[1] Diego Ugarte (RIKEN R-CCS), Shoji Takada (Department of Biophysics, Graduate School of Science, Kyoto University) and Yuji Sugita (RIKEN CPR, RIKEN R-CCS, RIKEN BDR). *Molecular dynamics simulation of large multi-component lipid bilayers in GENESIS using the iSoLF coarse-grained model.*

Abstract. Molecular dynamics (MD) simulation of lipidic systems is an active research topic due to the many biological phenomena that occur on cellular membranes. Applying this technique, however, presents some difficulties when studying processes that occur at large scale or require long simulation times with traditional all-atom models. Thus, coarse-grained (CG) models have been developed as one possible solution, trading off structural detail with speed. In this work, we have extended the implicit solvent lipid force field, iSoLF, to represent several lipid molecules. We have implemented it in the GENESIS MD software and evaluated its applicability to large multi-component membrane environments.

Keywords: Molecular Dynamics, Coarse-grained, GENESIS, Lipid bilayer

[2] Jaewoon Jung (RIKEN R-CCS, RIKEN CPR), Cheng Tan (RIKEN R-CCS), Chigusa Kobayashi (RIKEN R-CCS), Diego Ugarte (RIKEN R-CCS) and Yuji Sugita (RIKEN CPR, RIKEN R-CCS, RIKEN BDR). *Acceleration of residue-level coarse-grained molecular dynamics on Fugaku by efficient parallelization with improved load balancing.*

Abstract. Molecular dynamics (MD) is one of the powerful tools in investigating various biological phenomena. One of the main concerns in MD is the difficulties in reaching the physically meaningful time scale. Residue-level coarse-grained (CG) models can be an excellent way to extend the time scale despite insufficient accuracy. However, CG has yet to be tried for very huge-scale biomolecules because of difficulties in computational parallelization from nonuniform particle density. In this presentation, we suggest a new parallelization scheme of domain decomposition with good load balancing for CG. The new method is implemented in GENESIS MD software, which speeds up 10~100 folds of the existing programs for a huge system like chromatin on the Fugaku supercomputer.

Keywords: Molecular Dynamics, Coarse-grained model, parallelization

[3] Weitong Ren (RIKEN CPR), Hiromasa Yagi (RIKEN BDR), Takanori Kigawa (RIKEN BDR) and Yuji Sugita (RIKEN CPR; RIKEN BDR; RIKEN R-CCS). *Illuminating the correlation of substrate binding cooperativity and hPGK domain motion by atomistic MD simulations.*

Abstract. Phosphoglycerate kinase (PGK) is a well-conserved enzyme that catalyzes the reversible phosphoryl transfer from 1,3-bisphosphoglycerate (1,3-BPG) to MgADP and produces the MgATP and 3-phosphoglycerate (3PG) in the glycolytic pathway. Usually, the protein exists in two distinct states, termed open and closed states respectively. A large-scale open-to-closed conformational transition is required to bring the two substrates together in the right orientation and promote phosphoryl transfer. Recent NMR experiment in cells revealed the opposite cooperativity of substrates binding for human PGK in which the binding of 3PG and MgAMPPNP (a nonhydrolyzable analog to MgATP) show positive cooperativity while the binding of 3PG and MgADP exhibit negative cooperativity. In addition, the authors found that single point mutation D374N leads to positive binding cooperativity of 3PG and MgADP. But the molecular mechanism of such different cooperativities of substrates binding is still ambiguous. Also, it's uncertain how the mutation D374N changes the binding cooperativity of 3PG and MgADP. Meanwhile, it's still an open question whether the cooperativity of substrate binding is correlated with the protein allosteric domain motion. In this work, we intended to answer these questions through extensive atomistic molecular simulations of wild-type and D374N mutated hPGK ternary complexes. We identified undetected half-closed states of various hPGK ternary complexes from the simulations and found that the positive cooperativity of 3PG and MgADP binding to D374N variant is very likely related to the formation of the more compact half-closed state. Structural analysis of the half-closed conformations suggested that the Mg2+ coordination, the water-mediated inter-substrate interaction, and also the fine-tuned interaction between K215 and substrates contribute to the formation of the half-closed states. Furthermore, our simulations hinted that the half-closed states are very likely to present as intermediate states of the hPGK open-to-closed conformational transition. Based on all these results, we proposed that the substrate-regulated conformational change of hPGK proceeds via a two-step process involving first a thermal flucutation-driven quick transition from open to half-closed states followed by the substrates induced local structural rearrangment to the fully-closed state. The cooperativity of substrates binding at the half-closed states enables the sophisticated regulation of hPGK domain movement to retain high enzyme specificity and efficiency. Such a picture that substrate binding cooperativity and protein allosteric motion are correlated through the half-closed intermediate states could serve as a general mechanism for protein activity regulation.

Keywords: Molecular dynamics, Protein conformational transition, Substrate binding cooperativity

[4] Jorji Nonaka (RIKEN Center for Computational Science), Masaaki Terai (RIKEN Center for Computational Science), Masahiro Nakao (RIKEN Center for Computational Science), Keiji Yamamoto (RIKEN Center for Computational Science), Hitoshi Murai (RIKEN Center for Computational Science), Fumiyoshi Shoji (RIKEN Center for Computational Science) and Naohisa Sakamoto (Kobe University). *Large Data Visualization Environment on the Fugaku.*

Abstract. We have worked on large data visualization and analysis environment on the Fugaku taking into consideration some lessons learned from the K computer. In this poster, we will present the current visualization environment on the Fugaku, where users can execute traditional batch-based visualization, via CLI (Command Line Interface), on both compute nodes and pre/post environment. In addition, interactive visualization on GUI (Graphical User Interface)-based applications can also be executed on the pre/post environment via SSH port forwarding. This functionality can also be applied to execute client/server-based distributed visualization. To further facilitate this kind of usage, we took advantage of the Fugaku VPN service to get rid of the necessity for setting the SSH port forwarding and facilitating the use of client/server-based visualization. This VPN service also facilitates the connection to the license servers running on the Fugaku front-end servers for running commercial visualization applications such as the aforementioned AVS/Express. GNOME desktop-based remote visualization is also provided on the pre/post environment via VNC (Virtual Network Computing) and can benefit from the Fugaku VPN service. Although it is still in the beta test mode, with the introduction of the Fugaku Open OnDemand service, users are currently able to access an even easy-to-use remote desktop (Xfce) environment for carrying out visualization and analysis tasks on the Fugaku. Apart from those already existing visualization-oriented applications and tools for the traditional post-hoc visualization, we have also been working on the in-situ visualization, which has recently received increasing attention as a feasible and effective approach for the visualization and analysis of the ever-increasing scale and complexity of the simulation, especially on the leading-edge HPC systems. We have focused on optimizing the image-based in-situ visualization approach by reducing the number of generated image files, which we named as Smart In-Situ Visualization. We have worked on some underlying techniques for realizing smart in-situ visualization, and we are planning to practically apply the smart in-situ visualization onto a WRF-based in-situ visualization being developed under the RIKEN-MOST collaborative research activities.

Keywords: Visualization, In-Situ Processing, Fugaku Environment

[5] Ryo Kanada (RIKEN Center for Computational Science), Kei Terayama (Graduate School of Medical Life Science, Yokohama City University), Atsushi Tokuhisa (RIKEN Center for Computational Science) and Yasushi Okuno (RIKEN Center for Computational Science). *Efficient Conformational Sampling with an Adaptive Coarse-Grained Elastic Network Model using Bayesian Optimization.*

Abstract. Coarse-Grained simulation based on structure based model are not suitable for efficient sampling structures that depart significantly from the initial structure without any biased force. We developed a new adaptive CG elastic network model, in which the efficient parameter tuning for model is conducted by machine learning methods (Bayesian Optimization) based on short-time AA-MD of at most ns order. To evaluate the performance of adaptive CG-ENM, we applied the new methodology to several proteins such as glutamine binding protein (GBP) in the apo state. The results showed that the structural ensembles explored by adaptive CG-ENM could be significantly more diverse than those by conventional ENMs with enhanced sampling such as temperature replica exchange MD and long-time AA-MD.

