# 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 [25, 26, 28].



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 KMTHLIB 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 [1, 7], 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) [27].

### 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 algorithmi, 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. Base 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  $32 \times 10^6$  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 DSYGVD, 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 [16] and EPASA [17] as oral and poster presentations in 2015, respectively.



Figure 3.3: The execution time of DSYGVD 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 [3], and the divide and conquer part at EPASA2015 [14].

#### 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 [6].

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[4], and Communication Hiding (CH) technique was also presented at SIAM LA 2015 [19]. 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).



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 [2].

## 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 matrixvector multiplication (GEMV and SYMV) routines on CUDA [24, 29]. In this FY2015-2016, we have extended the study to other memory-bound linear algebra kernels [10, 5]. 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 threadblock 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 [8]. By eliminating waste data movement in the computation using the shorter floating-point formats, we expect



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.

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 [30].

### 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), 'Highperformance 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 threedimensional 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 torelance,

- 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 [23]. 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

## **Journal Articles**

- Jaewoon Jung et al. "Parallel implementation of 3D FFT with volumetric decomposition schemes for efficient molecular dynamics simulations". In: *Computer Physics Communications* 200 (2016), pp. 57–65.
- [2] Yoshiharu Ohi and Soichiro Ikuno. "Numerical Investigation of Electromagnetic Wave Propagation Phenomena by Three-Dimensional Meshless Time-Domain Method". In: *Plasma and Fusion Research* 10.3406072 (2015), pp. 3406072–1, 3406072–4.

# **Conference Papers**

- [3] Takeshi Fukaya and Toshiyuki Imamura. "Performance Evaluation of the Eigen Exa Eigensolver on Oakleaf-FX: Tridiagonalization Versus Pentadiagonalization". In: Proceedings of the Parallel and Distributed Processing Symposium Workshop (IEEE IPDPS, PDSEC 2015). 2015, pp. 960– 969.
- [4] Toshiyuki Imamura et al. "CAHTR: Communication-Avoiding Householder Tridiagonalization". In: Proceedings of ParCo2015, Advances in Parallel Computing. Vol. 27: Parallel Computing: On the Road to Exascale. 2016, pp. 381–390.
- [5] Toshiyuki Imamura et al. "Implementation and Evaluation of SYMV·GEMV kernels on multiple-GPU's". In: *IPSJ SIG Technical Reports HPC151*, Vol. 2015-HPC-151 (in Japanese). 2015.
- [6] Yosuke Kumagai et al. "Performance Analysis of the Chebyshev Basis Conjugate Gradient Method on the K Computer". In: Proceedings of PPAM2015, Lecture Notes in Computer Science (LNCS). Vol. 9573. 2016, pp. 74–85.
- [7] Seikichi Matsuoka et al. "Quality and Performance of a Pseudo-Random Number Generator in Massively Parallel Plasma Particle Simulations". In: Proceedings of Mathematics and Computations, Supercomputing in Nuclear Applications and Mote Carlo International Conference (M&C+SNA+MC2015). Vol. 2. 2015, pp. 923–935.
- [8] Daichi Mukunoki and Toshiyuki Imamura. "Automatic selection of the number of threads for memory-bound BLAS kernels on an NVIDIA GPU". In: IPSJ SIG Technical Reports HPC152, Vol. 2015-HPC-152 (in Japanese). 2015.
- [9] Daichi Mukunoki, Toshiyuki Imamura, and Daisuke Takahashi. "Fast Implementation of General Matrix-Vector Multiplication (GEMV) on Kepler GPUs". In: Proceedings of Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP 2015). 2015, pp. 642–650.
- [10] Daichi Mukunoki, Toshiyuki Imamura, and Daisuke Takahashi. "Proposal of a Reduced-length Floating Point Format". In: IPSJ SIG Technical Reports HPC150, Vol. 2015-HPC-150 (in Japanese). 2015.

[11] Susumu Yamada, Toshiyuki Imamura, and Machida Machida. "High Performance Eigenvalue Solver in Exact-diagonalization Method for Hubbard Model on CUDA GPU". In: *Proceedings* of *ParCo2015*, Advances in Parallel Computing. Vol. 27: Parallel Computing: On the Road to Exascale. 2016, pp. 361–369.

