The Eighth
AICS International Symposium

Roadmap on computer science and computational science in the future AICS and HPC communities

February 7—8, 2018, Kobe, JAPAN

Organized by:

RIKEN Advanced Institute for Computational Science

Supported by:

Priority issues (9 issues) on Post-K computer
High Performance Computing Infrastructure Consortium
Research Organization for Information Science and Technology (RIST)

The Eighth
AICS International Symposium

Roadmap on computer science and computational science in the future AICS and HPC communities

Program

■ Wednesday, February 5—6

AICS Youth Workshop

■ Wednesday, February 7 (Day 1)

09:00–17:00	**Registration**	
	Opening Session (Lecture Hall, 6th Floor)	
10:00–10:10	Opening Address: RIKEN AICS Overview Kimihiko Hirao (AICS)	
	Keynote talk	
10:10–10:50	Keynote talk: Future Hard- and Software Architectures for Large-Scale Data Assimilation and Inversion Roland Potthast (DWD/U. of Reading)	···6
10:50–11:20	**Coffee/Tea break**	
	Session Ia: Data Science	
11:20–12:00	Invited talk: Toward Exabyte ~ A Challenge of Energy Frontier Particle Physics Experiments Hiroshi Sakamoto (The University of Tokyo)	···7
12:05–12:20	**Group photo**	
12:20–13:30	**Lunch & Coffee/Tea break**	
	Session Ib: Data Science	
13:30–13:55	Contributed talk: A case of data analysis for efficient operation of datacenters Hiroya Matsuba (AICS)	···8
13:55–14:20	Contributed talk: Technical Challenges of Data Science regarding Computer Science and Applied Mathematics Toshiyuki Imamura (AICS)	···9
14:20–14:50	**Coffee/Tea break**	
	Session II: Software Management	
14:50–15:15	Contributed talk: An introduction to AICS Software Center Fumiyoshi Shoji (AICS)	···10
15:15–15:40	Contributed talk: Development of a software framework for parameter-space exploration, OACIS Yohsuke Murase (AICS)	···11
15:40–16:05	Contributed talk: Development of CFD frameworks for industrial applications: FrontFlow/red-HPC and CUBE Makoto Tsubokura (AICS)	···12
16:20–17:50	**Poster Session** (Seminar Room, 1st Floor)	
18:00–19:30	**Banquet** (Seminar Room, 1st Floor)	

■ Thursday, February 8 (Day 2)

09:00–15:00	**Registration**	
	Session IIIa: Hardware (Lecture Hall, 6th Floor)	
09:30–10:20	Invited talk: Accelerating Understanding: GPUs, AI, and the Future of Supercomputing in a Post-Moore's Law World Steve Oberlin (NVIDIA)	···14
10:20–10:50	**Coffee/Tea break**	
	Session IIIb: Hardware	
10:50–11:40	Invited talk: Update of Post-K Development Yutaka Ishikawa (AICS)	···15
11:40–13:30 (12:45–13:30)	**Lunch & Coffee/Tea break** **AICS and K computer Tour** (AICS Entrance Hall and Visitor Hall, 6th Floor)	
	Session IVa: Application 1 (Disaster Prevention)	
13:30–14:10	Invited talk: Integral porosity shallow water models for urban flood forecasting Ilhan Özgen (Technical U. of Berlin)	···16
14:10–14:35	Contributed talk: Connection between numerical weather simulation and hydrological simulation for disaster prevention and mitigation Satoru Oishi (AICS)	···17
14:35–15:00	Contributed talk: How does the regional climate assessment make more advance? Hirofumi Tomita (AICS)	···18
15:00–15:30	**Coffee/Tea break**	
	Session IVb: Application 2 (Material Science)	
15:30–16:10	Invited talk: Molecular dynamics simulations emerging as platform for the structural biology of large biomolecular assemblies Yibing Shan (D. E. Shaw Research)	···20
16:10–16:35	Contributed talk: Development of large-scale ab initio quantum chemical code based on massively-parallel sparse-matrix library Keisuke Sawada (AICS)	···21
16:35–17:00	Contributed talk: Application of DFT calculations for predicting the structural and electronic properties of materials Mohammad Khazaei (AICS)	···22
17:00–17:30	**Closing Discussion**	

Poster Session Program

Talk Session

(Day 1)

February 7

Future Hard- and Software Architectures for Large-Scale Data Assimilation and Inversion

Roland Potthast[1]
[1] *Deutscher Wetterdienst/University of Reading*

Abstract

The goal of the talk is to discuss emerging needs and techniques for analysing and assimilating huge data sets into dynamical models. Traditionally, the structure of modern supercomputers is optimized for throughput of particular differential equation based algorithms. With higher resolution of numerical models in all areas of science from geophysics and meteorology to medicine, neuroscience and biological models, this will remain a first critical success parameter for future hard- and software architectures.