Keywords: Coarse-Grained Model, enhanced sampling, Bayesian Optimization

[6] Wei Huang (RIKEN Center for Advanced Intelligence Project (AIP)), Peng Chen (National Institute of Advanced Industrial Science and Technology (AIST)) and Haonan Wang (University of Illinois Urbana-Champaign). *Deep Active Learning by Leveraging Training Dynamics.*

Abstract. Active learning theories and methods have been extensively studied in classical statistical learning settings. However, deep active learning, i.e., active learning with deep learning models, is usually based on empirical criteria without solid theoretical justification, thus suffering from heavy doubts when some of those fail to provide benefits in applications. In this paper, by exploring the connection between the generalization performance and the training dynamics, we propose a theory-driven deep active learning method (dynamicAL) which selects samples to maximize training dynamics. In particular, we prove that convergence speed of training and the generalization performance is positively correlated under the ultra-wide condition and show that maximizing the training dynamics leads to a better generalization performance. Further on, to scale up to large deep neural networks and data sets, we introduce two relaxations for the subset selection problem and reduce the time complexity from polynomial to constant. Empirical results show that dynamicAL not only outperforms the other baselines consistently but also scales well on large deep learning models. We hope our work inspires more attempts in bridging the theoretical findings of deep networks and practical impacts in deep active learning applications.

Keywords: Machine Learning, Active Learning, Deep Learning

[7] Chigusa Kobayashi (R-CCS), Hisham M. Dokainish (RIKEN CPR), Suyong Re (National Institutes of Biomedical Innovation, Health and Nutrition), Takaharu Mori (RIKEN CPR), Jaewoon Jung (R-CCS, RIKEN CPR) and Yuji Sugita (R-CCS, RIKEN CPR, RIKEN BDR). *Analysis of structural changes of multi-chain/multi-domain proteins.*

Abstract. Recent advances in the field of structural biology have revealed the structures of large proteins under various physiological conditions. In addition, Supercomputers like Fugaku and the high-speed algorithms that run on them have dramatically expanded the range of possible research for biomolecules in both time and space. It has become clear that correlated motions between domains and chains in many protein systems play an essential role in biological function. However, analyzing such extensive data and extracting the motions is still challenging. We propose an analysis method that uses the geometric average of domains to extract critical structural changes. We apply this method to the cryo-EM structures of the S-protein of COVID-19 and those from gREST simulations on Fugaku to identify significant structural changes.

Keywords: Structural changes, multi-chain/multi-domain proteins, Cryo-EM structures, Molecular Dynamics Simulation

[8] Ryutaro Tsuji (Tohoku University/RIKEN Center for Computarional Science), Ken-Ichi Ishikawa (Hiroshima University), Yasumichi Aoki (RIKEN), Yoshinobu Kuramashi (University of Tsukuba), Shoichi Sasaki (Tohoku University), Eigo Shintani (University of Tsukuba) and Takeshi Yamazaki (University of Tsukuba). *Precision computation of nucleon structure from lattice QCD with all-mode-averaging .*

Abstract. We present the results of nucleon structure studies measured in 2+1 flavor QCD with the physical light quarks in a large spatial extent of about 10 fm. Our calculations are carried out with the PACS10 gauge configurations generated by the PACS Collaboration with the six stout-smeared O(a) improved Wilson fermions and Iwasaki gauge action at beta=1.82 and 2.00 corresponding to the lattice spacing of 0.085 fm (coarser) and 0.063 fm (finer) respectively. When we compute nucleon two-point and three-point functions, the all-mode-averaging technique is employed to reduce statistical errors significantly without increasing computational costs. At both lattice spacings, we evaluate appropriate quantities representing the nucleon structure.

Keywords: Lattice QCD, Proton/Neutron structure, High-precision determination, First-principal calculation

[9] Tatsumi Aoyama (University of Tokyo), Issaku Kanamori (RIKEN R-CCS), Kazuyuki Kanaya (University of Tsukuba), Hideo Matsufuru (KEK), Yusuke Namekawa (Hiroshima University) and Keigo Nitadori (RIKEN R-CCS). *Benchmark result of Lattice QCD code set Bridge++ 2.0 on Fugaku.*

Abstract. Lattice QCD (LQCD) is a physics application to treat the interaction among quarks. It is one of typical HPC applications on massive parallel system. We have been developed a general purpose lattice QCD code set Bridge++ [1] and its new version contains an optimization for A64FX systems like supercomputer Fugaku. In this presentation, we show the benchmark results of Bridge++ on Fugaku. The bottleneck of LQCD application is solving linear equations, Dx = b, where fermion matrix D is a large sparse matrix and its operation is a stencil computation on four-dimensional space-time lattice. We apply iterative algorithms to solve this equation. Therefore, the performance of D multiplication is crucially important. The shape of matrix D is not unique and Bridge++ has implementation of several types of D that are widely used in the LQCD simulations. The benchmark result covers the performance of the following types of D: Wilson, Clover, Staggered, Domainwall,and their site even-odd preconditioned version. In the implementation, we adopt so-called Array of Structure of Array (AoSoA) data structure to use the SIMD feature of A64FX, and the lattice site degrees of freedom is vectorized. We use 2-dimensional tiling of the lattice sites for the SIMD vectorization. The kernel codes are written using the Arm C-Language Extension (ACLE). The communication to exchange the boundary data is overlapped with the bulk computations. More details of the implementation for Fugaku are found in [2] and [3]. As mixed precision schemes are often used in the iterative solvers, we implement the fermion matrix D in both double- and single- precisions. The performance of D multiplication is around 400 GFlops/node in the single precision. We observe a very good weak scaling up to 512 nodes, which is the largest benchmark we tried. We also observe a good weak scaling of iterative BiCGstab (or CG) solvers. As an exaample, we show the scaling of the Clover type D in Figure 1. References
[1] Lattice QCD code Bridge++, https://bridge.kek.jp/lattice-code/.
[2] Tatsumi Aoyama, Issaku Kanamori, Kazuyuki Kanaya, Hideo Matsufuru and Yusuke Namekawa, to appear PoS LATTICE2022 (2023) 284, https://doi.org/10.22323/1.430.0284.
[3] Issaku Kanamori, Keigo Nitadori and Hideo Matsufuru, to appear in International Conference on High Performance Computing in Asia-Pacific Region Workshops (HPCASIA-WORKSHOP 2023), February 27-March 2, 2023, Raffles Blvd, Singapore. ACM, New York, NY, USA, 10 pages. [doi:10.1145/3581576.3581610]

Keywords: Lattice QCD, stencil computation, Fugaku

[10] Yasumichi Aoki (RIKEN Center for Computational Science), ジスヌ ゴスワミ (RIKEN Center for Computational Science), Issaku Kanamori (RIKEN Center for Computational Science), Nakamura Yoshifumi (RIKEN Center for Computational Science) and Yu Zhang (RIKEN Center for Computational Science). *Reweighting of ensemables generated using monte carlo simulations in Lattice QCD.*

Abstract. The exploration of the structure of the phase diagram of strongly interacting matter in the temperature(T), chemical potential(μ_B) and quark mass (m) plane is an active field of research both in the experimental and theoretical nuclear physics communities with important phenomenological implications for heavy-ion collisions, early universe physics. To understand this complex phase structure, we generate huge number of ensembles using (Hybrid) monte carlo algorithm simulations and conduct measurements on those ensembles. This calculations require powerful, highly parallelized computer code as we need to perform multiple higher dimensional matrix-vector multiplications to generate a single ensemble with the help of powerful supercomputers. Thus, to obtain statistically significant results, a large number of ensembles must be generated, which can take several months or years. Reweighting is a technique for creating a new ensemble distribution from an existing one by adjusting one or more parameters. The process shifts the original distribution's histogram horizontally. For reweighting to be effective, there must be a significant overlap between the original and target histograms. In this poster, we will be using a specific type of reweighting called mass reweighting, on ensembles generated by the JLQCD collaboration that utilize 5-d Domain wall fermions \cite{MDWF:2023,MDWF:2023_2}. We will use the Bridge++: 2.0 \cite{Aoyama:2023jk} code base on the Fugaku supercomputer to calculate reweighting factors. We will first project the 5-d large sparse matrix to a 4-d large sparse matrix in order to perform reweighting. We will use realistic parameters and demonstrate when it is successful and when it fails. This is important because we only have a limited number of configurations available. Additionally, we will apply the reweighting method to real calculations and present the observables before and after reweighting.

