cost. The QM method is applied to the small region, the interface between the p- and n-type materials, while the remaining part is treated as MM. The method will give insights into the environmental effects on charge transport dynamics that occurs at the interface. We earnestly hope that NTChem will be an important tool leading the way toward a new frontier of computational molecular science.

7.5 Publications

Chapter 8

Computational Materials Science Research Team

8.1 Members

Seiji Yunoki (Team Leader)

Yuichi Otsuka (Research Scientist)

Shigetoshi Sota (Research Scientist)

Shixun Zhang (Postdoctoral Researcher)

Ahmad Ranjbar (Postdoctoral Researcher)

James S. M. Anderson (Foreign Postdoctoral Researcher)

Kazuhiro Seki (Special Postdoctoral Researcher, joint)

Sandro Sorella (Senior Visiting Scientist)

Takami Tohyama (Senior Visiting Scientist)

Yutaka Imamura (Visiting Scientist)

Susumu Yamada (Visiting Scientist)

Michele Casula (Visiting Scientist)

Keiko Matsuoka (Assistant)

8.2 Research Activities

The main purpose of our team is to understand and predict quantum states of condensed matter systems and solid state materials for the next generation highly functional materials and for the basis of future revolutionary technologies. More emphasis is put specially on strong many-body correlations that require the treatment going beyond the single-particle approximation. For this purpose, we develop many-body numerical methods, which include numerically exact diagonalization method, classical Monte Carlo and molecular dynamics methods, quantum Monte Carlo method, and density matrix renormalization group method. We also develop highly parallelized application software based on these numerical methods to efficiently perform large-scale simulations.

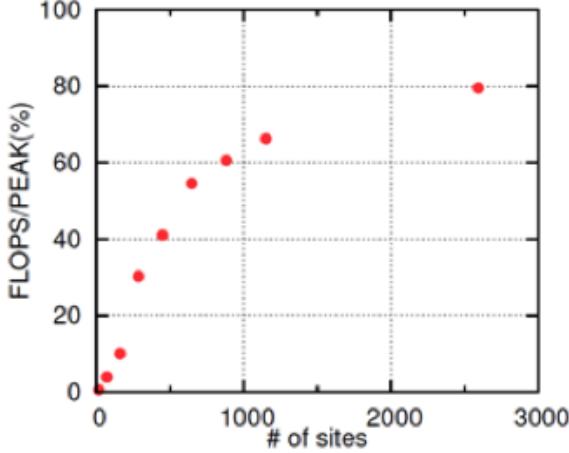


Figure 8.1: Peak performance of benchmark calculations for the Hubbard model with various system sizes.

8.3 Research Results and Achievements

8.3.1 Development of large-scale quantum Monte Carlo (QMC) simulations

The QMC method is one of the most powerful tools to investigate quantum many-body systems. There are several variants of QMC techniques depending on target systems. Among them, we have developed the determinant Monte Carlo method to study the interacting fermions on lattice, i.e., the Hubbard model, with a specific aim to calculate physical observables with a high degree of accuracy on quite large systems. The accuracy and the system size are inseparable in the lattice simulations as the energy resolution is roughly determined by the inverse of the linear size L of the system. Therefore, it is inevitable to perform simulations on very large clusters in order to accurately calculate physical observables, which demands huge computational resources. Since the numerical calculations are based mostly on linear algebraic operations such as matrix-matrix product and numerical orthogonalization, we can take advantage of the highly optimized numerical library on K computer. We have also optimized the delayed update algorithm to efficiently use the cache memory. We have confirmed that our simulation code achieves 80% of the peak performance in the single node calculation with increasing the system size (see Fig. 8.1). Since in this case one Slater determinant, representing one ensemble in the Monte Carlo sampling, can be stored in the single node, we are able to use up to 24,576 nodes with quite high efficiency of about 50% of the peak performance in total. We have utilized this code in the practical study of the two-dimensional (2D) Hubbard model on honeycomb lattice with the system size up to $N = 2,596$ sites, which is about 4 times larger than the previous study ($N = 648$) [Z. Y. Meng et al., *Nature* **464**, 847 (2010)] and represents currently the largest system size ever. The computational complexity scales as N^3 and thus it could have not been possible without K computer. We have also implemented more non-trivial parallelization for the case where one Slater determinant is distributed to a group of multiple nodes, in which the parallelized numerical library ScaLAPACK is employed to update the Slater determinant in each Monte Carlo sweep. With this implementation, it has become possible to perform the simulations with N up to 8,100 within reasonable CPU time.

8.3.2 Absence of spin liquid in the half-filled Hubbard model on the honeycomb lattice

First, we have applied our improved QMC code to elucidate the ground state phase diagram of the half-filled Hubbard model on the honeycomb lattice model (honeycomb lattice model), in which a gapped spin liquid (SL) phase was proposed previously [Z. Y. Meng et al., *Nature* **464**, 847 (2010)]. Since it is widely believed that not only strong quantum fluctuations but also geometrical frustrations are responsible for stabilizing a SL phase, the possibility of SL phase in the unfrustrated honeycomb lattice is rather surprising, and thus has been one of the most debated issues in recent years. The

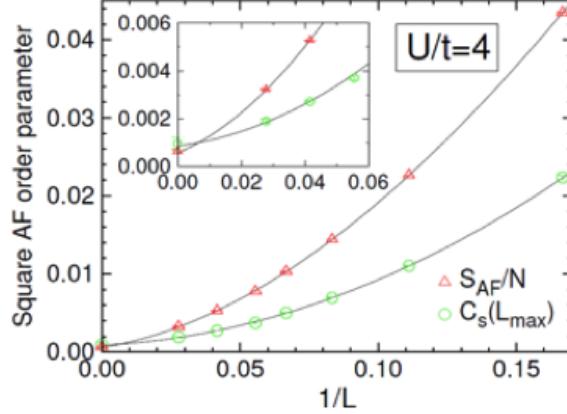


Figure 8.2: Finite size scaling of AF order parameter squared at $U/t = 4$.

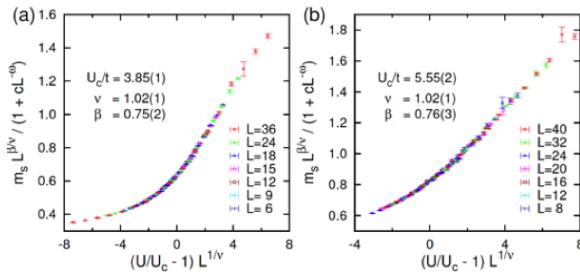


Figure 8.3: Data collapse fits for (a) honeycomb lattice and (b) π -flux models.

proposed SL phase in the previous report was claimed to exist for $3.4 < U/t < 4.3$ (U : Hubbard interaction, t : nearest neighbor hopping) as a spin-gapped insulating phase without any broken symmetry. We thus have first tried to clarify the existence of SL at $U/t = 4$, which corresponds to the middle of the proposed SL region. Taking a full advantage of K computer, we have performed the QMC simulations on the lattice with $N = 2L^2$ up to 2,596 sites. Figure 8.2 shows our results of the antiferromagnetic (AF) spin structure factors, S_{AF} , and spin-spin correlation functions at the maximum distance, $C_s(L_{max})$, at $U/t = 4$. The extrapolated values of both quantities are confirmed to be finite within statistical errors, indicating the AF long-range order. Complemented with simulations performed at other U/t , our results strongly support the conventional scenario that a single and direct phase transition occurs between semi-metal and antiferromagnetic Mott insulator with increasing U/t .

8.3.3 Quantum criticality of metal-insulator transition in 2D interacting Dirac electrons

Next, having established the continuous character of the transition, we have performed careful finite-size scaling analysis of our QMC data to obtain critical exponents of this metal-insulator transition. As shown in Fig. 8.3, the data of the staggered magnetization calculated on finite-size clusters are excellently collapsed into a universal function in the honeycomb lattice model and also the Hubbard model on square lattice with a magnetic flux π per plaquette (π -flux model). The latter model is known to have massless Dirac dispersions as in the honeycomb lattice model. It is turned out that the critical exponents are exactly the same within the statistical errors for the two lattice models, which strongly suggests that the metal-insulator transitions in these models belong to the same universality class. This class should be coincident with that in the Gross-Neveu model, a model extensively studied in the particle physics, since it has recently been recognized that the effective model for the 2D interacting Dirac fermions in the continuous limit is described by the Gross-Neveu model. Thus, we expect that our finding based on the unbiased large-scale QMC simulations have an impact also in an interdisciplinary field.

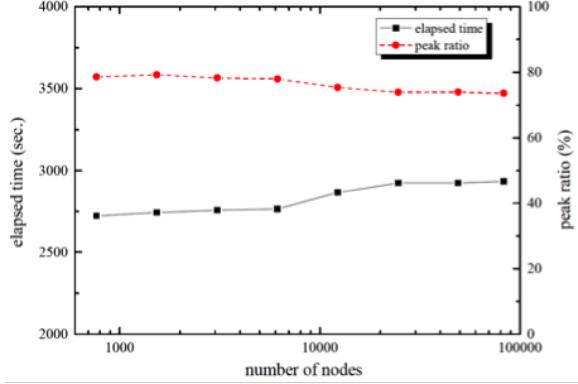


Figure 8.4: Performance of our massively parallel DMRG code on K computer.

8.3.4 Development of massively parallel density matrix renormalization group (DMRG) algorithms

The DMRG method was originally proposed by S. White in 1992. Although the DMRG method was soon recognized as one of the best numerically exact methods for strongly correlated many-body quantum systems, the application is severely limited mostly to one-dimensional (1D) systems because of the exponential increase of degrees of freedom in higher spatial dimensions. However, owing to high performance computers that are recently available, it has become realistic to apply the DMRG method even to 2D systems. We have been developing massively parallel DMRG algorithms to investigate strongly correlated quantum systems in two dimensions. We have already successfully developed our massively parallel DMRG code that employs many efficient parallelization techniques, and the first version of our massively parallel DMRG code has been opened in public for general users of K computer. A typical performance of the present version of our massively parallel DMRG code is summarized in Fig. 8.4. For this example, we have calculated the optical conductivity for the 1D Hubbard model, using K computer with up to 82,488 nodes. We have performed the dynamical DMRG calculations with the DMRG truncation number $m = 8,064$ (the number of bases kept during the calculation). The energy parallelization number is varied in such a way to correspond to the weak scaling. As shown in Fig. 8.4, we have achieved the extremely high peak performance ratio of 73.6% when 82,488 nodes are used on K computer, which is about 7.8 PFLOPS.

8.3.5 Magnetic excitations of the doped Hubbard model in two dimensions: Dynamical DMRG study

In order to disentangle magnetic excitations in the resonant inelastic X-ray scattering (RIXS) spectrum for electron-doped high-Tc cuprate $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_4$ (x : electron doping rate from half filling), we have performed the dynamical DMRG calculations using our massively parallel DMRG code on K computer. It has been reported experimentally that the peak position corresponding to the magnetic excitation in the RIXS spectrum shifts toward higher energy with increasing doping rate x [K. Ishii, et. al., Nat. Commun. **5**, 3714 (2014)]. To understand this unexpected doping dependence, we have calculated dynamical spin structure factor $S(\mathbf{q}, \omega)$ for one of the simplest models for the high-Tc cuprate, i.e., the 2D Hubbard model with long range hoppings on the square lattice. Using K computer, we were able to do the calculations for clusters up to 6x6 sites, which represents currently the largest dynamical calculations in the world for the Hubbard model away from the half filling. We have found that the lowest peak in the dynamical spin structure factor $S(\mathbf{q}, \omega)$ shift toward higher energy with electron doping, thus in good qualitative agreement with the RIXS observation.

8.3.6 Time-dependent DMRG method for real time quantum dynamics in two dimensions

Up to now, no reliable numerical method has existed for the real time dynamics of 2D strongly correlated quantum systems. We have challenged this issue by extending our massively parallel DMRG method combined with the kernel polynomial expansion. The previously proposed time-

dependent DMRG method employs the Suzuki-Trotter decomposition for the time evolution operator $e^{i\delta t \hat{H}}$. Although this scheme costs computationally quite small since the time-evolution operator is given a product of local operators, it is limited only to 1D systems. Therefore, we have proposed a new time-dependent DMRG method combined with the kernel polynomial expansion to calculate directly the time evolution operator in any spatial dimension. Employing Legendre polynomials $P_l(x)$, the time-evolved state at $T = t + \delta t$ is given as $|t + \delta t\rangle = \lim_{L \rightarrow \infty} \sum_{l=0}^{\infty} (-i)^l (2l + 1) j_l(\delta t) P_l(\hat{H}) |t\rangle$, where $j_l(x)$ is a spherical Bessel function. Since the special functions are satisfied with the three-term recursive formula, the time-evolved state is calculated recursively. In the practical calculation, we can truncate the number L of kernel polynomials to a finite value because the convergence of the kernel polynomial expansion is quite fast when δt is small. With our new algorithm, we were successfully able to perform the reliable real time dynamics simulations in two dimensions for the first time. We have investigated, for example, the quantum annealing for the 2D Ising model with the time-dependent transverse magnetic field, a relevant system for quantum information.

8.3.7 Development of ab-initio DMRG method for strongly correlated molecular systems

Ab-initio DMRG method is developed in the quantum chemistry to perform the full configuration interaction (full-CI) calculations. Traditionally, the exact diagonalization method is used in the full-CI calculations. However, the degrees of freedom for strongly correlated molecular systems are quite large and increase exponentially with the number of orbitals considered. Thus, the exact diagonalization method is severely limited to small molecules. To overcome this difficulty, we have been developing massively parallel ab-initio DMRG method for the full-CI calculations. The computational complexity depends strongly on the connectivity of the system, which relates to the area-low of the entanglement entropy of the quantum state, and thus the order of the orbitals is crucial for the efficient ab-initio DMRG calculations. We have developed a general scheme to give the most efficient order of the orbitals in the DMRG calculation by using the graph theory. This allows us to perform the ab-initio DMRG calculations without explicitly considering the orbital order. Our ab-initio DMRG method has been extended for the dynamical calculations using the dynamical DMRG method.

8.3.8 Magnetic Skyrmions in chiral magnets: Electrons coupled with classical spins

Spins in chiral magnetic materials such as MnSi with a non-centrosymmetric structure are modulated significantly by Dzyaloshinskii-Moriya (DM) interaction and thus exhibit complex spin texture. Notably, there exists a vortex-like spin configuration named Skyrmion that shows non-trivial topological properties such as topological Hall effect. Due to the potential application of Skyrmion in the future information storage, Skyrmion has attracted intensive attention in recent years. We have examined chiral magnetic structure including Skyrmion using Monte Carlo and molecular dynamics simulations.

The first class of systems studied is modeled by the double exchange model where the classical spins are coupled to the non-interacting electrons described by a simple tight binding model. To simulate this model, we have developed a semi-quantum Monte Carlo method, in which the electron degrees of freedom are numerically solved exactly for a give spin configurations and the spin degrees of freedom are treated using the importance sampling. Because the system size N has to be large enough to accommodate many Skyrmions with typically tens of nanometers, the exact diagonalization of electron degrees of freedom requires huge computation cost of $O(N^3)$. Therefore, we have adopted a kernel polynomial method (KPM) to expand the electron Green's function using Chebyshev polynomials, which reduces the computational complexity down to $O(N)$. Applying the KPM, we were able to simulate up to 10,000 electrons coupled with 10,000 classical spins. The Skyrmion phase diagram with temerature vs magnetic field is shown in Fig. 8.5, which is in excellent agreement with experimental observations.

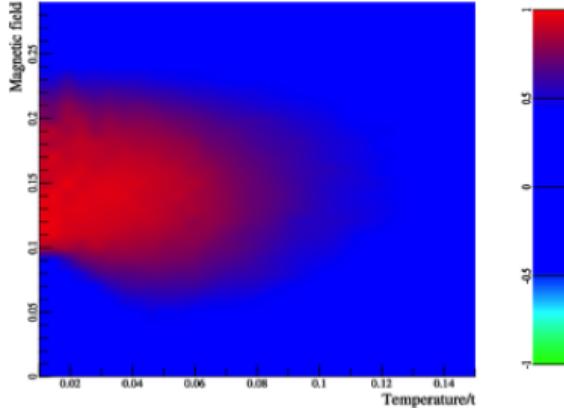


Figure 8.5: Normalized Skyrmion number distribution in terms of magnetic field (B/t) and Temperature (T/t).

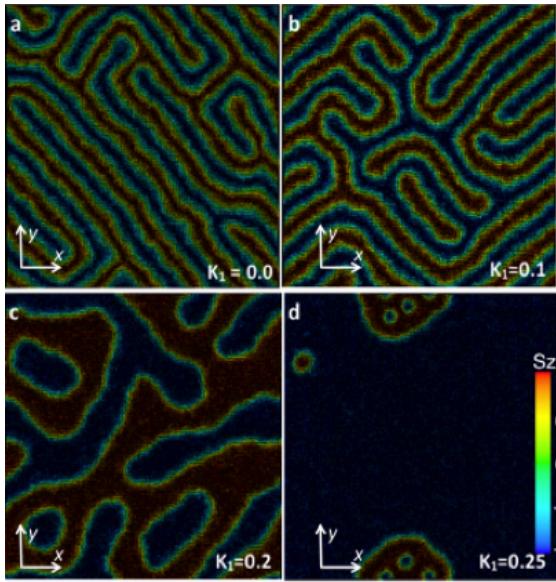


Figure 8.6: Spin structure of the 2D FM Heisenberg model with the easy axis spin anisotropy (K_1) but no external magnetic field.

8.3.9 Classical Monte Carlo and molecular dynamics simulations for chiral spin texture

In addition, we have developed classical Monte Carlo and molecular dynamics simulations for pure classical spin models. Because the spin interactions are only between the nearest neighbors, we can parallelize our codes very efficiently by separating spins into different sublattice groups, for which spins in each group can be modified or updated simultaneously. We were capable to simulate over 2 million spins for millions of MC sweeps or molecular dynamics time steps. We have studied the chiral spin texture of a 2D ferromagnetic (FM) Heisenberg model with the DM interaction and the easy axis spin anisotropy. As shown in Fig. 8.6 (a), when the magnetic field and the anisotropy are both absent, the stable spin structure is helical, where spins are aligned in strip configuration (the blue strips and red strips are separated by Bloch domain walls). The anisotropy breaks the stripe structure to form larger FM domains separated by Bloch wall, as shown in Figs. 8.6 (b) and (c). When the anisotropy becomes very strong, isolated Skyrmions emerge in a FM background, as shown in Fig. 8.6 (d). Our results explain why Skyrmions have been observed experimentally even without magnetic field in some materials with strong easy axis anisotropy.

8.4 Schedule and Future Plan

8.4.1 Large-scale quantum Monte Carlo (QMC) simulations for strongly correlated electrons

QMC is one of the most powerful methods for quantum simulations in condensed matter physics, especially when there is no notorious negative sign problem. We have developed a highly efficient determinant QMC code to perform the ground state calculation for the two-dimensional (2D) Hubbard model on the largest lattice sites ever using K computer. We plan to continue to develop the determinant based QMC method for strongly correlated electrons at zero temperature as well as at finite temperatures, to treat even larger lattice sites on exascale computers. One of the challenges for the large-scale simulations is to study strongly correlated electrons in three dimensions. We also plan to develop the first principles QMC for strongly correlated materials and molecules through international collaboration.

8.4.2 Density matrix renormalization group (DMRG) method for quantum dynamics in two dimensions

The DMRG method was originally proposed and is best performed for the ground state calculations of one-dimensional (1D) quantum systems. Owning to high performance computers that are available recently, it has become possible to apply the DMRG method not only to the ground state calculations in 2D quantum systems but also to the excited state calculations, thermodynamics, as well as real time dynamics. However, these dynamical calculations are limited mostly to 1D systems. This is partially because of the exponential increase of degrees of freedom to be kept for 2D systems but mainly because of the lack of reliable algorithms for thermodynamics and real time dynamics in two dimensions. By combining the kernel polynomial method, we plan to develop a new scheme based on the DMRG method for thermodynamics and real time dynamics in two dimensions. This will provide the first reliable quantum dynamics calculation for 2D strongly correlated quantum systems.

8.4.3 First-principles calculations for strongly correlated materials

Strongly correlated materials are promising as highly functional materials for future technological applications. However, the standard first-principles calculations based on the density functional theory (DFT) is not reliable because the strong many-body interactions require the treatment that goes beyond the single-particle approximation. We plan to develop new first-principles calculation schemes for strongly correlated materials. For strongly correlated solid state materials such as transition metal oxides, we apply the many-body cluster approximation scheme such as dynamical mean field theory (DMFT), variational cluster approximation (VCA), and cluster perturbation theory (CPT), combined with the first-principles calculations based on the DFT. These cluster approximation can treat the local electron correlation exactly within a cluster considered and therefore is apparently superior to the single-particle approximation. We also plan to apply directly the DMRG method as well as the tensor network states to develop an interdisciplinary research on strongly correlated molecules such as metalloprotein, in collaboration with Computational Molecular Science Research Team in AICS.

8.4.4 Tensor network scheme for strongly correlated quantum systems

Although the DMRG method is very powerful for 1D quantum systems, it meets the severe difficulty due to the exponential increase of degrees of freedom that have to be kept to guarantee the numerical accuracy when it is applied to quantum systems in higher spatial dimensions. This difficulty can be partially overcome by aggressively using high performance computers. However, the fundamental resolution has to rely on a radical revolution of the algorithm itself. Such a revolution has begun to happen when condensed matter physics meets quantum information. Based on the idea of information compression, a tensor network scheme has been recently proposed to represent a quantum many-body states with a product of tensors. There still exists a big challenge on how to treat the basic tensor operations highly efficiently in large scale calculations. We will address this issue to develop the renormalization group method based on tensor network states for strongly correlated quantum systems in higher dimensions.

8.5 Publications

Chapter 9

Computational Biophysics Research Team

9.1 Members

Yuji Sugita (Team Leader (Concurrent))*, **
Osamu Miyashita (Senior Research Scientist)
Jaewoon Jung (Research Scientist)
Chigusa Kobayashi (Research Scientist)
Yasuhiro Matsunaga (Research Scientist)
Hiromi Kano (Assistant)
Takaharu Mori (Research Scientist (Concurrent))**
Isseki Yu (Research Scientist (Concurrent))**
Raimondas Galvelis (Postdoctoral Researcher (Concurrent))**
Takao Yoda (Visiting Scientist) ***
Mitsunori Ikeguchi (Visiting Scientist)****
Naoyuki Miyashita (Visiting Scientist)
Michael Feig (Visiting Scientist)

* The main affiliation of these people is Laboratory for Biomolecular Function Simulation, Computational Biology Research Core, RIKEN Quantitative Biology Center.

** The main affiliation is RIKEN Theoretical Molecular Science Laboratory.

*** The main affiliation is Nagahama Bio Institute.

**** The main affiliation is Yokohama City University.

9.2 Research Activities

In this team, we have developed GENESIS(Generalized Ensemble Simulation System) for molecular dynamics simulations. The key features of GENESIS are that it is highly parallelized for K and other massively parallel supercomputers and that GENESIS contains a lot of enhanced conformational sampling methods and various molecular models for multi-scale and multi-resolution simulations. We have already open the code of GENESIS as free software under the license of GPLv2 and will update it every two year by adding new functions and optimizing the code into K or other computational platforms. These activities are necessary, in particular, for biological applications, since many interesting biological phenomena happen on the milliseconds or slower but current all-atom MD simulations cover only 1-10 microseconds on the general-purpose supercomputers or GPU clusters. We intend to spread GENESIS into academia as well as industries as a basic MD program that is useful for research and development.

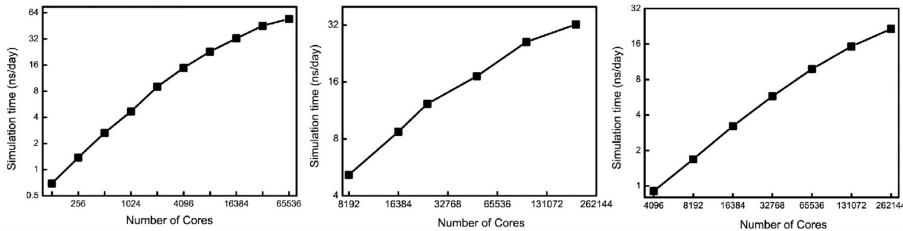


Figure 9.1: GENESIS performance of 1M (left), 8.5M (middle), and 28M (right) systems

9.3 Research Results and Achievements

9.3.1 Developement of GENESIS

We have already optimized GENESIS for large scale MD simulations on K computer. In the given fiscal year, we further optimized it by increasing parallel efficiency and enlarging the available number of processors. First, we make use of a multiple-program, multiple-data approach by separating computational resources responsible for real space and reciprocal space interactions. Second, we assign multiple time step integrator where time-consuming parts are skipped regularly based on the multiple-program and multiple-data approach. Our new implementation was tested on the K computer, and we could obtain very good performance results for big systems consisting of 1 million, 8.5 million, and 28 million atoms systems just increasing the parallel efficiencies. One MD cycle with the PME calculations for systems containing 1 million, 8.5 million, and 28 million atoms could be finished within 2.8 ms, 5.4 ms, and 8 ms (Figure 9.1).

9.3.2 Multi-resolution simulation methods for reactions couple with large conformational changes

Recently, experimental studies proposed that large conformational changes of proteins play important roles on biological functions. The conformational changes can originate as domain motions, where rigid structural units (domains) change their positions and/or orientations with respect to each other through flexible hinges or loops. It is difficult to investigate atomistic details of multi-domain proteins by experimental studies. In addition, it is still difficult to simulate using all-atom MD due to the slow time-scale. To overcome the difficulties, we have developed multi-resolution simulation method including the following three steps; 1. Analysis for “dynamic domains” and the magnitude of local domain motions in a protein through “Motion Tree”, a tree diagram that describes conformational changes in a hierarchical manner from two structures. (Koike et al., *J. Mol. Biol.*, 2014) 2. Development of a structure-based coarse-grained (CG) model enables a stable and efficient MD simulation from the information of domain motion obtained by “Motion Tree” [Kobayashi01]. The CG model provides a stable trajectory that is comparable to experimental studies and long-time all-atom MD simulations. 3. Performing sampling simulations with the CG model and investigate conformational changes in response to reactions in biological systems. We examine how many CVs are required to capture the correct transition-state structure during the open-to-close motion of adenylate kinase using a coarse-grained model in the mean forces string method to search the minimum free-energy pathway [Kobayashi02].

9.3.3 Systematic evaluation of collective variable choice for describing conformational changes of a protein

Collective variables (CVs) are often used in molecular dynamics simulations based on enhanced sampling algorithms to investigate large conformational changes of a protein. The choice of CVs in these simulations is essential because it affects simulation results, and impacts on the free-energy profile, the minimum free-energy pathway (MFEP), and the transition-state structure. Here, we examine how many CVs are required to capture the correct transition-state structure during the open-to-close motion of Adenylate Kinase using a coarse-grained model in the mean forces string method to search the MFEP. Various numbers of large amplitude principal components (PCs) are tested as CVs in the simulations. The incorporation of local coordinates into CVs, which is possible in higher

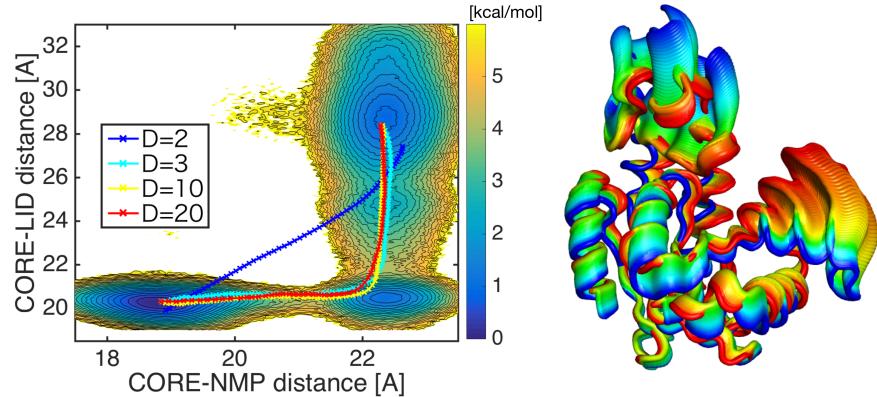


Figure 9.2: Free-energy landscape in the distances between domains' centers of mass. Lines indicate minimum free energy paths calculated in 2D (dark blue), 3D (light blue), 10D (yellow), and 20D (red) principal component spaces.

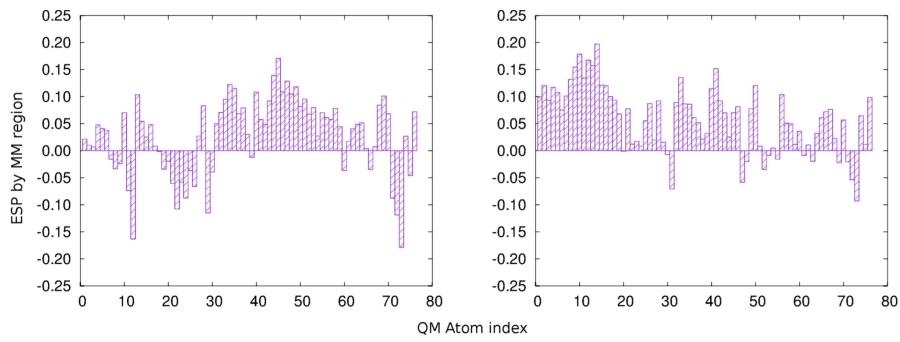


Figure 9.3: Electrostatic potential on QM atoms of the reactant state in the model crowding environment (left) and solution (right). QM atom index 43-55 corresponds to triphosphate moiety of GTP.

dimensional CV spaces, is important for capturing a reliable MFEP. The Bayesian measure proposed by Best and Hummer is sensitive to the choice of CVs, showing sharp peaks when the transition-state structure is captured. We thus evaluate the required number of CVs needed in enhanced sampling simulations for describing protein conformational changes (Figure 9.2 [Matsunaga02]).

9.3.4 Molecular crowding effect on GTP hydrolysis reaction in Ras-GAP complex

Macromolecular crowding effects have essential role in biomolecular system. Such effects have been extensively investigated experimentally, and also in classical Molecular Dynamics (MD) calculations. However, in the quantum chemistry level, those effects are not investigated due to the computational costs and methodological difficulties. In this study, we studied the molecular crowding effect on the GTP hydrolysis reaction in Ras-GAP complex by QM/MM RWFE method, which can take crowding effects into account with a reasonable computational cost. We modeled a crowding environment by adding 7 BSAs to the system as a crowder, and refined the reactant and transition states of the hydrolysis reaction by QM/MM RWFE method, where MD calculations were performed by GENESIS at K-computer. The structural difference around GTP were not significant between solution and crowding environment. However, there was a large difference in the electrostatic potential (ESP) imposed by the surroundings as shown in Figure 9.3. This large ESP change suggests that there must be significant differences in the free energy barrier between crowding and solution environments.

9.4 Schedule and Future Plan

So far, GENESIS has been optimized mainly on K computer. In this year or later, we consider other platforms than K, such as intel CPU cluster, nvidia GPU processor, and post K. Since these CPU (or GPU) architectures are quite different with each other, a single MD kernel does not work well for all the different computational platforms. So, GENESIS will have multiple kernels that are optimized to one of the computational platforms. The disadvantage of this approach is that we have more effort on programming, reducing potential bugs for each kernel, and so on. It should be hard task for our team, but there is no other good ways to improve the performance of GENESIS in multiple platforms.

We would like to simulate more and more large biological systems for investigating slow biological dynamics. For this purpose, we need to develop multi-scale and multi-resolution programs that are scalable on K or post-K computers. Currently, GENESIS/SPDYN is useful for all-atom MD simulations on these supercomputers, but does not show good performance on CG-modeling and simulations of biological systems due to the small number of particles and load-balance problems. We need a new program that is suitable for such CG-modeling and simulations by introducing a different parallelization scheme. Such new program, which we call CGDYN, will be developed soon.

Another important aspect is the introduction of quantum effect to investigate the chemical reactions in enzymes. Bond-formation or breaking can not be simulated by using classical force fields, but should be investigated by using ab initio Quantum theory. Considering the large system size in biological systems, only possible approach is to use QM/MM hybrid calculations. We have a basic QM/MM code for computing potential energies of QM/MM systems and optimizing the systems based on the hybrid QM/MM potential energy functions. We plan to extend the calculations for larger periodic boundary systems and to allow the reaction calculations in proteins.

9.5 Publications

Chapter 10

Particle Simulator Research Team

10.1 Members

Junichiro Makino (Team Leader)
Keigo Nitadori (Research Scientist)
Yutaka Maruyama (Research Scientist)
Masaki Iwasawa (Postdoctoral Researcher)
Ataru Tanikawa (Postdoctoral Researcher)
Takayuki Muranushi (Postdoctoral Researcher)
Natsuki Hosono (Postdoctoral Researcher)
Daisuke Namekata (Postdoctoral Researcher)
Miyuki Tsubouchi (Technical Staff)

10.2 Research Activities

We are developing particle-based simulation software that can be used to solve problems of vastly different scales.

Simulation schemes for hydrodynamics and structural analysis can be divided into grid-based and particle-based methods. In grid-based methods, the computational region is mapped to regular or irregular grids. Continuous distributions of physical values are represented by discrete values at grid points, and the governing partial differential equation is approximated to a set of finite difference equations.

In the case of the particle-based methods, physical values are assigned to particles, while the partial differential equation is approximated by the interactions between particles.

Both methods are widely used, and they have their advantages and disadvantages. The computational cost of grid-based schemes is generally lower than that of particle-based methods with similar number of freedoms. Thus, if a near-uniform grid structure is appropriate for the problem to be solved, grid-based methods perform better.

The advantage of the particle-based methods comes from the fact that they use "Lagrangian" schemes, in which the particles move following the motion of the fluid in the case of the CFD calculation. In the case of grid-based methods, we generally use "Eulerian" schemes, in which the grid points do not move.

There are three points in which the Lagrangian schemes are better than Eulerian schemes. One is that the Lagrangian schemes are, to some extent, adaptive to the requirement of the accuracy, since when a low-density region is compressed to become high density, Second one is that the timestep criteria are quite different. In the case of the Lagrangian schemes, the timestep is determined basically by local sound velocity, while in the Eulerian scheme by global velocity. Thus, if a relatively cold fluid is moving very fast, the timestep for the Eulerian schemes can be many orders of magnitude

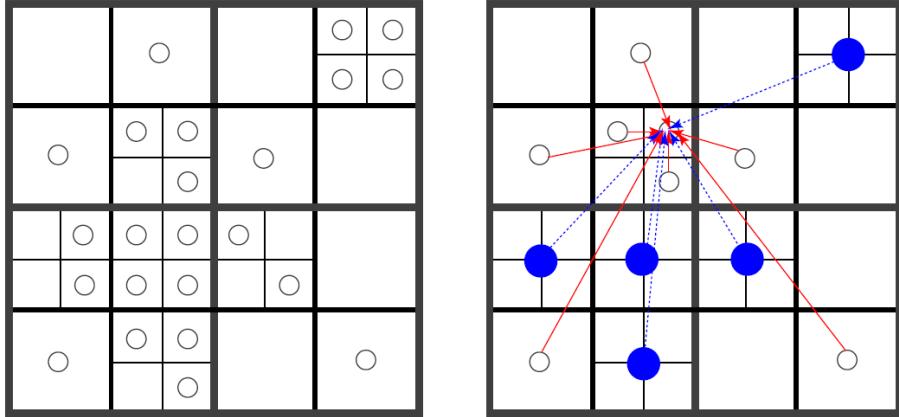


Figure 10.1: Basic idea of tree algorithm

shorter than that for Lagrangian schemes. Finally, in the case of fast-moving low-temperature fluid, the required accuracy would be very high for Eulerian scheme, since the error comes from the high velocity, while that error would be transferred to internal energy of the fluid element which is much smaller than that of the kinetic motion.

Of course, there are disadvantages of Lagrangian schemes. The primary one is the difficulty of construction of such schemes in two or higher dimensions. In the case of one-dimensional calculation, it is easy to move grid points following the motion of the fluid, but in two or higher dimensions, the grid structure would severely deform if we let the grid points follow the flow. Thus, we have to reconstruct the grid structure every so often. This requirement causes the program to become complex. Moreover, reconstruction of the grid structure (so called remeshing) means we lose numerical accuracy.

Particle-based methods "solve" this difficulty by not requiring any mesh. In particle-based methods, particles interact with its neighboring particles, not through some connection through grid, but through distance-dependent kernel functions. Thus, there is no need of remeshing. As a result, particle-based schemes are simple to implement, and can give reasonable results even when the deformation is very large. Another important advantage is that it is relatively easy to achieve high efficiency with large-scale particle-based simulation.

In the case of grid-based schemes, in order achieve some adaptivity to the solution, we have to use either irregular grid or regular grid with adaptive mesh refinement. In both cases, adaptivity breaks the regularity of the mesh structure, resulting in non-contiguous access to the main memory. In the case of the particle-based schemes, it does require some irregular memory access, but it is relatively straightforward to make good use of spacial locality, and thereby achieving high efficiency. Similarly, very high parallel performance can be achieved.

However, it has its own problems. In the case of the SPH method, it has been known that the standard scheme cannot handle the contact discontinuity well. It also require rather strong artificial viscosity, which results in very low effective Reynolds number.

Thus, in many fields of computational sciences, many groups are working on implementation of high-performance particle-based simulation codes for their specific problem.

One serious problem here is that, high-performance, highly-parallel simulation codes for particle-based simulations are becoming more and more complex, in order to make full use of modern supercomputers. We need to distribute particles to many computing nodes in an appropriate way, so that the communication between nodes is minimized and at the same time near-optimal load balance is achieved. Within each nodes, we need to write an efficient code to find neighbor particles, rearrange data structure so that we can make good use of the locality, make good use of multiple cores and SIMD units within each core.

Even for the case of very simple particle-particle interaction such as the Lenard-Jones potential or Coulomb potential, the calculation code tends to be very large, and since the large fraction of the code is written to achieve a high efficiency on a specific architecture, it becomes very hard to port a code which is highly optimized to one architecture to another architecture.

Our goal is to develop a "universal" software that can be applied to a variety of problems whose

scales are vastly different. In designing such universal software, it is important to ensure that it runs efficiently on highly parallel computers such as the K computer. Achieving a good load balance with particle-based simulation is a difficult task, since using a regular spatial decomposition method causes severe load imbalance, though this works well for grid-based software. Consequently, we have developed an adaptive decomposition method that is designed to work in a way that the calculation time on each node is almost the same, resulting in near-optimal load balance.

The strategy to develop such a universal software is as follows.

We first construct an highly parallel and very efficient implementation of the TreePM algorithm for gravitational N-body problem. This is actually not a completely new implementation, but the GreeM code developed by researchers of the Strategic Program for Innovative Research (SPIRE) Field 5 “The origin of matter and the universe. In collaboration with the Field 5 researchers, we improve the efficiency of the code and study the issues of the data structure, domain decomposition, load balance strategy etc.

In the second stage, we will develop a prototype of the parallel particle simulation platform. We will design the platform so that it can be used for multiple physical systems. In practice, we consider the following three applications as the initial targets.

1. Gravitational N-body simulation 2. Smoothed Particle Hydrodynamics 3. Molecular Dynamics

In the meantime, we will also investigate the way to improve the performance and accuracy of the current particle-based algorithms for hydrodynamics.

10.3 Research Results and Achievements

As we stated in section 2, we are working on the three major subtopics, in order to develop the universal platform for particle simulations.

In the following, we briefly describe the status of our research in each subtopic.

10.3.1 High-performance gravitational N-body solver.

We use the TreePM algorithm as the basic method for the evaluation of gravitational interaction between particles. TreePM is a combination of the tree method and the P³M (particle-particle particle-mesh) scheme. Figure 10.1 shows the basic idea of the tree algorithm. The space is divided into a hierarchical octree structure (quadtree in the figure). Division is stopped when a cell contains one or no particle. When we calculate the force on a particle, we evaluate the force from a group of particles, with size larger for more distant particles. In this way, we can reduce the calculation cost from O(N²) to O(N log N).

The tree algorithm is widely used, but when the periodic boundary condition is applied, we can actually use a more efficient efficient scheme, since we can calculate the long-range, periodic term using FFT. The P³M scheme has been used for such problem, but it has the serious problem that when the density contrast becomes high, the calculation cost increases very quickly. The TreePM scheme solves this difficulty by using the tree algorithm to evaluate the forces from nearby particles. Even when there are very large number of neighbor particles, the calculation cost does not increase much, since the calculation cost of the neighbor force is proportional to the logarithm of the number of neighbors.