However, the interplay of models with data, both for analysing and understanding the massive data sets available today with exponential increase over the coming 10 years, as well as for frequent initializing models for forecasting within a data assimilation cycle and the online adaptation and determination of model errors, biasses and model parameters demands the extension of plannig processes to an integrated system. The intense communication needed for the distribution and processing of large-scale data sets and observations is the second critical success parameter of future hard- and software architectures

Traditionally, the execution of programs is started at some point in time, large amounts of data are transfered from some hard drive or data base into the main memory of a HPC system, calculations are carried out and the result is saved on hard drive or in some database. Then, postprocessing algorithms and visualization tools read the data from the database and distribute the results. However, with both temporally and spatially dense observations and high-resolution model fields in rapid assimilation and forecasting systems, the traditional processing chain needs to be redesigned. This has strong implications for resilience and error tolerance of full systems, which usually depend strongly on a reasonable time-track of its past history. The full system design with the proper integration of storage, post-processing, product feeds, error tolerance on all of its layers and emergency mechanisms is a third critical success parameter for future hard- and software architectures.

We will discuss the HPC needs with particular reference to rapid forecasting systems at DWD, ranging from nowcasting and high-resolutional regional scale to global atmospheric systems, including the prediction of uncertainty of state estimates and forecasting based on ensemble data assimilation (EDA) and ensemble prediction systems (EPS).

Toward Exabyte ˜ A Challenge of Energy Frontier Particle Physics Experiments

Hiroshi Sakamoto[1]
[1] *International Center for Elementary Particle Physics, the University of Tokyo*

Abstract

The world largest particle accelerator LHC (Large Hadron Collider) has been in operation since 2009 and ATLAS and other three experiments have been accumulating unpreceded amount of experimental data, which is now approaching the Exabyte scale. In order to process the data, a globally distributed computing infrastructure, the worldwide LHC computing grid (WLCG), has been developed and deployed among more than 150 sites from 35 countries. Even only for the ATLAS experiment, 300PB of data are recorded and 500,000 CPU cores are used for their processing, on the WLCG infrastructure. The University of Tokyo has been contributing to the WLCG infrastructure from the beginning as a key site in the Asia-Pacific region. In the upgrade program of LHC, expected amount of data to be generated will be 100 times more than now and a substantial evolution is required to satisfy the requirement. Many R&D programs have been started to achieve the evolution. In this presentation, the overview of the LHC experiments and WLCG, the contribution of Tokyo, and the upgrade plan will be discussed.

Biography

Degree

Doctor of Science (May 1985, Kyoto University)

Research and Teaching

2001–Present	Professor, International Center for Elementary Particle Physics, The University of Tokyo Representative of the ATLAS regional analysis center in Tokyo
1993–2001	Associate Professor, Department of Physics, Kyoto University Leading Experimental particle physics group working for KEK-PS E162 Experiment, KEK-B Belle Experiment, LHC ATLAS experiment
1985–1993	Research Associate, Physics Division, National Laboratory for High Energy Physics (KEK) Data acquisition group of TRISTAN VENUS experiment

URL

`http://www.icepp.s.u-tokyo.ac.jp/~sakamoto/`

A case of data analysis for efficient operation of datacenters

Hiroya Matsuba[1]
[1] *HPC Usability Research Team, AICS, RIKEN*

Abstract

This talk will introduce a case of data analysis that aims for efficient operation of computer systems. In this work, we analyzed packet traces taken from network switches of a data center, which is used as an infrastructure of cloud computing and discovered the structure of applications running in the cloud environment. Then, we developed a new algorithm for a network load balancer that uses the structure information and calculates the best policy of packet distribution to minimize the utilization of the most loaded network link.

The key is that we proposed a method to handle sampled packet traces. Treating sampled traces is necessary because the number of packets forwarded by network switches is enormous and it is not feasible to collect all the packets. To handle the sampled traces, we designed the algorithms so that they are completely volume-based, not timing-based.

This work is valuable for data center operators because the application structure, which is otherwise the hidden information for infrastructure operators, can be utilized to optimize the resource allocation. I think this work is worth introducing to the supercomputer community because the idea of extracting application level knowledge from infrastructure system is applicable as well for efficient operation of supercomputers and because this is an interesting example where data analysis contributes to computer science, not vice versa.