Keywords: Monte carlo simulations, histogram, reweighting

[11] Masaaki Terai (RIKEN R-CCS). *Understanding Usage Trends of the Fugaku Pre/Post-Processing System and Improvements.*

Abstract. Historically, the pre/post-processing system has been installed as one of the peripheral servers of our supercomputer systems. Those servers have been designed to handle various pre/post-processing tasks (e.g., visualization, data conversion, and so on) to make input/output data, which are generated before/after performing users' calculations (e.g., parallel computation and simulation) on the main system. For the Fugaku, we procured the "Fugaku Pre/Post Environment" system, composed of nodes designed for specific purposes (with eight GPU nodes and two large memory nodes) and installed Slurm as the job manager. As the basic concept, high-performance node-parallel computations are recommended to be executed on the Fugaku main system. Therefore, the pre/post-processing system is intended to be used for versatile applications that use a small number of CPUs but hundreds of gigabytes of memory and GPUs. Also, while the main system employs an Arm-based processor, the pre/post-processing system is easy to run open-source software packages developed in the Intel architecture environment.

We have operated and maintained this pre/post-processing system since FY 2021, and currently an average of 80 unique users from around 30 unique groups per month have been actively using this system. Due to the limited number of nodes, the job queue waiting time basically tends to increase when a large number of jobs are submitted. During the busy season in FY 2021, the average waiting time exceeded 33 hours for the GPU nodes, and 30 hours for the large memory nodes due to the overcrowded job queues. Since the waiting time highly depends on the running time of other submitted jobs, in FY 2022, we attempted to improve the situation by changing the job queue configuration. Firstly, to reduce the maximum elapsed time of a job, we divided the resources into a short time job queue with a maximum of three hours and a long-time job queue with a maximum of 24 hours, and modified the default resources (e.g., number of CPUs, memory size, and elapsed time). In addition, to facilitate job scheduling optimization (e.g., backfill scheduling), we changed the operational policy that promotes explicitly specifying the required resources by a user at the job submission. If a user does not specify the resources, the user's job is assigned a minimum amount of resources (e.g., a single SMT/Core and elapsed time of up to 30 min.) as default.

As a result, in both the long/short-time job queues of the GPU nodes, we observed that the waiting time decreased or remained the same in most months. On the other hand, in large memory nodes, while we observed that the waiting time decreased in the short-time job queue, the waiting time remained the same or increased in the long-time job queue. The large memory system consists of two nodes, and each queue is assigned to only a single node. Therefore, the waiting time tends to be extremely long when a large number of jobs are submitted at once. To address the issue, we procured an additional large memory node and planned to become operational in FY 2022. This poster report on the current usage trends of the Fugaku pre/post-processing system, as well as recent efforts to improve job waiting time in the system.

Keywords: pre/post-processing, Slurm, Fugaku, operational data analysis

[12] Cheng Tan (RIKEN Center for Computational Science), Jaewoon Jung (RIKEN Center for Computational Science; RIKEN Cluster for Pioneering Research), Chigusa Kobayashi (RIKEN Center for Computational Science), Diego Ugarte La Torre (RIKEN Center for Computational Science), Shoji Takada (Kyoto University) and Yuji Sugita (RIKEN RCCS; RIKEN CPR; RIKEN BDR). *Implementation of Residue Level Coarse-Grained Models in GENESIS for Large-Scale Molecular Dynamics Simulations.*

Abstract. Residue-level coarse-grained (CG) models have become one of the most popular tools in biomolecular simulations to achieve modeling accuracy and computational efficiency. Unifying treatments of proteins and nucleic acids and efficient parallel computations are indispensable to exploring large-scale biological phenomena in molecular dynamics (MD) simulations with CG models. We implement several residue-level CG models in the GENESIS MD software, covering structure-based and context-based potentials for both well-folded biomolecules and intrinsically disordered regions. An amino acid residue in a protein is represented as a single CG particle centered at the Cα atom position. At the same time, a nucleotide in RNA or DNA is modeled with three beads. Then, a single CG particle depicts around ten heavy atoms in proteins and nucleic acids. The input data in CG MD simulations are treated as GROMACS-style input files generated from a newly developed toolbox, GENESIS-CG-tool. To optimize the performance in CG MD simulations, we employ multiple neighbor lists, each attached to a different nonbonded interaction potential in the cell-linked list method. We found that random number generations for Gaussian distributions in the Langevin thermostat are one of the bottlenecks in CG MD simulations. Accordingly, we parallelize the computations with the message-passing interface (MPI) to improve the performance on PC clusters or supercomputers. We simulate Herpes simplex virus (HSV) type 2 B-capsid and chromatin models containing more than 1,000 nucleosomes in GENESIS as examples of large-scale biomolecular simulations with residue-level CG models. This framework extends accessible spatial and temporal scales by multi-scale simulations to study biologically relevant phenomena, such as genome-scale chromatin folding or phase-separated membrane-less condensations. (C. Tan et al. PLoS Comput. Biol. 18, 4, (2022), e1009578 https://doi.org/10.1371/journal.pcbi.1009578)

Keywords: molecular dynamics, coarse-grained, LLPS, liquid-liquid phase separation, biomolecules

[13] Xin Lu (Iwate University) and Yiyu Tan (Iwate University). *Probability Density Function Estimation for a Transformed Normal Random Vector.*

Abstract. When a normal random vector is transformed by small fractional Hadamard powers, the probability density function (PDF) of the transformed random vector derived by the change-of-variable technique does not exactly fit its distribution. Whether the normal random vector is near or far from zero, the change-of-variable technique inevitably causes more numerical errors to occur in the PDF estimation. To reduce these errors, this paper proposes a method based on the Fourier transform to estimate PDFs, which allows the PDFs to more appropriately fit the transformed distributions.

Keywords: probability density function, change-of-variable technique, Fourier transform

[14] Daichi Mukunoki (RIKEN Center for Computational Science), Atsushi Suzuki (RIKEN Center for Computational Science) and Toshiyuki Imamura (RIKEN Center for Computational Science). *Multiple and Mixed Precision BLAS with C++ Template.*