In order to map the problem to the distributed-memory parallel computer such as the K computer, we adopted the approach to divide the space into domains and assign particles in one domain to one calculation node. We used the orthogonal recursive multisection method developed by the team leader some years ago. It is the generalization of the orthogonal recursive bisection (ORB), which has been widely used in many parallel implementations of the tree algorithm.

With ORB, we recursively divide space into two halves, each with the same number of particles. An obvious disadvantage of the ORB approach is that it can utilize the computing nodes of integral powers of two. Thus, in the worst case we can use only half of the available nodes.

The difference between the multisection method and the ORB is that with the multisection method we allow the divisions to arbitrary number of domains, instead of bisection. This would allow too many possible divisions. In our current implementation, we limit the number of levels to three, and make the numbers of divisions at all levels as close as possible. Thus, our domain decomposition is topologically a simple three-dimension grid. This fact makes the multisection method well suited to the machines with the 3D torus network like the K computer.

We have developed a "reference code" for gravitational N-body simulation on the K computer. This code is fairly well optimized for the K computer, and shows quite good scalability for even for relatively small-size problems. The asymptotic speed per timestep for large number of nodes is around 7ms. This speed is comparable to that of highly optimized molecular dynamics codes on K, even though our code is designed to handle highly inhomogenous systems.

We used this code as the reference implementation for more generalized particle simulation platform which will be described in the next subsection.

10.3.2 Particle Simulation Platform.

In FY 2014, We have completed and released Version 1.0 of the particle simulation platform, which we call FDPS (Framework for Developing Particle Simulator). In FY 2015, we have applied a number of improvements to FDPS.

The basic idea of FDPS is that the application developer (or the user) specified the way the particles interact with each other, and the rest is taken care by FDPS. Here, "the rest" includes domain decomposition and re-distribution of particles, evaluation of interactions between particles, including those in different domains (different MPI processes, for example).

In practice, there are many additional details the user should give. Consider a relatively simple case of particles interacting with softened $1/r$ potential. There are a number of small but important points one has to decide on. For example, what algorithm should be used for the interaction calculation? Even if we limit the possibilities to reasonably adaptive schemes for open boundary problems, we have the choice between Barnes-Hut tree and FMM. For both algorithms, there are many different ways to parallelize them on distributed-memory parallel computers. Also, there are infinitely many variations for the time integration schemes.

The base layer of FDPS offers the domain decomposition based on the recursive multisection algorithm, with arbitrary weighting function for the load balancing. It also offers the parallel implementation of interaction calculation between particles.

The domain decomposition part takes the array of particles on each node as the main argument. It then generates an appropriate domain for each node, redistribute particles according to their locations, and returns.

The interaction calculation part takes the array of particles, the domain decomposition structure, and the specification of the interaction between particles as main arguments. The actual implementation of this part need to take into account a number of details. For example, the interaction can be of long-range nature, such as gravity, Coulomb force, and interaction between computational elements in the boundary element method (BEM). In this case, the user should also provide the way to construct approximations such as the multipole expansion and the way to estimate error. The interaction might be of short-range nature, with either particle-dependent or independent cut-off length. In these cases, the interaction calculation part should be reasonably efficient in finding neighbor particles.

We have successfully implemented all of these functionalities in FDPS version 1.0. (<https://github.com/FDPS/FDPS>). Using FDPS, a gravitational N-body simulation code can be written in 120 lines, and that code is actually fully scalable even to full-node runs on K computer. For SPH calculations, we have also achieved similar scaling.

FDPS is implemented as a class template library in C++ language. It receives the class definition of particles and a function (or multiple functions in the case of complex interactions) to evaluate the interaction between particles. When a user program is compiled with the FDPS library, the class template is instantiated with the user-specified definition of the particle class. Thus, even though the FDPS library functions are generic ones not specialized to a particular definition of particles, it behaves as if it is a specialized one.

The measured performance of applications developed using FDPS is quite good. Both for gravity-only calculation and SPH calculation, weak-scaling performance is practically perfect, up to the full-node configuration of K computer. Moreover, the measured efficiency, in terms of the fraction of the peak floating-point performance, is also very high. It is around 50% for gravity-only calculation. For SPH calculations, at the time of writing the performance is around 10%.

In FY 2015, we have extended FDPS in several important directions. The first one is the improvement of the strong scaling. The algorithm used for the domain decomposition contains one serial bottleneck. The "sampling" algorithm used in FDPS 1.0 works well only when the average number of particles per MPI process is significantly larger than the total number of MPI processes.

We developed a new parallel algorithm, in which $O(p^{1/3})$ MPI processes are used to decompose the computational domain. Here p is the total number of MPI processes. Thus now the requirement for the number of particles is relaxed from larger than p to larger than $p^{2/3}$. Now we can achieve pretty good performance for around 1 billion particles, on the full nodes of K computer. Previously we need near 100 billion particles to achieve good efficiency.

The second one is the addition of new interface method to interaction calculation function, which allows efficient use of accelerator hardware such as GPGPU or Intel MIC. In order to achieve high performance on accelerators, it is important to pass a large chunk of work at one time. In order to achieve this goal, in the current version of FDPS the CPU creates the list of multiple interaction lists, and send all of them at once so that the overhead of the initialization of the accelerator would not become a bottleneck. This interface has been tested on NVIDIA GPGPUs as well as the PEZY-SC processor.

10.3.3 Improvements on SPH.

SPH (Smoothed Particle Hydrodynamics) has been used in many fields, including astrophysics, mechanical engineering and civil engineering. Recently, however, it was pointed out that the standard formulation of SPH has numerical difficulty at the contact discontinuity. The reason is that the formulation of the standard SPH requires that the density is differentiable, which is by definition not the case at the contact discontinuity.

We have been working on the possible solution on this problem. One approach is to reformulate SPH so that it does not use the density in the right-hand side of the equation of motion. We one way to achieve the density independence. We constructed an SPH scheme which uses artificial density-like quantity as the base of the volume estimator. It evolves through usual continuity equation, but with additional diffusion term. Thus, we can guarantee the continuity and differentiability of this quantity, except at the initial condition or at the moment when two fluid elements contact with each other. This scheme seems to work extremely well, and we are currently working on the way to extend this scheme so that it can handle free surface accurately.

We are also working on a completely different approach, in which we replace the SPH formulation to evaluate the gradient to other schemes. SPH has a known problem that its kernel estimate contains $O(1)$ error, since the summation of contributions from neighbor particles is not guaranteed to be unity. The reason why SPH uses this mathematically inconsistent formulation is to achieve symmetry and conservation. In SPH discretization, interaction between two particles is symmetric, which guarantees the conservation of linear and angular momenta. However, the use of SPH approximation resulted in rather low accuracy, which limits the reliability of the results obtained using SPH. We are experimenting with several different schemes which can achieve much higher accuracy, while losing some of the nice features of SPH such as the symmetry of interaction.

10.4 Schedule and Future Plan

We plan to improve the performance of FDPS further in FY 2015. In particular, we plan to extend the API so that the users of FDPS can easily use heterogeneous machines such as machines with GPGPUs or Intel MIC.

10.5 Publications

Chapter 11

Computational Climate Science Research Team

11.1 Members

Hirofumi Tomita (Team Leader)
Yoshiyuki Kajikawa (Senior Scientist)
Shin-ichi Iga (Research Scientist)
Seiya Nishizawa (Research Scientist)
Hisashi Yashiro (Research Scientist)
Sachiho Adachi (Research Scientist)
Yoshiaki Miyamoto (Special Postdoctoral Researcher)
Yousuke Sato (Postdoctoral Researcher)
Tsuyoshi Yamaura (Postdoctoral Researcher)
Ryuji Yoshida (Postdoctoral Researcher)
Hiroaki Miura (Visiting Researcher)
Mizuo Kajino (Visiting Researcher)
Hiroshi Taniguchi (Visiting Researcher)
Tomoko Ohtani (Research Assistant)
Keiko Muraki (Research Assistant)

11.2 Research Activities

Our research team conducts the pioneering research work to lead the future climate simulation. In order to enhance the reliability of climate model more, we have aimed to construct a new climate model based on the further theoretically physical principles. Conducting such a new model needs tremendously large computer resources. Therefore, it is necessary to design the model to pull out the capability of computers as much as possible. Recent development of supercomputers has a remarkable progress. Hence another numerical technique should be needed under the collaboration of hardware research and software engineering for the effective use on the future HPC, including the K computer and Post K computer.

For the above research purpose and background, our team is cooperating with the computational scientists in other fields and computer scientists. We enhance the research and development for the future climate simulations including effective techniques; we build a next-generation climate model.

The establishment of the above basic and infrastructure research on the K Computer is strongly required, because this research leads to the post K computer or subsequent ones in the future.

We highlight the following studies in this fiscal year.

1. Construction of a new library for climate study:

We have proposed the subject “Estimation of different results by many numerical techniques and their combination” as a synergetic research to MEXT in 2011 through the discussion with the Strategic 5 fields (SPIRE). We develop a new library for numerical simulation. The progress in development of SCALE is reported. NICAM-DC was imported to SCALE as a global dynamical core in this fiscal year. The two landmark papers of the SCALE are reported.

2. Grand challenge run for sub-km horizontal resolution run by global cloud-resolving model:

Another outstanding simulation of global model NICAM on the K computer, with super-high resolution (870m), has been done. We analyze the simulation in cooperation with the SPIRE3. We report the further comprehensive analysis of convection properties in the simulation.

3. Disaster prevention research in establishment of COE project:

Hyogo-Kobe COE establishment project has accepted 5 subjects in 2012. One of subjects is “the computational research of disaster prevention in the Kansai area”. In this subject, one of sub-subjects is “Examination of heavy-rainfall event and construction of hazard map”, which our team is responsible for. The tuning of physical properties focusing on the climatological precipitations and the preliminary result by direct downscaling are reported.

11.3 Research Results and Achievements

11.3.1 Construction of a new library for climate study

We are working on research and development of a library (named SCALE) for numerical models in fluid dynamical field especially in meteorological field. We examined feasibility of numerical scheme and methods for developing new ones which are suite on massive parallel computers especially the K computer. In order to validate the schemes and test their performance in atmospheric simulations, we have been developing an atmospheric regional model (named SCALE-RM) as a part of the SCALE library. The SCALE library and the SCALE-RM are currently available as open source software at our web site (<http://scale.aics.riken.jp/>). It is also installed on the K computer and is available for the K computer users as an AICS Software (<http://www.aics.riken.jp/en/kcomputer/aics-software.html>). In this year, we continued to develop components which are necessary for real atmospheric simulations; a boundary turbulence scheme, an urban canopy model, nesting system, and preprocessing tools. We also have improved the library and the model for better performance in both physical and computational aspects. As a remarkable feature, NICAM-DC (Nonhydrostatic ICosahedral Atmosphere Model Dynamical Core) was equipped to SCALE library as a global dynamical core.

The validation for the physical performance is an important issue as well as code development. In this fiscal year, we tuned the microphysical scheme, focusing on the one-moment bulk method (Tomita et al.2008). Although this scheme was used also in NICAM through the project SPIRE, it depends on the resolution and phenomena we can see. For this purpose, we conducted the series of systematic parameter tuning suitable to Japanese western region using the GSM data as the boundary condition and compared the precipitation with AMeDAS data. Figure 11.1 (a) and (b) shows results of hourly precipitation histogram from two typical parameter sets of the microphysics, which has been used in NICAM experiments (Miyakawa et al. 2014, Miyamoto et al.2013). After several key parameters were swept, we successfully tuned the parameters as shown in Fig.11.1 (c).

In this fiscal year, two landmark paper for SCALE was published. The first paper describes the proof-of-concept like study according to SCALE policy (Sato et al. 2015[Sato et al'2015]). The three microphysical schemes, the one-moment bulk, two-moment bulk, and spectral bin schemes were compared by sensitivity experiments in which the other components were fixed in SCALE-RM. Since SCALE is targeting to enable self-model inter-comparison easily, this paper is high significant as a SCALE reference paper. The other paper is about the model description of SCALE-RM dynamical core(Nishizawa et al. 2015[Nishizawa et al'2015]). In this paper, we reveals that the influence of the grid aspect ratio of horizontal to vertical grid spacing on turbulence in the planetary boundary layer (PBL) in a large-eddy simulation (LES). This paper gives a deep suggestion to meteorological LES. One key point is how the filter length be configured. It should be based on consideration of the

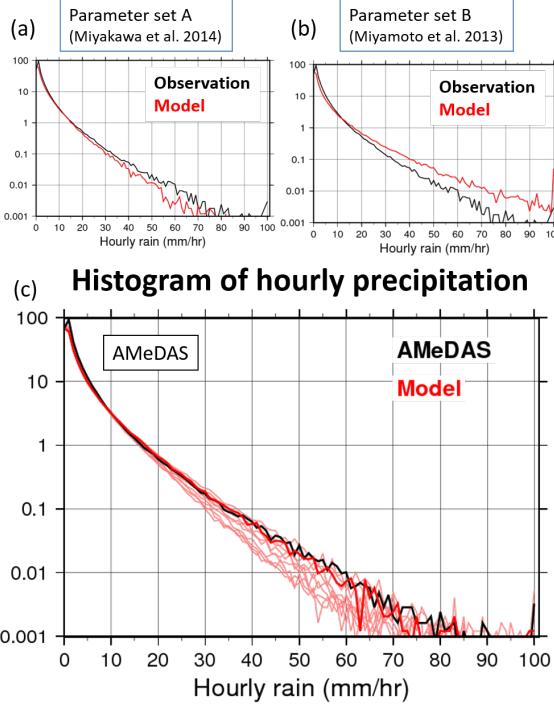


Figure 11.1: Results of microphysics tuning.

numerical scheme. We also confirmed necessity of a corrective factor for the grid aspect ratio into the mixing length. As shown in Fig.11.2, these remedy generates the theoretical slope of the energy spectrum; otherwise, spurious energy piling appears at high wave numbers.

We investigated also the computational performance of SCALE-RM from the viewpoint of strong scale. Figure 11.3 shows the results of the strong scaling experiments for SCALE-LES. The most time-consuming part is the dynamics, and its scaling factor tends to be saturated by decreasing the problem size. This degradation comes from the increasing ratio of the communication time against the computational time. On the other hand, the scaling of physics gives relatively ideal scaling. In addition, the I/O part is not a bottleneck. To obtain the faster calculation, we implemented several choices both for the temporal and spatial difference schemes. As a result, the longer time step can be obtained in a certain configuration that is 4th order Runge-Kutta scheme in time and 3rd order advection scheme in space.

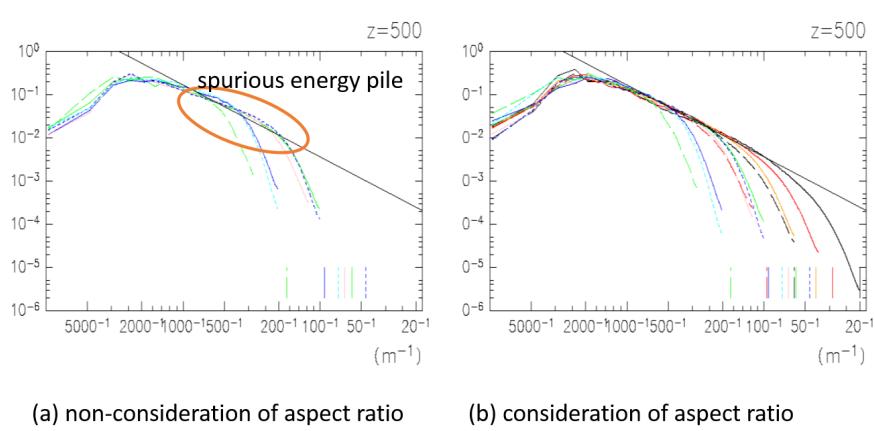


Figure 11.2: The kinetic energy spectrum.

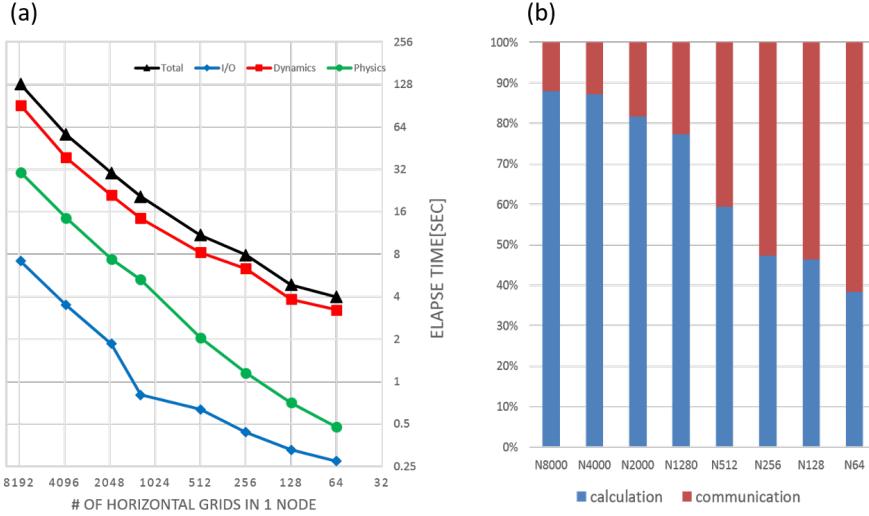


Figure 11.3: (a) Strong scale of SCALE-RM. (b) Ratio of communication to calculation.

11.3.2 Grand challenge run for sub-km horizontal resolution run by global cloud-resolving model

Using the K computer, we have succeeded in conducting the global simulation with the world's highest resolution, 870 m, which is published in the 2013 fiscal year (Miyamoto et al. 2013). In the fiscal year 2015, an additional analyses to reveal the differences in convection properties in various atmospheric disturbances has been done. We focused on the differences in convection under four representative cloudy disturbances: Madden-Julian Oscillation, Tropical Cyclones, Mid-latitudinal Lows, and Fronts (Miyamoto et al. 2015[Miyamoto et al'2015x]). In this fiscal year, we summarized the knowledge that we have obtained so far, as a review paper (Kajikawa et al. 2016[Kajikawa et al'2016]). We conducted further comprehensive analysis of the global-mean state and the characteristics of deep convection, to clarify the difference of the essential change by location and environment. By this paper, this project in our team collaborating with SPIRE was closed once. The subsequent collaboration project leads to the post-K priority project 4.

11.3.3 Hyogo-Kobe COE establish project

In this fiscal year, the boundary conditions inputted to our regional model SCALE-RE was selected. We decided to use the data from the global warming experiments by MRI-AGCM. Figure 11.4 (a) shows the target region. Figure 11.4 (b) shows the histogram of hourly precipitation of downscaling result, compared with AMeDAS in the present climate. The simulation period is from June to September in 10 years. Owing to the successful model tuning described in the previous section, the observation and model result indicate little difference. Figure 11.4 (c) gives comparison between the present and future climates, regarding to the precipitation intensity. The heavy rainfall event increases in the whole Japan area, while the frequency of heavy rainfall is not so changed in the Japanese western area. Figure 11.4 (d) shows an expected maximum rainfall intensity that stochastically occurs once twenty years. This result indicates that extreme rainfall occurs in the Kyushu area, but little change occurs in the Kansai area. However, we should note that this result was obtained from just one scenario and one GCM model output; we have to regard this result as one of possible ensembles. For the more reliable result, we should increase the number of scenario and the integration period.

11.4 Schedule and Future Plan

In the next year, we will continue to further develop, update and maintain the numerical library for the K computer (SCALE library). We also try to enhance the performance of each existing scheme. Especially, validation of cumulus parameterization is necessary. At the same time, we will work on

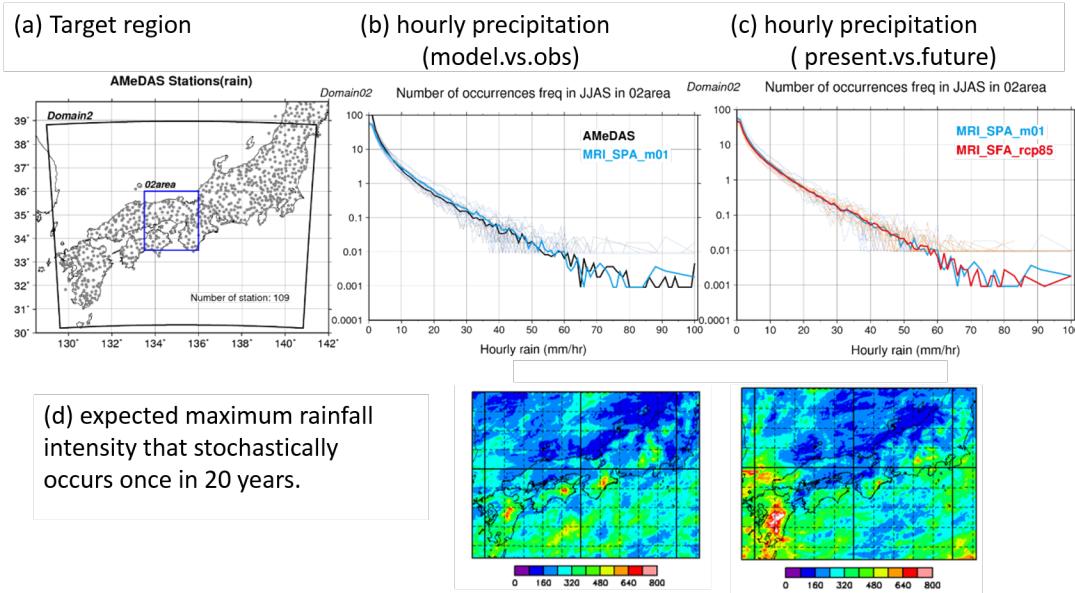


Figure 11.4: The result of the precipitation property in the Japan/Kansai region by the direct downscaling.

the following three projects in the collaboration with other team in AICS and the scientist in other institute.

1. On the Hyogo/Kobe COE establish project, we will continue the long-term climate simulation by using SCALE-RM to examine the heavy rainfall over Kobe city area. Several MRI-AGCM results for the future climate is used for increase of scenarios. We will obtain the geographical distribution of the frequency of heavy rainfall and evaluate it more precisely. For this purpose, the pseudo global warming method will be also employed.
2. We will also contribute to the CREST, Strategic Basic Research Programs “Innovating ”Big Data Assimilation” technology for revolutionizing very-short-range severe weather prediction” to develop the main climatological model in SCALE library. In collaboration with the Data Assimilation team in AICS, we developed a prototype of SCALE-LETKF (Local Ensemble Transform Kalman Filter). For this short range forecast, we will pursue both of the computational and physical performance.
3. We join the POST K Science priority project under the collaboration with JAMSTEC. NICAM-LETKF is one of the target applications. Our team will continue to develop and update that application.

11.5 Publications

Chapter 12

Complex Phenomena Unified Simulation Research Team

12.1 Members

Makoto Tsubokura (Team Leader)
Keiji Onishi (Postdoctoral Researcher)
Chung-gang Li (Postdoctoral Researcher)
Leif Niclas Jansson (Postdoctoral Researcher)
Rahul Bale (Postdoctoral Researcher)
Tetsuro Tamura (Visiting Researcher)
Ryoichi Kurose (Visiting Researcher)
Gakuji Nagai (Visiting Researcher)
Kei Akasaka (Visiting Researcher)

12.2 Research Activities

The objective of our research team is to propose a unified simulation method of solving multiple partial differential equations by developing common fundamental techniques such as the effective algorithms of multi-scale phenomena or the simulation modeling for effective utilization of the massively parallel computer architecture. The target of the unified simulation is supposed to be complex and combined phenomena observed in manufacturing processes in industrial circles and our final goal is to contribute to enhance Japanese technological capabilities and industrial process innovation through the high-performance computing simulation.

Most of the complex flow phenomena observed in manufacturing processes are relating to or coupled with other physical or chemical phenomenon such as turbulence diffusion, structure deformation, heat transfer, electromagnetic field or chemical reaction. While computer simulations are rapidly spreading in industry as useful engineering tools, their limitations to such coupled phenomena have come to realize recently. This is because of the fact that each simulation method has been optimized to a specific phenomenon and once two or more solvers of different phenomena are coupled for such a complicated target, its computational performance is seriously degraded. This is especially true when we utilize a high-performance computer such as K-computer. In such a situation, in addition to the fundamental difficulty of treating different time or spatial scales, interpolation of physical quantities like pressure or velocity at the interface of two different phenomena requires additional computer costs and communications among processor cores. Different mesh topology and hence data structures among each simulation and treatment of different time or spatial scales also deteriorate single processor performance. We understand that one of the keys to solve these problems is to adopt unified structured mesh and data structure among multiple simulations for coupled phenomena. As

a candidate of unified data structure for complicated and coupled phenomena, we focused on the building-cube method (BCM) proposed by Nakahashi[1].

[1]K. Nakahashi, High-Density Mesh Flow Computations with Pre-/Post-Data Compressions, Proc. AIAA 17th CFD Conference (2005) AIAA 2005-4876

12.3 Research Results and Achievements

12.3.1 Development of a unified framework for large-scale multiphysics problems

Based on the Building Cube Method (BCM), we have developed a unified solver framework CUBE (Complex Unified Building cube) for solving large-scale multiphysics problems. The framework has a modular design where CUBE provides a core library containing kernel functionalities e.g. a mesh, flow fields and I/O routines. Solvers are then developed on top of the kernel by connecting necessary kernel modules together, forming a solver pipeline, describing the necessary steps to solve a particular problem.

Load balancing is an essential component in today's large scale multiphysics simulations, and with an ever increasing amount of parallelism in modern computer architecture it is essential to reduce even the slightest workload imbalance. An imbalance could severely impact an application's scalability. Traditionally, load balancing is seen as a static problem, closely related to the fundamental problem of parallel computing, namely data decomposition. For a CFD simulation based on BCM, since each cubes contains the same amount of cells the goal is to evenly distribute the cubes among the available cores. However, such a decomposition assumes that the workload for each cube is uniform. For most cubes this is true, but for cubes which contain immersed bodies, combustion, chemical reactions, etc. the workload is slightly higher, which implies a workload imbalance. Therefore, to retain good scalability a load balancing method that balances the workload not only considering the BCM mesh, but also the additional workload from the immersed body, chemical reactions, etc., was developed.

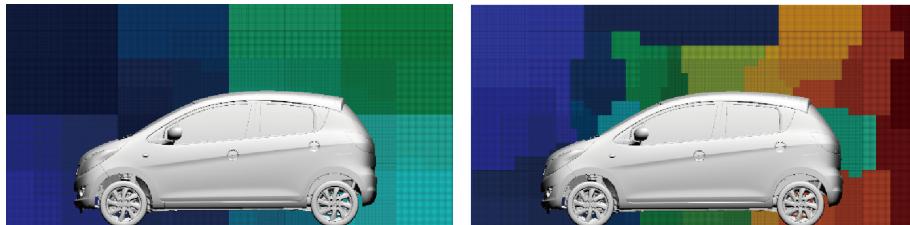


Figure 12.1: An example of load balancing with respect to the cost of evaluating the immersed geometry and the cost of computing the fluid cells, colored by MPI rank.

To evaluate the performance of the load balancer, we used CUBE to solve two different incompressible flow problems (full vehicle and a landing gear model) on the K computer. And, the total execution time for performing a fixed number of time steps for both an unbalanced (no load balancing) and a balanced case (using load balancing) on various numbers of cores are compared (Figure 12.2).

12.3.2 Development of a very large scale incompressible flow solver with a hierarchical grid system

The CUBE, name of our software framework, has been developed by conjunction with incompressible flow code which developed to realize the analysis in a real development process of industrial field on the massively parallel environment, including pre- and post-processing.

Industrial collaboration, MAZDA: We have utilized it for the complexed geometrical analysis using dirty CAD data received from automotive company. In last year, we have conducted basic aerodynamics validation using the vehicle geometry of MAZDA Motor Corporation. The conclusion was we need to improve an accuracy of drag force prediction. In this year, we have improved it conducting method survey of interpolation technique onto surface/volume, and the fundamental

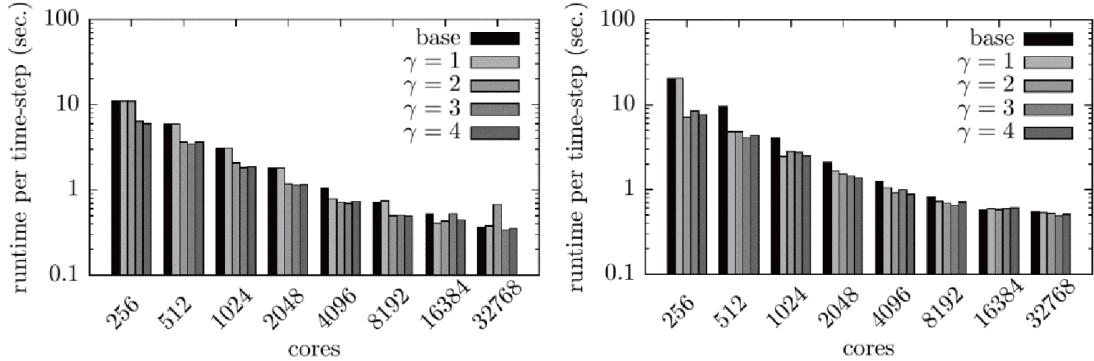


Figure 12.2: Comparison of runtime per time-step for balanced and unbalanced cases for the landing gear model (left) and the vehicle model (right)

investigation of approximate domain method on immersed boundary (IB). The difficulty was came from the uncertainty of front/back face of complicated geometry because the interpolation caused huge error if the search of face orientation has been missed. It was highly depends on the complexity of geometry and grid resolution. The method which is drawn by the analogy of Poisson solver providing a front/back face information based on flow solution has been developed. Then we could successfully get a reasonable absolute drag value (about 2% error), and drag delta between 2 different aerodynamic configurations. At the same time, we could successfully reproduce the characteristic total pressure flow field comparing to wind-tunnel measurement data.

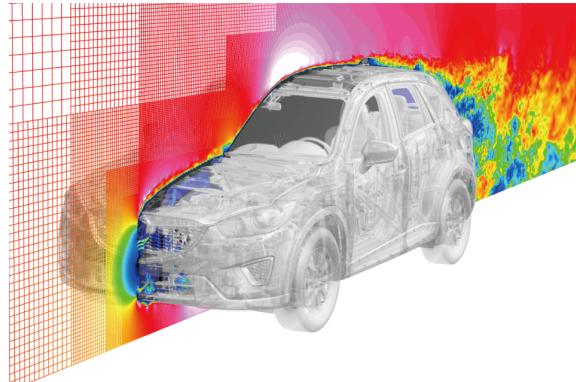


Figure 12.3: Overview of computational grid and flow field (MAZDA Motor Company).

Table 12.1: predicted drag on 2 configurations normalized by experimental base case.

C_d	Exp.	Sim.
Base	1.000	1.021
Aero	1.011	1.043
ΔC_d	1.15%	1.46%

Industrial collaboration, SUZUKI: The usability on the practical use of industrial application has been evaluated by discussing with professional engineers of automobile company. we have corroborated with SUZUKI MOTOR CORPORATION to do this work. We have provided CUBE to them with simple documents and training, and they have evaluated by their own way, and gave us their feedback report. The geometry preparation time has been accelerated from 35 hours using commercial software to 2 hours using CUBE framework, based on dirty CAD data. The estimated drag coefficient between 2 configuration had better agreement with measurement than commercial code. It shows CUBE is already ready for usage in the actual design process in term of the usability, human-cost, accuracy, and turn-around time. But, they had a strong request to improve calculation time because their process has a limitation to finish each job within 1 night whereas our method

requires 2 or 3 days. To say more, their method is based on Reynold averaged approach based on strong turbulence model which is known to have relatively large error, our method is based on the pure transient approach which generally requires 10 times larger calculation resource comparing to RANS. After the discussion, we have agreed to improve it in near future because it is very important to know the true needs on the field of industrial engineering. As the first step of that, we have developed a new function to enable us to use local refinement grids in the grid generation software. It can reduce the calculation load about from 1/5 to 1/10 for the vehicle case, to accelerate the solution. And the implementation and testing of the local refinement functionality for IB has been started in this year. We thinks it also can lead to the enhancement of effective multi-grid method (MG), or adaptive mesh refinement method (AMR) in the future development scope.

Both of MAZDA and SUZUKI could decide to promote the results inside their organization, and continue the current research activity using our software in FY2016 by submitting the application and accepted the industrial use project on K-computer.

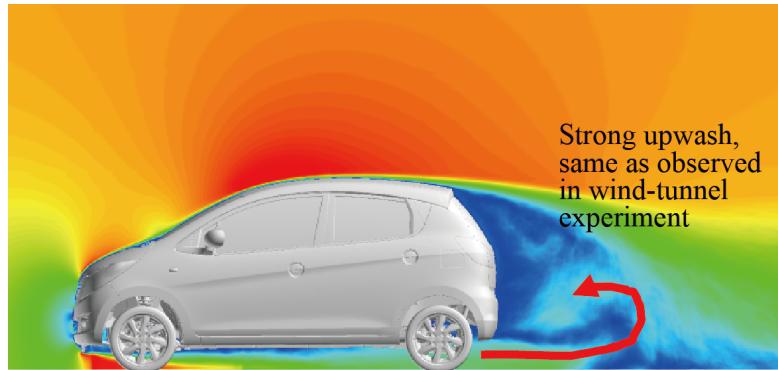


Figure 12.4: Overview of characteristic flow field (SUZUKI MOTOR CORPORATION).

Table 12.2: predicted drag on 2 configurations normalized by experimental base case.

Condition	CUBE	Commercial Code A
Finest grid size	2.0 [mm]	2.0 [mm] (layer 0.04[mm])
Num. of cells	400 million	53 million
Fluid Configuration	LES (Standard Smagorinsky)	DES (SST k- ω)
Num. of timesteps	145,000	4,000
CAD prepare	8 hours	8 hours
Pre processing time	2 hours	34.5 hours
Parallel num.	4,096 cores (K-computer)	512 cores (Intel Xeon)
Flow computation time	258 hours	8 hours
Post processing time	1.5 hours	1.2 hours
Error of C_d prediction	Applx. 10%	Applx. 11-16%
Error of ΔC_d prediction	8%	12%

Wind HPC consortium: At a research activity on Wind-HPC consortium which is organized by Tokyo Institute of Technology and several Japanese major construction companies, the detailed turbulence characteristics on wind canopy of actual urban area geometry that has housings, buildings, vegetation, and street, and so on, has been investigated. And, the academic case validation using square cylinder has been conducted. The results shows reasonable accuracy, so both of results has been published in the paper of architecture design. This research will continue on the FLAGSHIP 2020 project of priority # 4 regarding wind environment evaluation for building construction on severe climate condition through next years.

12.3.3 Development of unified compressible flow solver for unified low to moderate Mach number turbulence with hierarchical grid system

The Simulation of the low speed compressible turbulence is a key challenge for the industrial applications such as combustion, aeroacoustics and significant heat transfer phenomena. Roe scheme with a

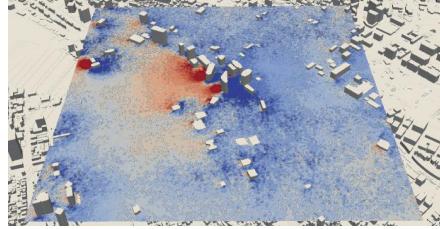


Figure 12.5: Pressure field in Shiodome area(50m Height)



Figure 12.6: Pressure field in Marunouchi area(50m Height)



Figure 12.7: Q-criterion ($Q=0.0015$) in Shiodome area



Figure 12.8: Q-criterion ($Q=0.0015$) in Marunouchi

low-Mach number fix [1] is adopted to tackle slow flows with variable densities. An immersed boundary method (IBM) for compressible flows with a fast, easy to implement and robust interpolation method is developed to handle the complex geometries.

Basic Validation with Academic case: Based on the experiments conducted by Jia and Gogos [2], a steady-state natural convection around a heated sphere under the condition of that, the Grashof number based on the radius of the sphere is 104, is conducted to validate the unified solver. Fig. 12.9(a) shows the contour of the velocity magnitude. The entrainment comes from the bottom of the sphere which is consistent with the description by [2]. Fig. 12.4 shows the temperature contour. Above the top of the sphere, higher temperature region is formed, which cause worse natural convection near the surface so the velocity in this region is quite low. Comparisons of the averaged Nusselt number (Nu), drag coefficients caused by pressure and viscous are tabulated in Table 12.3. The results are in good agreement with the experimental data and show the accuracy and availability of our program for dealing with the complex geometry and heat transfer problems. The present results have been published in [3].

The simulation of a sphere at $Re=104$ is performed to investigate the availability of the unified solver for higher Reynolds numbers. Fig. 12.11 shows the distribution of mean pressure coefficient. The result is well consistent with the [4] and the separation angle can be also accurately captured,

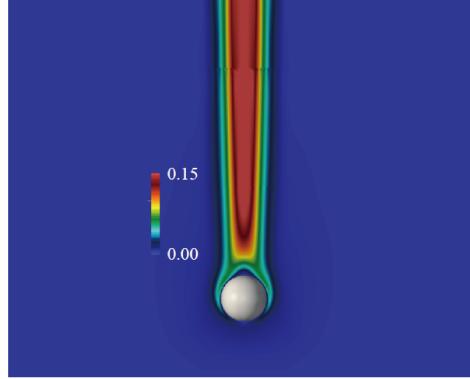


Figure 12.9: A heated sphere: velocity magnitude (m/s)

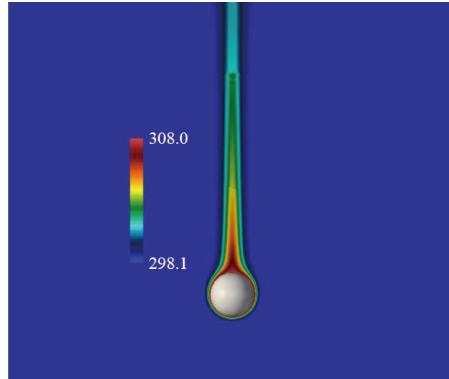


Figure 12.10: A heated sphere: temperature (K)

which is around 86 degrees.

Fig. 12.12 shows the Q criterion contoured by the magnitude of the velocity. The turbulence structures are mainly formed after the separated shear layers generated from the separation point. Besides, the transition from large to small turbulence structures can be also clearly observed in the wake region.

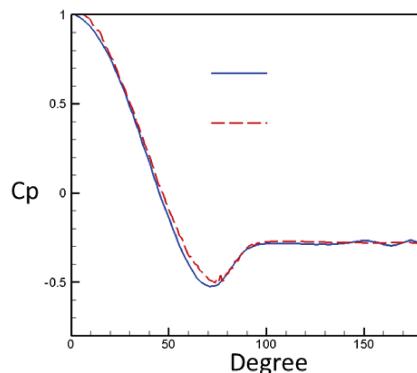
Figure 12.11: Sphere flow at $Re=10^4$: Distribution of mean pressure coefficient

Table 12.3: Comparison with existing experimental data

	Nu	$C_{D,p}$	$C_{D,u}$
Exp. [2]	8.74	0.46	0.62
Present	8.77	0.46	0.59

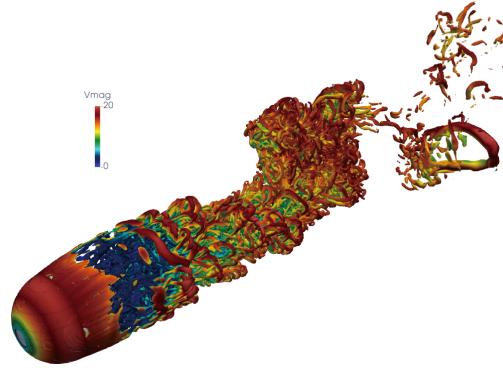


Figure 12.12: Sphere flow at $\text{Re}=10^4$: Q criterion

Finally, the simulation of the whole vehicle demonstrates the capability of the unified solver. Fig. 12.13 shows the contour of the velocity magnitude. The development of the boundary layer on the front window and roof can be clearly observed, which shows the capability of unified solver for handling the complex geometry. Besides, the flows penetrate the front of the car to the engine room is also obviously shown, which indicates that our immersed boundary can also treat non-watertight geometry. Fig. 12.14 shows the history of the drag and lift coefficients. After reaching the quasi steady state, the average values of them are good agreement with the experimental data. This is an indication that the unified solver is also able to obtain accurate results for this kind of practical application. In Fig. 12.15, the Q-criterion contoured by the magnitude of velocity is shown. The development of turbulent coherent structures near the wheel, mirror and side windows is well captured. In addition, the typical turbulent structures-hair pin can be also obviously observed on the roof.

- [1] F. Rieper, Journal of Computational Physics, 230 (2011) 5263-5287.
- [2] H. Jia, G. Gogos, Int. J. Heat Mass Transfer 19 (1996) 1603-1615.
- [3] C. Li, M. Tsubokura, Int. J. Heat Mass Transfer 75 (2016) 52-58.
- [4] C. George, S. Kyle, Physics of Fluids, 16 (2004) 1449-1466.