## **Invited** Talks

- [12] Toshiyuki Imamura. *Present and Future of the EigenExa library*. Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing (ATAT2016). 2016.
- [13] Daichi Mukunoki, Toshiyuki Imamura, and Daisuke Takahashi. Automatic Thread-Block Size Adjustment for Dense Matrix-Vector Multiplication on CUDA. Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing (ATAT2016). 2016.

## **Posters and Presentations**

- [14] Takeshi Fukaya and Toshiyuki Imamura. Performance evaluation of the divide-and conquer method in the EigenExa eigensolver. Poster presentation at International Workshop on Eigenvalue Problems: Algorithms; Software and Applications, in Petascale Computing (EPASA2015). 2015.
- [15] Takeshi Fukaya, Yusaku Yamamoto, and Toshiyuki Imamura. Performance Perspectives of a Dense Eigenvalue Solver on pos-Petascale computers. Poster presentation at HPCS2015 (in Japanese). 2015.
- [16] Yusuke Hirota and Toshiyuki Imamura. Divide-and-Conquer Method for Symmetric-Definite Generalized Eigenvalue Problems of Banded Matrices on Manycore Systems. Oral presentation at SIAM Conference on Applied Linear Algebra (SIAM LA15). 2015.
- [17] Yusuke Hirota and Toshiyuki Imamura. Performance of Divide-and-Conquer Method for Symmetric-Definite Generalized Eigenvalue Problems of Banded Matrices on Multicore and Manycore Systems. Poster presentation at International Workshop on Eigenvalue Problems: Algorithms; Software and Applications, in Petascale Computing (EPASA2015). 2015.
- [18] Toshiyuki Imamura. Numerical Libraries for post-Peta Scale Supercomputer Systems. Oral presentation at US-Japan Joint Institute for Fusion Theory (JIFT) Workshop on Innovations and co-designs of fusion simulations towards extreme scale computing. 2015.
- [19] Toshiyuki Imamura. Performance Analysis of the Householder Back-transformation with Asynchronous Collective Communication. Oral presentation at SIAM Conference on Applied Linear Algebra (SIAM LA15). 2015.
- [20] Daichi Mukunoki, Toshiyuki Imamura, and Daisuke Takahashi. Implementation and Performance Evaluation of 'MUBLAS', numerical linear algebra kernels facilitated with automatic selection mechanism of thread configuration on a GPU. Poster presentation at GTC Japan 2015 (in Japanese). 2015.
- [21] Yoshiharu Ohi and Soichiro Ikuno. Stability analysis on nodes arrangement in Meshless Time-Domain Method. Poster presentaion at 25th International TOKI Conference (ITC25). 2015.
- [22] Yoshiharu Ohi et al. Implementation of Numerical Software Framework, KMATHLLIB on K computer. Poster presentation at annual meeting of JSIAM 2015 (in Japanese). 2015.
- [23] Narimasa Sasa et al. Accumulated numerical error on a time-evolutional PDE problem by using FFT. Oral presentation at annual meeting of JSIAM 2015 (in Japanese). 2015.

## Patents and Deliverables

- [24] ASPEN.K2. http://www.aics.riken.jp/labs/lpnctrt/ASPENK2\_e.html.
- [25] *EigenExa*. http://www.aics.riken.jp/labs/lpnctrt/EigenExa\_e.html.
- [26] KMATH\_EIGEN\_GEV. http://www.aics.riken.jp/labs/lpnctrt/KMATH\_EIGEN\_GEV\_e. html.

- [27] KMATHLIB\_API. http://www.aics.riken.jp/labs/lpnctrt/KMATHLIB\_API.html.
- [28] KMATH\_RANDOM. http://www.aics.riken.jp/labs/lpnctrt/KMATH\_RANDOM\_e.html.
- [29] MUBLAS-GEMV. http://www.aics.riken.jp/labs/lpnctrt/ASPENK2\_e.html.
- [30] *SLFP*. http://www.aics.riken.jp/labs/lpnctrt/members/mukunoki/slfp-0.0.1.tgz.