Abstract. In this poster, we will present a Basic Linear Algebra Subprograms (BLAS) library, for multiple and mixed precision computation using C++ template, which is under development and called "tmblas" as an abbreviation templated mixed precision BLAS.
Numerical computations have been using only IEEE 754 binary32 (FP32) or binary64 (FP64) formats. Nowadays, however, various formats are introduced, e.g., binary16 (FP16) and binary128 (FP128) were added in IEEE 754-2008, and some environments support them, such as FP16 in A64FX, FP128 in POWER9 and in compiler emulation by GNU/LLVM. In addition, new low-precision formats are being implemented in hardware, mainly for use in deep learning, e.g. bfloat16 (BF16), TensorFloat-32 (TF32). Besides, mixed-precision computations, where several different precision are used in a single code, have also been investigated to improve computational speed and energy efficiency by using as low a precision as possible. However, many of the existing libraries can perform computations by one of FP32 or FP64, independently.
For BLAS, major implementations only provide conventional routines that support mono precision interfaces in FP32 or FP64. On the other hand, there have been some discussions about supporting multiple and mixed precision computations, and some extended implementations are developed. For example, MPBLAS in MPLAPACK supports multiple precision by combining existing high-precision arithmetic libraries such as MPFR (but the mixed-precision interface is not supported); XBLAS supports mixed precision interface by mixing different precision operands and supporting arithmetic precision higher than the operand precision (FP64 or 106-bit mantissa with double-double (DD) arithmetic); BLAS++ supports mixing different precision operands, and its implementation in a C++ template facilitates the introduction of arbitrary data types (although only FP32 and FP64 are instantiated for now).
We are developing a multiple and mixed precision BLAS to support combinations of precision for input/output arguments and internal arithmetic operations. Our implementation is based on BLAS++ code, with decoupling precision of operator and operands, which is realized by explicit description of operators with data type of the operands to exclude implicit conversions of data during operations. Now the code instantiates subroutines with FP16/BF16, FP32, FP64 and FP128/DD as input and output data and operations, which also support operations in one level higher precision than the data precision.
Since our idea is rather simple, we are expecting the development of tmblas should be a simple task. In reality, however, several challenges became apparent during development. For example, some routines require loop reordering and/or additional working arrays to support mixed precision operations. Such code changes and conversions between precision types for mixed-precision operations may introduce some bugs. Therefore, the code validation becomes an important issue. Since the current implementation aims a reference one without optimization, a way to develop optimized version on different processors and parallelized one in shared and distributed environments is also a major challenge for the future.
By this poster, we would like to discuss with researchers from different backgrounds the validity of our new BLAS design and to share the challenges in its development.

Keywords: BLAS, mixed-precision, multiple-precision

[15] Amarjit Singh (RIKEN, R-CCS) and Kento Sato (RIKEN R-CCS). *Research and Development of an Infrastructure for Big Data Collection, Analysis, and Utilization in Large Scale Research Facilities.*

Abstract. Big data is a term for vast datasets, more complicated structure for collecting, analyzing, and utilizing or visualizing for further processing. The process of research into huge amount of data are further considered for big data analysis. For big data computation, data needs to be analyzed and executed accurately. Due to the rapid growth of big data, it is required to be studied and evaluated from different types of data behavior.
In the big data, compression of information plays a vital role in it. Data compression is broadly used by the community. A variety of data has transferred for preprocessing, training, and analysis. Large intermediate I/O transfer can make these phases slower. To accelerate the series of the processes, different data compression algorithms can be applied on the intermediate data. But existing compression algorithms show a low compression ratio for such kind of data with evenly distributed values. The purpose of this research proposal is to further establish "Big Data Infrastructure" to realize efficient data collection, analysis, and utilization between the large synchrotron radiation facility (Spring-8)/ the X-ray free electron laser facility (SACLA) and the flagship supercomputer "Fugaku" The RIKEN's supercomputer "Fugaku", "Spring-8/SACLA" are positioned as leading-edge large-scale research facilities that support Society 5.0. It is essential to establish a new shared platform to promote the maintenance and utilization of these facilities, as well as to conduct research and development on this platform.
In this poster presentation, we introduce the background, the motivation and the plans of our on-going research project called research and development of an infrastructure for collecting, analyzing, and utilizing big data in large-scale research facilities. The objective of this project is to establish a "big data infrastructure between Spring-8/SACLA and Fugaku. The project consists of four sub-projects: 1) Data preprocessing platform: Data conversion and preprocessing at the hardware level using FPGAs to efficiently store experimental data obtained from sensors in temporary storage; 2) Data compression and transfer infrastructure: Development of data compression and transfer infrastructure for high-speed transfer of experimental data on temporary storage to Fugaku; 3) Data analysis infrastructure: To develop an infrastructure (workflow tools and deep learning framework) to efficiently analyze data collected on Fugaku; 4) Data utilization infrastructure: To build a data utilization infrastructure for the wide-ranging utilization of collected primary data and analysis results.
As a part of sub-project(2), we developed TEZip (Time evolutionary data compression tool) which targets time evolutionary data. The time evolutionary data is a collection of consecutive frame images. Video data is one of the examples. Time evolutionary data set is captured and transferred to internal and / or external computers for further analysis in Spring-8/SACLA. TEZip makes use of temporal localities appeared in time evolutionary data and applies a neural network technique to achieve higher compression ratio. PredNet is such a neural network which is trained to predict the future movement of pixels and objects in time evolutionary data. As per our project planned in FY2023, we are planning to further improve compression ratio of TEZip and propose new domain-specific compression algorithms specialized for X-ray CT data captured in Spring-8. Our poster presentation also present the status in sub-project(2).

Keywords: Big Data, Data Compression, Time Evolutionary Data, TEZip, Neural Networks

[16] Jumpei Mano (Dept. of Computer Science, School of Sci.&Tech., Meiji University), Takaaki Miyajima (Dept. of Computer Science, School of Sci.&Tech., Meiji University), Peng Chen (National Institute of Advanced Industrial Science and Technology), Mohamed Wahib (Processor Research Team, RIKEN R-CCS) and Kentaro Sano (Processor Research Team, RIKEN R-CCS). *A study on optimizing Back Projection Computation using FPGA.*

Abstract. X-ray CT (Computed Tomography) image reconstruction has become one of the essential medical technologies in today's healthcare. CT is a medical imaging technology that creates tomographic images from X-ray transmission acquired by a scanner. The FBP (Filtered Back Projection) method is a mainstream computational method in CT. It is a mathematical method to restore a cross-sectional image from measurement data. In CT, an image reflecting the distribution of attenuation coefficients can be obtained. The FBP method is a computational bound application since the computational complexity is O(N4). CPUs are the mainstream to perform the FBP method, and the power consumption of a typical portable CT device is 225-450W. Although GPUs can improve computing performance, the power consumption of GPUs is 300-400 W. So the power consumption of the portable CT device with GPUs will be about 1,000 W. Power consumption is another bottleneck for the potable CT device. Chen's implementation is chosen as a concrete example in this study, and a preliminary evaluation is conducted to understand the current bottleneck and consider speed-up for FPGAs. FPGAs can be more suitable than CPUs since FPGAs have been considered a low power consumption. It has been found that the bottleneck of this method is the memory access and computation problems associated with bilinear interpolation. The ultimate goal of our research is to speed up the method of FPGAs. We believe that FPGAs are more suitable than GPUs or CPUs for the FBP method.

Keywords: CT, FPGA, FBP

[17] Tan Yiyu (Iwate University), Xin Lu (Iwate University), Guanghui Liu (Cedars-Sinai Medical Center), Peng Chen (National Institute of Advanced Industrial Science and Technology) and Yusuke Tanimura (National Institute of Advanced Industrial Science and Technology). *High-Performance Sound Field Auralization .*

Abstract. Sound field auralization is compute- and memory-intensive. In sound field auralization, highly accurate room impulse response is critical to achieve precise auralization results. In this research, an FPGA-based high-performance sound field auralization system is investigated, in which a sound field renderer is provided to compute room impulse response accurately. When a sound space is a three-dimensional shoebox with dimension being 16m×8m×8m, after implemented by using an FPGA card DE10-Pro, the FPGA-based sound field renderer achieves 11 times performance gain over the software simulation carried out on a desktop machine with 512 GB DRAMs and a Xeon Gold 6212U processor running at 2.4 GHz in the case of the hardware-oriented FDTD algorithm being applied to analyze the room impulse response, even though the it runs at about 350 MHz.