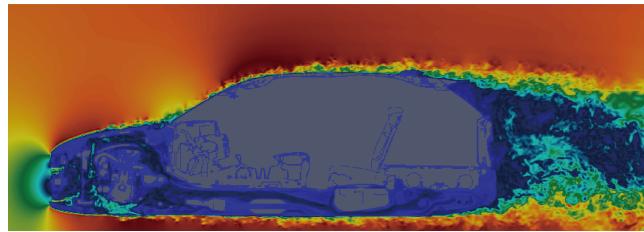


Figure 12.13: Vehicle Simulation: Snapshot of velocity magnitude

12.3.4 Development of high performance moving boundary solver for realistic motion

Aerodynamics of Vehicle in a turn: The aerodynamic performance and stability of a vehicle is strongly influenced by the crosswinds during cruise and while in turning maneuvers. It is difficult to simulate such real-world flow scenarios in wind tunnel experiments. Furthermore, it is also difficult to measure unsteady aerodynamic forces in wind tunnel experiments. Thus, it is desirable for numerical methods to be able to efficiently and accurately simulate such flow conditions. To this end, here, we present simulation of a vehicle (the complex full vehicle geometry discussed in the previous section) undergoing a turning motion, including wheel rotation and turn, chassis roll and turn, in a uniform flow. This simulation is the result of a collaborative work between Mazda Motor Corporation and RIKEN AICS. The detailed vehicle geometry and its motion data were provided by Mazda, and the simulation was carried out on the K-computer using CUBE. The Lagrangian-Eulerian approach

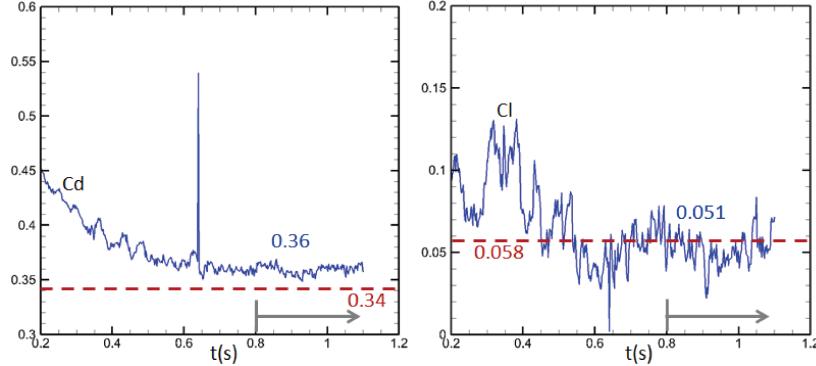


Figure 12.14: Vehicle Simulation: Time series of drag (left) and lift (right) coefficients

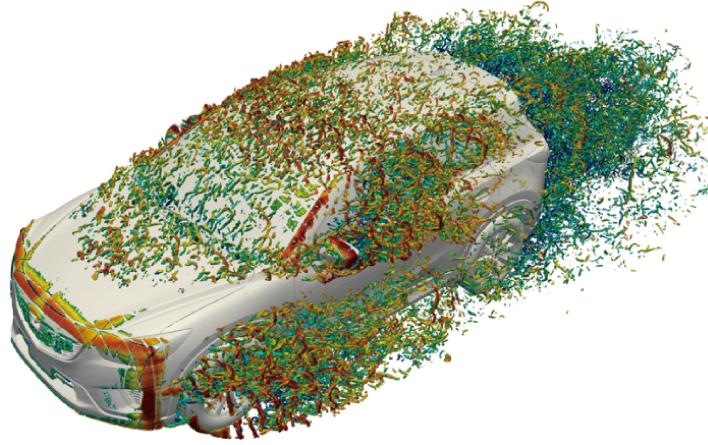


Figure 12.15: Vehicle Simulation: Snapshot of flow structures extracted by the Q-criterion

developed during the previous year was used for this simulation. As mentioned above all the vehicle motion, except linear translation, is imposed on the vehicle. If the linear translational motion was imposed on the vehicle, a fine mesh would be needed in the vehicle's path, which makes the mesh size excessively large. An alternate approach, where instead of imposing the linear translation on the vehicle it is imposed on the entire mesh, was used. In this approach the vehicle's center of gravity remains fixed relative to the mesh. So, the fine mesh is needed only in a small region around the vehicle instead of the region of the vehicle's path. This reduced the mesh size by a factor of 3-5. The results of the simulations are shown in Fig. 12.16.

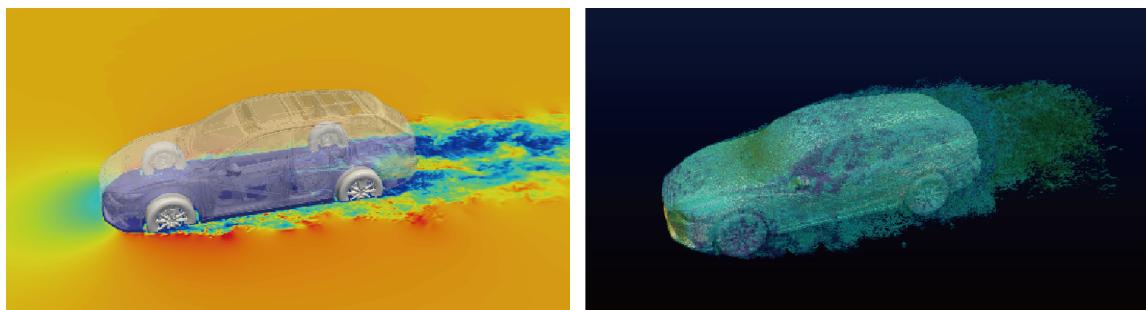


Figure 12.16: Flow field around a vehicle in a turn. (Left) Velocity magnitude on a horizontal plane. (Right) Iso surface of swirl.

Aerodynamic performance of a Ski jumper: Ski jump is a popular winter sport and is a part of winter Olympics. It is one of the sports in which Japan is competitive and has some of the best ski jumpers in the world. Ski Jump is a sport where aerodynamic interaction of the jumper and

air plays a key role in outcome of the sport. Minute changes in an athlete's posture can go a long way, literally. The distance covered by an athlete is strongly correlated to the drag and lift forces on the athlete while in air. And, these forces are greatly influenced by the athlete's posture. In collaboration with Prof. Keizo Yamamoto of Hokusyo University we investigated the aerodynamic performance of two of Japan's top ski jump athletes, Haruka Iwasa and Sara Takanashi. Through the unsteady aerodynamic simulation of the two ski jumpers we analyzed the evolution of forces during a short period before and after the jump from the ski ramp. During this period the jumper changes from a sitting posture to a standing posture. Our analysis revealed that the posture and motion of Haruka Iwasa lead to lower drag force and higher lift force compared to the forces on Sara Takanashi. This is consistent with the real-world performance record of the two ski-jumpers.

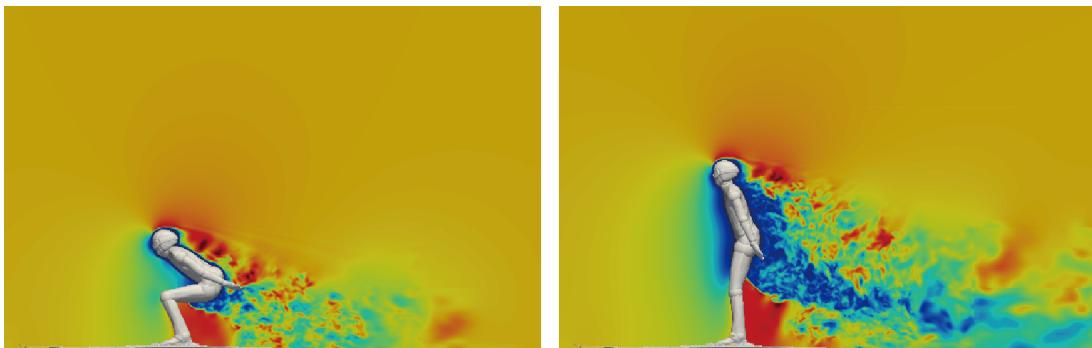


Figure 12.17: Evolution of flow around Ski jumper Haruka Iwasa during a jump. (Left) Starting posture of the jump. (Right) Final posture of the jump.

12.4 Schedule and Future Plan

(1) Five-year objectives and goals toward 2017

- Construction and development of the simulation technology for bringing out the performance of K-computer
- Proposal of the technological trend of HPC simulation toward EXA-scale

(2) Long-term objectives

- Establishment of the research and development center for industrial simulation technology
- Contribution to computer science by expanding the developed simulation technology to different fields

(3) Time schedule

12.5 Publications

	2012	2013	2014	2015	2016	2017
Proposal of the project		Interview to the industry and feasibility study Making specification list for the development				
Building light libraries			Library development Porting guideline Application development			
Development of the coupling algorithms for the PETA-scale computing			Development of the scaling algorithms Development of the coupling algorithms			
Validation studies			PETA-scale applications	Performance test of the post PETA-scale applications		

Chapter 13

HPC Programming Framework Research Team

13.1 Members

Naoya Maruyama (Team Leader)
Motohiko Matsuda (Research Scientist)
Shinichiro Takizawa (Research Scientist)
Mohamed Wahib (Postdoctoral Researcher)
Keisuke Fukuda (Research Associate)
Koji Ueno (Student Trainee)
An Huynh (Student Trainee)
Satoshi Matsuoka (Senior Visiting Scientist)
Tomoko Nakashima (Assistant)
Aya Motohashi (Assistant)

13.2 Research Activities

We develop high performance, highly productive software stacks that aim to simplify development of highly optimized, fault-tolerant computational science applications on current and future supercomputers, notably the K computer. Our current focus of work includes large-scale data processing, heterogeneous computing, and fault tolerance. A major ongoing project in our group will deliver a MapReduce runtime that is highly optimized for the intra- and inter-node architectures of the K computer as well as its peta-scale hierarchical storage systems. Another major project focuses on performance and productivity in large-scale heterogeneous systems. We also study high performance graph analytics on the K computer. Below is a brief summary of each project.

13.3 Research Results and Achievements

13.3.1 KMR

Improving Locality When Running MPI Programs as MapReduce Tasks

Although MapReduce systems can allocate tasks to nodes where their inputs reside to increase data locality for improving performance, these systems only target on tasks implemented as serial programs and do not consider running tasks implemented as parallel programs using MPI as Map or Reduce task. As many scientific applications are implemented using MPI and some application

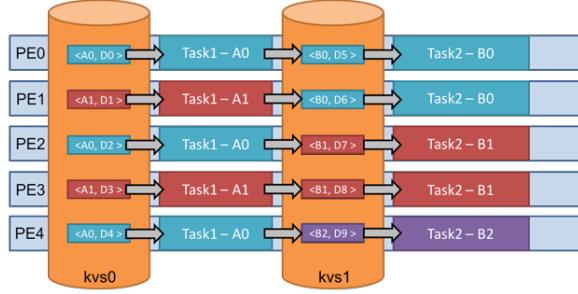


Figure 13.1: Execution flow of MPI program in MapReduce model

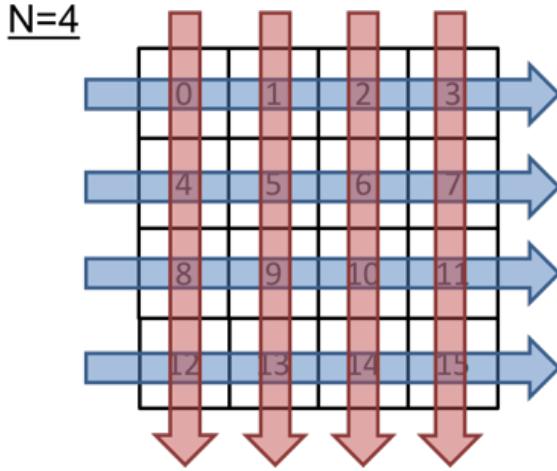


Figure 13.2: Calculation pattern of the benchmark program

workflows form ensemble execution patterns of such MPI programs, the workflows can be implemented easily and efficient data access can be achieved if a MapReduce system can allocate tasks implemented using MPI so that it can exploit data locality in them.

We proposed an extension of the execution model of MapReduce to achieve high performance when running MPI programs as Map/Reduce tasks. We model data to be processed as Key-Value as the traditional MapReduce model. However, to process the data, we propose a new `map` function which makes process groups where each process in a group has a key-value whose key is same as those of other processes in the group and applies a user-defined mapper, which is implemented using MPI, to the key-values using processes in each group. Figure 13.1 shows the execution flow.

To evaluate our proposal, we used $N \times N$ nodes of the K computer and compared performance of our method in which data access was performed locally and that of random data access. We used a synthetic benchmark program where each node has an individual data and which iterates the following two computation; the first computation groups N nodes in low direction and processes data on them, and the second groups N nodes in column direction and processes data on them as shownen in Figure 13.2. The result is shown in Figure 13.3. The horizontal axis is the amount of data on each node and the vertical axis is the relative performance of an iteration against random data access. As can be seen from the figure, the performance of our proposal improve as the number of nodes and the amount of data increase.

Skew-Tolerant Shuffling for Load Balancing for Reduce Operation

In a MapReduce program, the number of tasks in Map phase is defined by the number of split of the input data and that in Reduce phase is defined by the number of keys generated by the previous Map phase, and these numbers defines the maximum degree of parallelism of both phases. Though the former can be defined by users when executing the program, users can not control the latter because it depends on patterns of applications and input data. In contrast, as KMR statically defines number of processes used in both phase as the same value, we need to adopt load balancing techniques to

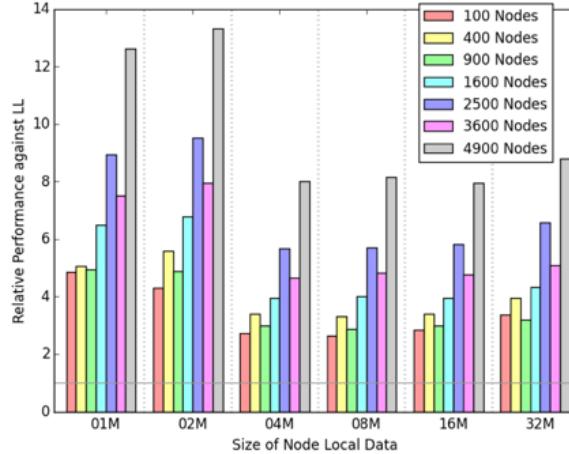


Figure 13.3: Experimental results

average loads between processes in both phase. Though balancing loads in Map phase can be left to users, a MapReduce system should support load balancing in Reduce phase as loads on processes heavily depend on the shuffle communication which is transparently performed by a system.

To balance loads on processes in Reduce phase, we proposed and implemented a shuffle algorithm that minimizes the skew of numbers of key-value pairs processed on each process [nishinaga'aics16]. Our method is based an existing algorithm called LEEN and is extended as follows:

- It uses a distributed algorithm.
- It reduces the number of search target keys (keys in key-value pairs) to reduce the cost of running the algorithm.

We compared the performance of our method and that of using a shuffle that randomly assigned key-values to processes by using a benchmark program that implemented k-means on top of MapReduce. As a result, time spent for executing Reduce task was reduced against the random shuffle as our proposal averaged the number of keys among processes. On the other hand, we also confirmed that the overall execution time increased due to the large cost of running the algorithm. We plan to reduce the algorithm execution cost as a future work.

Visualize MapReduce Task Execution

Widely used MapReduce systems, such as Apache Hadoop and Spark, have their own profiling and visualization tools to see the status of jobs. They help users to look for performance bottlenecks, to do debug and to optimize their programs. As KMR is implemented as one of an MPI library using the C language, we can use any profilers, such as gprof, Intel Vtune and FUJITSU profiler. However, as they target on low level events, such as memory access and function calls, they are not suitable for profiling task level events in a MapReduce program. We developed an event tracer for KMR that traced MapReduce operations, such as Map/Reduce tasks and Shuffle communication, and a visualization tool named KMRViz that displayed the traces in a GUI window.

The tracer traces KMR function calls and records times of start and end of a function and numbers of input and output key-value pairs of the function. To eliminate IO overhead for writing records during a program execution, we implemented the tracer so that it recorded profiles in-memory while execution and wrote them to files at the end of the program execution. Moreover, to eliminate the burden of using the tracer, we implemented it so that the tracer could be enabled just by setting an environment variable. KMRViz receives trace files generated by the tracer and displays them in time series by each process as the Figure 13.4. It is implemented using GTK+3 and users can zoom-in and out using mice and trackpads.

13.3.2 High Level Framework for High Performance AMR

Summary: Adaptive Mesh Refinement methods reduce computational requirements of problems by increasing resolution for only areas of interest. However, in practice, efficient AMR implementations

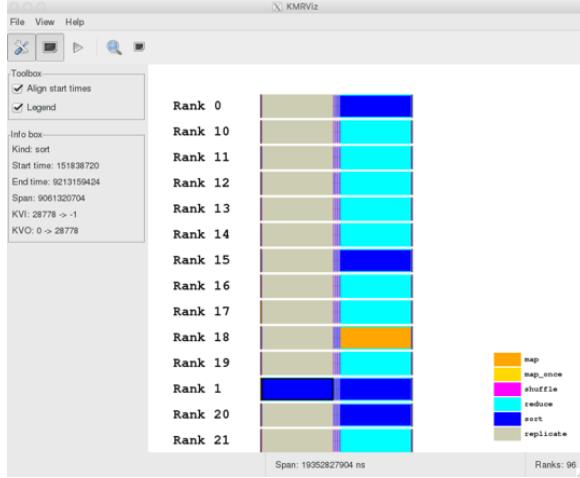


Figure 13.4: Screenshot of KMRViz

are difficult considering that the mesh hierarchy management must be optimized for the underlying hardware. Architecture complexity of GPUs can render efficient AMR to be particularly challenging in GPU-accelerated supercomputers. In this project, we present a high-level framework that can automatically transform serial uniform mesh code annotated by the user into parallel adaptive mesh code optimized for GPU-accelerated clusters. We show experimental results on three production applications. The speedups of code generated by our framework are comparable to hand-written AMR code while achieving good and weak scaling up to 1000 GPUs.

Motivation: Frameworks that provide support for GPU in AMR applications require the programmer to write his own versions of the target-optimized solvers. Moreover, there can be scalability limitations caused by the overhead of the CPU-GPU communication schemes in those frameworks. We present a high-level framework that specializes in enabling efficient and scalable structured AMR solutions to scientific applications running on GPU-accelerated systems [Wahib2015]. Our framework uses a high-level programming model that provides an architecture-neutral programming interface and adopts an AMR strategy that would eliminate any CPU-GPU communication schemes that can limit scalability. We base our framework on octree-based AMR implementation in which we use a distributed tree and adapt the mesh in a parallel fashion with minimum inter-node communication. We base our GPU implementation on a data-centric approach at which the CPU is specialized in managing the data structures representing the mesh hierarchy, while AMR-specific routines that operate on mesh application data are executed on the GPU. Hence keeping the mesh application data arrays on the GPU memory for the entirety of the simulation.

Programming model: The programming model is designed to be exposed in an architecture-neutral manner; the programmer has no knowledge of the underlying architecture. We provide the programmer with a set of C language directives to identify the stencil functions and data arrays in a logically fixed and uniform mesh implementation of solver(s). The programming interface enables the underlying compiler-based framework to statically analyze the solvers, construct the adaptive mesh hierarchy, automatically parallelize the mesh partition over distributed memory, and apply optimizations required for keeping the mesh application data, i.e., stencil arrays, in GPU memory throughout the simulation.

Optimizations: When an AMR code generated by our framework is executed on a GPU-accelerated cluster, the stencil and mesh adaptation kernels run on the GPU, while managing the octree data structures and load balancing is done on the CPU side. Since we pursue efficiency and scalability, code on both the CPU and GPU should be optimized. The stencil kernel and mesh adaptation kernels are memory-bound kernels that are optimized to use the shared memory of the GPU and maintain coalesced memory accesses. For load balancing, *intra-node* load balancing is applied by moving balancing the number of blocks equally among the GPUs. For the *inter-node* load balancing, we rely on a space fitting curve to decide on how to redistribute the blocks.

Implementation: Our framework consists of a compiler and runtime components. We generate executables optimized for GPU execution by leveraging the LLVM compiler infrastructure. Our compiler builds on the LLVM compiler infrastructure. First, we use the front end to analyze and

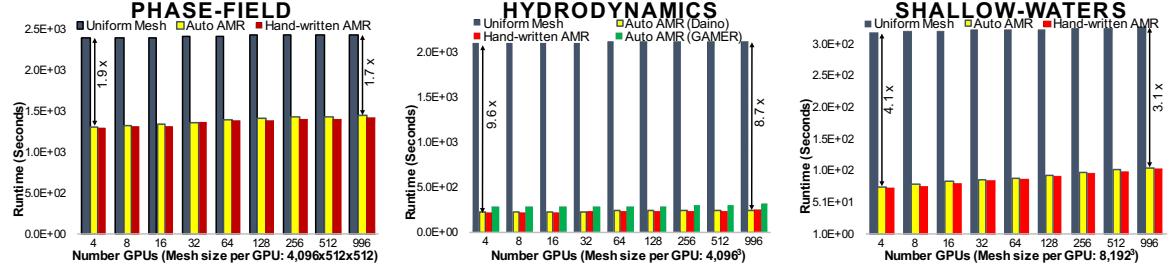


Figure 13.5: Weak scaling of uniform mesh, hand-written and automated AMR (GAMER-generated AMR included in hydrodynamic)

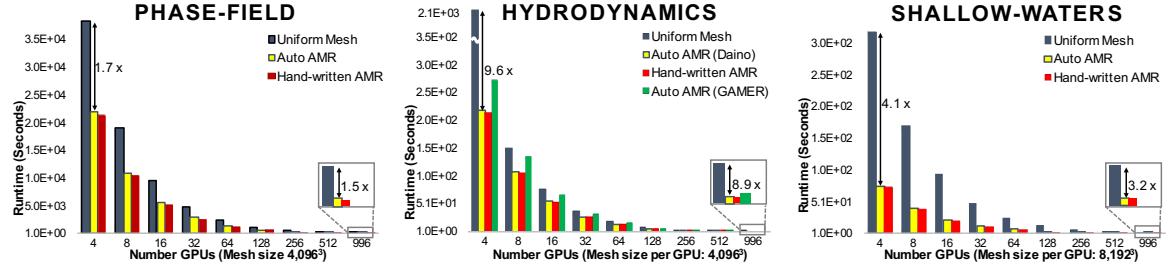


Figure 13.6: Strong scaling of uniform mesh, hand-written and automated AMR (GAMER-generated AMR included in hydrodynamic)

translate the stencil source code into GPU-optimized code in the form of LLVM Intermediate Representation (IR). Next, compiler passes are applied on the IR to add the AMR management code, which in turn make API calls to: a) the runtime API, and b) the GPU-optimized code generated by Nvidia back end code generator. Finally, LLVM IR is compiled and linked with the runtime libraries to generate an executable. The different stages of the compilation is managed by a shell script that the programmer invokes.

We have built two libraries into the runtime. First, the *AMR library* that encapsulates the AMR hierarchy management software. Second, the *communication library* that wraps the MPI runtime library to simplify data movement operations for the AMR driver.

Applications: We evaluate our framework with three production applications. *Phase-field simulation for dendritic growth:* We simulate 3D dendritic growth during solidification in a binary alloy using phase-field model. The computation requires a 2nd order finite difference scheme for space with 1st forward Euler-type finite difference method for time on a 3D uniform mesh. We use the code by Shimokawabe et. al as our reference implementation. *Hydrodynamics Solver:* We model a hydrodynamics application using 3D Euler equations. We explore a 2nd order directionally split hyperbolic schemes to solve the equation. We use the code by GAMER framework for the Hydrodynamics solver as our reference implementation. *Shallow-water Solver:* We model shallow water simulations by depth-averaging the Navier–Stokes equations. We use a numerical method based on a semi-discrete, 2nd order, central scheme. To advance the time step, we base our flux, max step, boundary condition, and time integration kernels on the model proposed by Saetra et. al.

Experimental Setup: We use the TSUBAME2.5 supercomputer at Tokyo Tech. Each node has two socket Intel Xeon X5670 2.93GHz CPUs (12 cores), and three Nvidia Kepler K20x GPUs with total 54GB and 18GB of system and GPU memory. The compute nodes are interconnected by dual QDR Infiniband networks with a full bisection-bandwidth fat-tree topology network. We use CUDA v7.0 Toolkit for GPU code and LLVM compiler infrastructure v3.8 for the framework. Single precision variables are adopted in all experiments for all applications. The hand-written and automated AMR versions of all applications use the data-centric AMR approach. All experiments used 16³ mesh block size and 2D CUDA thread blocks of size 16x16 threads. All test runs are collected for 100,000 time steps with a constant maximum of six refinement levels.

Results: In a weak scaling experiment, shown in Figure 13.5, the runtime for uniform mesh, hand-written AMR, and auto-generated AMR are compared. The following points are important to note. First, more than 1.7x speedup is achieved using our framework with 1000 GPUs for the phase-field simulation. This is a considerable improvement considering that the uniform mesh im-

plementation is a Gordon Bell prize winner for time-to-solution. Second, we included a comparison with the auto-generated AMR by GAMER framework for the hydrodynamics solver. The AMR code generated by our framework is faster than the code generated by GAMER, mainly because GAMER uses a pipeline to hide data movement latency while our framework uses a data-centric approach to avoid data movement altogether. Third, we achieve good scaling that is comparable to the scalability of the hand-written AMR code.

Figure 13.6 shows a strong scaling comparison for hand-written AMR and auto-generated AMR against uniform mesh implementation. The auto-generated AMR by our framework achieves runtime and scalability comparable to that of the uniform mesh implementation. However, when using more GPUs, reduction in speedup starts to occur as the management of AMR starts to dominate the simulation runtime.

13.3.3 High Performance Graph Analytics Study with Graph500

We have been working on performance analyses and optimization of the Graph500 benchmark with the K computer. Previously, our implementation of the benchmark achived 19585 GTEPS at the problem size of scale 40, which was ranked second in the Graph500 ranking updated in November 2014.

To further improve the performance, we mainly focused on its inter-node communication cost, and developed a series of optimizations that allowed us to develop a nearly 2x faster implementation. We extended the data structures and algorithm called Hybrid BFS, which is known to be effective small-diameter graphs, so that it scales to top-tier supercomputers with tens of thousands of nodes with million-scale CPU cores with multi gigabyte/s interconnect. With those optimizations, its performance when running on the K computer with 82,944 compute nodes and 663,552 CPU cores reached 38612 GTEPS, which was ranked first both at the two consecutive lists of July and November 2015. A paper describing the details of the algorithms and optimizations will be published in the future.

13.4 Schedule and Future Plan

13.4.1 KMR

We plan to extend our first research activity (Improve locality when running MPI programs as MapReduce tasks) to improve the representation of programs. Because the current version depends on key-value pairs as data exchanged between tasks as the usual MapReduce model but supports both serial and parallel MPI execution, we have to distinguish Map functions for these two cases and actually implemented them as different functions. It is obviously redundant as their objective is the same; mapping data. Moreover, users can not recognize the number of processes used in each MPI programs run as Map because they are unaware of locations of input key-value pairs and number of processes that hold the input key-value pairs.

To overcome these problems, we are developing a hybrid programming model of MPI message passing and MapReduce model. In this model, we model data exchanged between tasks as multi-dimensional array which can be split to process each chunk depending on user perspectives at run time. The location of chunks of the multi-dimensional array is also changeable at run time. A Map function in this model applies a user-defined function to each chunk of the multi-dimensional array using processes that hold the data. We are developing the model and applying it to application workflows that perform ensemble simulations.

As for KMRViz, we plan to support more KMR operations to trace and visualize them.

13.4.2 High Level Framework for High Performance AMR

Initially, we intend on testing expanding the framework to allow for user-defined boundary conditions and error estimation functions. We also intend on testing framework at larger scale using TSUBAME2.5 supercomputer (under a grant from JHPCN). Finally, we are closely working with collaborators on FLASH project at Chicago university (Dr. Anshu Dubey) to introduce an integration between our framework and FLASH

13.5 Publications

Chapter 14

Advanced Visualization Research Team

14.1 Members

Kenji Ono (Team Leader)
Jorji Nonaka (Researcher)
Mikio Iizuka (Researcher)
Chongke Bi (Postdoctoral Researcher)
Daisuke Sakurai (Postdoctoral Researcher)
Kazunori Mikami (Technical Staff)
Tomohiro Kawanabe (Technical Staff)
Seiji Fujino (Senior Visiting Scientist)
Naohisa Sakamoto (Visiting Scientist)
Masahiro Fujita (Visiting Technician)
Kentaro Oku (Visiting Technician)
Steve Petruzza (Student Trainee)
Yukiko Hayakawa (Assistant)
Keiko Matsuoka (Assistant)

14.2 Research Activities

Although the research activities of this team cover a broad range of the end-to-end simulation pipeline, including simulation, visualization and analysis, the main activities are centered on the study, design, and development of effective tools and mechanisms for visualization and analysis of large-scale parallel simulation results generated by the K computer. In this fiscal year, the team has worked on some core technologies necessary for building a production-level visualization and analysis framework. Taking into consideration the heterogeneous hardware infrastructure for post-processing, which includes the K computer itself, some post-processing servers, a visualization oriented GPU cluster, and local computer devices, we have worked on a framework named **HIVE** (**Heterogeneously Integrated Visual analytics Environment**) [2015:AVRTeam:HIVE]. Some of the core technologies, for enabling large data visualization and analysis, include a scalable parallel image compositing library for massively parallel rendering environments (**234Compositor**) [2015:Nonaka:JASSE, 2015:Nonaka:IPSJ-HPCS, 2015:Nonaka:HPCS], a visual data analysis mechanism for multivariate data (**Fiber Surface**), a performance monitoring library for scientific

applications (**PMlib**) [2015:AVRTeam:PMlib], a multi-display management library for building scalable cooperative workspace on tiled wall displays (**ChOWDER**) [2015:AVRTeam:Chowder], a parallel-in-time integration framework (**PinT**) based on *Parareal* algorithm, and a sparse modeling approach based data compression for *in-situ* visualization (**Sparse Modeling**).

Effective visualization and analysis of large-scale data sets generated by modern leading-edge supercomputers such as the K computer is widely recognized as a difficult and challenging problem. There is still a lot of discussion among researchers about this topic, and we can cite the Dagstuhl-style *Shonan Meeting Seminar* focusing on “*Big Data Visual Analytics*” [Ono:2015:Shonan], and also a panel discussion entitled “*Top Computational Visualization R&D Problems 2015*” [Ono:2015:SA15] during the *Symposium on Visualization in High Performance Computing*. Our research activities have focused on a sustainable long-term development lifecycle for the HIVE framework, trying to assure the easiness for enhancement, maintenance and support, and also mechanisms for absorbing the heterogeneity of the computer platforms normally found on modern HPC operational environments. In addition to these items, we have also considered the following topics: easy usability; customizability for enabling diverse visualization scenarios; interactivity for finding the best visualization parameters; and in-situ processing mechanisms.

Most of the results of aforementioned research and developemnt efforts have already been released as open source libraries and applications, which are continuously maintained and updated. Following are the full list of those libraries and applications: *PDMlib* [2015:AVRTeam:PDMlib]; *UDMlib* [2015:AVRTeam:UDMlib]; *HDMlib* [2015:AVRTeam:HDMlib]; *CDMlib* [2015:AVRTeam:CDMlib]; *Polylib* [2015:AVRTeam:Polylib]; *Cutlib* [2015:AVRTeam:Cutlib]; *LPTlib* [2015:AVRTeam:LPTlib]; *TextParser* [2015:AVRTeam:TextParser]; and *JHPCN-DF* [2015:AVRTeam:JHPCN-DF].

14.3 Research Results and Achievements

14.3.1 Visual Analytics Framework (HIVE)

We integrated some of the core technologies listed in the previous section, and developed a prototype of the HIVE framework focusing on the operational environment of the K computer. Considering the hardware heterogeneity of this environment, which includes the K computer itself, some post-processing servers, visualization oriented GPU cluster, and local computational devices, HIVE was designed to run on *SPARC64fx CPUs*, *x86 CPUs*, and *GPUs*. Figure 14.1 shows the Web-based User Interface(UI) designed for accessing the HIVE functionalities, and for small data sets it is possible to execute the entire visualization and analysis on the local computational side. For larger data sizes, this Web-based UI can be useful for preparing the visualization scenes to be used as a parameter during the batch job execution of large-scale parallel visualization processing. As shown in the Figure 14.1, HIVE adopted the sort-last parallel rendering approach, which requires the image compositing **process** right after the parallel rendering stage. It is worth noting that the scalability target of both parallel processing modules were the full node of the K computer.

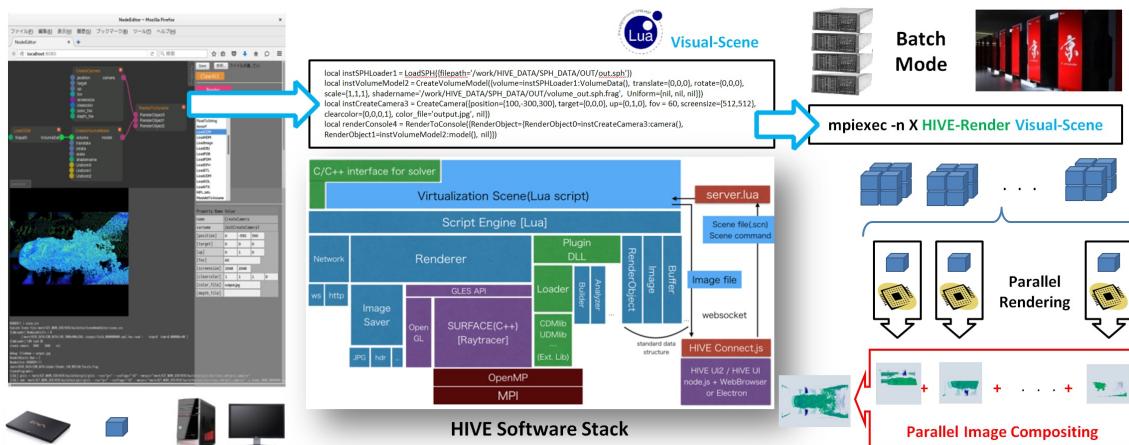


Figure 14.1: The software stack and the Web-based user interface of the HIVE framework designed for the visualization and analysis of large-scale simulation results generated from HPC platforms.

In order to enable the sustainable long-term development and maintenance life cycle of the HIVE framework, we focused on a minimum set of external libraries for avoiding the software dependency as much as possible. For instance, the **SURFACE** rendering module adopted the *GLES 2.0* (OpenGL for Embedded Systems Version 2.0) compatible *API* (Application Programming Interface). The *GLES* is a subset of the original cross-language and multi-platform *OpenGL API* for rendering 2D and 3D vector graphics. This compatibility facilitates the development of both CPU and GPU based implementations. HIVE has designed to have a minimum but necessary set of functionalities, and the **Plugin DLL** module facilitates the extensibility and maintainability of the HIVE system by separating the core modules and other external or third party modules. This feature is possible, in part, due to the adoption of the light-weight and multi-platform scripting language called **Lua**. This scripting language is also used for describing the entire dataflow of a given visualization scene (**Visual-Scene** in the Figure 14.1), which can be easily converted to a batch job script for executing directly on the K computer. Examples of the third-party modules already incorporated to the HIVE include the large-scale parallel image compositing module (**234Compositor**), and a set of large-scale data management libraries (**xDMlib**). Other modules which are in the process of integration include those for visual analysis: (**Fiber Surface**) and Performance Monitoring (**PMlib**).

14.3.2 234Compositor

The core rendering engine of the HIVE is a highly scalable *Raytracer* named **SURFACE** (Scalable and Ubiquitous Rendering Framework for Advanced Computing Environments), which adopted the *Sort-last parallel rendering* approach, that is, after the parallel rendering stage the entire set of rendered images should be gathered and merged into the final single image as shown in the Figure 14.1. To meet the performance and scalability demands of the SURFACE, we have developed the *234Compositor*, a hybrid MPI/OpenMP parallel image compositing library, based on the well-known *Binary-Swap* parallel image compositing algorithm. Although *Binary-Swap* has proven highly scalable, there exist some issues including the requirements for power-of-two number of processes, and the final sub-image collecting cost by using the *MPI_Gatherv* collective operation.

There already exist some extensions for the *Binary-Swap* to enable the use with non-power-of-two number of processes. One of these extensions is *Telescoping* method where the number of processes is gradually converted to the largest power-of-two number of processes, and was implemented on the *Ice-T* parallel image compositing library adopted by some parallel visualization applications such as *ParaView* and *VisIt*. Although it has proven efficient on specific supercomputer hardware architectures, we focused on a single-stage conversion approach to minimize the conversion overhead, and utilized the *3-2* and *2-1 Eliminations* techniques proposed by *Rabenseifner et al.* for MPI communications involving non-power-of-two odd number of processes. We extended the original algorithm for enabling the use to the even number of processes by applying the *234 Scheduling* [2015:Nonaka:HPCS], as shown in Figure 14.2. The communication pattern of the *3-2 Elimination* was also extended to enable the overlapping of communication and computation.

The final stage of the *Binary-Swap* image compositing algorithm is the gathering of image frag-

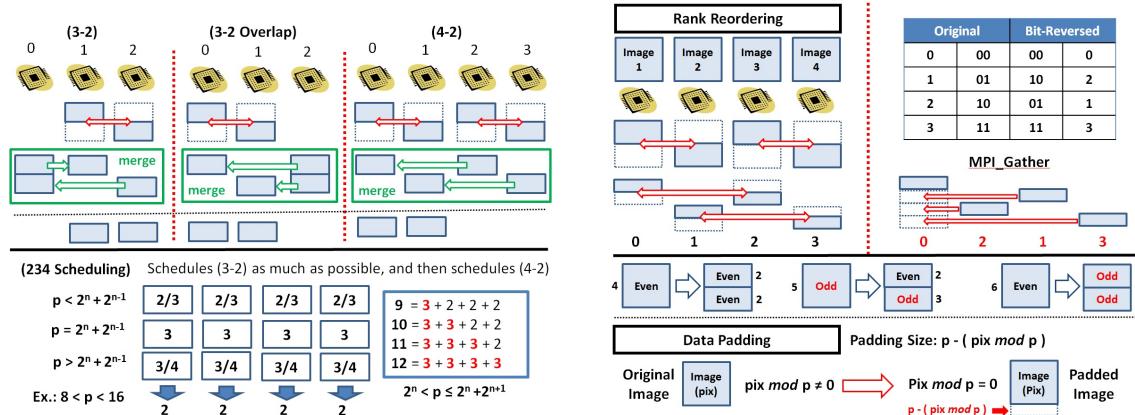


Figure 14.2: *3-2*, *3-2 Overlap* and *4-2 Eliminations* with the *234 Scheduling* mechanism (Left side). *MPI Rank Reordering* and *Data Padding* for enabling the use of *MPI Gather* (Right side)

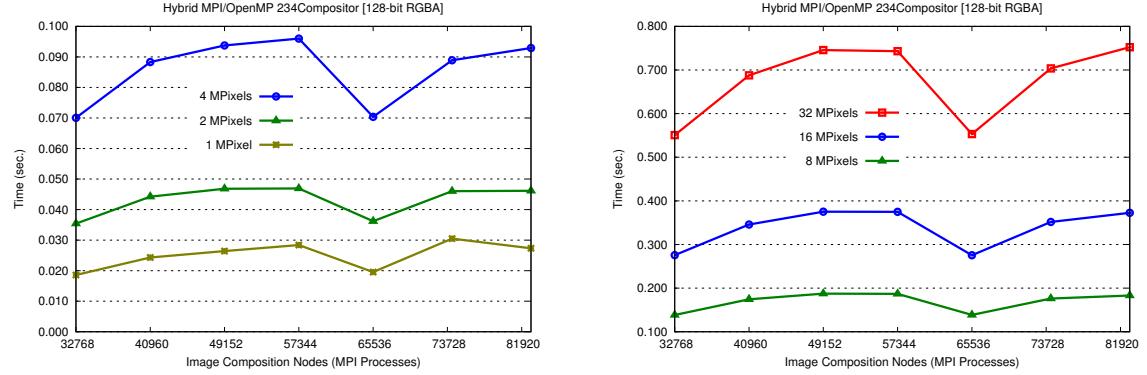


Figure 14.3: Scalability evaluation of the *234Compositor* for image sizes from 1 to 32 Mega Pixels (RGBA 128-bit) using up to the full node of the K computer.

ments, distributes among the image compositing nodes, to the master node and the assembly of the final image. In order to minimize the performance degradation when using a large number of node counts, we developed the *Multi-Step Image Composition* approach[2015:Nonaka:JASSE], which works by recursively dividing the composition nodes into smaller groups with a pre-defined threshold. We also worked on *MPI rank reordering* and *image data padding* techniques for enabling the use of *MPI_Gather*, in substitution to the traditional *MPI_Gatherv*, for optimizing the final image gathering stage[2015:Nonaka:IPSJ-HPCS]. Figure 14.3 shows the performance evaluation results of the *234Compositor*, which integrates the aforementioned techniques, and we could confirm the scalability up to the full node of the K computer when compositing 128-bit RGBA images with sizes varying from 1 to 32 MP (Mega Pixels).