Technical Challenges of Data Science regarding Computer Science and Applied Mathematics

Toshiyuki Imamura[1]
[1] *RIKEN AICS*

Abstract

Data Science is known as a new computational science field including experimental science, data assimilation, artificial intelligence, and so on. It becomes an ordinary terminology supporting science, engineering and social infrastructures, which are tightly linked to our social lives. To establish data science in the real world, we need to consider and interact with various technologies in computer science and applied mathematics.

As far as the author recognizes, the recent advancement in computer science fields such as cloud computing, deep-learning, low-power device and sensing technologies, distributed computing framework like Spark, DSL, and also the development of In-Situ, are solicited. Also, from the viewpoint of applied mathematics, high-order data (tensor) analysis, and sparse compression-and-analysis are innovated in recent years with the help of computational power. These technical issues are fused into a true driving force of development of new industry topics such as AI and auto-driving technologies. In this symposium, the author will present a summary of MUST-DO topics in the future activities in AICS with an analysis of personal perspectives to data science.

An introduction to AICS Software Center

Fumiyoshi Shoji[1]
[1] *RIKEN AICS*

Abstract

RIKEN AICS Software Center was founded in 2017 to promote software developed at AICS. The missions of RIKEN AICS Software Center are to develop and deploy high quality applications, libraries, programming tools, etc. (called "AICS Software") for many HPC platforms including the K computer and the successor and to support and collaborate with AICS Software users from various fields of science and engineering to improve and enhance the AICS software. In this talk, we introduce current activity and future plan of the RIKEN AICS Software Center.

Development of a software framework for parameter-space exploration, OACIS

Yohsuke Murase[1]
[1] *RIKEN Advanced Institute for Computational Science*

Abstract

We present an open-source software framework for parameter-space exploration, named OACIS (Organizing Assistant for Comprehensive and InteractiveSimulations), developed by discrete event simulation research team in AICS [1]. When we conduct numerical simulations in scientific research, we usually carry out many simulation jobs changing models and parameters by trial and error. This kind of trial-and-error approach often causes a problem of job management since a large amount of jobs are often created while we are exploring the parameter space. For instance, submitting one simulation job would require various tasks including logging in to the remote host, preparing a shell script, submitting jobs to a scheduler, checking the job status to wait for the completion, and copying the results to your local machine. Such repetitions are not only troublesome and tedious but prone to human errors. OACIS is a software framework to help researchers conduct their calculation in an efficient, reliable, and reproducible way to focus on the essential part of their research activities. In the presentation, after giving an overview on OACIS, we will share some use cases as well as how we have developed it and what we are doing for future.

References

[1] `https://github.com/crest-cassia/oacis`

Development of CFD frameworks for industrial applications: FrontFlow/red-HPC and CUBE

Makoto Tsubokura[1], Keiji Onishi, Rahul Bale, WeiHsiang Wang, and Koji Nishiguchi
[1] *RIKEN Advanced Institute for Computational Science*

Abstract

High-Performance Computing(HPC) frameworks for the fluid-structure interaction problems with complicated geometry have been developed, considering its application to industrial problems. Two frameworks are considered here; one is based on the unstructured mesh system (FrontFlow/red-HPC), and the other is based on the hierarchically structured grid system (CUBE).

For the fully unstructured finite volume method developed here, arbitrary Lagrangean-Eulerian (ALE) method together with solving the governing equations on the non-inertial reference of frame was adopted to track the complicated motion of rigid body in fluids. The method was specially tuned on the K computer, and so far it can treat up to ten billion unstructured meshes for the coupled fluid-solid motion problems. The advantage of this framework is its accurate prediction of aerodynamic forces acting on the complicated geometry by optimizing the unstructured meshes near the wall, especially in the case flow is in turbulence state. On the other hand its disadvantage is its workload required for the pre-processing including dirty CAD treatment and mesh generation. This problem in fact prevents its spread and utilization in industrial process.

To overcome this problem, alternative framework based on the hierarchically structured finite volume method was developed, in which both the fluid motion and structure deformation are solved in Eulerian manner. To achieve higher computational efficiency of parallelization and scaling on the massively parallel environment, Building Cube Method (BCM) proposed by Nakahashi[1] was adopted. In the method, numerical domain is first decomposed into cubic sub-domains based on the Octree method. Then the same number of numerical grids is allocated to each cubic subdomain. In the simulation framework, the solid surface with complicated geometry is represented by the immersed boundary method (IBM). In the fluid-structure interaction problems with structure surface in motion, accurate representation of the immersed body is indispensable. Thus Lagrangian description for tracking the moving solid body surface is adopted in the Eulerian framework of solving fluid and structure motions. The parallel scalability of the numerical method and the efficacy of the load balancer were evaluated through simulation with up to 32,768 cpu cores on the K-computer. So far, the framework can handle maximum of tens of billions of numerical meshes using hundreds of thousands of CPU cores on the K-computer.