Keywords: FPGA, Sound field auralization, Sound field rendering

[18] Yasumitsu Maejima (RIKEN Center for Computational Science) and Takemasa Miyoshi (RIKEN Center for Computational Science). *A Control Simulation Experiment for August 2014 Severe Rainfall Event Using a Regional Model.*

Abstract. Torrential rainfall is a threat in the modern society. To predict severe weather, convection resolving numerical weather prediction (NWP) is effective. This study explores a Control Simulation Experiment (CSE) aimed at controlling precipitation amount and locations to potentially prevent catastrophic disasters by simulating different scenarios of interventions of small perturbations taking advantage of the chaotic nature of dynamics. In this study, we perform a CSE using a regional model SCALE-RM for a severe rainfall event which caused catastrophic landslides and 77 fatalities in Hiroshima, Japan on August 19 and 20, 2014.
We perform a 1-km-mesh, hourly-update, 50-member observing system simulation experiment (OSSE) for this rainfall event initialized at 0000 UTC August 18. This provides the initial conditions for a 6-hour ensemble forecast initialized at 1500 UTC August 19. To create small perturbations to change the nature run, we take the differences of all model variables between an ensemble member having the heaviest rain and another ensemble member having the weakest rain. Moreover, we normalize the perturbations so that the maximum wind speed is 0.1 m s-1. In this preliminary CSE, we try to control the heavy rainfall by giving the perturbations to the nature run in the OSSE at each time step from 1500 UTC to 1600 UTC on August 19, although the perturbations for all variables at all grid points are something beyond human's engineering capability. In the nature run, 6-hour accumulated rainfall amount from 1500 UTC to 2100 UTC reaches 210 mm at peak. By contrast, the rainfall amount decreases to 118 mm in the CSE. We plan to apply limitations to the perturbations.

Keywords: Data assimilation, Numerical weather prediction, Meso-scale meteorology

[19] Keijiro Fujita (Kobe University), Naohisa Sakamoto (Kobe University), Takanori Fujiwara (Linkoping University), Toshiyuki Tsukamoto (RIKEN Center for Computational Science) and Jorji Nonaka (RIKEN Center for Computational Science). *Log Data Visual Analytics System for Analyzing Cooling System Behavior of HPC System.*

Abstract. The size and complexity of supercomputer systems and their power and cooling facilities have continuously increased, thus posing additional challenges for long-term and stable operations. Supercomputers are shared computational resources and usually operate with different computational workloads at different locations (space) and timings (time). It is worth noting that an Operational Data Analytics (ODA) framework has been implemented on the Fugaku, which allows for the continuous monitoring, archiving, and analysis of the gathered data from the system and sensors. There is also a publicly available Grafana-based dashboard (https://status.fugaku.r-ccs.riken.jp/) for verifying part of the gathered data. This poster focuses on the analysis of the archived data to gain knowledge for ameliorating the operation of the HPC system and its facility. We have also focused on the vast amount of data gathered from the K computer operational period trying to better understand the HPC system's heat generation and cooling behavior, as well as the possible correlation of the generated heat with critical hardware failures. For the latter purpose, a visual analytics framework named MulTiDR was developed to enable interactive exploration of the log data from the K computer system. A single-day log data is composed of 1,163 different measurements, from each of the 864 compute racks, sampled every 5 minutes, which will correspond to 1.4 billion elements for a single day. It has proven highly useful for the detailed analysis of log data from a short period of time (up to a few days). To enable visual analysis over a longer period of time, we developed a dimensionality reduction-based visual analytics system to capture characteristic spatiotemporal features and behaviors of the cooling system during the operation. We believe that this system can complement the MulTiDR system by narrowing down the period of time and spatial locations for further analysis. Figure 1 shows a case study using year-long log data (April 1, 2016 to March 31, 2017) from the K computer system and its facility. The selected log data correspond to 361-day operations after excluding the scheduled maintenance period (from October 6 to October 9). We analyze four different types of temperatures (CPU, Water, AirIn, AirOut temperatures) measured on each of 864 compute racks. From this case study, we observed that this visual analytics system can comprehensively handle temporal and spatial features of the HPC log data, and enables us to select the time and space, where the system behaved in a characteristic way. It is worth mentioning that the obtained results and assumptions we made require further analyses for validation, and there are some future works for further ameliorating the proposed system. As such future works, we also expect to apply to the log data generated from the Fugaku environment.

Keywords: Visualization, Visual Analytics, Log Data, HPC Facility, Cooling System

[20] Yu Zhang (RIKEN R-CCS), Yasumichi Aoki (RIKEN), Shoji Hashimoto (KEK), Issaku Kanamori (RIKEN Center for Computational Scinence), Takashi Kaneko (KEK) and Yoshifumi Nakamura (RIKEN). *Finite temperature QCD phase transition with 3 flavors of Möbius domain wall fermions from Lattice QCD .*

Abstract. We investigate the finite temperature QCD phase transition with three degenerate quark flavors using Möbius domain wall fermions. To explore the order of phase transition on the lower left corner of Columbia plot and if possible, to locate the critical endpoint we performed simulations at temperatures around 181 and 121 MeV with lattice spacing a= 0.1361(20) fm corresponding to temporal lattice extent Nt=8,12 with varying quark mass for two different volumes with aspect ratios Ns/Nt ranging from 2 to 3. By analyzing the volume and mass dependence of the chiral condensate, disconnected chiral susceptibility and Binder cumulant we find that there is a crossover at renormalized quark mass around 44 MeV in the MS bar scheme at 2 GeV for T~181 MeV, At temperature 121 MeV, the binder cumulant suggests a crossover at renormalized quark mass around 3.7 MeV in the MS bar scheme at 2 GeV, although a study of volume dependence would be important to confirm this.

Keywords: QCD phase transition, Möbius domain wall fermions, Lattice QCD

[21] Peng Chen (National Institute of Advanced Industrial Science and Technology (AIST)), Mohamed Wahib (RIKEN Center for Computational Science), Wei Huang (RIKEN Center for Advanced Intelligence Project), Yiyu Tan (Iwate University) and Yusuke Tanimura (National Institute of Advanced Industrial Science and Technology (AIST)). *High-performance Tomographic Imaging.*

Abstract. Tomographic imaging is a widely used technology that requires intensive computations. We will introduce our activity on high-performance image reconstruction using GPU-accelerated supercomputers. We take advantage of the heterogeneity of GPU-accelerated systems by overlapping the filtering computation and back-projection on CPUs and GPUs, respectively. We propose a novel decomposition scheme and reconstruction algorithm for distributed image reconstruction. This scheme enables arbitrarily large input/output sizes, eliminates the redundancy arising in the end-to-end pipeline and improves the scalability by replacing two communication collectives with only one segmented reduction. More specifically, we propose a distributed framework for high-resolution image reconstruction on state-of-the-art GPU-accelerated supercomputers and implement the proposed decomposition scheme in a framework that is useful for all current-generation tomographic imaging devices.

Keywords: HPC, Image Reconstruction, GPU

[22] James Taylor (RIKEN R-CCS Data Assimilation Team), Arata Amemiya (RIKEN Center for Computational Science), Yasumitsu Maejima (RIKEN), Shigenori Otsuka (RIKEN) and Takemasa Miyoshi (RIKEN). *Ensemble analysis of 30-minute precipitation forecasts using the real-time SCALE LETKF numerical weather prediction system using Fugaku.*

Abstract. We performed a demonstration of the first ever real-time numerical weather prediction (NWP) system in Summer 2021 during the Tokyo Olympic and Paralympic Games to provide up-to the minute extended 30 min precipitation forecasts for the Tokyo metropolitan region. During the demonstration, the system performed 30-s update cycling for a 500-m grid using observations from a new-generation multi-parameter phased array weather radar using priority access to the Fugaku supercomputer. This study examines the ensemble forecasts for a convective squall line that occurred during the demonstration, which brought heavy rainfall to parts of northern Tokyo and was well predicted by the rapid refresh NWP system. In addition, we compare the ensemble mean forecasts for this event to nowcasts performed with a simple advection model to highlight the advantage of the NWP system to predict heavy rainfall at short lead times.