14.3.3 Fiber Surface

With our *fiber surface GUI* (Fig. 14.4), a user can analyze multivariate data to quickly extract 3D features such as vortex structures in meteorology data and objects in CT scans. When the user is familiar with the popular isosurface analysis, it is not difficult to understand the concept of fiber surface – it is a straight-forward generalization of isosurface. The GUI is integrated into HIVE so that the user can analyze large-scale datasets in various computational architectures. In fact, multivariate datasets are often the output of simulations in supercomputers like the K computer. The user explores data by querying data values of interest. The query is as simple as drawing a polygon in a scatterplot. Our GUI then visualizes the queried samples in the 3D coordinate as 2D polyhedra. Thanks to fiber surface, the data size becomes magnitudes smaller compared to the original 3D data. This means that once the fiber surface is extracted, even computers with weaker performance can visualize the simulation outputs efficiently. The GUI resembles to traditional isosurfacing tools, while respecting today's GUI design for multifield exploration. The GUI also offers a scripting environment in order to support flexible data analysis.

14.3.4 Development of PMlib (Performance Monitoring Library)

PMlib is an open source software library designed for monitoring the computational performance of scientific applications. PMlib has designed to support hybrid parallelism where the parallel distributed memory and shared memory computing are exploited by the applications through specific libraries such as MPI and OpenMP. This library is suitable for monitoring computational workloads of applications which can change dynamically over the time and space, depending on their initial conditions and state variables, such as computing particle trajectories over the subdomains in parallel fluid simulations.

The performance statistics for the computational workload such as floating point operations per second (flops) and memory load/store operations per second (bandwidth) can either be defined as the algorithmic workload, that is, theoretical requirements decided by the source program, or as the actually executed system workload. The latter varies depending on the system conditions including the compiler optimization, prefetch strategy and data movement on hierarchical cache components, and numbers of pipeline stages for arithmetic operations.

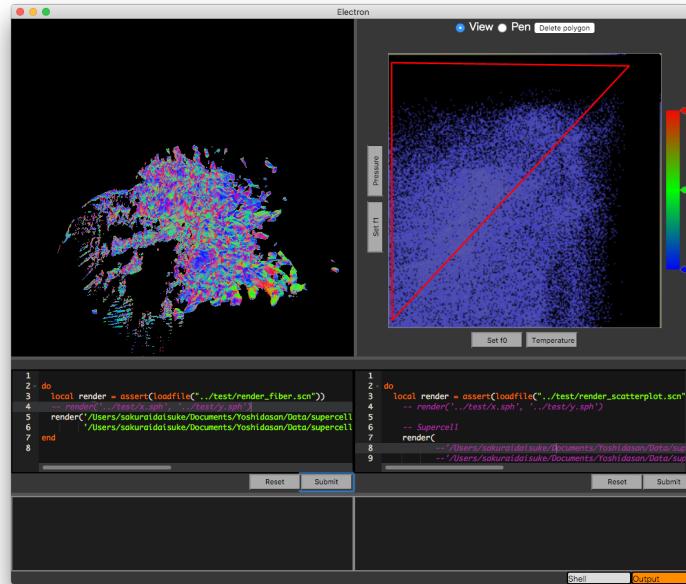


Figure 14.4: Fiber surface GUI visualizing water cohesion in the atmosphere (simulated with SCALE, which is developed by the Computational Climate Science Research Team of RIKEN AICS).

This library allows the users to select the best policies to obtain the statistics during runtime executions. The statistics policies can be explicitly declared by the users as API arguments inside the source program, or can be automatically acquired from the processor built-in hardware counters. PMlib will produce the report in text format at a post processing phase as shown in Figure 14.5.

```
Parallel Mode: Hybrid (8 processes x 2 threads)
The environment variable HWPC_CHOOSER=FLOPS is provided.
Total execution time          = 2.270602e+00 [sec]
Exclusive sections statistics per process and total job.
Inclusive sections are marked with (*)
```

Section	Label	call	accumulated time[sec]	avr	avr[%]	sdv	avr/call	avr	sdv	speed	[hardware counter flop counts]
First section	:	1	9.086e-01	30.29	5.52e-01	9.086e-01		5.909e+09	7.25e+08	6.50 Gflops	
Second section(*)	:	1	1.996e+00	66.54	4.18e-01	1.996e+00		5.752e+09	3.94e+08	2.88 Gflops(*)	
Subsection X	:	1	7.349e-01	24.50	1.89e-01	7.349e-01		6.586e+09	5.32e+08	8.96 Gflops	
Subsection Y	:	1	6.123e-01	20.41	1.26e-01	6.123e-01		5.896e+09	3.49e+08	9.63 Gflops	
Sections per process			2.256e+00	-Exclusive CALC sections-				1.839e+10		8.15 Gflops	
Sections total job			2.256e+00	-Exclusive CALC sections-				1.471e+11		65.22 Gflops	

Figure 14.5: Output example of PMlib.

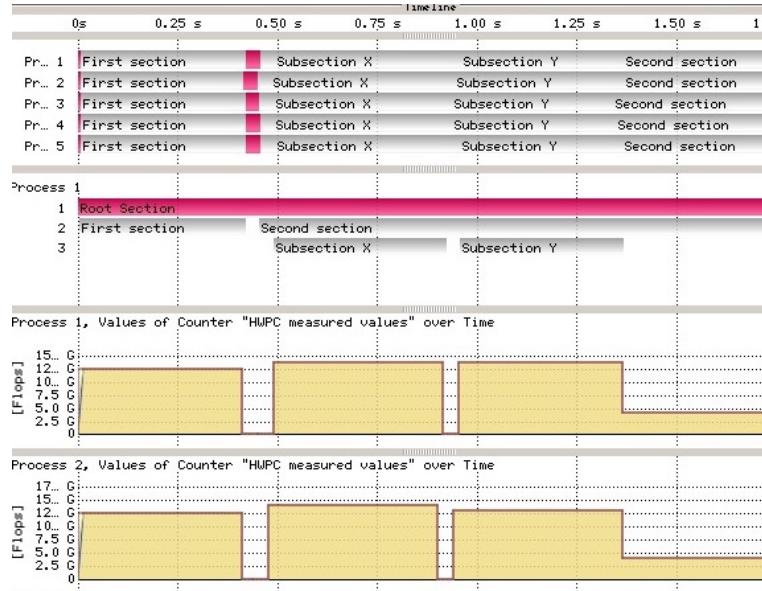


Figure 14.6: Statistical view of PMlib's output.

In this fiscal year, the development efforts for PMlib was focused on the following items:

- Fortran API support
- Multiple MPI group support
- Addition of example programs and update of the user document
- Addition of an optional sorted output
- Initial study of a graphical interface for the Open Trace Format

14.3.5 ChOWDER

ChOWDER (*Cooperative Workspace Driver*) is a Web browser based multi-monitor controller system designed to assist collaborative work among multiple users and applications. The development of *ChOWDER* has been conducted by the financial support of a cross-ministerial Strategic Innovation promotion Program (SIP) grant. *ChOWDER* delivers the functionality of virtual display driver for tiled wall display, and can be used for building scalable cooperative workspace environment for designers and engineers.

As shown in Figure 14.7, *ChOWDER* system is composed of the following three subsystems:

- Client Controller
- Client Display
- Display Server

Client Controller is used for positioning the client displays on the user defined virtual display space. The controller is also used for positioning the displaying objects inside the virtual display space, via Web-based User Interface (UI) since *Client Controller* works inside the Web browser.

Client Displays are the physical displays where the contents in the virtual display space are actually displayed. Figure 14.7 shows an example of 8 *Client Displays* forming a single virtual display, and the user can make a huge virtual display space by simply adding new *Client Displays* to these existing *Client Displays*. These subsystems can work on a variety of devices based on the following Operational Systems: Windows, Mac, Linux, iOS, and Android. *Display Server* controls the communications between *Client Controller* and *Client Displays*, and manages the display objects to be displayed on *Client Displays*. *Display Server* was implemented by using the *node.js*, *socket.io*, *websockets*, and *redis* technologies. As shown in the Figure 14.7, *Display Server* possesses an API for cooperative behavior, which enables a seamless communication with HIVE system, that is, the graphics output of the HIVE can be sent to *ChOWDER*'s *Client Displays*.

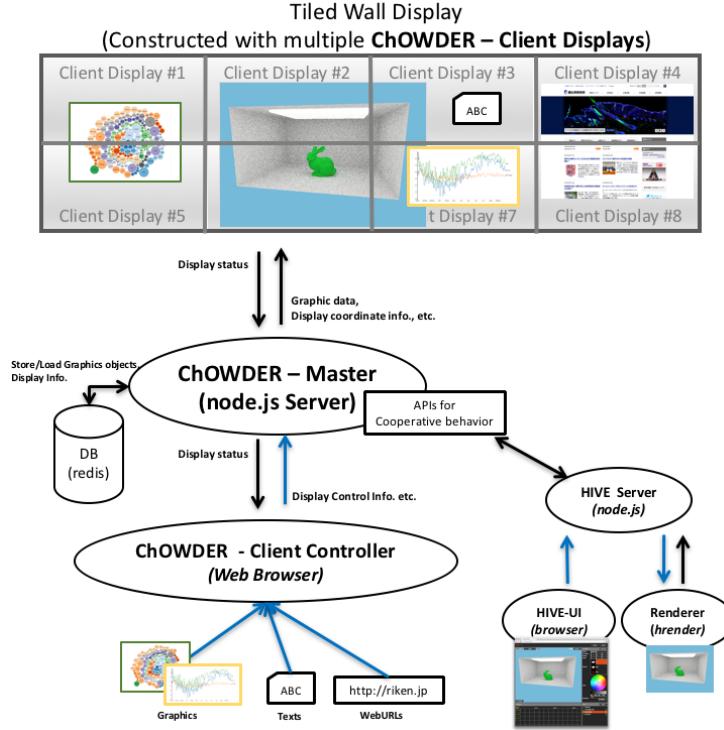


Figure 14.7: System diagram of ChOWDER.

14.3.6 Parallel-in-Time Integration

Parallel-in-time integration (PinT) is desired for effectively exploiting the concurrency of the post peta-scale parallel computing environments which are expected to increase the currently achievable degree of concurrency. The study of PinT advances rapidly after the *Parareal method* (Fig. 14.8(a)) was proposed by Lions at the beginning of the 2000's. The Parareal method has the characteristic that it is easy to reuse existing codes and it doesn't demand large memory capacity. In other words, Parareal method provides us a general and practical framework to describe parallel-in-time integration algorithms for a variety of simulation codes. Therefore, we have investigated a general-purpose PinT framework, which employs the *Parareal* method, to reduce the developer efforts for describing the PinT codes (Fig. 14.8(b)).

We have investigated and found that *Parareal* method works well for parabolic PDEs, but does not work so well for hyperbolic PDEs. Therefore, we performed some investigations on the trends and prospects of PinT methods, and then we set a research direction for the PinT study. As the next steps, we will intensively study the Reduced basis methods (Fig. 14.9(a)) which use the sparse modeling technique, phase alignment methods (Fig. 14.9(b)), adjoint-based methods which use symmetrization of the time-integrator and Parareal/SDC hybrid methods.

14.3.7 Sparse Modeling

Large-scale simulation usually needs several days, or even several months to be executed on current supercomputers. It is strongly desired by researchers to observe the visualization results of such kind of simulations in real time for analysis and adjustment of initial simulation parameters. However, due to the limitation of memory size, currently most of such kinds of simulation datasets are analysed as post processing such as visualization. It is usually necessary to execute these large-scale simulations several times for selecting a set of good initial parameters for simulations and visualizations, and this leads to a time consuming process for scientists and engineers. In this research, we investigated a sparse modeling-based interactive *in-situ* visualization method, in order to assist the users to visualize the simulation results along with the running simulation. By analysing these visualization results in real-time, the users will be able to adjust the simulation parameters and restart the simulation without the need for waiting to the conclusion of the entire simulation. As a result, it becomes possible to avoid the repetitive execution of the entire simulation just for the parameter selection

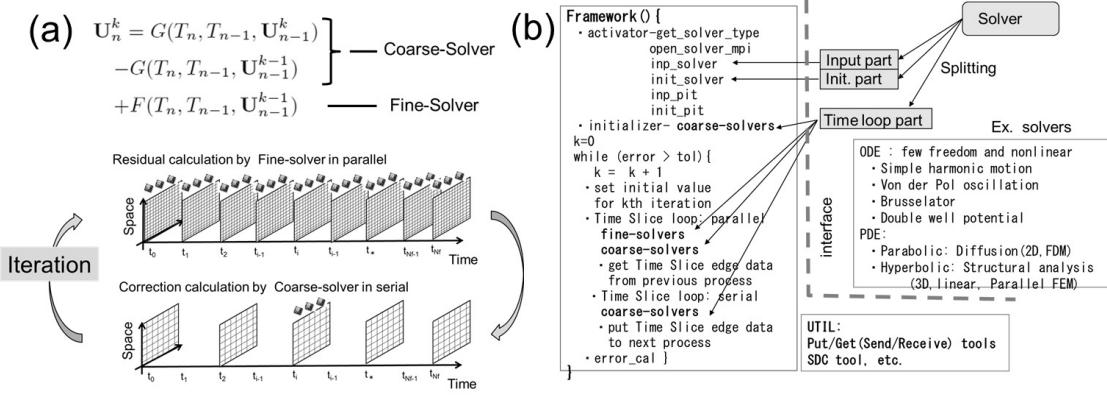


Figure 14.8: *PinT* framework: (a) *Parareal* algorithm, (b) Prototype of a *PinT* framework.

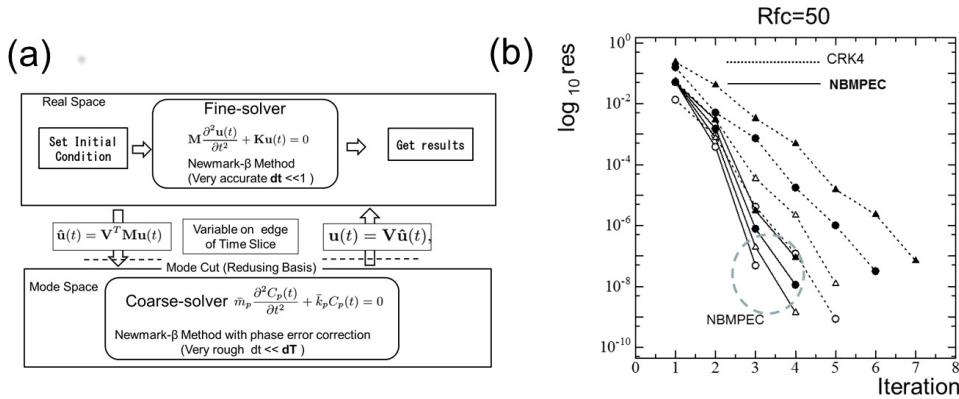


Figure 14.9: *PinT* methods: (a) *Parareal* process based on Reduced basis methods, (b) Convergence of *Parareal* method using *Newmark- β* method with phase error correction(NBMPEC) vs CRK4(Classical 4th-order Runge-Kutta method). Number of cycle = 100 (○), 1,000 (△), 10,000 (●), 100,000 (♠).

process. The Proper Orthogonal Decomposition (POD) algorithm is used to design a sparse model for getting a good compression ratio for assisting the *in-situ* visualization. We also investigated a data sharing approach between the simulation code and the compression code to obtain the best performance. The visualization framework shown in Figure 14.1 is employed for the rendering and interactive visualization. HIVE is a hardware independent parallel visualization system with a scalable rendering capability.

14.4 Schedule and Future Plan

The next big step for the *HIVE* framework is its transition from the prototype stage to a solid and robust production-level system. During the early stage of the development, we have already held an initial presentation through the AICS Open Source Workshop Program to present an overview, and at the same time, to obtain a feedback from the potential users. In the next fiscal year, we are planning to take more advantage of this kind of AICS internal events, as well as other external events, not only for disseminating *HIVE* system, but also to verify and discuss potential future enhancements based on the real-world requirements and needs. Focusing on the AICS internal demands, we are collaborating with some AICS teams involved in the computational science fields for utilizing their real simulation results. In this initial stage, we are using some datasets from the Computational Climate and Computational Fluid Dynamics(CFD) simulations, and have investigated some directions for further enhancements. We are planning to gradually enlarge this group of collaboration teams for

aggregating new functionalities based on real world demands. The loosely coupled modular approach adopted by the HIVE system greatly facilitates the aggregation of new functionalities as native or third-party modules. Our initial priorities are: **Fiber Surface**, **FFV/C[2015:AVRTeam:FFVC]**, **PBVR**, and **PIDX**. In the next fiscal year, we are planning to aggregate the large-scale **PBVR** (Particle-Based Volume Rendering) functionality being developed under the Grant-in-Aid for Scientific Research (KAKENHI) in collaboration with Kyoto University and Kobe University. **PIDX** is a data format for streaming based visualization developed by the University of Utah, which our team has a signed MOU (Memorandum of Understanding). In the next fiscal year, we will accept a graduate student from this university as a student trainee, in order to investigate the potential of this approach on the K computer environment and a future aggregation to *HIVE* system.

14.5 Publications

Chapter 15

Data Assimilation Research Team

15.1 Members

Takemasa Miyoshi (Team Leader)
Koji Terasaki (Research Scientist)
Shigenori Otsuka (Postdoctoral Researcher)
Keiichi Kondo (Postdoctoral Researcher)
Shunji Kotsuki (Postdoctoral Researcher)
Guo-Yuan Lien (Postdoctoral Researcher)
Takumi Honda (Postdoctoral Researcher)
Yasumitsu Maejima (Research Associate)
Africa Perianez Santiago (Research Associate)
Hazuki Arakida (Technical Staff)
Gulanbaier Tuerhong (Technical Staff)
Juan J. Ruiz (Visiting Scientist)
Shinichiro Shima (Visiting Scientist)
Masahiro Sawada (Visiting Scientist)
Shu-Chih Yang (Visiting Scientist)
Stephen G. Penny (Visiting Scientist)
Masaru Kunii (Visiting Scientist)
Kozo Okamoto (Visiting Scientist)
Michiko Otsuka (Visiting Scientist)
Marimo Ohhigashi (Research Assistant)
Yukie Komori (Assistant)
Rie Deguchi (Assistant)

15.2 Research Activities

The Data Assimilation Research Team (DA Team) was launched in October 2012 and is composed of 18 research and technical staff including 7 visiting members as of March 2016. Data assimilation (DA) is a cross-disciplinary science to synergize computer simulations and real-world data, based on statistical mathematics and dynamical systems theory. As computers advance and enable precise simulations, it will become more important to compare the simulations with real-world data. DA Team performs cutting-edge research and development on advanced DA methods and their wide applications, aiming to integrate computer simulations and real-world data in the wisest way. Particularly, DA Team tackles challenging problems of developing efficient and accurate DA systems for “big simulations” with real-world “big data” from various sources including advanced sensors. The specific foci include 1) theoretical and algorithmic developments for efficient and accurate DA, 2) DA methods and applications by taking advantage of the world-leading K computer and “big data” from new advanced sensors, and 3) exploratory new applications of DA in wider simulation fields. These advanced DA studies will enhance simulation capabilities and lead to a better use of high-performance computers including the leading-edge K computer.

In FY2015, we continued on the ongoing data assimilation research in the following aspects: 1) theoretical research on challenging problems, 2) leading research on meteorological applications, 3) optimization of computational algorithms, and 4) exploratory research on wider applications. We also continued close collaborations with several research teams within the AICS Research Division. We have made substantial progress on the following research items:

Theoretical research

- A paper on the discrete Bayesian optimization approach to find optimal ensemble sizes in a multi-model ensemble Kalman filter (EnKF) was published (1 paper published).
- A new local particle filter method to treat non-Gaussian PDF was explored (1 paper submitted).

Leading research on meteorological applications

- We have successfully run the largest-ever ensemble DA experiments with 10,240 samples for the global atmosphere using real observations with the real-case Nonhydrostatic ICosahedral Atmospheric Model (NICAM) (1 paper published, press release on November 11, 2015).
- Impact of localization in ensemble Kalman filter was investigated with 10,240 samples using an intermediate AGCM (1 paper submitted).
- Non-Gaussian statistics in the global atmospheric dynamics were investigated with the 10,240-sample ensemble Kalman filter using an intermediate AGCM.
- Multi-scale data assimilation was investigated based on 10,240-member ensemble Kalman filter using an intermediate AGCM and NICAM.
- A DA system for Advanced Microwave Sounding Unit (AMSU)-A radiance data was developed with NICAM-LETKF.
- A DA system for JAXA’s Global Satellite Mapping of Precipitation (GSMaP) data was developed with NICAM-LETKF.
- A high resolution experiment using NICAM-LETKF system assimilating the conventional, AMSU-A, and GSMAp observations has been performed on the K computer.
- The spatial and inter-channel observation error correlations of AMSU-A were investigated with the NICAM-LETKF system.
- An earlier study on the assimilation of global precipitation data with the low-resolution NCEP GFS model were summarized and published (2 papers published).
- A paper on the new quality control algorithm for the Osaka Phased Array Weather Radar (PAWR) was published (1 paper published).
- The LETKF system with the SCALE model (SCALE-LETKF) was developed and improved in collaboration with Computational Climate Science Research Team. Several new functions, including PAWR assimilation, Himawari-8 satellite data assimilation, relaxation-to-prior-spread (RTPS) method, offline-nested domains, were implemented with the SCALE-LETKF system.
- A near-real-time regional weather analysis and forecast system based on the SCALE-LETKF has been set up and run on the K computer. It has produced weather analyses and 5-day forecasts every 6 hours for more than 9 months.

- The “Big Data Assimilation” experiments for a local severe rainstorm case were performed with the SCALE-LETKF system, assimilating the PAWR data. The results were compared with the previous NHM-LETKF experiments and the nowcasting experiments (1 paper submitted).
- A project-wide paper for “Big Data Assimilation” with the first results of NHM-LETKF experiments was accepted for publication (1 paper accepted).
- Model output statistics have been investigated using machine learning algorithms and deep learning algorithms.
- Convective predictability was investigated by performing breeding experiments. Dependency on the model resolution was investigated.
- A precipitation nowcasting system was developed to take advantage of the dense and frequent PAWR data (1 paper published).
- A space-time extrapolation system for GSMAp with DA was developed using LETKF. The system has been running in real time since January 2016 (1 paper accepted in June 2016).
- A new Himawari-8 observation operator for SCALE-LETKF was developed and tested with a tropical cyclone case.
- Himawari-8 may capture clouds at an earlier stage of convective development before a radar captures large raindrops. This potential advantage was explored.
- The potential usage of Himawari-8 observation for estimating microphysics parameters through DA was explored.
- A 4-dimensional NHM-LETKF system was developed to investigate a fail-safe workflow and the relationship between the DA window length and forecast accuracy.
- A series of DA experiments for a sudden local severe rainstorm case in Kobe on September 11, 2014 was performed to investigate the predictability.

Computational optimization

- Huge-jobs, computing the “Big Data Assimilation” problem with the SCALE-LETKF, as big as near full nodes of the K computer were performed to measure the computational time and scalability of the code.

Wider applications

- A particle filter was applied to a dynamical vegetation model known as the SEIB-DGVM (Spatially-Explicit, Individual-Based Dynamic Global Vegetation Model). Uncertainties in the state variables and the parameters were greatly reduced by assimilating satellite based Leaf Area Index.
- Land surface DA system was developed with SiBUC (Simple Biosphere including Urban Canopy) model and LETKF.
- A global crop calendar was estimated with the satellite-sensed vegetation index (1 paper published).
- Impacts of satellite-based solar radiation data on land surface simulations were estimated (1 paper published).

Several achievements are selected and highlighted in the next section.

15.3 Research Results and Achievements

15.3.1 Big Ensemble Data Assimilation in Numerical Weather Prediction

Taking advantage of the K computer, we have successfully run the largest-ever ensemble DA experiments with 10,240 samples for the global atmosphere with the real-case Nonhydrostatic ICosahedral Atmospheric Model (NICAM) (Miyoshi et al. 2015) using the NICAM-LETKF (Local Ensemble Transform Kalman Filter) system developed in FY2014 (Terasaki et al. 2015). This research result was highlighted by RIKEN Press Release on November 11, 2015. The samples size of 10,240 is two orders of magnitude larger than the typical choice of about 100. The computational cost for the ensemble DA is proportional to the cubic power of the sample size, and active collaborations with AICS Large-scale Parallel Numerical Computing Technology Research Team (PI: Dr. T. Imamura) played an essential role in accelerating the computation by a factor of 8 using an eigenvalue solver “EigenExa” optimized for the K computer. The large sample size reduces the sampling error (Fig.

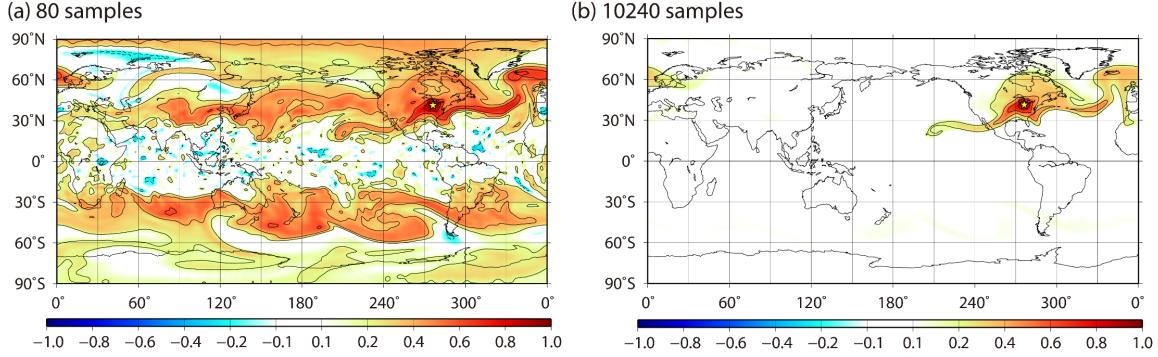


Figure 15.1: Horizontal maps of real atmosphere autocorrelations with (a) 80 samples and (b) 10240 samples. Adopted from Fig. 5 of Miyoshi et al. (2015), ©IEEE 2015.

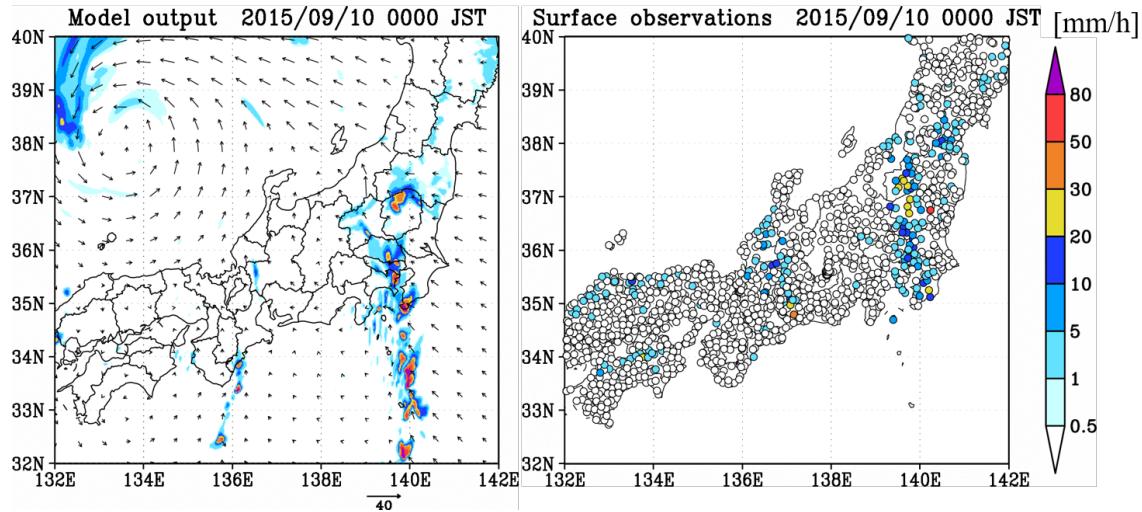


Figure 15.2: 1-hour accumulated rainfall amount at 0000 JST, September 10, 2015, when River Kinugawa had flooded and caused severe disasters, from a model simulation initiated at 1200 JST, September 8 (left) and surface rain-gauge observation (right) from NTT Docomo environmental sensor network.

15.1), and we discovered potential long-range correlations up to about 7,000 km (Fig. 15.1 b). This suggests potential use of faraway observations to improve numerical weather prediction (NWP), although we usually assume that the impact of observations is limited within a range up to 4,000 km or so. The large ensemble DA experiments would provide fundamental datasets to improve our knowledge on the flow-dependent error statistics including non-Gaussian and multi-scale structures, and would help develop advanced approaches for non-Gaussian and multi-scale DA, the topics at the center of theoretical DA research.

15.3.2 Near-real-time Implementation of Regional Numerical Weather Prediction (NWP)

The LETKF-based DA system has been newly developed with the regional NWP model ‘‘SCALE’’ in collaboration with Computational Climate Science Research Team (PI: Dr. H. Tomita). Real-world observation data are available from the US National Centers for Environmental Prediction (NCEP) in near real time, by about 3-hour delay from the real time. We have implemented the near-real-time 5-day NWP using the SCALE-LETKF system, and have been running continuously from May 7, 2015. Figure 15.2 indicates an example of a forecast in a case of the worst disaster in 2015 by flooding of River Kinugawa, Tochigi, Japan. The line-type system is well simulated. By continuously running SCALE-LETKF analysis and forecasts, we accumulate experiences and verification samples, which will be very useful for further improvements of DA and NWP system developments.

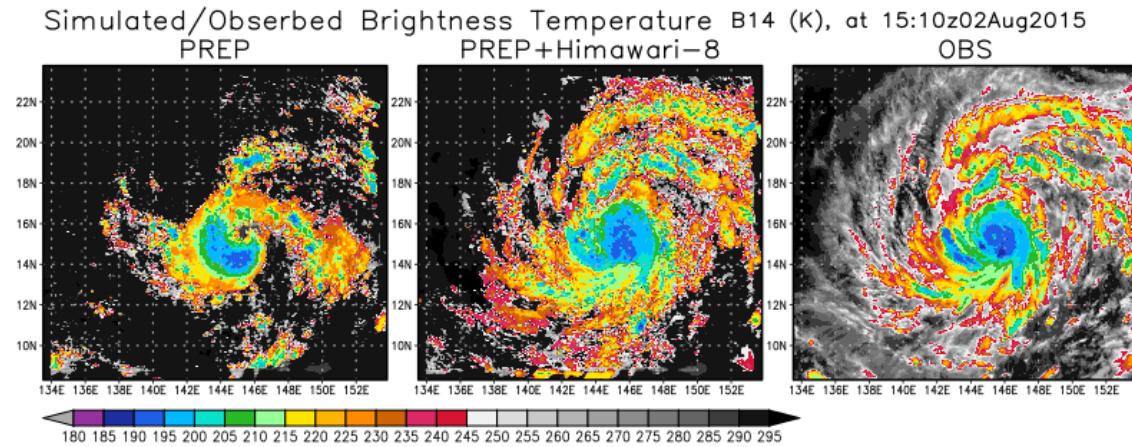


Figure 15.3: Geostationary satellite Himawari-8 band-14 brightness temperature (K) at 1510 UTC, August 2, 2015 (right) and simulated images with/without assimilating Himawari-8 data (center/left).

15.3.3 Effective Use of Satellite Big Data

Developing DA methods for effective use of various observation data is important. Satellite remote sensing provides a relatively uniform coverage of a broad area of the globe, and plays an essential role in NWP. However, the observed quantities are not direct to the prognostic variables of NWP models, and the data volume and variety keep increasing rapidly as sensor technology advances. Therefore, satellite DA requires algorithmic and methodological developments, and has become a major field in meteorology. We have been actively working on the research on effective assimilation of satellite-based precipitation data including the GPM (global precipitation measuring mission) core satellite data in collaboration with JAXA (Japan Aerospace Exploration Agency), microwave radiances from low-earth orbit satellites, and big data from the new-generation geostationary satellite Himawari-8 in collaboration with JMA (Japan Meteorological Agency) Meteorological Satellite Center and Meteorological Research Institute. Figure 15.3 shows an example of the impact of assimilating Himawari-8 infrared radiances in the case of Typhoon Soudelor 2015. The outer rainbands north of the vortex are greatly enhanced by assimilating Himawari-8 and becomes much closer to the observation.

15.3.4 New Application with Individual-based Dynamical Vegetation Model

As a new explorative application of DA, we have been working on land-surface and vegetation applications. We have successfully applied DA to the Spatially-Explicit Individual-Based Dynamic Global Vegetation Model (SEIB-DGVM) for the first time. The individual-based model simulates individual plants explicitly; some trees may die, while new trees may emerge. Therefore, the model prognostic variables such as tree height and root depth may change time to time and place to place. Typical DA methods assume the phase space to be predefined and never change. Alternatively, we applied a SIR (Sequential Importance Resampling) particle filter. The results were encouraging. Figure 15.4 shows the results with the real satellite-based MODIS (Moderate Resolution Imaging Spectroradiometer) LAI (Leaf Area Index) observations at a single location. It is not surprising that the uncertainties of the simulated LAI are reduced significantly (Fig. 15.4 a), but we also find that the uncertainties of unobserved model parameters and variables are reduced significantly (Fig. 15.4 b, c).

15.3.5 Theoretical and algorithmic developments

Theoretical DA research is an important part of DA Team's scope. Here, we highlight some of the major theoretical developments in FY2015. A paper on the discrete Bayesian optimization approach to find optimal ensemble sizes in a multi-model ensemble Kalman filter (EnKF) was published (Otsuka and Miyoshi 2015). Potential impact of assimilation order of observations in serial EnKF was investigated (Kotsuki et al., manuscript in preparation). A new local particle filter method

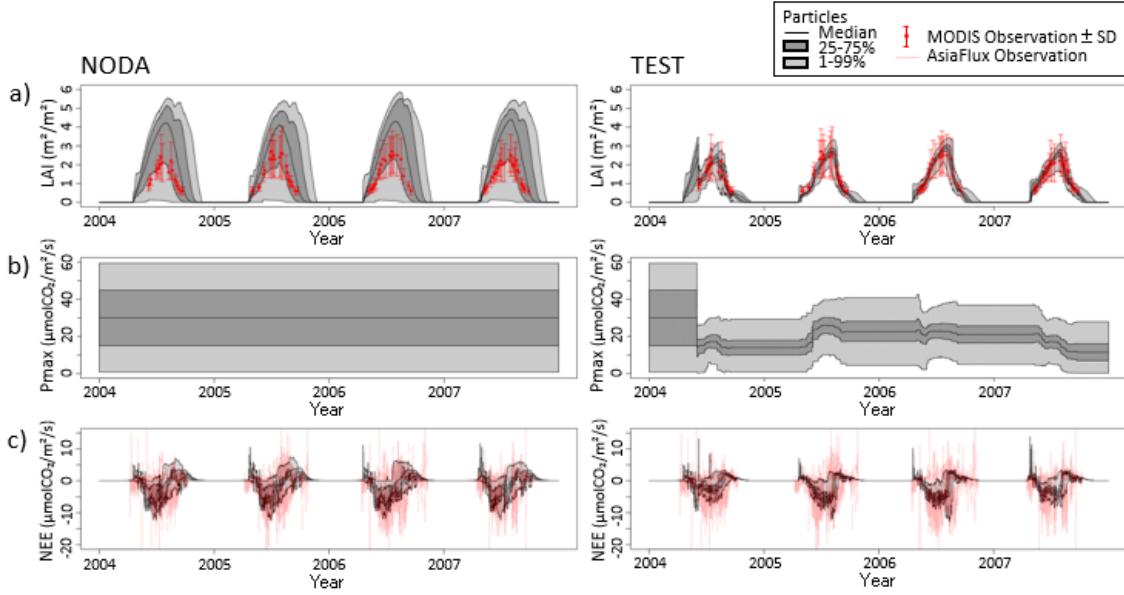


Figure 15.4: 4-year time series of (a) LAI, (b) a model parameter (maximum photosynthesis rate), and (c) a model variable (net ecosystem exchange) for the cases without DA (left) and with DA (right). Adopted from Arakida et al. (2016, *Nonlin. Processes Geophys. Discuss.*).

to treat non-Gaussian PDF was explored (Penny and Miyoshi 2016, *Nonlin. Processes Geophys. Discuss.*).

15.4 Schedule and Future Plan

In FY2015, DA team had one additional full-time research staff, and the team will grow further in FY2016. DA team has been exploring various aspects of DA including theoretical problems, meteorological applications, and wider applications, with large-scale computing to fully utilize “Big Data.” We have very strong projects in weather forecast applications supported by the JST CREST program, JAXA, and MEXT FLAGSHIP 2020 project. We will continue our efforts in each topic in FY2016.

“Big Data Assimilation” (BDA) is one of the major activities that we have developed in FY2015. The prototype system showed promising results, but the computing speed with the K computer did not meet the requirement for real-time implementation. Also, the physical performance for the 30-minute forecast of precipitation patterns needs to be improved. In FY2016, we will continue to work on the development of the BDA system to further improve the computational and physical performances.

In FY2016, we also plan to enhance close collaborations with experts in mathematics, sensor technology, and various application fields. Enhancing collaborations with other AICS Research Teams will also be beneficial.

15.5 Publications

Journal Articles

- [1] Lien, G.-Y., T. Miyoshi, and E. Kalnay: “Assimilation of TRMM Multisatellite Precipitation Analysis with a low-resolution NCEP Global Forecast System”, *Mon. Wea. Rev.*, Vol.144, p.643–661, 2016.
- [2] Lien, G.-Y., E. Kalnay, T. Miyoshi, and G. J. Huffman: “Statistical properties of global precipitation in the NCEP GFS model and TMPA observations for data assimilation”, *Mon. Wea. Rev.*, Vol.144, p.663–379, 2016.

- [3] Otsuka, S., G. Tuerhong, R. Kikuchi, Y. Kitano, Y. Taniguchi, J. J. Ruiz, S. Satoh, T. Ushio, and T. Miyoshi: “Precipitation nowcasting with three-dimensional space-time extrapolation of dense and frequent phased-array weather radar observations”, *Weather and Forecasting*, Vol.31, p.329–340, 2016.
- [4] Dillon, M. E., Y. G. Skabar, J. Ruiz, E. Kalnay, E. A. Collini, P. Echevarría, M. Saucedo, T. Miyoshi, and M. Kunii: “Application of the WRF-LETKF Data Assimilation System over Southern South America: Sensitivity to model physics”, *Weather and Forecasting*, Vol.31, p.217–236, 2016.
- [5] Hattori, M., J. Matsumoto, S. Ogino, T. Enomoto, and T. Miyoshi: “The Impact of Additional Radiosonde Observations on the Analysis of Disturbances in the South China Sea during VPREX2010”, *SOLA*, Vol.12, p.75–79, 2016.
- [6] Sluka, T. C., S. G. Penny, E. Kalnay, and T. Miyoshi: “Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation”, *Geophys. Res. Lett.*, Vol.43, p.752–759, 2016.
- [7] Sueyoshi T., K. Saito, S. Miyazaki, J. Mori, T. Ise, H. Arakida, R. Suzuki, A. Sato, Y. Iijima, H. Yabuki, H. Ikawa, T. Ohta, A. Kotani, T. Hajima, H. Sato, T. Yamazaki, A. Sugimoto: “The GRENE-TEA model intercomparison project (GTMIP) Stage 1 forcing data set”, *Earth Syst. Sci. Data*, Vol.8, p.1–14, 2016.
- [8] Miyoshi, T., M. Kunii, J. Ruiz, G.-Y. Lien, S. Satoh, T. Ushio, K. Bessho, H. Seko, H. Tomita, and Y. Ishikawa: “Big Data Assimilation” Revolutionizing Severe Weather Prediction”, *Bulletin of the American Meteorological Society*, 2016.
- [9] Otsuka, M., M. Kunii, H. Seko, K. Shimoji, M. Hayashi, K. Yamashita: “Assimilation Experiments of MTSAT Rapid Scan Atmospheric Motion Vectors on a Heavy Rainfall Event”, *Journal of the Meteorological Society of Japan. Ser. II*, Vol.93, 4, p.459–475, 2015.
- [10] J. Ruiz and M. Pulido: “Parameter Estimation Using Ensemble Based Data Assimilation in the Presence of Model Error”, *Mon. Wea. Rev.*, Vol.143, p.1568–1582, 2015.
- [11] Seko, H., M. Kunii, S. Yokota, T. Tsuyuki, and T. Miyoshi: “Ensemble experiments using a nested LETKF system to reproduce intense vortices associated with tornadoes of 6 May 2012 in Japan”, *Progress in Earth and Planetary Science*, Vol.2:42, 2015.
- [12] Kotsuki S., H. Takenaka, K. Tanaka, A. Higuchi and T. Miyoshi: “1-km-resolution land surface analysis over Japan: Impact of satellite-derived solar radiation”, *Hydrological Research Letters*, Vol.9(1), p.14–19, 2015.
- [13] Sawada, M., T. Sakai, T. Iwasaki, H. Seko, K. Saito and T. Miyoshi: “Assimilating high-resolution winds from a Doppler lidar using an ensemble Kalman filter with lateral boundary adjustment”, *Tellus A* 2015, Vol.67, p.23473, 2015.
- [14] Chang, C.-C., S.-C. Yang and C. Keppenne: “Applications of the mean re-recentering scheme to improve typhoon track prediction: A case study of typhoon Nanmadol (2011)”, *JMSJ Ser. II*, Vol.92, 6, p.559–584, 2015.
- [15] Miyoshi, Takemasa; Kondo, Keiichi; Terasaki, Koji: “Big Ensemble Data Assimilation in Numerical Weather Prediction”, *COMPUTER*, Vol.48, p.15–21, 2015.
- [16] Otsuka, Shigenori; Miyoshi, Takemasa: “A Bayesian Optimization Approach to Multimodel Ensemble Kalman Filter with a Low-Order Model”, *MONTHLY WEATHER REVIEW*, Vol.143, p.2001–2012, 2015.
- [17] Kotsuki, S.; Tanaka, K.: “SACRA - a method for the estimation of global high-resolution crop calendars from a satellite-sensed NDVI”, *HYDROLOGY AND EARTH SYSTEM SCIENCES*, Vol.19, p.4441–4461, 2015.