Keywords: Data assimilation, Convective weather forecasts, radar observations

[23] Rakesh Teja Konduru (Data Assimilation Research Team, RIKEN Center for Computational Science, Kobe), Jianyu Liang (Data Assimilation Research Team, RIKEN Center for Computational Science, Kobe) and Takemasa Miyoshi (Data Assimilation Research Team, RIKEN Center for Computational Science, Kobe). *High-frequency microwave satellite radiances data assimilation using NICAM-LETKF in the OSSE framework.*

Abstract. This study investigates the impact of high frequency, such as 3-hourly and 1-hourly satellite microwave radiances, in global atmospheric data assimilation. To understand the impact of such a high-frequency satellite radiances data assimilation, we designed an observing system simulation experiment (OSSE) using the global NICAM-LETKF system at 56 km horizontal resolution. A free run was conducted with the NICAM model and treated as the reference (Nature) for the OSSE experiments. With the NICAM-LETKF system, we conducted five experiments, without data assimilation (NoDA), with only conventional data assimilation but not satellite radiances (NoSat), 6-hourly (6H), 3-hourly (3H), and 1-hourly (1H) satellite clear-sky radiances assimilation. The results showed that satellite microwave radiances assimilation improved the forecast of air temperature and wind over the global ocean compared to NoSat experiments. With the increase in the assimilation frequency of the satellite radiances, the air temperature and winds showed improvement in their representation over the ocean but degraded over land. Over the ocean, microwave radiances assimilation improved the typhoon eyewall wind intensities and its structure for 1H satellite radiances assimilation compared to 6H. These improvements in the wind intensities are prominent during the landfall stage of the typhoon. Forecasting landfall storms' strong winds is essential for disaster prevention and mitigation.

Keywords: NICAM-LETKF, OSSE, Microwave radiances, convection, data assimilation, Tropical cyclones, Satellite data, High frequency data

[24] Shun Ohishi (RIKEN Center for Computational Science (R-CCS) Data Assimilation Research Team), Takemasa Miyoshi (RIKEN Center for Computational Science (R-CCS) Data Assimilation Research Team) and Misako Kachi (JAXA Earth Observation Research Center). *LETKF-based Ocean Research Analysis (LORA).*

Abstract. Various atmosphere and ocean analysis datasets have been produced by operational centers and research institutions and are used for a variety of geoscience studies. In the Pacific region, multiple high-resolution ocean analysis datasets are now available, but no ensemble Kalman filter (EnKF)-based high-resolution analysis datasets exist to the best of our knowledge.
We have developed an eddy-permitting EnKF-based ocean data assimilation system in the western North Pacific (WNP) and Maritime Continent (MC) regions, in which satellite and in-situ observations are assimilated at a 1-day interval, shorter than a typical 5-day interval for the existing EnKF-based systems. We have demonstrated that the combination of three schemes [incremental analysis updates (IAU; Bloom et al. 1996), relaxation-to-prior perturbation (RTPP; Zhang et al. 2004), and adaptive observation error inflation (AOEI; Minamide and Zhang 2017)] results in significant improvement of geostrophic balance and analysis accuracy. With the recent enhancement of high-performance computing resources, we are now able to integrate high-resolution systems sufficient to resolve fronts and eddies. Therefore, this study creates high-resolution EnKF-based ocean analysis datasets in the WNP and MC regions and compares the accuracy with the existing analysis and observational datasets [JCOPE2M (Miyazawa et al. 2017) and AVISO (Ducet et al. 2000), respectively]. The dataset constructed in this study is referred to as LORA representing the Local ensemble transform Kalman filter (LETKF)-based Ocean Research Analysis. The validation results show that the accuracy of surface horizontal velocity in the LORA is lower in the subtropical region but higher in the mid-latitude region especially along the Kuroshio Extension (KE) compared with the JCOPE2M, although the accuracy in the AVISO is the best. The accuracy of sea surface temperatures in the LORA is better than the JCOPE2M almost everywhere in the domain. However, the validation results using an independent buoy south of the KE indicate that higher temperature accuracy in the LORA than the JCOPE2M might be limited to a vertical extent shallower than 300 m depth likely because of the prescribed vertical localization scale of 100 m. In summary, the LORA would be useful for a variety of geoscience research as well as fisheries, marine transports, and marine environment consultants.

Keywords: Ensemble Kalman filter, Ocean data assimilation, LORA

[25] Keita Watanabe (Kobe university), Naohisa Sakamoto (Kobe university), Jorji Nonaka (RIKEN R-CCS) and Yasumitsu Maejima (RIKEN R-CCS). *Large Ensemble Data Analysis Using Angular-based Edge Bundled Parallel Coordinates Plot.*

Abstract. Large-scale ensemble simulations have become widely used in various fields of science and engineering, especially in meteorological and climate science. Ensemble simulation executes multiple simulations with perturbed initial conditions, where each simulation is represented as an ensemble member (or member), and the set of simulation outputs is called ensemble data. Effective analysis of such ensemble data plays an important role in advancing the understanding of the underlying simulation phenomena [1]. It is worth noting that domain experts in meteorology have historically used basic statistical functions, such as the mean and variance. However, in such aggregated data analysis, it becomes difficult to evaluate the simulation behavior among the members as well as in comparison to the true state such as the observational data from PAWR radar. Therefore, individual analysis of each of the members becomes highly desirable. This poster focuses on the analysis of member and variable information at a certain time of the ensemble data. Different ensemble data visualization methods that enable the overview of the members and variable information for a pre-determined time have already been proposed so far. However, the drawback of these techniques is that the number of side- by-side plots proportionately increases with the number of members, thus making it difficult to analyze the ensemble data possessing a large number of members. To solve this problem, we developed a visual analytics tool (Figure1) based on Angular-based Edge Bundled Parallel Coordinates Plot (APCP) [2] to enable the analysis of variable information in addition to the member information. This visual analytics tool (Figure1) consists of four views. First, the APCP view (Figure1 (a)), which overviews variable information for multiple members. APCP which simplifies the visualization by using representative lines connected through scatter plots with angular distribution information of the line segments. APCP can show an overview of the entire ensemble dataset while maintaining the correlation information between adjacent variables. The second view (Figure1 (b)) is the Angular Distribution Plot (ADP). This view is the same as the scatter plots inserted between the axes of the APCP and supports the user's choice of which members to select. The third view (Figure1 (c)) is Binning Parallel Coordinates Plot (BPCP). In BPCP view, all variables for the selected member of interest will be displayed. BPCP allows users to select variables of interest. Users can also interactively change the drawing range. The fourth view (Figure1 (d)) is Cross Sectional Plot (CSP). CSP view shows the spatial distribution of the user selected variable. This view shows the horizontal section of the volume data corresponding to the member selected in APCP view. Users are able to interactively navigate through the cross sectional views by manipulating the slider presented in the upper right position. In this poster, we will present a case study by using a meteorological ensemble simulation of a sudden torrential rainfall that occurred at Kobe city on September 11, 2014 [3], and was carried out with 100 members on the supercomputer Fugaku. The simulation data has grid points of 160 × 160 with 500m horizontal resolution, 48 layers with 350m mean vertical resolution, and a total of 11 variables. In this simulation, data assimilation with a 30-second update cycle was performed by using phased array weather radar data from 08:00 JST (Japan Standard Time) to 08:30 JST. Following the initial data assimilation stage, a 15-minute numerical weather forecast, starting from 08:15 JST, was carried out and generated simulation output at every 30 seconds. The case study uses 8:30 as the time of interest, and the water vapor mixing ratio (qv) and cloud water mixing ratio (qc) as the variables of interest for the visual analysis by using our proposed APCP-based visual analytics tool.

Keywords: Human centered computing, Visualization, Visualization techniques, Human-centered computing, Visualization application domains, Visual analytics

[26] Lin Li (Prediction Science Laboratory, CPR, RIKEN) and Takemasa Miyoshi (Prediction Science Laboratory, CPR, RIKEN). *Controlling chaos through intermittent chaos enhancement.*

Abstract. The sensitivity to initial conditions in chaotic dynamical systems makes them unpredictable, yet the same sensitivity implies highly efficient controllability. Such controls have been utilized in stabilizing e.g., turbulent convections, cardiac fibrillations, and chemical reactions. However, existing methods such as the OGY method rely on bi-directional perturbations of system parameters and may not be applied to many practical control tasks where parameters are only allowed to be perturbed uni-directionally. Here in this study, we train an intelligent controller through reinforcement learning, by which we can control the chaotic Lorenz-63 system by temporarily increasing the level of chaos. This is a counter-intuitive task that has not been attacked before. We then open the black box of the intelligent controller and undercover its control principle — unlike previous OGY methods that stabilize existing unstable periodic orbits, this new controller creates new stabilized periodic orbits. This study sheds light on chaos control tasks where the perturbation is constrained to a single direction.

Keywords: chaos control, reinforcement learning, orbit stabilization

[27] Ken Iwata (Kobe University), Naohisa Sakamoto (Kobe University), Jorji Nonaka (RIKEN R-CCS) and Chongke Bi (Tianjin University). *Camera Path Estimation for In-Situ Visualization.*

Abstract. In-situ visualization has drawn widespread attention in recent years due to the continuous increase in the scale and complexity of numerical simulations. It is worth noting that a variety of in-situ visualization approaches have already been proposed, and we have focused on the so-called ``image-based approach'', where Cinema is probably the main representative. In this approach, a large number of images are generated from a large set of omnidirectional cameras to compensate for the lack of interactivity during the detailed post-hoc visual analysis. Therefore, the users are required to search through a large number of output images to find those images that will be useful in gaining scientific knowledge. Some approaches based on autonomic computing have already been proposed to minimize this problem. For instance, Yamaoka et al. proposed an adaptive time-step sampling approach for in-situ visualization. In this approach, the time interval for the visualization is adaptively determined based on changes in the state of the simulation, which is estimated by using kernel density estimation and Kullback-Leibler divergence. As another example, Marsaglia et al. proposed an approach to identify user-preferred camera positions. In this approach, camera positions are evaluated by using a combination of entropy-based metrics. They also proposed a trigger-based approach to determine when to optimize camera positions. In this poster, we will present a camera path estimation approach based on information entropy for facilitating image-based in-situ visualization. To implement this approach we utilized Kyoto Visualization System (KVS) which is an open-source C++ library for developing scientific visualization applications. In our experiments, we used the OpenFOAM simulation code provided by our collaborators. We also conducted some practical experiments on the Fugaku to investigate the running behavior in a highly parallel environment.

Keywords: visualization, in-situ visualization, autonomic computing

[28] Yuichi Otsuka (RIKEN Center for Computational Science), Kazuhiro Seki (RIKEN Center for Quantum Computing) and Seiji Yunoki (RIKEN Center for Computational Science). *Preparing the Gutzwiller wave function on a quantum computer.*

Abstract. We propose a scheme for implementing the Gutzwiller wave function, which is known as one of the most basic variational wave functions for strongly correlated systems, on a quantum computer. We design the quantum circuit based on the discrete Hubbard-Stratonovich transformation, where the Gutzwiller factor can be expressed as a linear combination of single-qubit rotation gates. The rotation angles are determined by the Gutzwiller variational parameter and the auxiliary-field directions. To perform the sum over the auxiliary fields, we propose two methods which have complementary features. The first method uses the linear-combination-of-unitaries circuit which allows one to probabilistically prepare the Gutzwiller wave function, while the second method uses the importance sampling to estimate observables similarly to the quantum Monte Carlo method. In addition to classical emulations of the proposed methods, we perform quantum simulations using a real quantum device (ibmq_manila), demonstrating that the proposed method can reproduce the exact ground-state energies of the two-site Hubbard model within the error bars.

Keywords: quantum computing, quantum circuit, quantum many body systems, classical simulation

[29] Masateru Ohta (RIKEN). *AI prediction system using data prepared by high performance computing: Hydration states around proteins.*

Abstract. An important element when creating an AI system is the data used for learning. However, there is not always enough data to build AI. Data relating to drug discovery are particularly scarce, sometimes with dozens of activity values for a particular assay. One strategy to overcome this situation is to use high performance computing (HPC) to create data and build an AI system by making it learn. In this poster, I will introduce the AI prediction system for the hydration states around a protein, which is very important when a ligand binds to a protein, as an example of creating learning data with HPC and building an AI system. The hydration states of 3,706 proteins as a probability of the existence of a water molecule, g(r), around the protein were calculated by HPC using 3D-RISM theory. [1] A g(r) value is calculated on each grid point around the protein. Information on the protein environment such as a 3D occupancy of protein atoms around each g(r) value was used as descriptors. An AI prediction system which estimates g(r) values around the protein quickly and accurately was established by letting AI learn the descriptors of the protein environment as explanatory variables and each g(r) value as the target variable. [2] As a deep learning method, we used U-Net, a subspecies of 3D-Convolutional Neural Networks. I will discuss the details on the day.

[1] Yoshidome, T.; Ikeguchi, M.; Ohta, M. Comprehensive 3D-RISM analysis of the hydration of small molecule binding sites in ligand-free protein structures. J. Comp. Chem. 2020, 41, 2406-2419

[2] Kawama, K.; Fukushima, Y.; Ikeguchi, M.; Ohta, M.; Yoshidome, T. gr Predictor: A Deep Learning Model for Predicting the Hydration Structures around Proteins. J. Chem. Inf. Model. 2022, 62, 4460–4473

Keywords: Deep learning, Artificial intelligence, High performance computing, Hydration, 3D-RISM, 3D-CNN, U-net

[30] Kazuyoshi Ikeda (RIKEN), Yugo Shimizu (Keio University/RIKEN), Tomoki Yonezawa (Keio University/RIKEN), Masateru Ohta (RIKEN), Masanori Osawa (Keio University) and Teruki Honma (RIKEN).
Development of AI Models Using Open Medicinal Chemistry Data and Its Application to Medium-Sized Molecules.

Abstract. The presence of large, high-quality training data is essential for developing good AI models. In early-stage drug discovery, public data resources such as ChEMBL and Puchem are useful. However, there can be a considerable distance in chemical spaces between compounds used in actual drug discovery and those contained in public databases. As an example of this context, we present the reality of using AI models derived from public data to search for inhibitors targeting protein-protein interactions (PPIs). For medium-sized molecules, considered one of new modalities, we acquired new experimental data and then developed AI models to improve initial prediction accuracy. In addition, we have developed a new PPI library database (called DLiP) consisting of novel synthetic compounds selected from diverse commercial libraries. A total of 32,647 PPI-related compounds are registered in DLiP. It includes 15,214 novel synthesized compounds with molecular weights ranging from 450 to 650 and the known compound data associated with 105 PPI targets extracted and integrated from DLiP. It also contains 17,433 active and inactive compounds. In addition, we extracted data from this database and tested various AI models that predict essential properties of oral drugs, such as target activity, solubility, and membrane permeability. Experimental validation revealed that the predicted compounds have unique chemical structures and possess physicochemical properties suitable for binding to protein-protein interfaces. In this presentation, we will discuss the effectiveness of our method and its extensibility for future applications to large-scale virtual libraries.

Keywords: AI, Large-scale data, Drug Discovery, Database

[31] Kota Takeda (Riken center for computational science) and Takashi Sakajo (Kyoto University).
Computing the invariant measure of the N-vortex problem on the unit sphere.