- [18] Miyazaki, S.; Saito, K.; Mori, J.; Yamazaki, T.; Ise, T.; Arakida, H.; Hajima, T.; Iijima, Y.; Machiya, H.; Sueyoshi, T.; Yabuki, H.; Burke, E. J.; Hosaka, M.; Ichii, K.; Ikawa, H.; Ito, A.; Kotani, A.; Matsuura, Y.; Niwano, M.; Nitta, T.; Oishi, R.; Ohta, T.; Park, H.; Sasai, T.; Sato, A.; Sato, H.; Sugimoto, A.; Suzuki, R.; Tanaka, K.; Yamaguchi, S.; Yoshimura, K.: “The GRENE-TEA model intercomparison project (GTMIP): overview and experiment protocol for Stage 1”, Geoscientific Model Development, Vol.8, p.2841–2856, 2015.
- [19] Ruiz, Juan J.; Miyoshi, Takemasa; Satoh, Shinsuke; Ushio, Tomoo: “A Quality Control Algorithm for the Osaka Phased Array Weather Radar”, SOLA, Vol.48–52, 2015.

Conference Papers

None.

Invited Talks

- [1] T. Miyoshi, “Data assimilation toward big data and post-peta-scale supercomputing: a personal perspective”, International Workshop on Theoretical Aspects of Ensemble Data Assimilation for the Earth System, Les Houches, France, 5–10 April 2015.
- [2] T. Miyoshi, “Numerical Weather Prediction: Chaos, Predictability, and Data Assimilation”, ICTS Data Assimilation Program organized by CMI, Hong Kong Baptist University, 2015/4/20.
- [3] T. Miyoshi, “Data Assimilation Toward Big Data and Post-Peta-Scale Supercomputing: A Personal Perspective”, ICTS Data Assimilation Program organized by CMI, Hong Kong Baptist University, 2015/4/20.
- [4] 三好建正, “スパコンを使った最先端の天気予報研究～「京」でゲリラ豪雨に挑む～”, 第27回日本学術振興会産学協力研究委員会「水の先進理工学」第183委員会研究会, 東京, 4/22/2015.
- [5] K. Terasaki, “Applying the Local Transform Ensemble Kalman Filter to the non-hydrostatic atmospheric model NICAM”, AICS Café, Kobe, 15 May 2015.
- [6] T. Miyoshi, “ポスト「京」による天気予報革命”, HPCS2015 (ハイパフォーマンスコンピューティングと計算科学シンポジウム), 東京, 5/19/2015.
- [7] T. Miyoshi, “「ビッグデータ同化」でゲリラ豪雨に挑む”, 第43回メソ気象研究会, 東京, 5/20/2015.
- [8] T. Miyoshi, “ビッグデータ同化による天気予報革命”, 日本気象学会春季大会, 公開気象講演会, つくば, 5/24/2015.
- [9] T. Miyoshi, “‘Big Data Assimilation’ Revolutionizing Severe Weather Forecasting”, JpGU Annual Meeting (Japan Geoscience Union), Makuhari, Japan, 5/26/2015.
- [10] T. Miyoshi, “Satellite Data Assimilation: A Perspective”, JpGU Annual Meeting (Japan Geoscience Union), Makuhari, Japan, 5/28/2015.
- [11] Takemasa Miyoshi, “Covariance Localization and Inflation”, 14th CAS-TWAS-WMO Forum Data Assimilation Summer School, Beijing, 2 July 2015.
- [12] Takemasa Miyoshi, “Data Assimilation toward Big Data and Post-peta-scale Supercomputing: A Personal Perspective”, 14th CAS-TWAS-WMO Forum Coupled Data Assimilation Symposium, Beijing, 7 July 2015.
- [13] Shu-Chih Yang, “Adjusting ensemble for EnKF assimilation: applicatiops to severe weather prediction”, 14th CAS-TWAS-WMO Forum Coupled Data Assimilation Symposium, Beijing, 7 July 2015.
- [14] T. Miyoshi, “‘Big Data Assimilation’ revolutionizing numerical weather prediction”, 2nd Mini-Symposium on Computations, Brains and Machine, Wako, Japan, 14 July 2015.

- [15] T. Miyoshi, “Satellite Data Assimilation: A perspective”, the Weather-Chaos meeting, College Park, USA, 8/21/2015.
- [16] T. Miyoshi, “Data Assimilation toward Big Data and Post-peta-scale Supercomputing: A Personal Perspective”, OneNOAA Science Seminar Series, College Park, USA, 8/24/2015.
- [17] Koji Terasaki, “Some aspects of the computation of the 3D normal-mode functions”, MODES workshop, 26–28 August 2015.
- [18] T. Miyoshi, “Numerical Weather Prediction: Chaos, Predictability and Data Assimilation”, iTHESS Colloquium, Wako, Japan, 15 September 2015.
- [19] T. Miyoshi, “Big Data, Supercomputing, and Data Assimilation”, Atmospheric and Oceanic Science Departmental Seminar Series, Maryland, USA, 7–9 October 2015.
- [20] 三好建正, “「ビッグデータ同化」でゲリラ豪雨に挑む”, サイエンティフィック・システム研究会 科学技術計算分科会 2015年度会合, 神戸, 2015/10/28.
- [21] 三好 建正, “(基調講演) 京, ビッグデータ, ポスト京へ”, 日本気象学会2015年度秋季大会シンポジウム「スーパーコンピューティングと気象学」, 京都, 2015/10/28–30.
- [22] T. Miyoshi, “Big Data, Supercomputing, and Data Assimilation”, KISTI, Daejeon, Korea, November 4, 2015.
- [23] T. Miyoshi, “Big Data Assimilation Revolutionizing Numerical Weather Prediction”, Workshop Big Data and Environment, Buenos Aires, Argentina, 10–13 November 2015.
- [24] T. Miyoshi, “Big Data Assimilation’ Revolutionizing Weather Prediction”, NCU, Taiwan, 24 November 2015.
- [25] T. Miyoshi, “Big Data Assimilation’ Revolutionizing Weather Prediction”, CWB, Taiwan, 25 November 2015.
- [26] T. Miyoshi, “Big Data Assimilation’ Revolutionizing Weather Prediction”, NTU, Taiwan, 26 November 2015.
- [27] 三好建正, “天気予報、データ同化、気象防災におけるAI活用の展望”, 理研シンポジウム「理研科学者が拓く“AI (革新知能)”」, 和光, 2015/11/28.
- [28] T. Miyoshi, Keynote “Big Data Assimilation’ revolutionizing numerical weather prediction”, Symposium on ”The Frontier of Numerical Weather Prediction”, Beijing, China, 11 December 2015.
- [29] 三好建正, “「ビッグデータ同化」でゲリラ豪雨に挑む”, 最新気象レーダが拓く安心・安全な社会2015, 京都, 2015/12/24.
- [30] T. Miyoshi, “Big Data Assimilation’ Revolutionizing Numerical Weather Prediction”, International Workshop on the Variations of East Asian Monsoon, Guangzhou, China, 9–11 January 2016.
- [31] Takemasa Miyoshi, “Ensemble-based Data Assimilation of TRMM/GPM Precipitation Measurements”, JAXA Joint PI meeting of Global Environment Observation Mission 2015 (PMM panel session), Tokyo, 20 January 2016.
- [32] 三好 建正, “「ビッグデータ同化」でゲリラ豪雨に挑む”, ソフトウェアジャパン2016 ITフォーラム : JST科学技術振興機構, 東京, 2016/2/4.
- [33] Takemasa Miyoshi, K. Kondo, K. Terasaki, M. Kunii, J. Ruiz, G.-Y. Lien, S. Satoh, T. Ushio, H. Tomita, Y. Ishikawa, K. Bessho and H. Seko, “Big data assimilation’ revolutionizing numerical weather prediction”, Third International Workshop on Tokyo Metropolitan Area Convective Study for Extreme Weather Resilient Cities (WMO/WWRP, TOMACS/RDP), Tokyo, 4–5 February 2016.
- [34] 三好建正, “「ビッグデータ同化」でゲリラ豪雨に挑む”, 気象衛星センター技術談話会, 東京, 2016/2/24.

- [35] Takemasa Miyoshi, "Big Data Assimilation: Toward post-peta-scale supercomputing", Blueprints for Next-Generation Data Assimilation Systems, Boulder, 8–10 March 2016.
- [36] 三好 建正, "全球モデルNICAM及び領域モデルSCALEを使った衛星観測データ同化", GSMAp及び衛星シミュレータ研究集会, 名古屋, 2016/3/17–18.
- [37] 三好建正, "観測ビッグデータを生かすデータ同化の未来", ポスト「京」重点課題④ 第1回シンポジウム「観測ビッグデータを活用した気象と地球環境の予測の高度化」, 東京, 2016/3/29.

Posters and Presentations

- [1] Shima, S., K. Hasegawa, and K. Kusano, "Preliminary numerical study on the cumulus-stratus transition induced by the increase of formation rate of aerosols", EGU, Vienna, Austria, 14 April 2015, European Geosciences Union General Assembly 2015, Vienna, Austria, 12–17 April 2015.
- [2] Shima, S., "Particle-based and probabilistic methods for warm-rain cloud microphysics", Workshop on Eulerian vs. Lagrangian methods for cloud microphysics, Warsaw, April 20–22, 2015.
- [3] Wang X., Yokozawa M., Arakida H., Mori K., Ise T., Kondo M., Uchida M., Kushida K., and Toda M., "Simulating a role of fire disturbance on the soil carbon storage of boreal forest and tundra ecosystems in Alaska". The Fourth International Symposium on the Arctic Research, Toyama, 29 April 2015.
- [4] 寺崎康児, 三好建正, 小楢峻司, "NICAM-LETKFの開発: GSMAp降水量及び衛星輝度温度の同化実験", GSMAp研究会, 2015年5月20日.
- [5] 小楢峻司, 寺崎康児, G.-Y. Lien, 三好建正, E. Kalnay, "NICAM-LETKFを用いたGSMAp降水量の同化実験", 気象学会2015年度春季大会, つくば, 2015年5月24日.
- [6] 近藤圭一・三好建正, "10240メンバーアンサンブルデータ同化実験による大気の非ガウス性の検証", 日本気象学会2015年度春季大会, つくば, 2015年5月24日.
- [7] 大塚成徳, G. Tuerhong, 菊地亮太, 北野慈和, 谷口雄亮, J. Ruiz, 三好建正, "フェーズドアレイ気象レーダを用いた三次元降水時間補外実験", 日本気象学会2015年度春季大会, つくば, 2015年5月24日.
- [8] 前島康光・國井勝・瀬古弘・前田亮太・佐藤香枝・三好建正, "局地的豪雨の予測における稠密な地上観測データ同化の効果", 日本気象学会2015年度春季大会, つくば, 2015年5月24日.
- [9] 寺崎康児, 三好建正, "NICAM-LETKFを使った衛星輝度温度データ同化", 日本気象学会2015年度春季大会, つくば, 2015年5月24日.
- [10] 森淳子, 斎藤和之, 宮崎真, 山崎剛, 伊勢武史, 荒木田葉月, 羽島知洋, 飯島慈裕, 町屋広和, 末吉哲雄, 矢吹裕伯, GTMIPグループ, "北極陸域モデル相互比較とサイト間差異—物理・物質循環過程に着目してー", Japan Geoscience Union Meeting 2015, 幕張, 2015年5月25日.
- [11] Kunii, M., J. Ruiz, G.-Y. Lien, T. Ushio, S. Satoh, K. Bessho, H. Seko, and T. Miyoshi, "30-second-update ensemble Kalman filter experiments using JMA-NHM at a 100-m resolution(水平解像度100m のNHMを用いた30秒サイクルデータ同化実験)", Japan Geoscience Union Meeting 2015, 幕張, 2015年5月26日.
- [12] Wang X., Yokozawa M., Arakida H., Mori K., Ise T., Kondo M., Uchida M., Kushida K., and Toda M., "Simulating soil carbon dynamics in Alaskan terrestrial ecosystems", Japan Geoscience Union Meeting 2015, Chiba, 26 May 2015.
- [13] 荒木田葉月・三好建正・伊勢武史・島伸一郎, "動的植生モデルSEIB-DGVMを用いた葉面積指数に基づくデータ同化実験", Japan Geoscience Union Meeting 2015, 幕張, 2015年5月26日.
- [14] Okamoto, K., A. Perianez, and M. Kazumori, "Prospect of future geostationary satellite observations for numerical weather prediction (数値予報における将来の静止衛星観測の展望)", Japan Geoscience Union Meeting 2015, Makuhari, 24–28 May 2015.

- [15] Dillon, M. E., J. Ruiz, and T. Miyoshi, "Impact of including AIRS observations in the WRF-LETKF system over South America (in Spanish)", Meeting of the Argentinean Meteorological Society, Mar del Plata, Argentina, 26–29 May 2015.
- [16] Saucedo, M., J. Ruiz, and T. Miyoshi, "Impact of increasing the horizontal resolution of a regional forecast and analysis system based on an Ensemble Kalman Filter: a real case study (in Spanish)", Meeting of the Argentinean Meteorological Society, Mar del Plata, Argentina, 26–29 May 2015.
- [17] Yang, S.-C., T. Miyoshi, K. Kondo, and S.-H. Chen, "Improving extreme precipitation prediction in Taiwan with a regional EnKF", 2015 US-Taiwan Extreme precipitation and weather workshop, 30 May 2015.
- [18] Kondo, K. and T. Miyoshi, "Multi-scale Localization Compared with 10240-member Ensemble Kalman Filter Using Real Observations", 14th CAS-TWAS-WMO Forum Coupled Data Assimilation Symposium, Beijing, 5 July 2015.
- [19] Terasaki, K. and T. Miyoshi, "Assimilating AMSU-A brightness temperature in NICAM-LTKF System", 14th CAS-TWAS-WMO Forum Coupled Data Assimilation Symposium, Beijing, 5 July 2015.
- [20] Kotsuki, S., K. Terasaki, G.-Y. Lien, T. Miyoshi, and E. Kalnay, "Ensemble Data Assimilation of GSMAp precipitation into the Nonhydrostatic Global Atmospheric Model NICAM", 14th CAS-TWAS-WMO Forum Coupled Data Assimilation Symposium, Beijing, 7 July 2015.
- [21] 三好建正, “「ビッグデータ同化」の技術革新の創出によるゲリラ豪雨予測の実証”, ビッグデータ応用領域会議, 東京, 2015/08/01.
- [22] Wang, X., M. Yokozawa, H. Arakida, K. Mori, T. Ise, M. Kondo, M. Uchida, K. Kushida, and M. Toda, "Simulating soil carbon dynamics in Alaskan terrestrial ecosystems", AOGS 2015, Singapore, 2–7 August 2015.
- [23] 三好建正, 寺崎康児, 小槻峻司, 大塚成徳, 佐藤正樹, 富田浩文, "GSMApとNICAMを生かした新たな降水プロダクトに向けて", 2015年度第2回GSMAp研究会, 京都, 2015年9月1日.
- [24] 小槻峻司, "NICAM-LETKFを用いたGSMAp降水量の同化実験", 2015年度第2回GSMAp研究会, 京都, 2015年9月1日.
- [25] 寺崎康児, 三好建正, 小槻峻司, "NICAM-LETKFの開発: GSMAp降水量及び衛星輝度温度の同化実験", 2015年度第2回GSMAp研究会, 京都, 2015年9月1日.
- [26] Lee, J.-Y., T. Imamura, K. Sugiyama, T.-M. Sato, and Y. Maejima, "Radiative forcing by the Venus clouds and its effects on atmospheric dynamics in the equatorial region", Comparative Climates of Terrestrial Planets II, Moffett Field, CA, 8–11 Sept 2015.
- [27] Satoh, S., F. Isoda, H. Hanado, T. Ushio, S. Otsuka, and T. Miyoshi, "Vertical Motion and Growth of Precipitation Measured by Dual Phased Array Weather Radar Every 30 Seconds", American Meteorological Society 37th Conference on Radar Meteorology, Norman, OK, 14 September 2015.
- [28] Ruiz, J. K., L. Vidal Sr., P. Maldonado, S. Suarez Ruiz, P. Salio, Y. Garcia Skabar, Y. Garcia Skabar, A. C. Saulo, S. W. Nesbitt, E. Kalnay, and T. Miyoshi, "Local Ensemble Transform Kalman Filter experiments using radar observations: A case study over central Argentina". American Meteorological Society 37th Conference on Radar Meteorology, Norman, OK, 17 September 2015.
- [29] Otsuka, S., G. Tuerhong, J. J. Ruiz, S. Satoh, T. Ushio, and T. Miyoshi, "Three-dimensional precipitation nowcasting with rapid and dense phased array weather radar observations", American Meteorological Society 37th Conference on Radar Meteorology, Norman, OK, 17 September 2015.
- [30] 荒木田葉月・三好建正・伊勢武史・島伸一郎, "動的植生モデルSEIB-DGVMを用いた衛星LAIのデータ同化", 第1回生態系データ同化に関する研究会, 神戸, 2015年9月26日.

- [31] Satoh, S., T. Ushio, H. Tomita, Y. Ishikawa, K. Bessho, H. Seko, and T. Miyoshi, “Innovating ‘Big Data Assimilation’ technology for revolutionizing very-short-range severe weather prediction”, oral, October 6, 2015, CREST Big Data Joint Meeting, Sendai.
- [32] Satoh, S., T. Ushio, H. Tomita, Y. Ishikawa, K. Bessho, H. Seko, and T. Miyoshi, “Innovating ‘Big Data Assimilation’ technology for revolutionizing very-short-range severe weather prediction”, poster, October 6, 2015, CREST Big Data Joint Meeting, Sendai.
- [33] Otsuka, S., S. Kotsuki, and T. Miyoshi, “Global precipitation nowcasting by spatiotemporal extrapolation with data assimilation”, RIKEN-UMD Data Assimilation Conference 2015, Maryland, USA, 7–9 October 2015.
- [34] Arakida, H., T. Miyoshi*, T. Ise, and S. Shima, “Data assimilation experiments with MODIS LAI observations and the dynamic global vegetation model SEIB-DGVM”, RIKEN-UMD Data Assimilation Conference 2015, Maryland, USA, 7–9 October 2015.
- [35] Terasaki, K. and T. Miyoshi, “Assimilating AMSU-A brightness temperature in NICAM-LETKF system”, RIKEN-UMD Data Assimilation Conference 2015, 7–9 October 2015.
- [36] Kotsuki, S., K. Terasaki, G.-Y. Lien, T. Miyoshi, and E. Kalnay, “Ensemble Data Assimilation of GSMAp precipitation into the Nonhydrostatic Global Atmospheric Model NICAM”, RIKEN-UMD Data Assimilation Conference, Maryland, 7 October 2015.
- [37] Honda, T., S. Otsuka, G.-Y. Lien, J. Ruiz, and T. Miyoshi, “Exploring Himawari-8 and phased array weather radar observations simulated from a sub-kilometer numerical model simulation”, RIKEN-UMD Data Assimilation Conference 2015, Maryland, 7 October 2015.
- [38] Kondo, K. and T. Miyoshi, “Impact of localization in ensemble Kalman filter with 10240 members”, RIKEN-UMD Data Assimilation Conference 2015, Maryland, 7–9 October 2015.
- [39] Lien, G.-Y., T. Miyoshi, S. Nishizawa, R. Yoshida, H. Yashiro, S. Adachi, and H. Tomita, “The SCALE-LETKF system and its early applications”, RIKEN-UMD Data Assimilation Conference 2015, Maryland, 7–9 October 2015.
- [40] 寺崎康児, 近藤圭一, 小槻峻司, 三好建正, “NICAM-LETKFの開発状況”, ポスト京サブ課題A研究連絡会, 2015/10/20, 東京.
- [41] 大塚成徳・三好建正, “データ同化手法を用いたGSMAp降水データの時空間補外実験”, 日本気象学会2015年度秋季大会, 京都, 2015年10月28日.
- [42] 近藤圭一・三好建正, “LETKFを用いたDual Localization法と10240メンバーによる誤差相関の比較”, 日本気象学会2015年度秋季大会, 京都, 2015年10月28日.
- [43] 荒木田葉月・三好建正・伊勢武史・島伸一郎, “植生モデルSEIB-DGVMへの観測LAIデータの同化実験”, 日本気象学会2015年度秋季大会, 京都, 2015年10月28日.
- [44] 荒木田葉月・三好建正・伊勢武史・島伸一郎, “個体ベースモデルSEIB-DGVMのデータ同化: その課題と展望”, 統合的陸域圏研究連絡会, 京都, 2015年10月28日.
- [45] 寺崎康児, 三好建正, “NICAM-LETKFシステムを使った高解像度データ同化実験”, 日本気象学会2015年度秋季大会, 京都, 2015年10月28日.
- [46] 小槻峻司, 寺崎康児, G.-Y. Lien, 三好建正, E. Kalnay, “NICAM-LETKFを用いた全球降水マップGSMApデータの同化実験”, 日本気象学会2015年度秋季大会, 京都, 2015年10月28日.
- [47] 前島康光・國井勝・瀬古弘・吳宏堯・佐藤香枝・三好建正, “ビッグデータ同化システムを用いた局地的豪雨発生過程のシミュレーション”, 日本気象学会2015年度秋季大会, 京都, 2015年10月28日.
- [48] Lien, G.-Y., T. Miyoshi, S. Nishizawa, R. Yoshida, and H. Tomita, “Development of the SCALE-LETKF for high-resolution data assimilation”, 日本気象学会2015年度秋季大会, 京都, 2015年10月28日.

- [49] 本田匠, Guo-Yuan Lien, Juan Ruiz, 三好建正, 足立幸穂, 吉田龍二, 富田浩文, “SCALEを用いた豪雨シミュレーションの雲微物理パラメータ依存性”, 日本気象学会2015年度秋季大会, 京都, 2015年10月28日.
- [50] 大塚道子, “気象衛星ひまわり高頻度大気追跡風のデータ同化実験”, 日本気象学会2015年度秋季大会, 京都, 2015年10月28日.
- [51] 寺崎康児, “NICAM-LETKFにおける衛星輝度温度及び降水量同化”, 第17回非静力学モデルに関するワークショップ, 那覇, 2015年12月1–2日.
- [52] 前島康光, “4次元NHM-LETKFによる局地的豪雨の予測精度と計算時間の比較”, 第17回非静力学モデルに関するワークショップ, 沖縄, 2015年12月1–2日.
- [53] 大塚成徳・三好建正, “積雲対流の予測可能性に関する100mメッシュのブリーディング実験”, 第17回非静力学モデルに関するワークショップ, 那覇, 2015年12月2日.
- [54] 三好建正, “準リアルタイム数値天気予報実験から学ぶこと”, 第17回非静力学モデルに関するワークショップ, 沖縄, 2015年12月1–2日.
- [55] 寺崎康児, “NICAM-LETKFの開発状況”, NICAM開発者会議, 箱根, 2015年12月7日.
- [56] Kotsuki, S., K. Terasaki, G.-Y. Lien, T. Miyoshi, and K. Eugenia, “Ensemble Data Assimilation of GSMAp precipitation into the Nonhydrostatic Global Atmospheric Model NICAM”, JAXA Joint PI meeting, Tokyo, 20 January 2016.
- [57] Terasaki, K., “Data assimilation experiments with GSMAp and AMSU-A data using NICAM-LETKF system”, JAXA Joint PI meeting, Tokyo, 20 January 2016.
- [58] 近藤圭一・三好建正, “「京」にて現実大気の世界最大規模アンサンブルデータ同化に成功—天気予報シミュレーションの精度向上へ—”, スーパーコンピュータの今とこれから, 東京, 2016/1/29.
- [59] 近藤圭一・三好建正, “Multi-scale Localization Compared with 10240-member Ensemble Kalman Filter Using Real Observations”, 第6回データ同化ワークショップ, 横浜, 2016年2月1日.
- [60] 大塚成徳・小槻峻司・三好建正, “データ同化手法を用いたGSMAp降水データの時空間補外実験”, 第6回データ同化ワークショップ, 横浜, 2016年2月1日.
- [61] 前島康光, 国井勝, Juan J. Ruiz, 吳宏堯, 佐藤香枝, 三好建正, “孤立積乱雲に伴う局地的豪雨の予測に向けた稠密観測データ同化のインパクト”, 第6回データ同化ワークショップ, 横浜, 2016年2月1日.
- [62] 近藤圭一・三好建正, “「京」にて現実大気の世界最大規模アンサンブルデータ同化に成功—天気予報シミュレーションの精度向上へ—”, 第29回国立研究開発法人理化学研究所と産業界との交流会（理研と親しむ会）, 東京, 2016/2/15.
- [63] Kotsuki, S., K. Terasaki, G.-Y. Lien, T. Miyoshi, and E. Kalnay, “Ensemble Data Assimilation of GSMAp precipitation into the Nonhydrostatic Global Atmospheric Model NICAM”, The 6th AICS International Symposium, Kobe, 22 February 2016.
- [64] Otsuka, S., S. Kotsuki, and T. Miyoshi, “Global precipitation nowcasting by spatiotemporal extrapolation with data assimilation”, The 6th AICS International Symposium, Kobe, 22–23 February 2016.
- [65] Honda, T., G. Y. Lien, S. Nishizawa, R. Yoshida, S. A. Adachi, K. Terasaki, K. Okamoto, H. Tomita, K. Bessho, T. Miyoshi, “Assimilating Himawari-8 Brightness Temperature: A Case Study on Typhoon Soudelor (2015)”, The 6th AICS International Symposium, Kobe, 22–23 February 2016.
- [66] Lien, G.-Y., T. Miyoshi, S. Nishizawa, R. Yoshida, H. Yashiro, H. Tomita, and G. Tuerhong, “A near-real-time weather analysis and forecast system based on the SCALE-LETKF”, The 6th AICS International Symposium, Kobe, 22–23 February 2016.

- [67] Arakida, H., T. Miyoshi, T. Ise, S. Shima, and S. Kotsuki, “Data assimilation experiments with MODIS LAI observations and the dynamic global vegetation model SEIB-DGVM”, The 6th AICS International Symposium, Kobe, 22–23 February 2016.
- [68] Terasaki, K., S. Kotsuki, and Takemasa Miyoshi, “High resolution data assimilation experiments with GSMAp and AMSU-A”, The 6th AICS International Symposium, Kobe, 22 February 2016.
- [69] Kondo, K. and T. Miyoshi, “Multi-scale Localization Compared with 10240-member Ensemble Kalman Filter Using Real Observations”, the 6th AICS International Symposium, Kobe, 22 February 2016.
- [70] Miyoshi, T., “Big Data Assimilation project progress”, CREST International Symposium on Big Data Application, Tokyo, 4–5 March 2016.
- [71] 小楢峻司, 寺崎康児, G.-Y. Lien, 三好建正, E. Kalnay, “NICAM-LETKFを用いた全球降水マップGSMApのアンサンブルデータ同化実験”, GSMAp研究集会, 名古屋, 2016年3月17日.
- [72] Arakida, H., T. Miyoshi, T. Ise, S. Shima, and S. Kotsuki, “Data assimilation experiments with MODIS LAI observations and the dynamic global vegetation model SEIB-DGVM”, The 63rd Annual Meeting of Ecological Society of Japan (ESJ63), Sendai, 22 March 2016.
- [73] 上原均, 八代尚, 寺崎康児, “重点課題④のコデザインについて”, ポスト「京」重点課題④第1回シンポジウム「観測ビッグデータを活用した気象と地球環境の予測の高度化」, 東京, 2016/3/29.

Patents and Deliverables

None.

Chapter 16

Computational Chemistry Research Unit

16.1 Members

Kimihiko Hirao (Unit Leader)
Jong-Won Song (Research Scientist)
Yukio Kawashima (Research Scientist)
Sachiko Kikumoto (Assistant)
Yumeno Kusuhara (Assistant)
Takao Tsuneda (Senior Visiting Scientist)
Bun Chan (Visiting Researcher)

16.2 Research Activities

Electronic structure calculations are now indispensable for understanding chemical phenomena. Density functional theory (DFT), which is simple, conceptual, and applicable to large systems, has emerged as a powerful computational tool to tackle molecular electronic structures. DFT efficiently calculates electronic structure with high accuracy, and its algorithm is suitable for parallel computing. DFT now plays an important role in applications of molecular science running on the K Computer. However, conventional DFT could not describe important properties such as van der Waals interaction and charge-transfer excitation, which are essential for accurate calculations for large-scaled molecular systems.

We have developed long-range corrected density functional theory (LC-DFT), which overcomes the drawbacks of conventional DFT mentioned above. LC-DFT also succeeded in describing induced/response properties. Recently, we found that LC-DFT obtains accurate energies of highest occupied molecular orbital (HOMO) and lowest occupied molecular orbital (HOMO). This indicates that prediction of chemical reactions can be done by LC-DFT calculations. The development of LC-DFT had a large impact in theoretical chemistry and the number of researches based on LC-DFT is growing intensively. However, LC-DFT has difficulty in describing photochemical reactions. Photochemical process includes avoided crossing among electronic states, spin-forbidden transitions, and states with high and low spins. These properties cannot be calculated accurately by LC-DFT. Moreover, the HF exact exchange, which remedies the shortcoming of the exchange functional in conventional DFT, requires large computational effort for real systems, which is the bottleneck for large-scale calculation.

The objective of our project is to establish LC-DFT to be a standard electronic structure theory by expanding its capability. We feature new developments of photo- and electro-chemical reaction theories and its high-speed computational algorithms for using on next-generation supercomputer “K”, and the elucidations of significant reaction mechanisms and the designs of new functional

materials in photo and electrochemistry. We also aim to increase reliability of electronic structure calculation by improving the accuracy of LC-DFT.

16.3 Research Results and Achievements

16.3.1 Long-Range Corrected Density Functional Theory with Linearly-Scaled Hartree-Fock Exchange Using a Two-Gaussian Operator

Hybrid density functionals have become a main quantum chemical tool for the calculation of energies and properties of molecular systems since its development in 1993. The Hartree-Fock (HF) exchange introduced in hybrid functionals remedy the shortcomings of density functional theory (DFT). Development of long-range corrected hybrid scheme for density functional theory, which follows a decade later, widened the applicability of the hybrid functional further. The introduction of the error function HF exchange operator in DFT calculation increased performance on orbital energy, excitation energy, non-linear optical property, barrier height, and so on. Nevertheless, the high cost associated with the evaluation of HF exchange integrals remain as a bottleneck for the broader and more active applications of hybrid functionals to large molecular and periodic systems. We proposed a very simple yet efficient method for the computation of long-range corrected hybrid scheme. It uses a modified two-Gaussian attenuating operator instead of the error function for the long-range HF exchange integral. As a result, the two-Gaussian HF operator, which mimics the shape of the error function operator, reduces computational time dramatically (e.g., about 14 times acceleration in C diamond calculation using periodic boundary condition) and enables lower scaling with system size, while maintaining the improved features of the long-range corrected density functional theory.

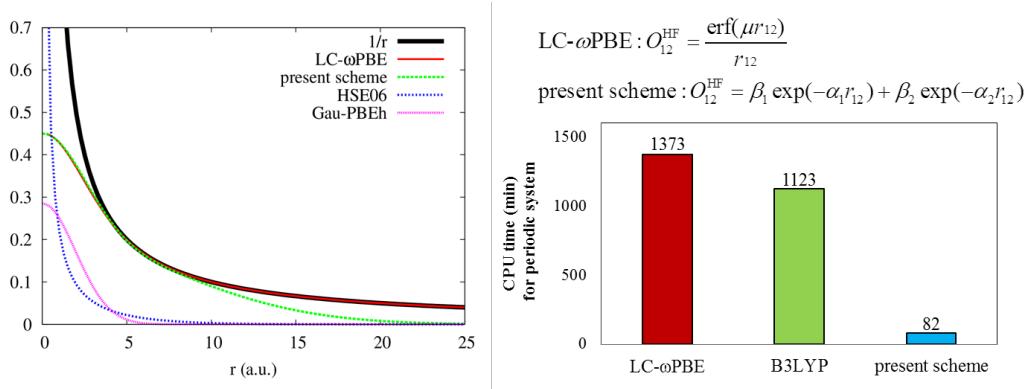


Figure 16.1: Functions of the HF exchange operator (left) and CPU time comparison of the new proposed method vs conventional methods (right).

16.3.2 Development of efficient algorithm for Gaussian Hartree-Fock exchange operator

Hybrid density functionals have succeeded in increasing the accuracy of density functional theory (DFT) by including Hartree-Fock (HF) exchange. Nowadays, most electronic structure calculations for isolated systems employ hybrid density functionals. Inclusion of HF exchange lead to improvement of electronic structure calculations for extended systems as well. The accuracy of bandgap, which is a major factor determining electronic conductivity in extended systems, increased dramatically. However, the computational cost of HF exchange inhibits the use of hybrid density functionals for electronic structure calculations for extended or large-scale systems. We previously developed an efficient screened hybrid functional called Gaussian-Perdew-Burke-Ernzerhof (Gau-PBE) [Song et al., J. Chem. Phys. 135, 071103 (2011)], which is characterized by the usage of a Gaussian function as a modified Coulomb potential for the Hartree-Fock (HF) exchange. We found that the adoption of a Gaussian HF exchange operator considerably decreases the calculation time cost of periodic systems while improving the reproducibility of the bandgaps of semiconductors compared to previously developed well-known methods. We present a distance-based screening scheme here that

is tailored for the Gaussian HF exchange integral that utilizes multipole expansion for the Gaussian two-electron integrals. We found the new multipole screening scheme saves the computational cost for the HF exchange integration by efficiently decreasing the number of integrals of the near field region without incurring substantial changes in total energy. In our assessment on the periodic systems of seven semiconductors, the Gau-PBE hybrid functional with a new screening scheme is 1.2 times faster than our previous implementation, and 2.1 times faster than the well-known HSE06 method.

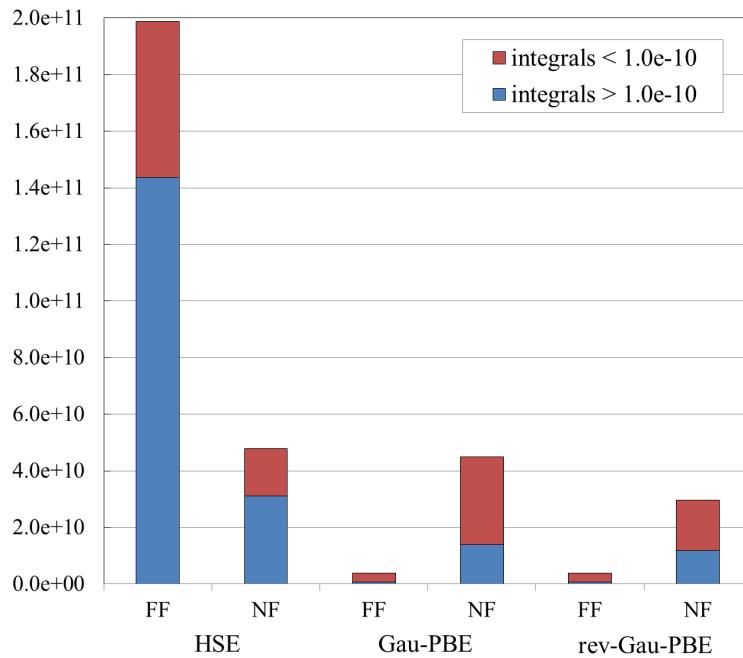


Figure 16.2: The numbers of calculated HF exchange integrals of HSE, Gau-PBE, and rev-Gau-PBE in C diamond.

16.3.3 Critical Assessment of Same-Spin Correlation in OP Correlation Functional

In the present study, we have investigated two significant features of the OP correlation functional, namely the incorporation of the exchange functional into itself, and the inclusion of only opposite-spin (OS) effects. To explore the latter feature, we have compared OP with B95 and a new functional introduced in the present study – the OPB method that combines OP with the same-spin (SS) component of B95. In general, we find that B95 and OPB perform comparably. Our comparisons of the various Density Functional Theory (DFT) procedures suggest that the incorporation of a meta-GGA (e.g., TPSS) into OP and OPB does not necessarily lead to a chemically more accurate procedure than the use of a related GGA (e.g., PBE). An important finding is the more notable (and somewhat more consistent) improvement in performance with the incorporation of SS correlation, particularly for longer-range chemical properties. Nonetheless, on average across our test sets of over 800 systems, the difference between the performances of OP versus B95 or OPB is not exceedingly large. By drawing a parallel between these DFT methods and the wavefunction scaled-MP2-type methods, we reason that one can further develop the OP functional, and perhaps a wider range of correlation functionals by combining it with the technique of range separation.

16.3.4 Probing Fullerene Formation by Supercomputers

Fullerenes are nano-sized carbon materials studied intensively due to its wide applicability such as a silver bullet for HIV, cosmetics, and superconductive devices. However, its precise value of "heat of formation", which is a fundamental property to understand how materials form and change,

was not yet known. We have carried out large-scale computational quantum chemistry calculations on the K computer with NTChem software to obtain heats of formation for C₆₀ and some higher fullerenes with the DSD-PBE-PBE/cc-pVQZ double-hybrid density functional theory method. Our best estimated values are 2520.0 ± 20.7 (C₆₀), 2683.4 ± 17.7 (C₇₀), 2862.0 ± 18.5 (C₇₆), 2878.8 ± 13.3 (C₇₈), 2946.4 ± 14.5 (C₈₄), 3067.3 ± 15.4 (C₉₀), 3156.6 ± 16.2 (C₉₆), 3967.7 ± 33.4 (C₁₈₀), 4364 (C₂₄₀) and 5415 (C₃₂₀) kJ/mol. Using the convergence behavior for the calculated per-atom heats of formation, we obtained the formula $\Delta_f H$ per carbon = $722n^{-0.72} + 5.2$ kJ/mol (n = the number of carbon atoms), which enables an estimation of $\Delta_f H$ for higher fullerenes more generally. A slow convergence to the graphene limit is observed, which we attribute to the relatively small proportion of fullerene carbons that are in “low-strain” regions. We further propose that it would take tens, if not hundreds, of thousands of carbons for a fullerene to roughly approach the limit. Such a distinction may be a contributing factor to the discrete properties between the two types of nanomaterials. During the course of our study, we also observe a fairly reliable means for the theoretical calculation of heats of formation for medium-sized fullerenes. This involves the use of isodesmic-type reactions with fullerenes of similar sizes to provide a good balance of the chemistry and to minimize the use of accompanying species.

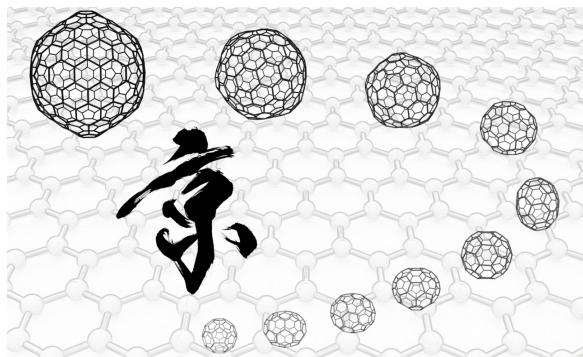


Figure 16.3: Illustration of fullerene systems calculated on the K computer.

16.4 Schedule and Future Plan

We will continue our effort to expand the capabilities of LC-DFT. First, we will develop the order-N calculation algorithm of LC-DFT to calculate large molecular systems quantitatively with much less computational time. We have gained insight on how to reduce the time-consuming exact exchange calculation and we will apply our knowledge to the algorithm development. We will then apply this algorithm to excited state calculations on time-dependent density functional theory (TDDFT). We will also develop open-shell spin-orbit TDDFT to calculate molecular systems including metal atoms. Furthermore, we will develop a new method to calculate the nonadiabatic coupling among different electronic states, and carry out nonadiabatic coupling calculations based on TDDFT to reproduce photochemical reactions comprehensively. We are also planning to develop methods for solid-state calculations.

16.5 Publications

Chapter 17

Computational Disaster Mitigation and Reduction Research Unit

17.1 Members

Muneo Hori (Unit Leader)

Hideyuki Ohtani (Research Scientist)

Jian Chen (Research Scientist)

Kohei Fujita (Research Scientist)

17.2 Research Activities

Computational disaster mitigation and reduction research unit is aimed at developing advanced large-scale numerical simulation of natural disasters such as an earthquake, tsunami and heavy rain, for Kobe City and other urban areas in Hyogo Prefecture. Besides for the construction of a sophisticated urban area model and the development of new numerical codes, the unit seeks to be a bridge between Science and Local Government for the disaster mitigation and reduction.

Our research unit addressed the following research objects in this fiscal year:

- 1) Construction of next generation hazard map for Kobe City. For two scenarios of Nankai Trough Earthquake presumed by National Government, the unit constructed next generation hazard map for Kobe City. Unlike conventional hazard map, the next generation hazard map is based on large scale numerical simulation of the physical processes of seismic wave propagation and seismic structural responses. The map achieves highest spatial resolution as well as higher rationality. Urban area models of Kobe City for underground and man-made structures are used in the physical simulation, and it is these models that determine the quality of the simulation; a more accurate model makes a more reliable estimation of earthquake hazard and disaster. Computational Disaster Mitigation and Reduction Research Unit develops a system that automatically constructs the urban area models using available data resources such as commercial Geographical Information System and governmental data. The system is designed to have high flexibility and expandability so that it can be used for various urban area to which suitable data resources are available.
- 2) Development of system for liquefaction occurrence estimation. Due to complicated processes, the liquefaction occurrence ought to be estimated by using empirical equations that use engineering indices of soil layers of a target site. Computational Disaster Mitigation and Reduction Research Unit constructs a system that estimates the liquefaction occurrence using numerical simulation of the processes of liquefaction, i.e. coupling of soil deformation and underground water flow induced by ground motion. The system succeeds to automatically generate an underground model for 10,000 sites in Kobe City using available boring data and to simultaneously estimate the liquefaction occurrence for a measured or presumed ground motion.
- 3) Estimation of lifeline damage induced by earthquake. Lifeline, or a network of buried pipelines of energy, water and communication, is an essential part of an urban area. Lifeline is damaged by ground deformation or strain, rather than ground motion such as acceleration or velocity, and hence

three-dimensional analysis of underground structure responses for a given earthquake is needed to estimate lifeline damage induced by an earthquake. Using large scale finite element method, Computational Disaster Mitigation and Reduction Research Unit is developing a system that estimate lifeline damage, numerically simulating ground deformation of a target area.

17.3 Research Results and Achievements

17.3.1 Construction of next generation hazard map for Kobe City

National Government has studied a possible scenario of Nankai Trough Earthquake. Two large and extremely large scenarios have been announced for preparation of earthquake disaster. A distribution of possible ground motion index (such as seismic index or peak ground acceleration/velocity) and residential building damage is reported; this distribution is computed by using conventional method of empirical equations. The distribution is summarized as a hazard map for relatively large "mesh" or "grid" of more than 500 m or town-wise number of damaged houses.

Computational Disaster Mitigation and Reduction Research Unit constructs next generation hazard map for Kobe City, using the two presumed scenarios of National Government; see Fig. 17.1. Unlike the conventional hazard map, the next generation hazard map takes advantage of large scale numerical simulation of the two physical processes, namely, the seismic wave propagation/amplification and the seismic structural responses. Numerical analysis methods for the physical processes are tuned for K computer, and the unit develops underground models for the seismic wave propagation/amplification process and a set of residential buildings the number of which exceeds 100,000 for Kobe City.

The next generation hazard map has much higher spatial resolution of possible earthquake hazard and damage; for instance, the next generation hazard map is able to compute the seismic response of a residential house by inputting ground motion of the building's site. The map is more rational than the hazard map since it is based on the simulation of the physical processes. Many ground motions could share the same ground motion index, but they surely produce different seismic response for various structures. This is a reason why the simulation of the physical processes of the seismic wave propagation/amplification and the seismic structural responses is made for seismic design.

We have to mention that while the next generation hazard map has higher spatial resolution and rationality, it is the urban model that determines the quality of the hazard map. That is, a more accurate urban model of underground and man-made structures makes a more reliable estimation of earthquake hazard and disaster. The data that are used to construct the urban area model has limited quality and quantity, and hence the models are not perfect. A possible range of earthquake hazard and damage could be estimated by ensemble simulation that uses a numerous set of urban area models which are randomly generated in view of the uncertainty of the data.

The automated construction of an urban area model is a key element of the next generation hazard map. Computational Disaster Mitigation and Reduction Research Unit develops a system that automatically constructs the urban area models, making use of all available data resources such as commercial Geographical Information System and governmental data managed by various sections of local government; see Fig. 17.2. The system is designed to have high flexibility and expandability, as it takes full advantage of object oriented and aspect oriented programming; the use of aspect oriented programming is especially important since the system has to deal with data of similar data structures but distinct attributes. Hence, it can be used for other urban areas to which suitable data resources are available.

17.3.2 Development of system for liquefaction occurrence estimation

Liquefaction is occurred through complicated processes of soil and underground water which are influenced by non-linear mechanical properties of soil, permeability of soil, and underground water pressure. When the coupling of soil deformation and underground water flow increases underground water pressure locally, liquefaction is induced. Hence, the liquefaction occurrence ought to be estimated by using simple empirical equations for a large urban area. The equations use engineering indices of soil layers of a target site which are obtained by analyzing boring hole data.

Computational Disaster Mitigation and Reduction Research Unit constructs a system that estimates the liquefaction occurrence by using numerical simulation of the complicated processes of liquefaction. The simulation uses a model of a target site which consists of a few soil layers and

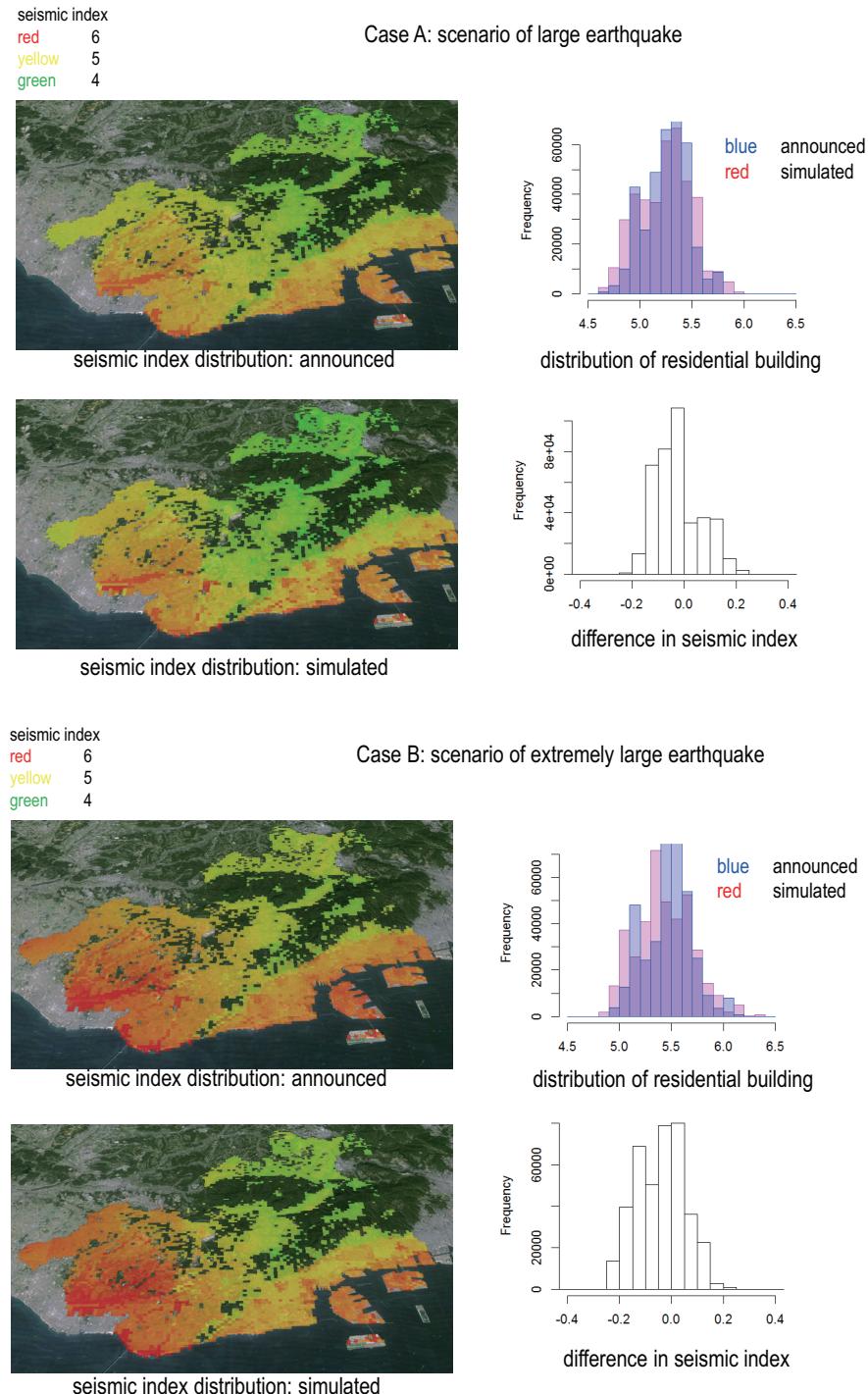


Figure 17.1: Example of next generation hazard map for Kobe City

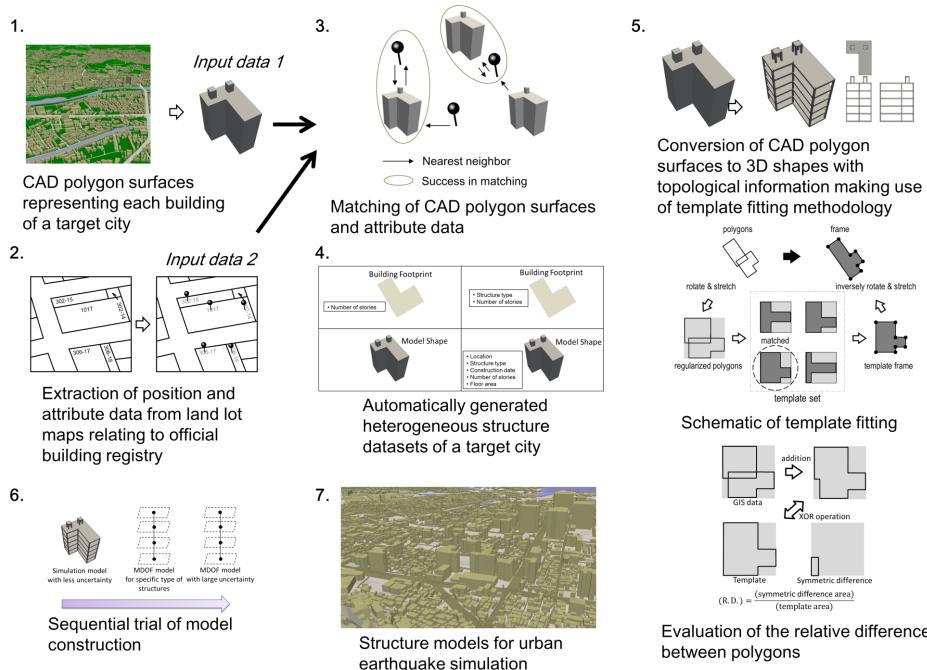


Figure 17.2: Platform of processing data resources to generate urban area model

underground water. The coupling of soil deformation and underground water flow for a given ground motion is computed by using a non-linear finite element method; see Fig. 17.3. The non-linear analysis of soil deformation is essential, and the finite element method specially tuned for this non-linear analysis is implemented in K computer

The developed system has two key features. First, the system is able to automatically generate an underground model of soil and underground water. There are a set of boding hole data for 10,000 sites in Kobe City, and the system succeeds to construct the model for all the site. Second, the system is able to simultaneously carry out numerical simulation of the liquefaction occurrence for all the sites. Each site is analyzed independently from each other, even though the same input ground motion, measured or presumed, is used. Suitable parallel I/O is implemented into the system.

17.3.3 Estimation of lifeline damage induced by earthquake

Lifeline is a dense network of pipelines that carry energy (electricity and gas), water and sewage, and communication, in an urban area. The pipelines are mostly buried, and, in general, has higher seismic safety compared with structures on the ground. However, the lifeline plays a fundamental role of supporting various activities in an urban area, and a damage of one spot could induce malfunctioning of a certain portion of the network.

A buried pipeline of the lifeline is damaged by ground deformation which is quantified in terms of strain; structures on the ground is damaged by ground motion which is quantified in terms of acceleration or velocity. Hence, three-dimensional analysis of underground structure responses is essential to estimate lifeline damage induced by an earthquake. Moreover, there is no device which measures ground deformation; it is possible to compute ground strain by distributing seismographs, but measuring strain along the pipeline remains a difficult task for the distributed seismographs. The numerical simulation of ground deformation is a unique solution for the estimate lifeline damage.

Using a largest scale finite element method that is implemented in K computer, Computational Disaster Mitigation and Reduction Research Unit is developing a system that estimate lifeline damage of an urban area of a few kilo meter. Numerically simulating ground deformation in this area is used to estimate candidates of the lifeline damage induced by ground motion; see Fig. 17.4.

While natural frequency of man-made structure is in the range of a few Hz, it is a standard practice to measure and analyze seismic response in this temporal resolution. A buried pipeline, however, is not damaged by being shaken but by being pulled or pushed, which could be a faster

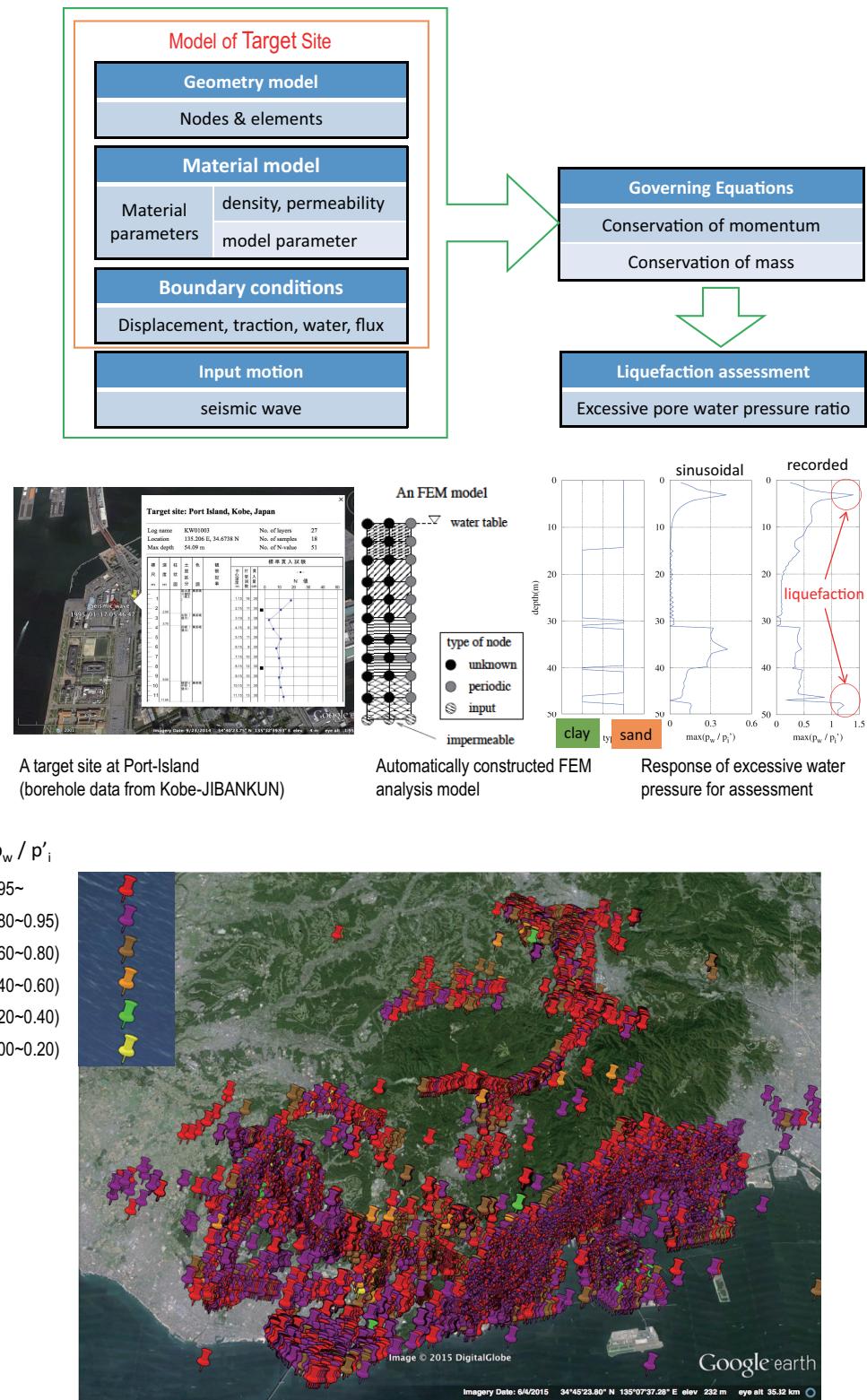


Figure 17.3: System of liquefaction occurrence estimation

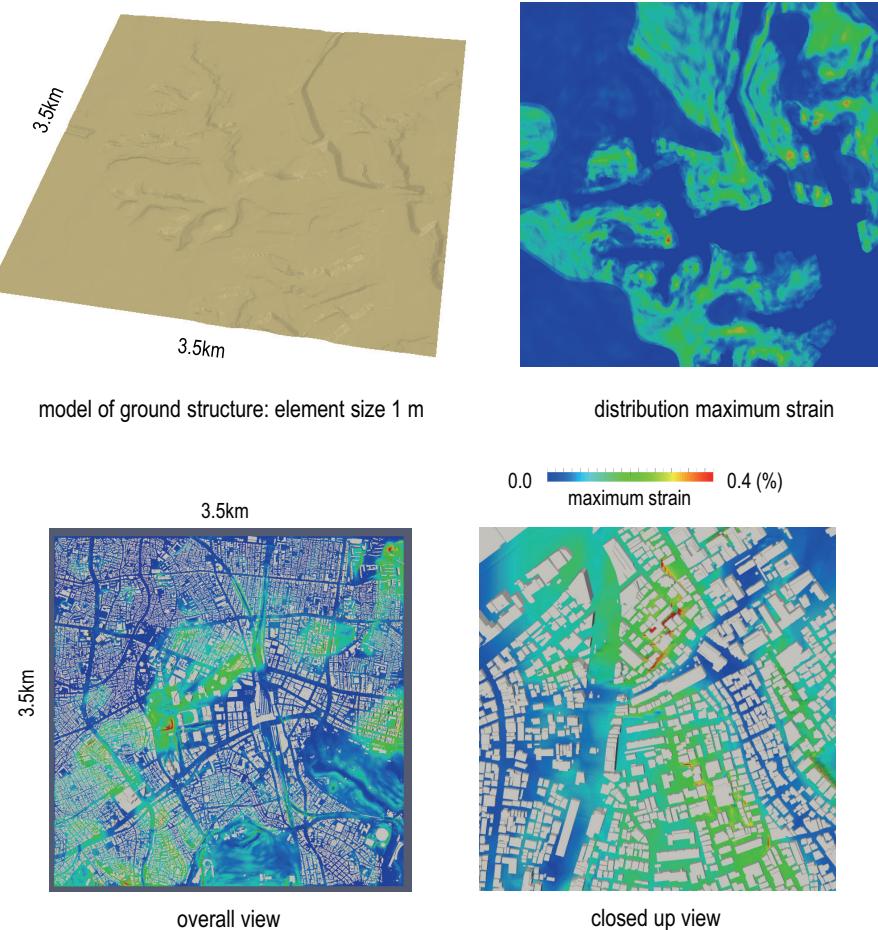


Figure 17.4: Example of ground deformation distribution and induced lifeline damage

phenomenon than vibration. We thus need to collaborate with experiments in which failure of buried pipeline induced by ground motion is observed.

17.4 Schedule and Future Plan

The fiscal year of 2016 is the final year of the first phase of the unit; the extension of the unit is determined. We plan to make a next-generation hazard map for a few cities in Hyogo Prefecture, using the system that processes available data resources. Collecting such data is essential, since the system is completed; sustainable improvement is made for the system. Collaboration of regional universities is strengthened, especially one with University of Hyogo. We also plan to combine tsunami simulation to ground motion simulation.

The objectives of the unit research are summarized as follows: 1) Construction of next generation urban area model for a few cities in Hyogo Prefectures; 2) development of system for liquefaction occurrence estimation; and 3) smart visualization of numerical simulation results of earthquake hazard and disaster.

17.5 Publications

Chapter 18

Computational Structural Biology Research Unit

18.1 Members

Florence Tama (Unit Leader)

Atsushi Tokuhisa (Research Scientist)

Miki Nakano (Post-doctoral researcher)

Sandhya Tiwari (Post-doctoral researcher)

Sachiko Kikumoto (Assistant)

Yumeno Kusahara (Assistant)

18.2 Research Activities

Biological molecular complexes such as proteins and RNAs are of great interest in the area of molecular biology as they are involved in cell replication, gene transcription, protein synthesis, regulation of cellular transport and other core biological functions. Those systems undergo large conformational transitions to achieve functional processes. Therefore characterization of structures of these macromolecular complexes is crucial to understand their functional mechanisms, and play an important role in the development of new drugs to treat human disease.

Experimentally, X-ray crystallography has been the primary tool to study protein conformations, providing high-resolution structures. Cryo electron microscopy (EM) has provided, although at lower-resolution, critical information on structure and dynamics of large biological molecules. More recently, efforts like in Riken/SPring 8 have focused on developing intense X-ray free-electron laser (XFEL) light sources, which offer a new possibility to image single biological macromolecules. Since crystallization is not necessary for such a protein structure analysis, it would be possible to investigate the structure of macromolecular complexes and proteins under various physiological conditions or to observe elementary steps of a biochemical function. However, at the current experimental condition, it cannot achieve atomic level resolution such as obtained by X-ray crystallography.

Computationally, methods have been developed to predict structures from low-resolution data such as cryo-EM either using rigid body fitting or flexible deformations of known atomic structures. In addition, even when structures of the molecules are unknown, atomic models can be predicted using homology modeling and ab initio predictions. While ab initio prediction still remains difficult for large proteins, success in predicting small proteins have been observed. Finally, algorithms to analyze protein/proteins interactions also have shown success in predicting proteins complexes.

Our research focuses on the development of computational tools to study biological systems, more specifically to help in their 3D structural determination using various experimental techniques and to analyze their potential interactions with small molecules in order to design new drugs. The ultimate line of our interdisciplinary research is to bring experimental data as obtained from X-ray, cryo-EM and XFEL with development and applications of computational tools through the K

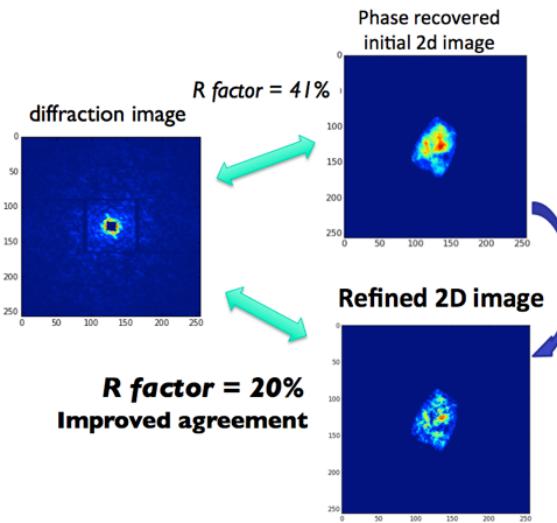


Figure 18.1: Refinement of 2D raw data

computer to acquire knowledge on the structure of a physiologically important protein complexes that are unattainable with existing experimental techniques, and to contribute to development of drug design and medical treatment in collaboration with pharmaceutical companies.

18.3 Research Results and Achievements

18.3.1 Refinement 2D raw data

Following discussions with experimental groups, in this fiscal year, we have been working on refinement of real 2D data. Diffraction patterns from X-ray Free Electron Laser is approximately the Fourier transform of 2D projection image of the target, however the detectors can observe only the intensity and phase cannot be determined. Reconstruction of 2D projection images requires the phase, and a common approach is "phase reconstruction", in which numerical algorithm is used to estimate the phase. To further improve agreement (R-factor) with experimental data, we developed a method that utilizes Limited memory Broyden Fletcher Goldfarb Shanno B Nonlinear Optimizer and Basin hopping algorithm. Such refinement will allow for a more accurate interpretation of the real 2D data derived from XFEL experiments (see Figure 18.1)

18.3.2 3D reconstruction from XFEL data

In order to restore the 3D real structure of the molecule from the diffraction patterns obtained by XFEL experiments, computational algorithms are necessary as one needs to estimate the laser beam incidence angles to the molecule and retrieve the phase information. We are developing a program package for XFEL analysis based on XMIPP, which is commonly used for image processing of single-particle 3D cryo electron microscopy. Since XMIPP is designed to work with 2D data in real space, some of the routine were modified to deal with 2D diffraction data in Fourier space. The overall protocol used for 3D reconstruction is shown in Figure 18.2.

The protocol was tested using simulated diffraction images from the crystal structure of a translation termination complex formed by the Thermus thermophilus 70S ribosome bound with release factor RF2 (PDB ID = 4V67). The algorithm is able to estimate relative orientations of the diffraction images. The resulting reconstructed 3D structure is shown in Figure 18.3.

18.3.3 Fast image filtering procedure

For current XFEL experiments, the number of diffraction data is limited. Therefore, we have been developing new approaches to obtain structural and dynamical information by combining such low-resolution and limited information from XFEL data with other data, such as X-ray structure and

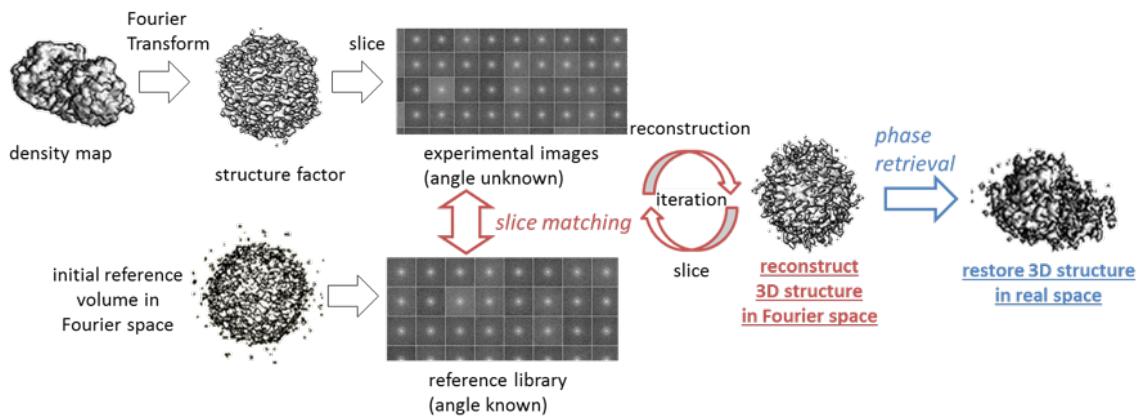


Figure 18.2: 3D reconstruction from XFEL

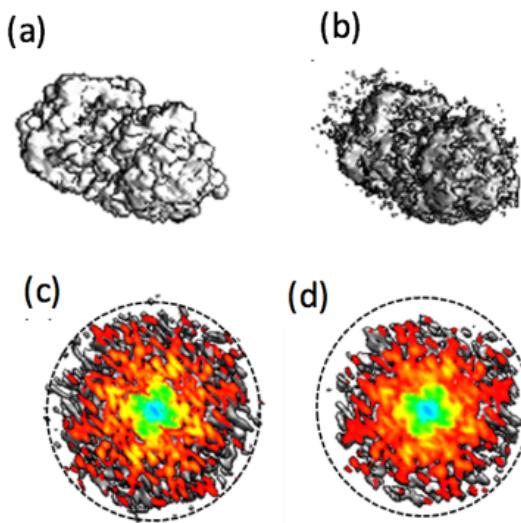


Figure 18.3: 3D XFEL reconstruction for the ribosome. (a) Electron density map of the ribosome created from the X-ray structure. (b) Electron density map obtained using the protocol described in Fig 18.2. (c) and (d) are cross section views of the structure factor (structure in Fourier space) for (a) and (b) respectively

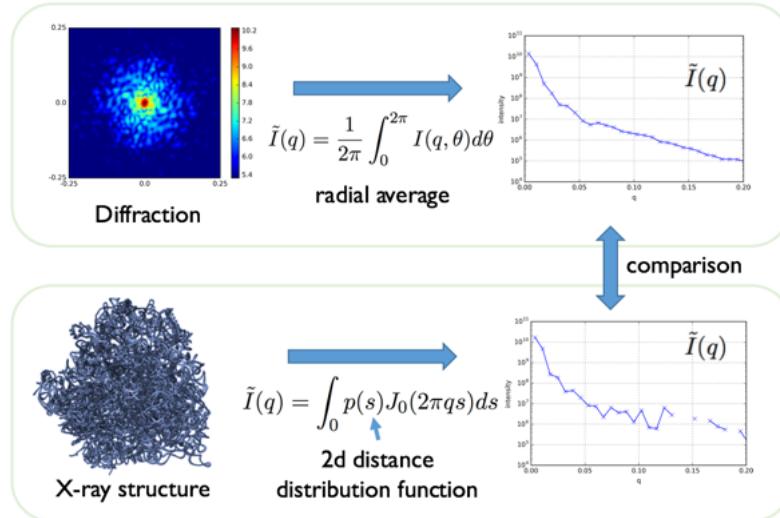


Figure 18.4: Fast image filtering procedure

computational tools. We published algorithms to identify a conformation, among several candidates, that agrees with XFEL data.

To perform ab initio modeling, a large number of conformations need to be explored to find a good model that agrees with the data. Thus it is important that the agreement between the models and the target diffraction can be evaluated quickly. In our previous study, diffraction patterns were first simulated from the candidate models, and then the simulated diffraction patterns were compared with the target pattern. While this approach is accurate, it is also time consuming; the simulation of X-ray diffractions is mathematically complex and 2D image comparisons need to be performed for different alignment angles.

To speed up this process, we have been developing an algorithm to filter out unpromising models using information reduction technique. In this approach, from the original 2D target image, 1D radial profile (average over circles) is calculated. Then for each model, corresponding 1D profile is calculated and compared against the target 1D profile. Such 1D profile can be quickly calculated directly from mass distribution, and 1D profile comparison is also fast. Since 1D approach is an approximation, this is used as a preliminary filtering process to reject the models with low 1D score. Then, we evaluate more precise similar score for the remaining models using 2D images to select the best models. Recently, we developed a new equation that can increase the sensitivity of the filtering:

$$\tilde{I}(q) = \int_0^{\infty} p(s) J_0(2\pi q s) ds \quad (18.1)$$

where $I(q)$ is the 2D profile used for comparison, J is a Bessel function, and $p(s)$ is 2D distance distribution function. This new formula replaces conventional Debye equation used for SAXS analysis and more accurate for 2D diffraction image analysis such as used here. We have tested this approach for a biological molecule to confirm the validity of this formalism (see Figure 18.4)

18.3.4 DNA solvation properties

Under physiological conditions, where various biomolecules and other components are present, DNA strands may adopt many different structures in addition to the canonical B-form duplex. Such effects of molecular crowding on thermal stabilities of DNA structures maybe associated with the properties of the water molecules around the DNAs.

Using MD simulations, we studied the thermodynamic properties of water molecules around a hairpin duplex and a G-quadruplex to understand how cosolutes affect the thermal stability of DNA structures. Our findings suggest that differences in the hydration shell structure around DNAs are one factor that affects the thermal stabilities of DNA structures under the crowding conditions. In addition, thermodynamic properties of water molecules around single- and double-stranded DNAs were investigated. Upon the formation of double-stranded DNA, thermodynamics features of the water molecules is primarily changed in the minor groove. Free energies of water molecules around

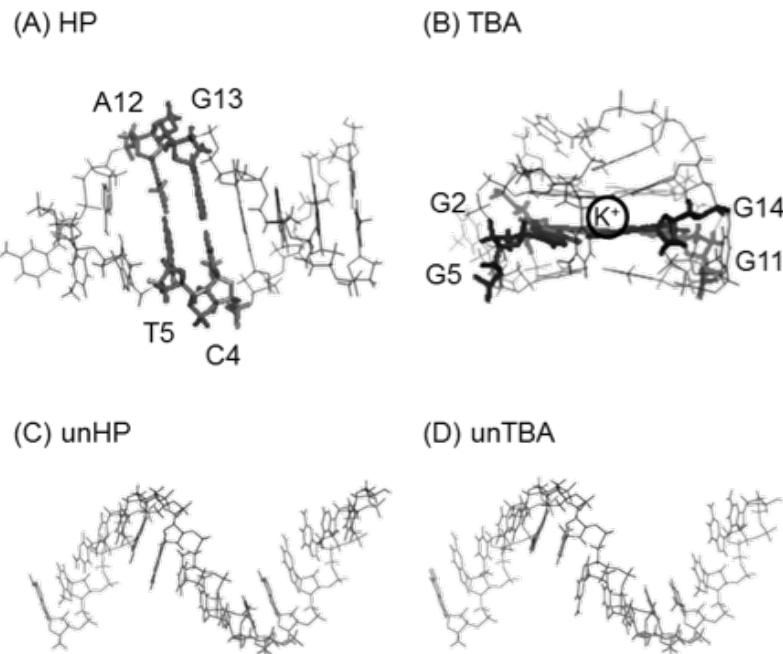


Figure 18.5: Different forms of DNA

double-stranded DNA are lower than those around single-stranded DNA even in the second and third hydration shells

18.4 Schedule and Future Plan

We are planning to continue to develop tools to analyze XFEL data to obtain structural information of biological molecules. In particular, we intend to improve the 3D reconstruction protocol proposed for XFEL data by providing more informed initial reference volume. To achieve such a goal, we plan to maintain a database of shapes and associated diffraction patterns found in biological molecules. In addition, efficiency of the protocols in terms of speed will be addressed. We will also apply our tools to experimental data. In parallel, we will continue the development of algorithms to extract information from a limited number of low-resolution data, since only limited experiments might be feasible for some systems even with the development of experimental techniques. Ab-initio modeling of sample shapes based on the use of simplified representations of the biological molecules as well as optimization procedure will be further developed taking advantage of the newly fast filtering scheme.

As experimental data from cryo-EM and for XFEL will continue to grow in numbers, analysis of such big dataset will increase the necessity of high performance computing. We aim to utilize K and post-K to break the limitation of current processing power to obtain new level of structural information of biological complexes from EM and XFEL data. For this goal, we plan to develop softwares to analyze data to obtain structural models that can utilize computers in different sizes such as cluster and supercomputer. By sharing the software and results from structural modeling with other research institutes, we aim to contribute to the structural biology community.

As our research focuses on developing computational tools to analyze low-resolution experimental data, we intend to establish collaborations with experimental groups in Japan and abroad in order to study structure, function and dynamics of biological molecules.

18.5 Publications

Part II

Operations and Computer Technologies Division

Overview of Operation and Computer Technologies Division

Operation and Computer Technologies Division of RIKEN AICS are responsible for operation and enhancement of the K computer and the facilities. Operation and support of High Performance Computing Infrastructure (HPCI) system are also a part of our missions. Operation and Computer Technologies Division has four teams. The missions and members of the teams are shown as follows:

- Facility Operations and Development Team (Team Head: Toshiyuki Tsukamoto)
Missions: Operation and Enhancement of the facility for the K computer
Members: Technical Staff(6)
- System Operations and Development Team (Team Head: Atsuya Uno)
Missions: Operation and Enhancement of the K computer
Members: R&D Scientist(5), Technical Staff(2)
- Software Development Team (Team Head: Kazuo Minami)
Missions: Operation and Enhancement of the K computer software and application tuning
Members: R&D Scientist(6)
- HPCI System Development Team (Team Head: Manabu Hirakawa)
Missions: Operation and Enhancement of HPCI system and logistics of the HPCI activity
Members: R&D Scientist(1), Senior Visiting Scientist(2), Visiting Scientist(1)

Overview of activities in JFY2015

Three and a half years have passed since the start of the K computer official operation. (Actually five years have also passed since the start of early access which was a service for limited user.) Although we suffered from many troubles at the beginning of the operation, many of those have already been fixed and the recent system operation is relatively stable. However we experienced a rapid increase of irregular down time caused by the file system failures in JFY2014. To fix them, we investigated the causes and developed some workarounds to reduce the impact of such a system failures. Then in JFY2015 we achieved shorter irregular down time and higher system availability than that of JFY2014. On the other hand, at the end of the term, a terrible job congestion occurred and waiting time of jobs rapidly increased. We have already investigated the causes of the event and taken some measures for usage in JFY2016.

In terms of facility operation, we achieved to improve Power Usage Effectiveness (PUE) which is a major metric to evaluate an energy efficiency of IT center etc. For the improvements we tried to improve a power generation efficiency and reduction of power consumption for cooling. Actually, we found that by a comparison between high and middle output level cases, the power generation efficiency of the gas turbine power generator could be improved more than 30%. We also achieved 40% power saving for air handlers compared with that of in JFY2012 by optimizing some operational parameters such as number of active handlers, blowout temperature, number of fans, etc.

In JFY2015, we could have many opportunities to present our research activities. A major achievements are listed as follows:

1. Kiyoshi Kumahata and Kazuo Minami. HPCG Performance Improvement on the K computer. HPCG BoF at The International Conference for High Performance Computing Networking Storage and Analysis 2015 (SC15) Austin. 2015.: This presentation showed a tuning techniques to enhance the performance of HPCG benchmark on the K computer.

2. Kazuo Minami. Power Consumption Reduction Effort for the K computer. Exhibits at The International Conference for High Performance Computing Networking Storage and Analysis 2015 (SC15), Austin. 2015.: This presentation introduced a stack of operational improvements for reduction of the power consumption of the K computer.
3. Fumiyo Shoji et al. “Long term failure analysis of 10 peta-scale supercomputer”. In: In HPC in Asia session at ISC2015, Frankfurt, Germany, July 12-16. 2015.: This presentation reported a failure analysis of the K computer and awarded as the Best Poster Award.

Chapter 19

Facility Operations and Development Team

19.1 Team members

Toshiyuki Tsukamoto (Team Head)

Masami Kurokawa (Technical Staff)

Mitsuru Tanaka (Technical Staff)

Hiroyuki Takitsuka (Technical Staff)

Satoru Matsushita (Technical Staff)

Katsuyuki Tanaka (Technical Staff)

Humio Tsuda (Technical Staff)

Mitsuo Iwamoto (Technical Staff)

19.2 Research Activities

The K computer facilities have many features not found at other supercomputer sites. These include an expansive and pillar-free computer-room, power supply system that consists of a co-generation system (CGS) and a high-speed current-limiting circuit breaker without uninterruptible power supply (UPS), distribution boards installed not on computer-room walls but under a raised floor, extremely quiet and high-efficiency air conditioners, and water-cooling system for CPUs featuring precise temperature control. To ensure stable working of K computer and its peripherals, the facility operations and development team (FODT) of the operations and computer technologies division, RIKEN AICS is responsible for operation and enhancement of the facilities. Furthermore, FODT conducts research on the advanced management and operations of AICS' facilities. One of the most serious problems is rapid and substantial increase in electricity prices since 2011. Therefore, we investigate the most suitable driving conditions for AICS facilities to achieve effective cost reduction. Another problem is increased power consumption by AICS. The use of electricity by AICS is strictly limited by a contract between AICS and the local electric supply company. However, recently, the facility's power consumption exceeded the contract limit. This matter is important because the company requires us to accept a raise in the upper/lower power limit, which amounts to an increase in electricity cost. To prevent this problem, we have investigated a method to control K computer's power consumption by using emergency job stopping together with the system operations and development team and the software development team of operations and computer technologies division, RIKEN AICS.