Abstract. n organization of a large vortex is a well-known phenomenon in two-dimensional turbulence. In 1949, Onsager introduced the dynamics of N-point vortices with circulations called the N-vortex problem. Then, he explained that a large vortex structure is organized in a statistically steady state of the N-vortex problem at a negative temperature on a bounded domain. The straightforward way to obtain the statistically steady state is taking an average of the long-time evolution of N-vortex dynamics. This takes much computational cost because long-time evolution contains evaluations of interactions among N vortices in each step. The vortex distribution obtained in this way corresponds to the invariant measure of the N-vortex dynamics for fixed energy, called the microcanonical Gibbs measure in terms of statistical mechanics. On the bounded domain in two-dimensional space such as a rectangular and circular domain, the invariant measure is already computed numerically, and the organization of large vortices is observed at a negative temperature.
However, the invariant measure of the N-vortex problem on the sphere is not yet computed. In this study, we compute the invariant measure on the sphere by a Monte Carlo method. To apply a Monte Carlo method, we consider the canonical Gibbs measure instead of the microcanonical Gibbs measure. The canonical Gibbs measure is the invariant measure defined for fixed temperature and is written as an exponential function of linear combinations of conserved quantities of the dynamics, while the microcanonical Gibbs measure is a level set of the conserved quantities. Hence, the canonical Gibbs measure is more suitable for Monte Carlo methods. For the sake of computational efficiency, we use the Markov Chain Monte Carlo (MCMC) method named Hamiltonian Monte Carlo (HMC).
MCMC is widely used to approximate high-dimensional integrals by generating samples from the target distribution. In the MCMC procedure, samples are successively generated by a Markov chain that converges to the target distribution, but all MCMC methods do not necessarily converge quickly. HMC is an MCMC method that utilizes Hamiltonian dynamics to realize efficient sampling. Theoretically, the efficiency of HMC is characterized by its convergence property, called geometric ergodicity. This estimates the convergence rate of the probability distribution of samples generated by the Markov chain associated with HMC. Owing to the geometric ergodicity of HMC, we can generate low-correlated samples and obtain high accuracy in Monte Carlo integration with small samples. So far, the geometric ergodicity of HMC is established only on flat space. In this study, we have established the geometric ergodicity of HMC on compact manifolds including the sphere.
By applying the HMC method for the N-vortex problem on the unit sphere, we approximate the flow in the statistically steady state at a negative temperature state and observe the vortex concentration.

Keywords: Turbulence, Monte Carlo, HMC

[32] Enzhi Zhang (Hokkaido University), Mohamed Wahib (RIKEN Center for Computational Science) and Masaharu Munetomo (Hokkaido University). *Training Knowledge Inheritance in Neural Network Through Deep Q-Learning.*

Abstract. After the weights of neural networks are updated at each step, the older weights are discarded. In this paper, we propose a method called Training Knowledge Inheritance (TKI) to use the knowledge about the progression of weight and loss data in reducing overfitting and improving the generalization in the later stages of training. We reformulate the traditional gradient optimization problem as a reinforcement learning task. In particular, by treating the trainable weight space as an environment, the learning rate as action, and the validation accuracies as the rewards, we train a Q-network (supervisor) to learn the discounted future validation accuracy and guide the later training of another network (student). We conduct the experiments on the MNIST, CIFAR-10, and CIFAR-100 datasets with naive dense neural networks and ResNet-56. Our results show that TKI could rediscover learning rate schedule rules similar to previous works, including increasing, decaying, and cyclical repeating.

Keywords: supervised learning, reinforcement learning, neural network, overfitting, learning rate scheduler

[33] Miwako Tsuji (RIKEN R-CCS). *RPC-task based Programming Environment toward the Cooperative Computation of HPC and Quantum Computers.*

Abstract. Quantum computers (QC) are systems based on the principles of quantum theory. Quantum computers have been expected to play an important role in the fields where classical computers should show little growth. On the other hand, large areas should require significant computational capabilities of supercomputers. Here, we present a new programming environment to support the cooperative computation of quantum computers and supercomputers. The proposed method offloads some kernels in an application to quantum computers and supercomputers based on their characteristics using a remote procedure call (RPC).

We adopt a workflow-based approach over quantum computers and supercomputers such as supercomputer Fugaku. An application consists of several tasks, some of which run on a supercomputer, while others may exploit a quantum computer's performance. Quantum tasks are not offloaded to the quantum computer directly, but they should be offloaded to a server called "near-QC" system. Then, the near QC system compiles the code to the quantum circuit and executes the kernel on the quantum computer.

Our framework will also support quantum simulators running on a classical computer as well as a real quantum computer. Mainly we are planning to support a large-scale state vector quantum simulator on a large-scale system, a middle-scale state vector quantum simulator on a single (or few) node(s), andtensor flow simulator on an external large-scale GPU system. We propose an API to use different simulators in a unified way, i.e. a single source code can be used to execute different quantum computer simulators and a real quantum system in the future.

Keywords: quantum computer, remote procedure call, Fugaku

[34] Konstantinos Anagnostopoulos (National Technical University of Athens), Takehiro Azuma (Setsunan University), Kohta Hatakeyama (High Energy Accelerator Research Organizaiton (KEK)), Mitsuaki Hirasawa (Istituto Nazionale di Fisica Nucleare), Jun Nishimura (KEK, SOKENDAI), Stratos Papadoudis (National Technical University of Athens) and Asato Tsuchiya (Shizuoka University). *Emergence of an expanding (3+1)-dimensional spacetime in the type IIB matrix model.*

Abstract. The type IIB matrix model, also known as the IKKT model, is a promising candidate for the non-perturbative formulation of the string theory. Its Lorentzian version, in which the indices are contracted using the Lorentzian metric, has a sign problem stemming from e^{iS} in the partition function (where S is the action). It has turned out that the Lorentzian version is equivalent to the Euclidean version, in which the SO(10) rotational symmetry is spontaneously broken to SO(3), under the Wick rotation as it is. This leads us to add the Lorentz-invariant mass term to the Lorentzian version of the type IIB matrix model. The model we study involves a sign problem, and we perform numerical simulations based on the complex Langevin method, a stochastic process for complexified variables. We discuss the possibility of the emergence of the (3+1)-dimensional expanding universe.

Keywords: string theory, complex Langevin method, sign problem, matrix model

[35] Andrès Rubio Proaño (Riken R-CCS) and Kento Sato (Riken R-CCS). *Power Consumption Metric on Heterogeneous Memory Systems.*

Abstract. The architecture of supercomputers over the years has evolved to support different need in aplications that seek to solve some human concerns. Heterogeneity role nowadays is important in processors and also in the memory-storage system. In processors, we can observe CPUs, GPUs and other accelerators coexisting. In the same fashion, different kinds of memory have appeared over the years, fulfilling some gaps in the memory-storage continuum. E.g., high bandwidth memory (HBM), that is embedded on the processor package, coexist mainly with dynamic random Access memory (DRAM) into Intel Xeon Phi Processors or Knight Landing (KNL). Non-volatile memory (NVM), that can be found with DRAM into the 2nd Generation Intel Xeon Scalable Processors. Nowadays, the upcoming Intel Sapphire Rapids support HBM inside the processor package, DRAM through the memory bus, and also it supports disaggregated memory by the Compute Express Link (CXL) that in principle allows to connect HBM, NVM and DRAM on it.
The task of developers when programming new applications or adapting the existing ones requires full knowledge of the memory system and without a specific strategy it can be very complicated depending on the conditions in which the applications are required to run. Today, for developers is pertinent to prepare their applications so that they adequately face at least the main hybrid memory system (HMS) setups. For that reason we consider that every developer should at least understand HMSs in terms of simple and easy metrics such as: bandwidth, latency, capacity, data persistence, power consumption, etc. Especially, it is essential to know how much memory power applications are going to use in a given memory system. It is vital in situations where executions need to be performed with minimal power consumption mode, or when we need to balance power consumption and performance.
In this poster presentation, we focuses on understanding and giving a perspective on how to analyse memory energy consumption metric over different HMS setups. We consider, identifying and exposing the memory system in the simplest manner developers could access. E.g., a memory system with DRAM and NVM can be exposed as different NUMAs in some systems and their access implies binding the applications process to the kind of memory required. Then, we have selected a some memory-intensive applications that should be profiled. Profiling depends on the expertise of developers and also tools can give more or less information depending on their capabilities. In our case, Intel Performance Counter Monitor (PCM) enables the possibility to get some performance counters related to memory power consumption in between others related to bandwidth. Also we used Linux Perf profiler tool to retrieve relevant information related to cache misses and verify if applications are behaving as a memory-intensive application. The final objective when analysing the power consumption metric is to be able to give a certain ordering for which the developer can look for a memory with very low consumption, as she/he could look for one that allows her/him to have a balance between the performance of the applications and the consumption of memory power. In addition to this analysis, we have sought to provide developers with an early HMS memory power prediction model, which allows getting an idea of the possible consumption of their application towards a given HMS.

Keywords: heterogeneous memory, profiling, memory-power, NUMA