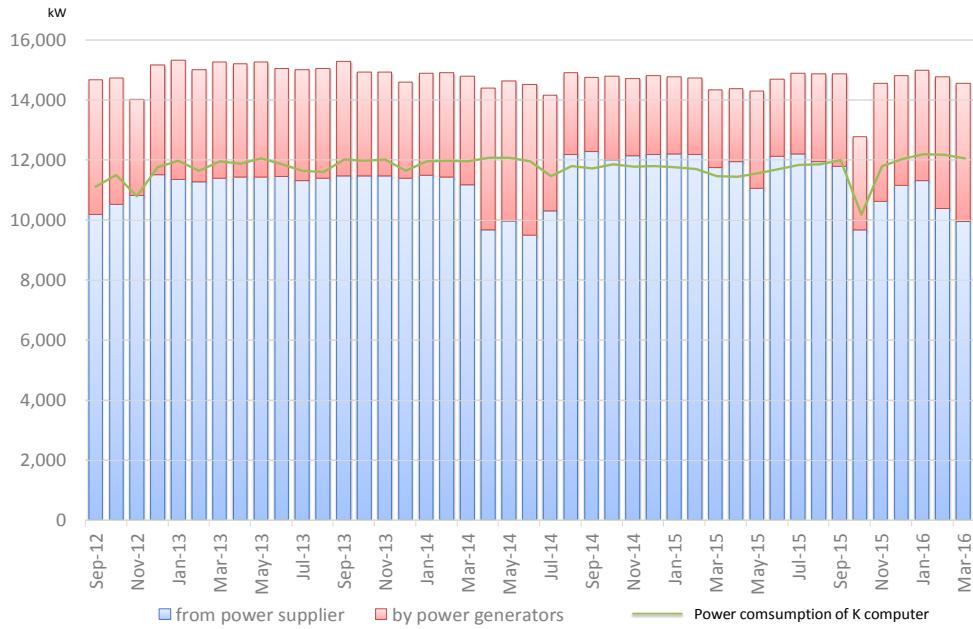


Figure 19.1: Monthly power supply and K computer power consumption

19.3 Research Results and Achievements

19.3.1 Optimum operation of electric power

Figure 19.1 shows the monthly total power supply and the power consumption of K computer from September 2012 to March 2016. The status of power supply, which consists of commercial power purchased from a supply company and the power generated by CGS.

The power consumption of AICS is almost synchronized with that of K computer. The power consumption of AICS is nearly 15,000 kW on average, and the power consumption of K computer accounts for approximately 80% (12,000 kW) of AICS' total consumption. As shown in Figure 19.1, AICS' electric power supply consists of commercial and CGS power. There are two CGS systems in AICS, and they are used by turn for two weeks at a time. Therefore, at least one CGS is always in use. Commercial electric power is covenanted at about 12,500 kW, and power consumption was approximately 12,000 kW (annual average), which corresponds to approximately 90% load factor. To minimize the cost, we try to optimize the ratio of commercial and CGS electricity. To investigate the optimized conditions that minimize the sum of electricity and gas cost, we determined the costs of several ratios of commercial electricity to CGS electricity. We also constructed a model to describe energy flow of the electric power supply and cooling system. Then, we performed computer simulation using the model and actual operating data. In near future, we intend to clarify the cost-optimized conditions that contribute toward reducing costs.

19.3.2 Improvements to power usage effectiveness (PUE)

We have continued to work on improvements for the effective use of electricity. PUE is a well known indicator of the effectiveness of electricity use. To improve PUE, we have attempted to optimize the operation of the air-conditioning system since FY2013. Figure 19.2 indicates a change in the annual average power consumption of K computer (including the peripheral devices) and cooling equipment. After FY2013, the power consumption of K computer has been almost flat at approximately 11800 kW, but the power consumption of the equipment decreased gradually from FY2013 to FY2015. Accordingly, the PUE of AICS improved to 1.356 in FY2015 from 1.447 in FY2012, thus contributing to reduction in electricity cost.

In FY2013, we reduced the electricity cost of air-conditioners by reducing the number of working air-conditioners. Total cooling performance was maintained by lowering the air temperature. We

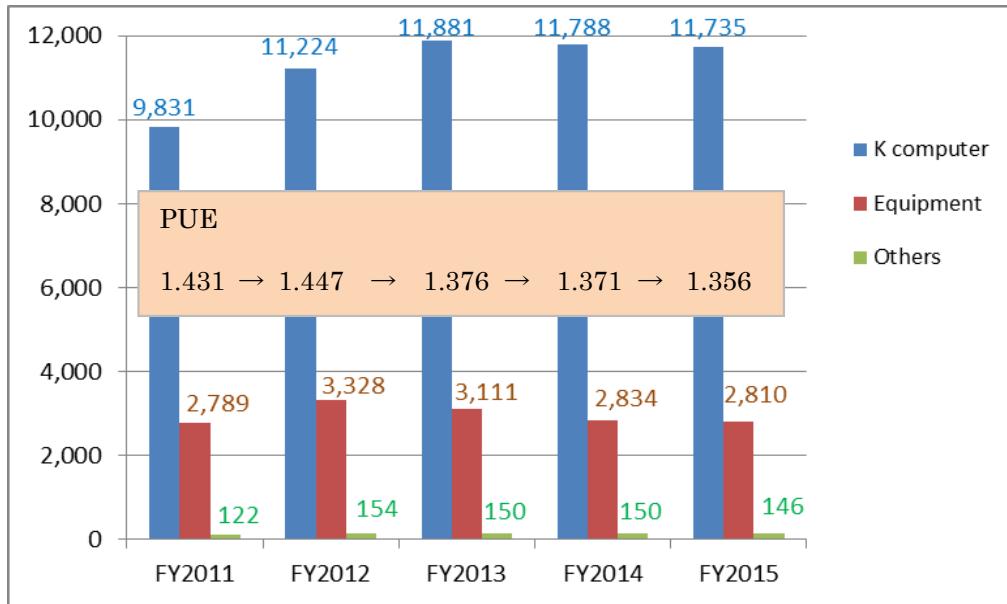


Figure 19.2: Trend in annual average electric power consumption

could achieve a reduction of 217 kW in power consumption. In FY2014, we focused on the fault-tolerance feature of the air-conditioning equipment. Each air-conditioner has two motors for fault-tolerance. We found that if one of the two motors could be stopped, airflow could be maintained at approximately 60%. Thus, we reduced power consumption by a further 277 kW in FY2014, and by 24 kW in FY2015.

19.4 Schedule and Future Plan

We will continue to improve the advanced management and operation of AICS facilities and contribute to the user service of K computer. We will work on reducing costs by investigating and applying the most suitable driving condition to all of the electric power supply and cooling equipment. Furthermore, we will improve electric power control of the entire AICS facility to prevent overshooting the contracted power demand with the system operations and development team.

19.5 Publications

Chapter 20

System Operations and Development Team

20.1 Members

Atsuya Uno (Team Head)

Hitoshi Murai (Research & Development Scientist)

Motoyoshi Kurokawa (Research & Development Scientist)

Keiji Yamamoto (Research & Development Scientist)

Fumio Inoue (Research & Development Scientist)

Yuichi Tsujita (Research & Development Scientist)

Mitsuo Iwamoto (Technical Staff)

Katsufumi Sugeta (Technical Staff)

20.2 Research Activities

K computer, a distributed-memory parallel computer comprising 82,944 computing nodes, has played a central role in the High Performance Computing Infrastructure (HPCI) initiative granted by the Ministry of Education, Culture, Sports, Science and Technology. The HPCI has achieved the integrated operation of the K computer and other supercomputer centers in Japan and has enabled seamless access from user machines to a cluster of supercomputers that includes the K computer. Moreover, the HPCI has provided large-scale storage systems that are accessible from all over Japan.

The System Operations and Development Team (SODT) has conducted research and development on the advanced management and operations of the K computer. While analyzing operational statistics collected during shared use, the SODT has improved the system configuration, including aspects involving job scheduling, the file system, and user environments. As an example, achieving higher system utilization is very difficult, because the K computer must simultaneously process various sizes and types of jobs. The SODT has responded flexibly to user requests and analyzed operational status, thereby realizing a high level of utilization of approximately 75% in the fiscal year 2015 (FY2015). Moreover, the SODT has developed tools that improve the usability of the K computer. The SODT has also helped users manage the K computer and utilize the K computer resources effectively by improving the system software. Note that this support was achieved together with the software development team.

In FY2015, we primarily implemented improvements to operational issues. In particular, we addressed the performance degradation issue of the file system and the long waiting times of some jobs. In addition, we implemented a performance monitoring system for the global file system (GFS) and local file system (LFS). As for electric power use, we developed an automatic emergency job-stopping method and implemented a system to estimate future electric power needs.

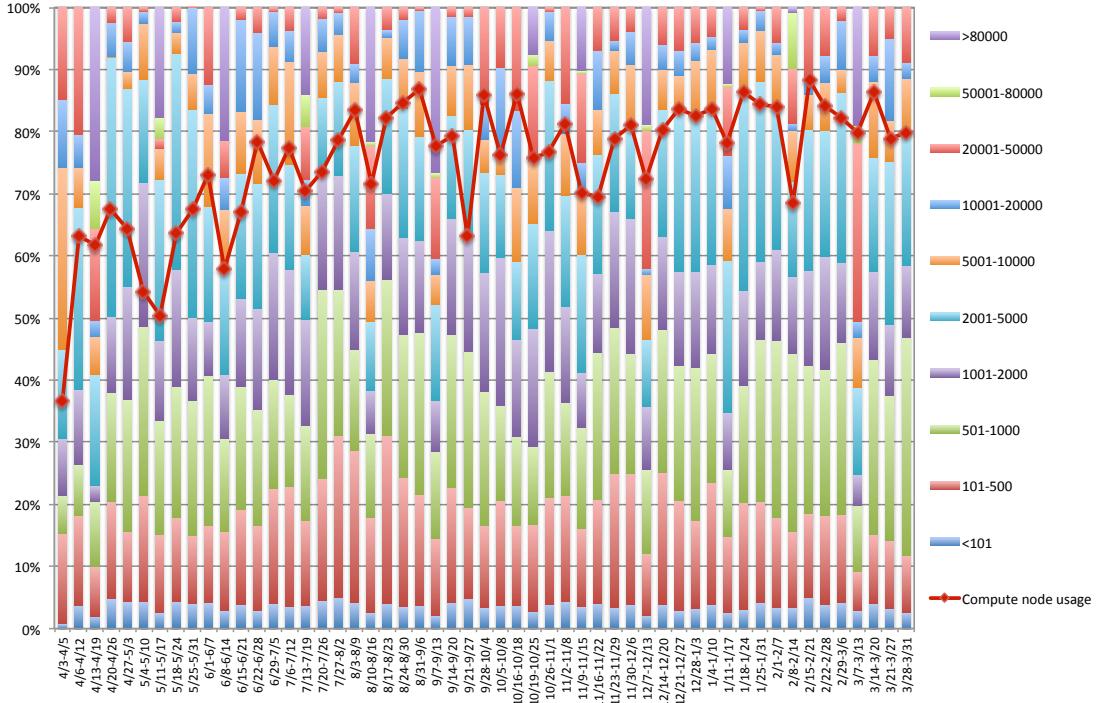


Figure 20.1: Resource usage in FY2015

20.3 Research Results and Achievements

Figure 20.1 shows resource usage details for FY2015. As shown in the figure, we achieved approximately 75% node usage, which is the same as FY2014. Each project had appropriate computing resources for a year, and these resources were divided into the following two terms : (1) from April to September and (2) from October to March. In FY2015, node usage during the first term was 71%, whereas the usage during the second term was 80%. In contrast, node usage per term in FY2014 was 79% and 72%, respectively. Because many new projects were initiated in FY2015, utilization of the first term (especially in April and May) decreased.

20.3.1 Analysis and improvements of operational issues

Shortening job waiting times

Figure 20.2 shows the average waiting times of large jobs (i.e., 12289–36864 nodes) in both FY2014 and FY2015. To consume remaining computing resources before they expired, users tended to submit many jobs at the end of each term, the average waiting times in September and March were substantially longer than those of other months. In FY2014, job congestion occurred in August and September; however, in the second term of FY2015, they occurred from December. In addition, average waiting times for FY2015 were longer than those of FY2014. We therefore analyzed job scheduling and found that the following two factors impacted long waiting times.

1. Influence of higher-priority jobs (prior/semiprior job)

From FY2014, we provide a priority use system by which users can run a job at a higher priority than normal jobs. Figure 20.3 shows the number of higher-priority jobs per month. As is evident in the figure, the number of priority jobs in FY2015 was greater than that of FY2014. Normal priority jobs, especially, large-scale jobs, are impacted by these higher-priority jobs. In February 2016, we therefore changed the region used by higher-priority jobs to a region that may not inhibit the execution of normal jobs.

2. Influence of jobs specified in excess of node LFS quotas

Users specify LFS quotas when a job is submitted. There were many jobs specified with excess node LFS quotas in FY2015. Figure 20.4 shows the number of occurrences of LFS space

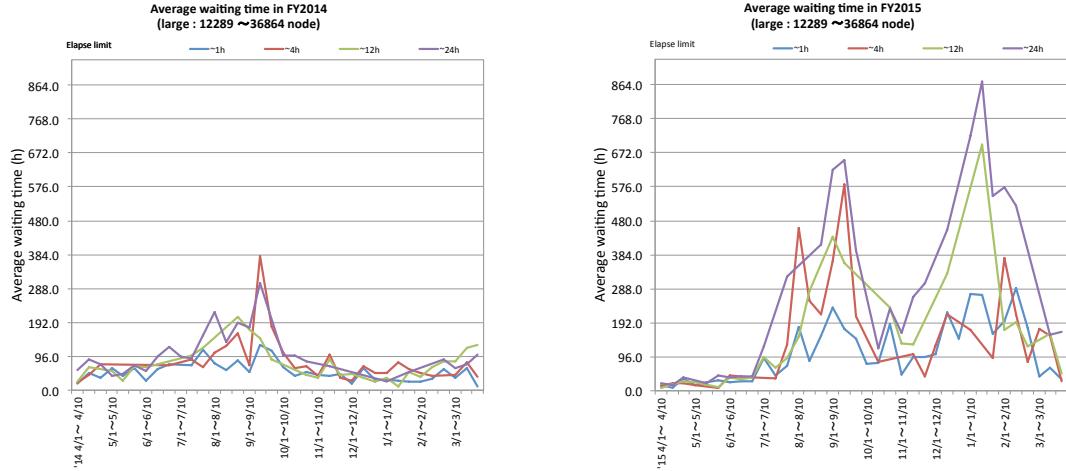


Figure 20.2: Average waiting times in FY2014 and FY2015

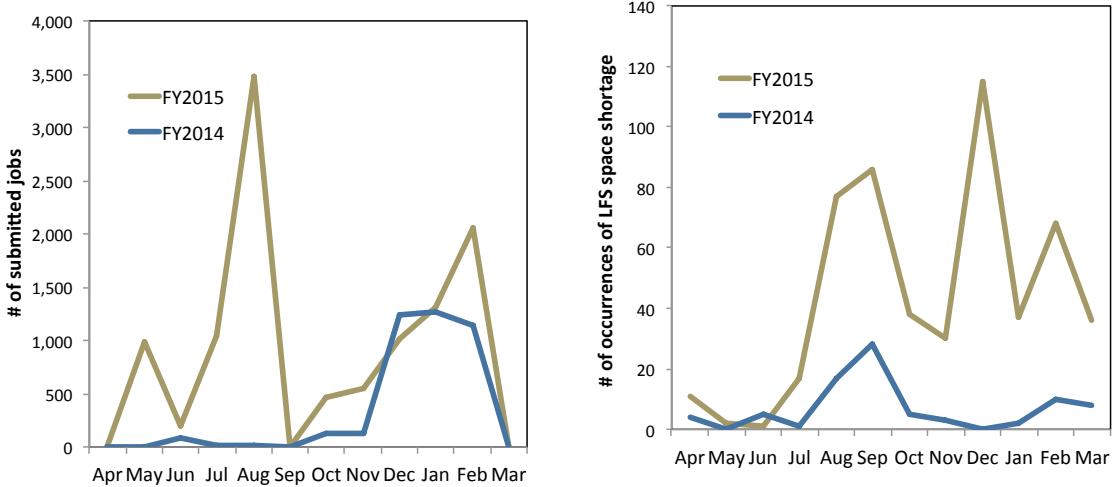


Figure 20.3: Number of submission priority jobs

Figure 20.4: Number of occurrences of LFS space shortages

shortages in the 10,000 or more nodes of a job. Comparing Figure 20.2 and Figure 20.4, we note a correlation between waiting times and LFS space shortages. We have therefore requested a review of node LFS quotas in order not to occurrence of LFS space shortages. In addition, we designed a system to monitor and detect LFS space shortages.

Response degradation problems in the GFS

Temporary reductions in the response rates of the GFS have occurred several times. The major factor is the load concentration on the GFS caused by large-scale staging. Because this bandwidth is filled with access to very large files, access to other files was inhibited.

We analyzed job scheduling and found that the following two factors impacted response times. We therefore performed the following three improvements. Through these improvements, load balancing was achieved and response time improved.

- Automatic striping of stage-out file
- Processing I/O thread allocation setting for access from the front-end node and staging system
- Changing the default stripe count on the GFS

Performance monitoring of the GFS and LFS

We started to collect performance data regarding the GFS and LFS to monitor load. We use these data to detect failures and analyze each job's I/O performance. Every 10 min, our monitoring system collects and records the size of read and write operations. Similarly, our system collects the number of metadata operations, such as open, close, getattemp, etc. In the future, we will evaluate the job's I/O performance using these gathered performance data.

Disclosure of information to the user

We constructed a webpage to present the following information on the K user portal.

- Resource use history, including job and storage information
- Detailed performance information of each job
- Login history
- Performance information on jobs executed by pre/post processing nodes

20.3.2 Power consumption problem

The K computer's power consumption exceeded the given limit several times during FY2013; this is an important problem because it forces us to increase the contracted upper limit for power, thereby increasing costs, which cannot be ignored. From FY2014, to prevent this problem, we performed a preliminary review that estimated the power consumption of each job, thereby enabling us to control the K computer's overall power consumption. Moreover, we investigated an emergency job-stopping method based on the estimated power consumption of each job in case power consumption again exceeds the given limit. In FY2015, we primarily worked on improvements to our emergency job-stopping method and algorithm to predict power consumption a few hours into the future.

Improvements to our emergency job-stopping method

In FY2014, the staff of the facility monitored and stopped jobs when power consumption exceeded given limits. Because this method was manual, it took time to stop a job and was prone to human error. In FY2015, we built a management system for the K computer that can read power information of the facilities. When this system detects excess power use, it automatically stops the current job. Immediate response times are required when excess power is demanded, thus automated job stopping led to further protection. In FY2014, we used a simple approach, i.e., we selected the largest job in terms of the number of nodes as the job to be stopped.

In FY2015, we evaluated other selection methods that take into account the power consumption, number of nodes and elapsed time of the jobs. Selecting jobs to stop is a combinatorial optimization problem; thus, we used a genetic algorithm to solve the problem. In FY2016, we plan to operate and evaluate this emergency job-stopping system.

Power consumption prediction

To prevent excess power consumption, the cogeneration system (CGS) is useful to temporarily increase power supply. Two CGS of 5 MW are operated alternately in AICS. If two CGSs are operated simultaneously when excess power is expected, we can avoid the excess power. Because CGS takes 1–2 hours to initiate, we must expect excess power to be demanded in a few hours. We abandoned the full-time operation of two CGSs, because fuel costs increasesd. We examined our prediction method regarding power fluctuations after a few hours. According to job statistics, many users execute jobs repeatedly with the same number of nodes. Moreover, 80% of users executed jobs with at most nine patterns in terms of the number of nodes.

We therefore implemented a power forecast system that estimates the power of a scheduled job using power history of already-executed jobs. Figure 20.5 presents an example of predicted power usage at 10:30 AM, showing data from 12 hours before until 12 hours after 10:30 AM. In other words, the center of the figure represents now. The past shows the record of power, while the future shows predicted power usage. The dashed lines in the figure indicate the upper and lower limits of the predicted power plus statistical prediction errors. In FY2016, we plan to improve prediction



Figure 20.5: Power Consumption Forecast at 10:30 AM

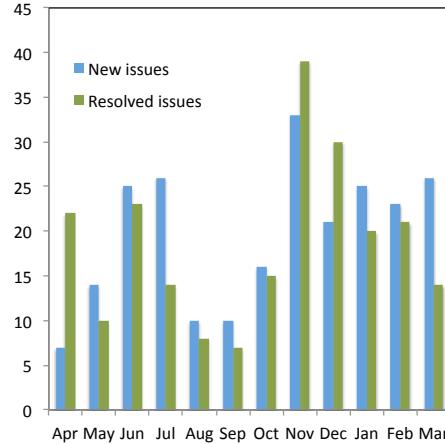


Figure 20.6: Number of issues addressed in FY2015

accuracy by using machine learning and evaluate our emergency job-stopping system. In addition, we plan to study optimal operation methods of the CGS using our prediction results.

20.3.3 User support

The K computer had approximately 170 groups and 1,900 users in FY2015. The total number of HPCI users and AICS researchers were approximately 1,600 and 300, respectively. The number of daily active users was approximately 150.

We supported users through the K support desk, providing users with technical information on the K computer, including information regarding its system environment, system tools, software libraries. In addition, we performed user registrations, failures investigation and software installation, etc. Our consulting services were offered together with the software development team. Figure 20.6 present the number of issues addressed in FY2015, showing the number of new issues in FY2015 to be approximately 230; the number of resolved issues was approximately 220. The number of new issues in FY2013 and FY2014 were approximately 230 and 170, respectively.

20.4 Schedule and Future Plan

In FY2015, we primarily performed improvements on operational issues. We analyzed the operational status, identified a variety of issues, and made improvements in these areas. To implement these improvements, we constructed a log-gathering environment and a database. In the next fiscal year, we are scheduled to publish the compiler, which supports the new standard. We continue to improve the user environment and provide user support. Moreover, we continue to address the K computer's power consumption problem.

20.5 Publications

Chapter 21

Software Development Team

21.1 Members

Kazuo Minami (Team Head)

Masaaki Terai (Research & Development Scientist)

Akiyoshi Kuroda (Research & Development Scientist)

Hitoshi Murai (Research & Development Scientist)

Kiyoshi Kumahata (Research & Development Scientist)

Kengo Miyamoto (Research & Development Scientist)

Yoshito Kitazawa (Research & Development Scientist)

21.2 Research Activities

Collaboration between the K computer system and its applications is key to creating innovative results. To maximize application and system performances and usability, the Software Development Team is conducting the following activities: (1) efforts to maintain the stable operation of the K computer and the support for users to obtain good outcome; (2) research into the high level operational efficiency of the K computer; and (3) research to improve application performances.

21.3 Research Results and Achievements

21.3.1 Operations of the K Computer

Improvement of the Operation Environment

- (1) Establish the rules for installing and maintaining open source software (OSS)

The installed software requires continuous maintenance because software compiled in an older language environment of the K computer may have trouble in the current newer language environment owing to discrepancies between the compiling and running environments.

Formerly, the maintenance rules for installed software were not sufficiently established. Therefore, we needed to judge whether maintenance for previously installed software was necessary when a trouble occurred. In FY2015, to improve usability, we established overall rules concerning OSS management.

- (2) Study on runtime fluctuations

Runtime fluctuations are differences in the runtime of an application that occur for some reason on the system despite the application being carried out under the same execution conditions and in the same manner. These fluctuations disturb efficient operations on the K computer. Therefore, we need to study runtime fluctuations and resolve them.

In FY2015, we encountered an application with a runtime fluctuation. A detailed analysis, e.g., using a profiler, special compile options, and monitoring values in the calculation, found an increased number of executed instructions and irregular floating point values when the runtime increased.

This irregular floating point value is called the “de-normalized number” and is defined in IEEE754. Usually, floating point calculations should be executed by the hardware; however, in this case, the calculation of the de-normalized number was carried out by the software. Therefore, an increased number of instructions were executed and more time was consumed when the de-normalized number was included in a program. We confirmed this phenomenon using a simple test program. Currently, we are investigating whether this is the actual cause of the runtime fluctuation in the application.

(3) Checking the system performance

When system software, such as operating systems, compilers, and libraries, are updated, we need to check if the update has any side effects on the execution of the applications. Therefore, we regularly test the performance. In FY2015, we had two consecutive updates of the language environments, K-1.2.0-18 and K-1.2.0-19, and we checked each environment using the check suite, which is a tool to check the comprehensive system performance.

1) K-1.2.0-18

- According to the result of the check suite, we detected performance degradations.
- As a result of an investigation, we determined that the cause of the degradations was not the language environment, which included the compilers and the libraries. Therefore, K-1.2.0-18 was formally released to all users.

2) K-1.2.0-19

- According to the result of the check suite, the performance was the same as the old environment. It was then formally released to all users.

(4) Release of the MPI performance data

In FY2015, we checked the performance of the major communication functions in the MPI library, measured in FY2014, and released it as a tutorial to all users as a reference to the communication performance. We researched trends in the performance using the Tofu specific algorithm, message size, node shape, and environment variables. Using the results, we described the trends in the performance in the tutorial. Further, we provided the data of the measured results [**MPI’HPC**].

(5) Evaluation of the FX100 System

We compared the performance of the K computer and FX100 via a collaboration with RIKEN ACCC. Using the check tool for the K computer, we evaluated the calculation and communication performances. Because the communication performance of the inner node was lower than that of the outer node, there were cases where the performance of the collective communication was low [**HOKUSAI’FX100**].

Development of Tools

(1) K-scope: The Fortran and C source code analysis tool

K-scope is a Fortran and C source code analysis tool that uses intermediate codes exported by the front-end of the Omni compiler. To find bottlenecks (e.g., multiple nested loops) from source codes, performance engineers can use K-scope prior to rewriting source codes to improve performance. In FY2015, to improve the usability of the tool, we extended the keyword search feature in the C version of the tool. In addition, we developed a feature that extracts a kernel code (a tiny program including a few subroutines or loops to quickly evaluate performance) from Fortran source codes. We have distributed K-scope since FY2012 as part of the AICS software. Furthermore, in collaboration with Kobe University in Japan, we developed the Eclipse plugin to visualize Fortran source code structure on the framework.

(2) Creating simple methods to analyze application performance

SPARC64 VIIIfx has more than 100 hardware counters for performance measurement. To collect detailed profiling data of an application, it was previously necessary to execute the profiler several times. In FY2015, we developed simple profile tools that are each executed once. These tools are the basic performance analysis and the time breakdown analysis. Using these tools, the user can collect profiling data by executing the profiler once. We will release the tools to users.

User Support

We provide user support services from the K support desk in collaboration with the system operations and development team. The K support desk's role is to solve any queries from the users, which come from both RIKEN AICS users and general users indirectly via the Research Organization for Information Science and Technology (RIST). RIST is an official organization that is engaged in consultation services for the K computer general users. We are in charge of compilers, profilers, programming support tools, debuggers, numerical libraries, an MPI library, user manuals, and tutorial documents.

The K support desk accepted 756 technical inquiries, requests, and problem reports in FY2015. The staff members of the K support desk held weekly meetings to check the progress towards solving these issues.

21.3.2 Improvement of the operational efficiency and power consumption reduction

The effective use of computing resources for the operation of the K computer is important. Improving the job performance to improve the operational efficiency and to reduce the power consumption is necessary. Here, operational efficiency improvement means saving computational resources. We constructed a system to collect information concerning jobs and to analyze the jobs to improve their operational efficiency.

Development of a collection system for the job information

A comprehensive statistical analysis to improve the operational efficiency is difficult because massive amounts of job information (e.g., flops, memory throughput, and I/O etc.) are separately managed in several different servers. In FY2015, to gather the statistical information from the separate databases, we performed a feasibility study and constructed a proto-system. This system transparently collects the information and does not require additional tasks (e.g., profilers) for the users. We estimate that the constructed database will reach 1GB in a year. Furthermore, the database will continue to expand with the incorporation of additional information. Moreover, we developed a set of tools for job analyses to efficiently analyze these big data. Using the system, we will publish this information to users via a portal site.

Analysis of job information

To improve the operational efficiency of the K computer, we analyzed the statistical information extracted from the database. Via the analysis, we expect to find problems in terms of improving operations. In addition, based on the profiling data, we can illustrate the relation between the application performance and power consumption [POWER·ACS] [POWER·AICS] [POWER·AXIES]. In FY2015, we determined the dominant power consumption factors on the K computer. One of the factors is the standby power consumption, which is a constant wastage regardless of whether an application is running. In other words, the power consumption is a two-storied structure: a basic part (standby power) and an increments part (active power). Therefore, a reduction in the execution time via performance improvement will contribute to saving computational resources. Even though the power consumption per second is increased by application tuning, the total power consumption will be reduced by decreasing the dominant standby power, as shown in Fig. 21.1.

Therefore, to improve the operational efficiency and reduce the power consumption, we plan to measure the actual application performance including the production runs and improve them based on the following metrics: 1) memory performance, 2) calculation performance, 3) I/O performance, and 4) communication performance. In particular, the memory and calculation performances are

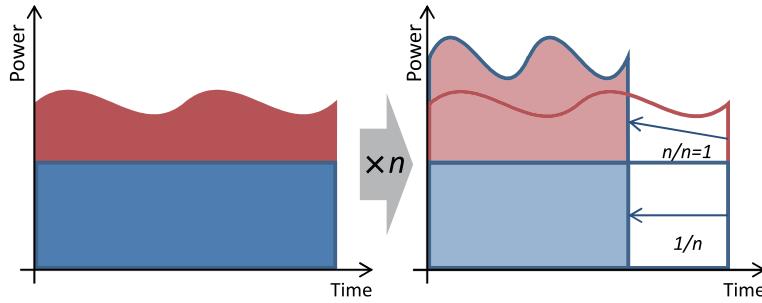


Figure 21.1: Relation between power consumption and application performance. The blue region indicates standby power and the red region is the power that is proportional to the performance.

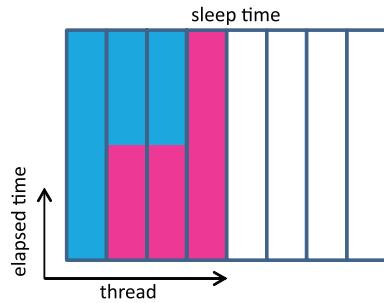


Figure 21.2: The concept of sleep time. The blue region is the CPU working time, the purple region is the waiting time caused by communication and imbalance, and the white region is the time spent not running. We took the sum of the purple and white regions to be the sleep time.

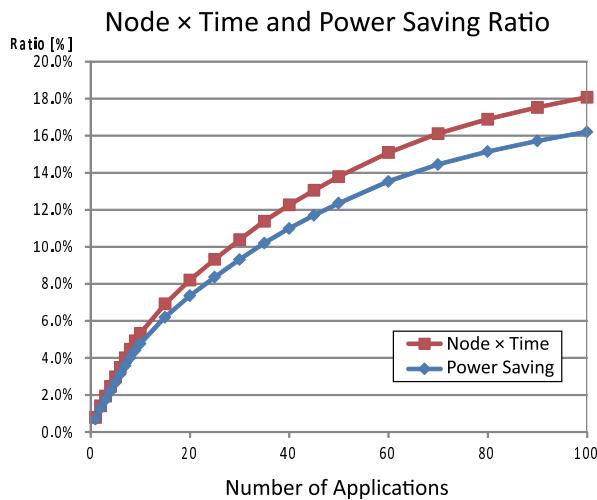


Figure 21.3: Relation between the number of evaluation applications and the calculated savings in resources and power consumption reduction.

major factors in power consumption. These factors can be analyzed from the hardware counter information. In addition, we found that the I/O performance has an imbalance. The amount of communication traffic was evaluated using a function of the operating system. However, the communication time is difficult to analyze because it is not constructed from the collected data. Therefore, to address this problem, we introduced sleep time, which includes the communication and waiting time due to the imbalance of the calculations (Fig. 21.2).

A job that requires large computational resources occupies a greater proportion of the total system resources. We employed a weighted average using the compute node time product and classified the jobs into fields, such as Earth science and materials science. Using this analysis, with the average performance in each field, we estimate that a 5.3% improvement in the operational efficiency would result from improving the top 10 major applications using computational resources on the K computer. In addition, an 8% improvement is estimated when improving the top 20 applications [POWER'SC15] (Fig. 21.3). A large usage improvement can be expected by evaluating just part of the applications. In addition, we found that the computational resources spend huge amounts time in sleep time due to communication and load imbalances in these applications.

21.3.3 Research and Development of Techniques to Improve the Computational Performance of Application Programs

HPCG performance improvement on the K computer

The high-performance conjugate gradient (HPCG) benchmark program has been proposed as a realistic performance metric for supercomputers. HPCG measures the speed at which symmetric sparse linear system equations are solved using the frequently used conjugate gradient method with a multi-grid symmetric Gauss-Seidel smoother.

We evaluated and improved the performance of HPCG on the K computer using an early version prior to its official announcement. HPCG showed good parallel scalability; however, its single CPU performance was very low. Therefore, we tuned the single CPU performance using several techniques, such as memory layout improvement and multithreading. Consequently, we obtained a speed-up of approximately 19 times. The K computer marked 0.461 PFLOPS using 82,944 nodes and ranked second on the ranking of HPCG. This rank was higher than that of the Titan supercomputer at Oak Ridge National Laboratory, which has a higher LINPACK score than the K computer.

The new version of the HPCG code was released in SC15. In this version, the method for calculating the GFLOPS value was modified. Therefore, we are now attempting to obtain a high HPCG score with this new version. In the new version, the time required for creating the problem matrix is included in the total runtime for calculating the GFLOPS value. Thus, we have shortened the time required to create the problem matrix. In addition, the effect of increasing the iteration steps for convergence due to coloring is more significant in the new version. Therefore, the running parameters we previously used are no longer suitable and we are investigating additional efficient running parameters.

Support for Improving the Computational Efficiency of Full-node Application Programs

(1) ELSEs

To analyze the quantum behavior of a billion atoms in an organic polymer electronic device, we optimized a simulation code based on order- N method with the full system of the K computer, in collaboration with the Hoshi laboratory at the Tottori University. The hottest kernel of the code was a four-nested loop with short iteration numbers for matrix-vector multiplication. Its performance was 1% of the K peak performance. We tried several matrix data storing formats, e.g., CRS and ELL, and their corresponding implementations. As a result, it was found that the performance of this kernel could reach 16% of the K peak performance.

In a performance evaluation with strong scaling on the full compute nodes, we achieved 72% in parallel efficiency and minimized the time-to-solution (less than 10^2 s in elapsed time) to realize practical research. Based on this achievement, we submitted a paper for the ACM Gordon Bell Prize in SC16.

Furthermore, the knowledge obtained in the collaboration contributed to improvements in the profiler and environment on the K computer.

(2) DMRG

This application approximately calculates the electronic states in a crystal using the density matrix renormalization group method for a strongly correlated quantum system. We supported tuning and measuring the detailed performance of this application. There are two parameters. One is the number of sites to decide the size of a system, and the other is the DMRG truncation number to decide the precision. In the ground state calculation for a 1D system, which is a model on a 16-site cluster with a truncation number of $m = 8064$, we achieved a peak performance of more than 70% using all 82,944 nodes on the K computer. In addition, we calculated the dynamics in 2D, using a model on a 24×18 site cluster with a truncation number of $m = 4032$, on the K computer.

(3) PHASE

We have continually supported the first-principles density functional theory calculation application PHASE, to be developed by NIMS to apply for the ACM Gordon Bell Prize in SC15, with further tuning and measurements. This fiscal year, we further improved its performance via a parameter optimization search. The application performance improved to 21.17% (2.25 PFLOPS) this fiscal year from 18.6% (1.97 PFLOPS) last fiscal year. In a strong scaling measurement using approximately 4,000 atoms, whose scalability was achieved on 36,864 nodes last fiscal year, we achieved scalability to 82,944 nodes. This means that we can calculate a structural relaxation calculation problem that required 51 days using 48 nodes within 8 hours. We were able to simulate long time relaxation using a number of compute nodes that was much larger than the number of atoms.

Speeding Up a Kernel Program to Multiply a Double-double Precision Matrix by a Double Precision Vector

We improved the computational efficiency of a one-process kernel program to multiply a double-double precision matrix by a double precision vector. The elapsed time of the kernel program was reduced to less than 1/3 on 8 cores of the K computer and 1/8 on 12 cores of an FX100 system in a case where the matrix size was 24 rows and 393,216 columns.

We stopped the unrolling of a loop of the matrix row number, which was the outermost loop, to reduce cache line conflicts by decreasing the number of data streams. We abolished the unnecessary double-double precision extension of the vector so that the thrashings resulting from prefetches on the L1D cache were reduced. For the inner product procedure contained in the row number loop, we adopted a tournament sum scheme so that the SIMD arithmetic units would effectively work.

Graph-based Economic Simulation

To study behavior in economic networks, a graph-based algorithm was proposed. The algorithm propagates financial stress in a network based on actual transactions among domestic companies. To perform massive parameter-sweep calculations, we needed to reduce the turnaround time in a graph traversal. In FY2015, based on a prototype code in C++, we developed new code in C with OpenMP. In addition, we employed a MapReduce framework to quickly deploy workloads to compute nodes on the K computer. Using the framework on the bulk-job mode, the turnaround time was reduced to 192 s from more than 9 hours (175 times faster). This research project was conducted in collaboration with University of Hyogo.

Evidenced-based Performance Tuning

Typical performance improvements for an application comprise two steps: finding a bottleneck using profilers and rewriting the source code. These steps are a form of engineering; however, the performance is strongly influenced by rewriting of the source code. Eventually, we often use an ad-hoc solution instead of engineering with sufficient investigation. Therefore, even if we have tips and know-how regarding performance optimizations for an application, it is difficult to systematically marshal this knowledge and reuse it for other similar applications. In FY2015, to perform a feasibility study to classify source code behavior with supervised learning, we collected more than 50 open source projects including more than 100 loops from GitHub to create training data from public resources. This study was conducted in collaboration with the HPC Usability Research Team in RIKEN AICS.

21.4 Schedule and Future Plan

The Software Development Team will continue the following activities in the next fiscal year: (1) efforts concerning the stable operation of the K computer and support for users to obtain good outcomes; (2) research into the high level operational efficiency of the K computer; and (3) research to improve application performances. We will evaluate the detailed behavior of applications contributing to the improvement of the operational efficiency in cooperation with developers.

21.5 Publications

Chapter 22

HPCI System Development Team

22.1 Members

Manabu Hirakawa (Team Leader)
Etsuya Shibayama (Senior Visiting Scientist)
Osamu Tatebe (Senior Visiting Scientist)
Manabu Higashida (Visiting Scientist)
Hiroshi Harada (Research & Development Scientist)

22.2 Research Activities

We are involved in the technical development of an innovative high performance computing infrastructure (HPCI) that will connect the major Japanese supercomputers installed in universities and research institutes, including the K computer, through a network. The computing environment realized by HPCI will meet the diverse needs of its intended users. Furthermore, the high-speed network will enable the supercomputers, shared storage, and other functionalities of each HPCI system provider to operate as a common HPCI. To this end, we are working on the technical side of the system operation, the technical coordination of the facilities comprising HPCI system, and the operation of the HPCI Shared Storage (HPCI-SS). The HPCI-SS is a large-scale distributed file system comprising 4 meta-data and 40 file-data servers. The HPCI-SS forms the data-sharing infrastructure of the HPCI led by the Ministry of Education, Culture, Sports, Science and Technology. It provides a 22.5 PB single-view “Gfarm” network file system that is open to all HPCI users and supports scalable I/O performance in a geographically widely distributed environment. The HPCI-SS meta-data servers are located in RIKEN AICS and the University of Tokyo (U-Tokyo), while the file-data servers are located in AICS, U-Tokyo, and the Tokyo Institute of Technology (Tokyo Tech). HPCI-SS also supports single sign-on authentication (GSI-AUTH), by which users can exchange their computational results between their local storage systems and HPCI-SS without additional authentication. HPCI-SS has developed into a key infrastructure for the management and sharing of computational data in HPCI projects.

22.3 Research Results and Achievements

22.3.1 Technological Planning and Coordination

For the Technological Planning and Coordination processes listed below, we offer total technical support across the entire HPCI system operation and coordination for other HPCI system providers that are in the same operating environment.

- Oversee discussion tables for HPCI system providers.
- Conduct investigations and countermeasures for technical problems that arise in HPCI system operations.

- Review software improvements that are related to the entire HPCI system.

We maintain the HPCI system operation which needs to accommodate various requests from users of those services. In addition we work with the University of Tokyo, in the National University Corporation to maintain the HPCI shared storage system. We provide data sharing storage for the community and computational resources for the pre-post processing of data.

As shown in Figure 22.1, the computational resources in the HPCI system are provided by the supercomputer centers of nine universities, RIKEN AICS and the Japan Agency for Marine-Earth Science and Technology (JAMSTEC). The Institute of Statistical Mathematics (ISM) has provided its resources since July 2014.

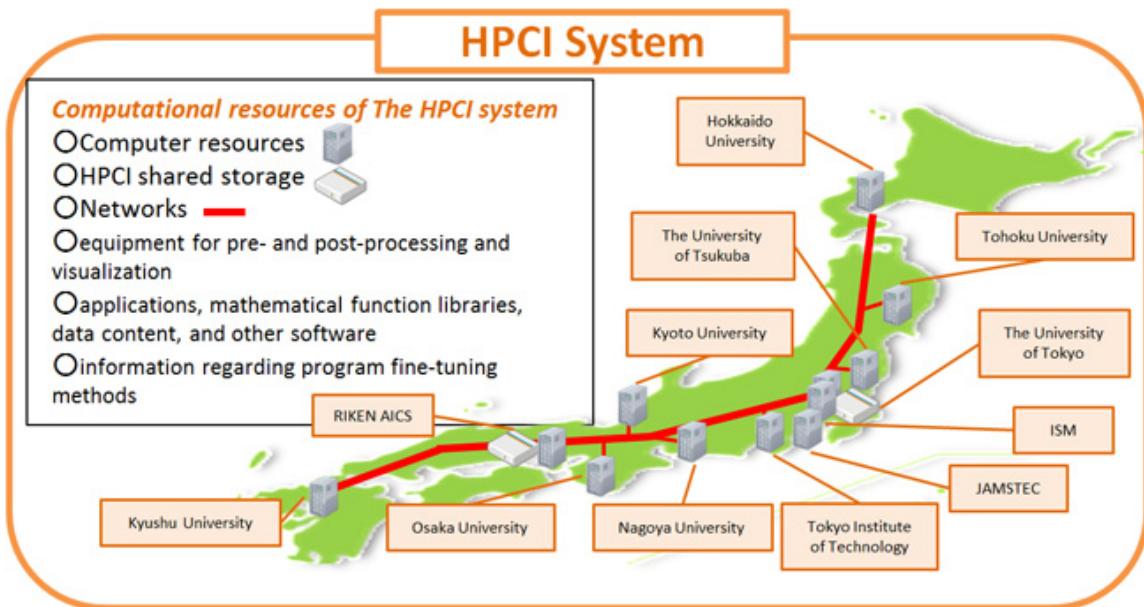


Figure 22.1: Overview of the HPCI Structure.

In FY2015, we coordinated and managed meetings for HPCI system providers and other relevant bodies.

- We hosted three meetings of the HPCI Cooperation Service Committee.
- We hosted ten meetings of the HPCI Cooperation Service and Operation Working group.
- We hosted three meetings of the HPCI Shared Storage Operation team.

22.3.2 HPCI System Monitoring

HPCI system monitoring allows us to monitor the status of the computing resources provided by each HPCI system provider, the operational status of the network, and the bandwidth and throughput performance. The HPCI system monitoring has the following duties:

- To investigate the causes of and response measures to technical problems that occur during the operation of the HPCI system.
- To research improvements in software related to the operation of the overall HPCI system.
- To operate and maintain the HPCI shared storage site (the West site);
- To operate and maintain the management of the Login Gateway and the HPCI IdP of AICS (one of the HPCI system provider);
- To develop the HPCI online application system;

- To screen institutions seeking to participate in the shared operation (i.e., seeking to provide computing resources to the HPCI system) to check whether they satisfy the technical requirements.

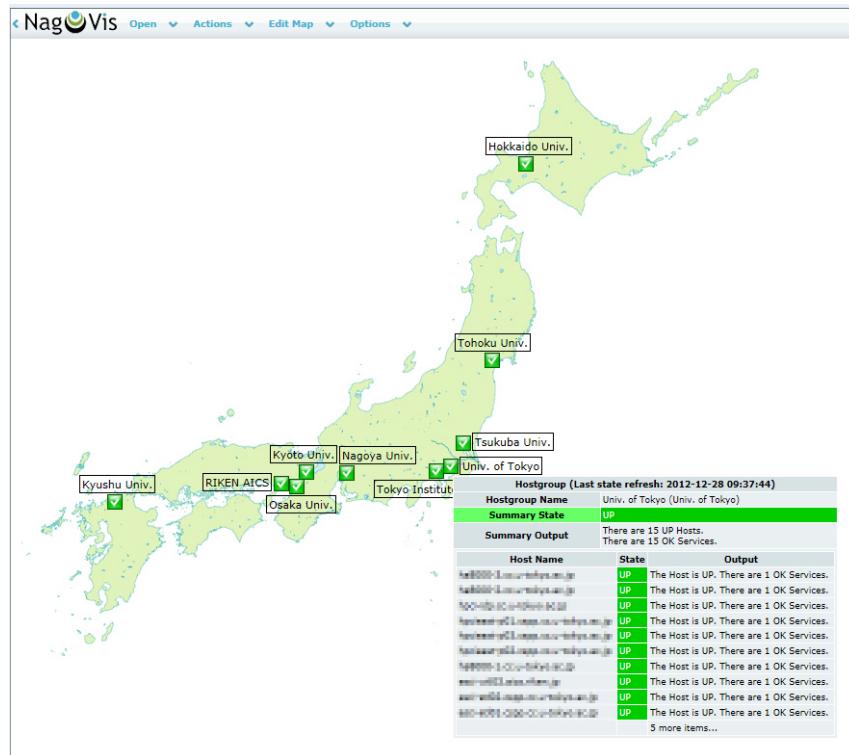


Figure 22.2: The overall Nagios/Nagvis monitoring system.

22.3.3 HPCI Shared Storage System

In FY2015, HPCI-SS was used by 70 HPCI projects and the total requested storage size was nearly two times the physical storage size. To cope with this demand, we determined a new resource allocation policy as follows.

- The maximum size of each project is set to 1.6 PB.
- The average number of file replicas is set to 1.5 (compared to 2 in FY2014).

Figure 22.4 shows the HPCI-SS disk usage from FY2014 to FY2015. As shown in these graphs, even though the HPCI-SS file usage is increasing, the disk usage is decreasing. This is the result of the new allocation policy.

Periodic Data Integrity Check for Data Corruption

HPCI-SS with its Gfarm file system supports data integrity checks. A file digest is calculated at the storage node before the data is written to storage and is stored in metadata. A file digest is also calculated when reading the entire data in a file. Data corruption can be detected by comparing the metadata digest to the file data digest via the gfspooldigest command. At the AICS site, data integrity checks are done on a weekly basis. All the created and updated files' data integrity checks can be completed within a week. The results of the periodic data integrity checks are managed on the HPCI web page.

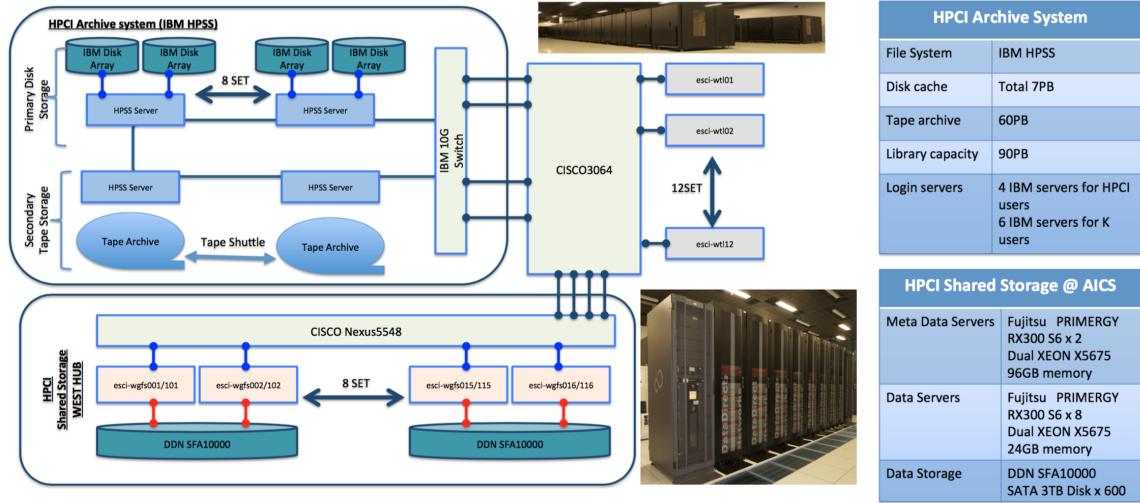


Figure 22.3: An overview of the HPCI Shared Storage and HPCI Archive Systems.

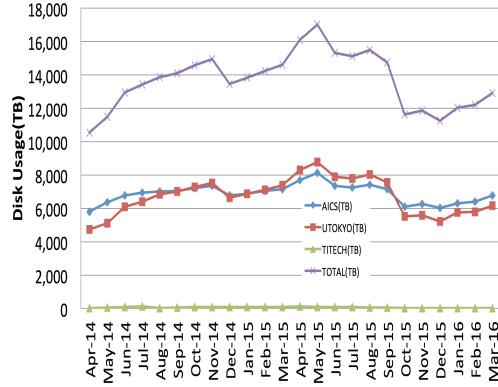


Figure 22.4: HPCI Shared Storage disk usage.

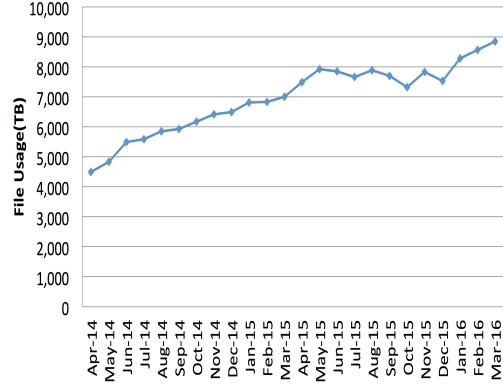


Figure 22.5: HPCI Shared Storage file usage.

Operational guidelines based on ITIL

To improve the standard level of our HPCI-SS system operations, we decided to introduce the ITIL methodology. First, we analyzed our operations using the ITIL methodology, and many improvements were made. Then we examined the service level, the quality of operation, and our relationships with our partners. Finally, a new HPCI-SS operational guideline was established. To satisfy our operational guideline, we evaluated our members' skills, documentations, personnel distributions, and management roles, in addition to multiple other points.

22.3.4 User Support

In collaboration with the U-Tokyo and Tokyo Tech teams, the HPCI System Development Team has developed additional user support system, such as HPCI user management and Gfarm consulting services.

- 700 or more scientists were using HPCI-SS by the end of FY2015. User support is provided through the HPCI Help desk and the hpci-tech@aics.riken.jp mailing list.
- Technical information concerning HPCI-SS and HPCI-AS, including the system configuration, manuals, tips, services, and maintenance schedules, are available through the HPCI CMS website, Figure 22.8.

定期ファイル検査 (gfspooldigest) 管理表 updated by Jun Ebihara (2016/06/14)															
スプール	2015/5 月	6月	7月	8月	9月	10月	11月	12月	2016/1月	2月	3月	4月	5月	6月	
esci-epgfd101	21	22,29	6,13,20,27	3,9,17,24,31	9,14,21,28	5,12,19,26	4,16,23,30	3(全数チェック),7,14,21,27	4,11,18,25, (29)	1,8,15,22,29	7,14,21,28	4,11,18,25	1,9,16,23,30	6,13	
esci- epgfd102	21	22,29	6,13,20,27	3,9,17,24,31	9,14,21,28	5,12,19,26	4,16,23,30	7,14,21,27	4,11,18,25, (29)	1,8,15,22,29	7,14,21,28	4,11,18,25	1,9,16,23,30	6,13	
esci-epgfd103	21	22,29	6,13,20,27	3,9,17,24,31	9,14,21,28	5,12,19,26	4,16,23,30	7,14,21,27	4,11,18,25, (29)	1,8,15,22,29	7,14,21,28	4,11,18,25	1,9,16,23,30	6,13	
esci-epgfd104	21	22,29	6,13,20,27	3,9,17,24,31	9,14,21,28	5,12,19,26	4,16,23,30	7,14,21,27	4,11,18,25, (29)	1,8,15,22,29	7,14,21,28	4,11,18,25	1,9,16,23,30	6,13	
esci-epgfd105	21	22,29	6,13,20,27	3,9,17,24,31	9,14,21,28	5,12,19,26	4,16,23,30	7,14,21,27	4,11,18,25, (29)	1,8,15,22,29	7,14,21,28	4,11,18,25	1,9,16,23,30	6,13	
esci-epgfd106	21	22,29	6,13,20,27	3,9,17,24,31	9,14,21,28	5,12,19,26	4,16,23,30	7,14,21,27	4,11,18,25, (29)	1,8,15,22,29	7,14,21,28	4,11,18,25	1,9,16,23,30	6,13	
esci-epgfd107	21	22,29	6,13,20,27	3,9,17,24,31	9,14,21,28	5,12,19,26	4,16,23,30	7,14,21,27	4,11,18,25, (29)	1,8,15,22,29	7,14,21,28	4,11,18,25	1,9,16,23,30	6,13	
esci-epgfd108	21	22,29	6,13,20,27	3,9,17,24,31	9,14,21,28	5,12,19,26	4,16,23,30	7,14,21,27	4,11,18,25, (29)	1,8,15,22,29	7,14,21,28	4,11,18,25	1,9,16,23,30	6,13	

Figure 22.6: History of the data integrity checks.

(6) HPCI-SS [†]															
項目番号	内容	作業頻度	近藤	芝野	尾崎	金山	公式関連ドキュメント	ローカルマニュアルの有無							
SS-01	Logwatch確認	毎日	○	×	○	主		有							
SS-02	Tripwire検知確認	毎日	○	×	○	主		有							
SS-03	Zabbix Alertメール確認	毎日	○	×	○	主		無							
SS-04	gfspooldigest実行・完了確認(管理表更新)	毎日	○	×	×	主		有							
SS-05	共用ストレージ運用部会 議事録作成(要旨・書き起こし担当)	毎四半期	△	×	○	主		有							
SS-06	共用ストレージ運用部会 議事録作成(書き起こし担当)	毎四半期	○	×	○	○		有							
SS-07	共用ストレージ運用部会 会議準備?	不定期	×	×	×	×		無							
SS-08	富士通 月例報告書のアップロード?	毎月	×	×	×	主		無							
SS-09	共用ストレージメンテナンススケジュール 作成	毎月	×	×	×	主		無							
SS-10	ホスト証明書・サービス証明書更新作業	毎年	○	×	×	主		有							
SS-11	ローカルアカウント登録作業: ログインノード(esci-wtl{01..04})	不定期	○	×	×	主		有							
SS-12	ローカルアカウント登録作業: ログインノード-「京」ユーザ向け-(escli-wtl{05..10})	不定期	○	×	×	主		有							
SS-13	ローカルアカウント削除作業: ログインノード(escli-wtl{01..10})	不定期	×	×	×	主		有							
SS-14	ローカル*テスト*アカウント登録作業: ログインノード(escli-wtl{01..10})	不定期	×	×	×	主		無							
SS-15	Gfarmアップデート	毎月	○	×	×	主		有							
SS-16	消失・破損ファイル確認	不定期	○	×	×	主		有							
SS-17	ログインノードのデフォルトシェルの変更	不定期	△	×	△	主		無							
SS-18	課題検討会議 月例報告資料作成	毎月	○	×	×	主		無							

Figure 22.7: The HPCI-SS skill map.

- The HPCI-SS usage for each project is conveyed to project leaders in our monthly reports.
- Technical consulting services on HPCI-SS are provided by Prof. Tatebe, U-Tokyo, and Tokyo-Tech.

22.3.5 HPCI Archive System for long-term data archives

By the end of FY2015, a total of 4.4 PB of computational data from 43 HPCI research projects was archived in the HPCI Archive system. In addition, the computational data of 8 expired research projects were archived to comply with a Research Organization for Information Science and Technology (RIST) request.

Figures 22.9 – 22.11 show the total usage, file-counts, and average-size of the HPCI Archive system. In addition, Figures 22.12 and 22.13 illustrate the file counts and size distributions on the HPCI Archive System.

As shown in the graphs, the total usage and file counts are increasing, and the average file size is decreasing. Large numbers of small files can cause performance degradation. Therefore, the average file size must be carefully monitored. Currently, the number of files smaller than 100 MB occupy 9.89 % of the archive.



Figure 22.8: An HPCI-SS user support page.

22.3.6 HPCI Data Analyzer West

Another HPCI computational resource is the 88-node Linux cluster called “Data Analyzer West (DAW)”. As shown in Figure 22.14, the DAW’s 88 computational nodes and 2 servers are connected via an InfiniBand network. An Intel compiler, OpenMPI, mvapich2, and GridEngine job scheduler are also available in the DAW environment. In FY2015, this cluster was made available to two HPCI research projects. Figure 22.15 plots the monthly execution node-time occupied by each project and Figure 22.16 shows the job consumption rate. DAW was used for a total of 163% of the allocated node-time in FY2015.

In FY2015, two HPCI research projects were chosen to use DAW, and the total allocated node-time was approximately 40% of DAW capacity. It is calculated that most of the computational nodes remained in an idle state and that electric power was wasted. To reduce the power consumption of idle nodes, we introduced an automatic idling stop operation to the DAW computational nodes in September 2015. In automatic idling stop operation, idle computational nodes are shutdown automatically, and when jobs that need more nodes are entered in the scheduler queue, down state nodes are booted up to execute the job. Figure 22.17 shows the DAW electric power consumption; the average monthly electric power consumption was 17472 kwh in normal operation and 7876 kwh in idling stop operation. As a result of the idling stop operation, we achieved an energy saving of 55%.

22.4 Schedule and Future Plan

22.4.1 Considerations for the second HPCI term

HPCI is a common base infrastructure that promotes innovative computational science technology in Japan. It is necessary to maintain and develop this function in cooperate with the user communities and the HPCI system providers in the future. In the second term, we need to maximize our achievements, not merely by continuing the improvements and operations of the first term but

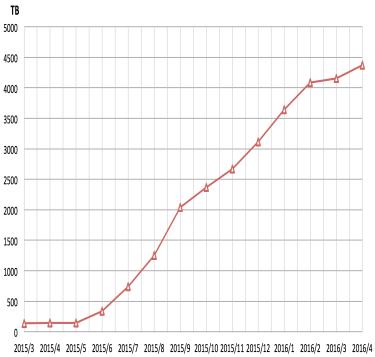


Figure 22.9: Usage of the HPCI Archive System.

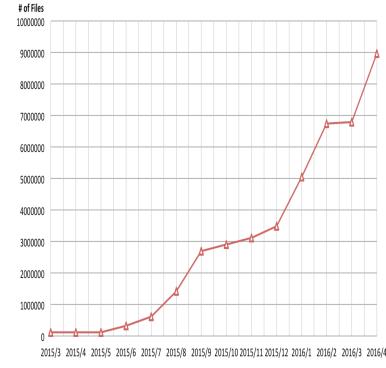


Figure 22.10: Number of HPCI Archive System files.

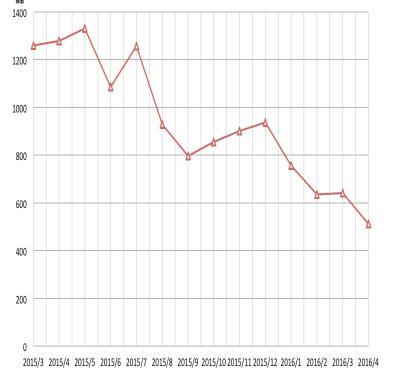


Figure 22.11: Average file size in the HPCI Archive System.

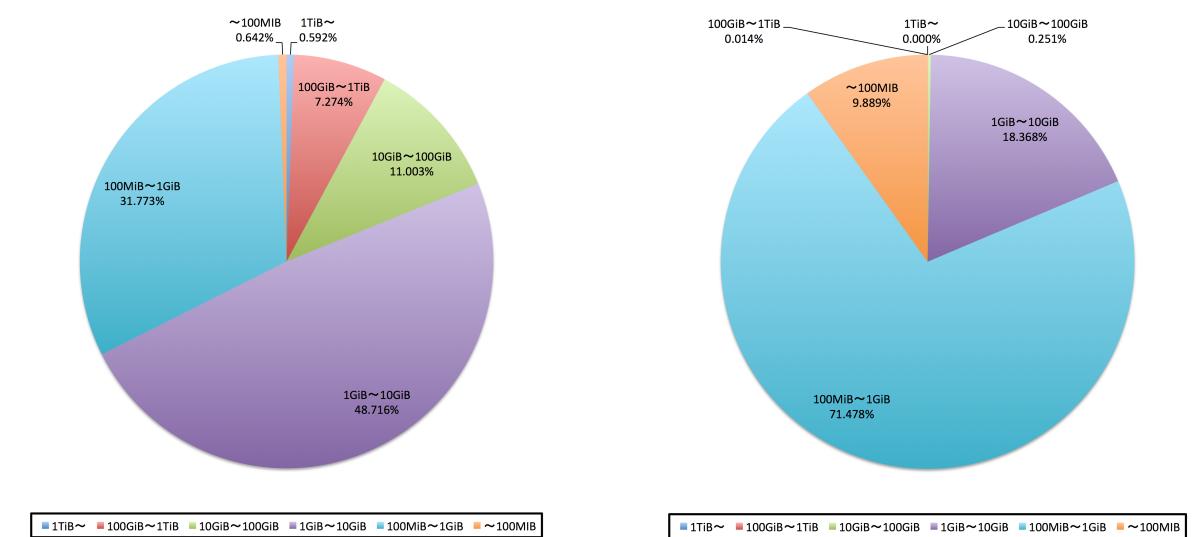


Figure 22.12: File size distribution in the HPCI Archive System.

Figure 22.13: File count distribution in the HPCI Archive System.

also by accommodating the agendas that appeared in the first term and are newly occurring in the second term.

22.4.2 Optimization of HPCI system operations

In the coming second term, the big challenge is how to maximize our achievements and offer them to the public even though the computational resource structure is changing dynamically and the budget has tightened. We need to optimize the operations based on our experiences in the first term. In particular, compared to the first term, we need to comprehensively understand the contents of the implementations within each organization to objectively see a comprehensive picture of an operation and evaluate it in a timely fashion, and to relay these findings to the operation. Therefore, in the second term, to achieve the most relevant operational improvements, it is necessary to reaffirm the relationship between the HPCI Consortium, Japanese government, and the HPCI system providers and to implement the PDCA cycle more appropriately for HPCI operations.

22.4.3 Technological Planning and Coordination Affairs

In the first term of the Technological Planning and Coordination Affairs, the HPCI cooperation service committee was established and operated as a place to make decisions concerning common

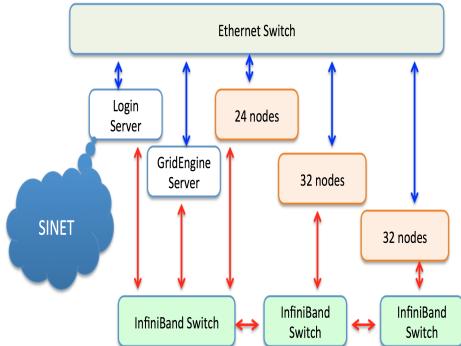


Figure 22.14: The components of DAW.

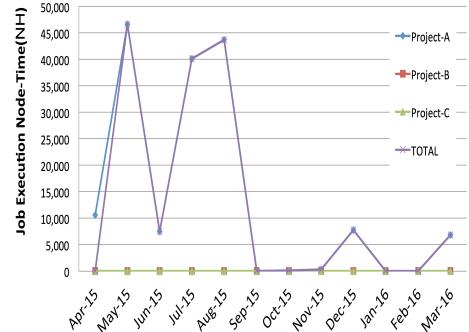


Figure 22.15: The monthly execution node-time for DAW.

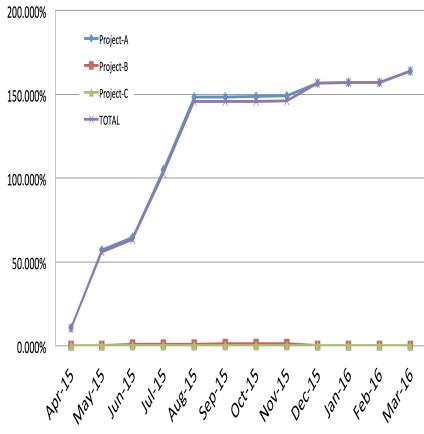


Figure 22.16: DAW job consumption rate.

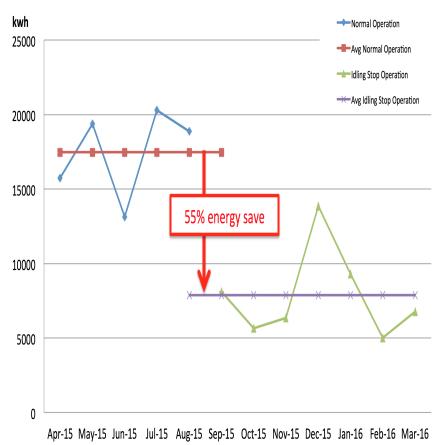


Figure 22.17: DAW electric power consumption.

services with members of the project implementing agency and the HPCI system providers. It is important to develop the role of this committee and to use it as a place to summarize and share information conducive to the PDCA cycle. For example, information that should be summarized, includes information concerning the amount, usage plans, and progress of resources, excepting the individual information of users for each project; the status of the achievements and problems of HPCI operation affairs; and the opinions and requests from users through the help desk that each project implementing agency received. For the HPCI cooperation service committee, to achieve the most relevant HPCI operation, it is important to share such information and promote cooperation with the HPCI Consortium and government for the planning and checking of the PDCA cycle.

Part III

Flagship 2020 Project

Chapter 23

Flagship 2020 Project

23.1 Project Overview

The Japanese government launched the FLAGSHIP 2020 project ¹ in JFYI 2014 whose missions are defined as follows:

- Building the Japanese national flagship supercomputer, the successor to the K computer, which is tentatively named the post K computer, and
- developing wide range of HPC applications that will run on the post K computer in order to solve the pressing societal and scientific issues facing our country.

RIKEN accepted a request by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to assume the responsibility of being the institution responsible for the research and development of the post K computer. In the summer of 2014, the Japanese Government selected nine Priority Issues to be targeted by the post K computer, and formulated project organizations that would be responsible for research toward solving them. The government also selected four research areas which are named Exploratory Challenging Issues to be developed. Table 23.1 lists the nine Priority Issues to be tackled by the FLAGSHIP 2020 project and institutions selected to lead the research for solving them. Table 23.2 lists four Exploratory Challenging Issues and selected institutions.

RIKEN AICS is in charge of co-design of the post K computer and development of application codes in collaboration with the Priority Issue institutes, as well as research aimed at facilitating the efficient utilization of the post K computer by a broad community of users. Under the co-design concept, AICS and the selected institutions are expected to collaborate closely. For the post K computer system development, in September 2015, Fujitsu Ltd. was selected to produce the basic design. A number of university partners are partially involved with the design.

23.2 AICS Teams

As shown in Table 23.3, four development teams are working on post K computer system development with the FLAGSHIP 2020 Planning and Coordination Office that supports development activities. The primary members are listed in Section 23.6.

The Architecture Development team designs the architecture of the post K computer in cooperation with Fujitsu and designs and develops a productive programming language, called XcalableMP (XMP), and its tuning tools. The team also specifies requirements of standard languages such as Fortran and C/C++ and mathematical libraries provided by Fujitsu.

The System Software Development team designs and specifies a system software stack such as Linux, MPI and File I/O middleware for the post K computer in cooperation with Fujitsu and designs and develops multi-kernel for manycore architectures, Linux with light-weight kernel (McKernel), that provides a noise-less runtime environment, extendability and adaptability for future application demands. The team also designs and develops a low-level communication layer to provide scalable, efficient and portability for runtime libraries and applications.

¹ FLAGSHIP is an acronym for Future LAcency core-based General-purpose Supercomputer with HIgh Productivity.

Table 23.1: Priority Issues

Health and longevity	1. Innovative computing infrastructure for drug discovery	RIKEN Quantitative Biology Center, and other institutions
	2. Personalized and preventive medicine using big data	Institute of Medical Science / the University of Tokyo, and five other institutions
Disaster prevention / Environment	3. Integrated simulation systems induced by earthquake and tsunami Earthquake	Research Institute, the University of Tokyo and 4 other institutions
	4. Meteorological and global environmental prediction using big data	JAMSTEC / Center for Earth Information Science and Technology of Japan, and three other institutions
Energy issues	5. New technologies for energy creation, conversion / storage, and use	Institute for Molecular Science / National Institute of Neural Science, and eight other institutions
	6. Accelerated development of innovative clean energy systems	School of Engineering / the University of Tokyo, and eight other institutions
Industrial competitiveness enhancement	7. Creation of new functional devices and high-performance materials	The Institute of Solid State Physics / the University of Tokyo, and eight other institutions
	8. Development of innovative design and production processes	Institute of Industrial Science / the University of Tokyo, and six other institutions
Basic science	9. Elucidation of the fundamental laws and evolution of the universe	Center for Computational Science / Tsukuba University, and seven other institutions

Table 23.2: Exploratory Challenges

10. Frontiers of basic science - challenge to the limits -	Institute for Material Research / Tohoku University, and nine other institutions
	Tokyo Woman's Christian University, and three other institutions
	School of Engineering / the University of Tokyo
11. Construction of models for interaction among multiple socioeconomic	RIKEN, Advanced Institute for Computational Science and twelve other institutions
	Tokyo University of Science and three other institutions
12. Elucidation of the birth of exoplanets (Second Earths) and the environmental variations of planets in the solar system	Kobe University Graduate School of Science and eight other institutions
13. Elucidation of how neural networks realize thinking and its application to artificial intelligence	Okinawa Institute of Science and Technology Graduate University and 4 other institutions
	The Research Center for Advanced Science and Technology / the University of Tokyo

Table 23.3: Development Teams

Team Name	Team Leader	Number of Members
Architecture Development	Mitsuhisa Sato	XX
System Software Development	Yutaka Ishikawa	XX
Co-Design	Junichiro Makino	
Application Development	Hirofumi Tomita	XX

Table 23.4: Target Applications

Program	Brief Description
GENESIS	Molecular dynamics simulation for protein in solvent (MD)
Genomon	Human Genome processing (Genome alignment)
GAMERA	Earthquake simulation system (FEM multigrid method)
NICAM+LETKF	Weather prediction system using Big data (structured grid & ensemble Kalman filter)
NTChem	Ab-initio molecular electronic structure (RI-MP2 method)
ADVENTURE	Computational mechanics system for large scale analysis and design (unstructured grid)
RSDFT	Ab-initio simulation based on density functional theory (real space DFT method)
FFB	FEM flow analysis around the complex geometry (Large Eddy Simulation)
LQCD	Lattice QCD simulation (structured grid Monte Carlo)

The Co-Design team leads to optimize architectural features and application codes together in cooperation with AICS teams and Fujitsu. It also designs and develops an application framework, FDPS (Framework for Developing Particle Simulator), to help HPC users implement advanced algorithms.

The Application Development team is a representative of nine institutions aimed at solving Priority Issues. The team figures out weakness of target application codes in terms of performance and utilization of hardware resources and discusses them with AICS teams and Fujitsu to find out best solutions of architectural features and improvement of application codes.

23.3 Co-design with applications from Priority Issues

During the development stage, emphasis will be placed on co-design concepts that allow the system and application development processes to work harmoniously together so that the societal and scientific issues can be solved effectively and early achievements can be attained. Since increasing power consumption is a critical issue in the design of the next-generation large-scale supercomputer, it will be important for the co-design to make trade-offs between energy/power, cost, and performance by taking application characteristics into consideration. It should be noted that we design the entire system using a proprietary processor and interconnect, in collaboration with Fujitsu, rather than choosing existing components to construct the system.

Table 23.4 shows a set of target applications provided from each area of the nine Priority Issues. System design is being performed using these applications as a target in the co-design process.

23.4 Target of System Development and Basic Design

The post K's design targets are as follows:

- A one hundred times speed improvement over the K computer is achieved in maximum case of some target applications. This will be accomplished through co-design of system development and target applications for the nine Priority Issues.
- The maximum electric power consumption should be between 30 and 40 MW.

In 2014, the basic design of the system was completed. The AICS decided to adopt a general-purpose many-core architecture rather than using accelerators, such as graphics processing units (GPUs), in order to support a wider range of applications. At the time of writing, Fujitsu has just announced that the node processor is based on an ARM architecture with Fujitsu HPC extensions that was invented for the K computer. To ensure the performance of large-scale parallel applications remained compatible with the K computer, we choose a similar topology for the interconnect, Tofu interconnect (6D mesh/torus).

The major components of system software shown in Figure 23.1 have been decided to be developed. These are summarized as follows:

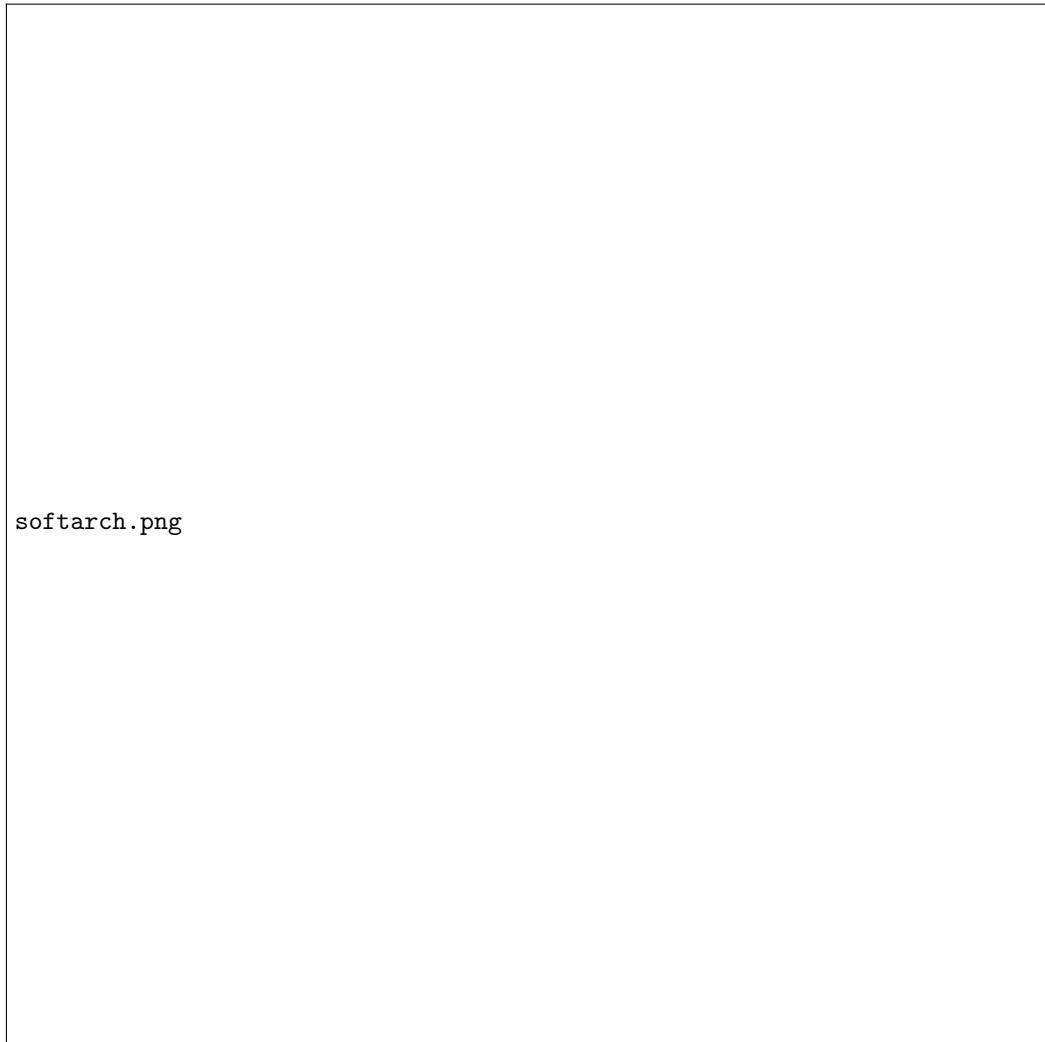


Figure 23.1: Software Architecture

- Highly productive programming language, XcalableMP
XcalableMP (XMP) is a directive-based PGAS language for large scale distributed memory systems that combines HPF-like concept and OpenMP-like description with directives. Two memory models are supported: global view and local view. The global view is supported by the PGAS feature, i.e., large array is distributed to partial ones in nodes. The local view is provided by MPI-like + Coarray notation.
- Domain specific library/language, FDPS
FDPS is a framework for the development of massively parallel particle simulations. Users only need to program particle interactions and do not need to parallelize the code using the MPI library. The FDPS adopts highly optimized communication algorithms and its scalability has been confirmed using the K computer.
- MPI + OpenMP programming environment
The current de facto standard programming environment, i.e., MPI + OpenMP environment, is supported. Two MPI implementations are being developed. Fujitsu continues to support own MPI implementation based on the OpenMPI. RIKEN is collaborating with ANL (Argonne National Laboratory) to develop MPICH, mainly developed at ANL, for post K computer.
- New file I/O middleware
The post K computer does not employ the file staging technology for the layered storage system. The users do not need to specify which files must be staging-in and staging-out in their job scripts in the post K computer environment. Asynchronous I/O and caching technologies are fully utilized in the post K computer in order to provide transparent file access with better performance.
- Application-oriented file I/O middleware
In scientific Big-Data applications, such as real-time weather prediction using observed meteorological data, a rapid data transfer mechanism between two jobs, ensemble simulations and data assimilation, is required to meet their deadlines. Keeping a file I/O interface, such as netCDF, direct data transfer between two jobs without involving a storage system is being designed and implemented.
- Multi-Kernel for manycore architectures
Multi-Kernel, Linux with light-weight Kernel (McKernel) is being designed and implemented. It provides: i) a noiseless execution environment for bulk-synchronous applications, ii) ability to easily adapt to new/future system architectures, e.g., manycore CPUs, a new process/thread management, a memory management, heterogeneous core architectures, deep memory hierarchy, etc., and iii) ability to adapt to new/future application demands, such as Big-Data and in-situ applications that require optimization of data movement.

It should be noted that these components are not only for post K computer, but also for other manycore-based supercomputer, such as Intel Xeon Phi.

23.5 Schedule

At the time of writing, the new development plan has been announced by the government. As shown in Figure 23.2, the design and prototype implementations will be done before the end of 2019, and the system will be deployed after this phase. The service is expected to start public operation at the range from 2021 to 2022.

23.6 Members

Primary members are only listed.

23.6.1 System Software Development Team

Yutaka Ishikawa (Team Leader)

Masamichi Takagi (Senior Scientist)



Figure 23.2: Schedule

Atsushi Hori (Research Scientist)
Balazs Gerofi (Research Scientist)
Masayuki Hatanaka (Research & Development Scientist)
Takahiro Ogura (Research & Development Scientist)
Tatiana Martsinkevich (Postdoctoral Researcher)
Fumiyo Shoji Research & Development Scientist
Atsuya Uno Research & Development Scientist
Toshiyuki Tsukamoto Research & Development Scientist

23.6.2 Architecture Development

Mitsuhisa Sato (Team Leader)
Yuetsu Kodama (Senior Scientist)
Miwako Tsuji (Research Scientist)
Hidetoshi Iwashita (Research & Development Scientist)
Jinpil Lee (Postdoctoral Researcher)
Tetsuya Odajima (Postdoctoral Researcher)
Hitoshi Murai (Research Scientist)
Toshiyuki Imamura (Research Scientist)

23.6.3 Application Development

Hirofumi Tomita (Team Leader)
Yoshifumi Nakamura (Research Scientist)
Hisashi Yashiro (Research Scientist)
Seiya Nishizawa (Research Scientist)
Hiroshi Ueda Research (Scientist)
Yukio Kawashima (Research Scientist)
Naoki Yoshioka (Research Scientist)
Yiyu Tan (Research Scientist)
Soichiro Suzuki Research & Development Scientist
Kazunori Mikami Research & Development Scientist

23.6.4 Co-Design

Junichiro Makino (Team Leader)
Keigo Nitadori (Research Scientist)
Yutaka Maruyama (Research Scientist)
Takayuki Muranushi (Postdoctoral Researcher)

23.7 Publications