Talk Session

(Day 2)

February 8

Accelerating Understanding: GPUs, AI, and the Future of Supercomputing in a Post-Moore's Law World

Steve Oberlin[1]
[1] *CTO, Accelerated Computing, NVIDIA*

Abstract

For decades, numeric simulation using supercomputers has enabled a deeper exploration of the makeup of our universe and how it works. The advance of CPU-based computational capability has slowed dramatically, however, as shrinking silicon semiconductors encounter fundamental physical limits. The natural architectural response to slowing Moore's Law is to increase parallelism, and the invention of massively-parallel GPU-based accelerated computing has led to the wide availability of extraordinarily efficient and general-purpose parallel computation capability. Ten years of success in HPC ideally positioned GPU-based accelerated computing to power a revolution in Deep Learning and AI, and now researchers are applying data-driven machine learning techniques to HPC problems with dramatic results that promise to redefine supercomputing architecture and application workflow, perhaps even rendering traditional approaches to achieving exascale obsolete.

This talk will review the technology trends and architecture factors that have so favored GPU-accelerated computing for HPC and AI, discuss how the fundamental architecture differences between GPUs and CPUs enable a sustainable future of performance enhancement despite waning Moore's Law technology improvements, explore the remarkable efficiency and performance potential at the intersection of AI and HPC, and imagine how the largest supercomputers in the world, scientific and technical applications, and how people use them may be forever disrupted as a result.

Biography

Steve Oberlin is the Chief Technology Officer for Accelerated Computing at NVIDIA.

Steve Oberlin's large-scale computing technology career has spanned over 30 years, launched in 1980 at Cray Research bringing up CRAY-1 supercomputer systems. Starting in 1981, he worked for Seymour Cray as a designer and project engineer on the CRAY-2 and CRAY-3 supercomputers. In 1988, he led early massively parallel processing research at Cray that ultimately led to his role as the chief architect of the CRAY T3D MPP and its successor, the CRAY T3E. He holds 15 architecture and design patents for the T3D and T3E.

Steve was VP of Hardware Engineering in Chippewa Falls, Wisconsin, for Cray/SGI from 1996 until early 1999, responsible for hardware development and support of all Cray products and their follow-ons.

Steve left SGI to found Unlimited Scale, Inc., in July of 2000, and spent the next 13 years creating new cloud computing infrastructure management and intelligent resource optimization technologies for start-up Cassatt and CA Technologies.

Steve returned to HPC at NVIDIA in November 2013. As CTO for Accelerated Computing, he is responsible NVIDIA's Tesla roadmap and architecture.

Update of Post-K Development

Yutaka Ishikawa[1]
[1] *RIKEN AICS*

Abstract

The next flagship supercomputer in Japan, replacement of the K supercomputer and thus we call it post-K computer, is being designed to be operated in early 2020s. Its node architecture and interconnect are Armv8-A + SVE and a 6-D mesh/torus network, respectively. A three level hierarchical storage system will be installed with compute nodes. The system software developed in the post-K supercomputer includes a novel operating system for general-purpose manycore architectures, low-level communication and MPI libraries, and file I/O middleware. After introducing an overview of the post-K architecture, we will present the current status of the system software development.

Biography

Yutaka Ishikawa is in charge of developing the post K supercomputer. From 1987 to 2001, he was a member of AIST (former Electrotechnical Laboratory), METI. From 1993 to 2001, he was the chief of Parallel and Distributed System Software Laboratory at Real World Computing Partnership. He led development of cluster system software called SCore, which was used in several large PC cluster systems around 2004. From 2002 to 2014, he was a professor at the University Tokyo. He led a project to design a commodity-based supercomputer called T2K open supercomputer. As a result, three universities, Tsukuba, Tokyo, and Kyoto, obtained each supercomputer based on the specification in 2008. He was also involved with the design of the Oakleaf-PACS, the successor of T2K supercomputer in both Tsukuba and Tokyo, whose peak performance is 25PF.

Integral porosity shallow water models for urban flood forecasting

Ilhan Özgen[1,2]
[1] *Technische Universität Berlin, Berlin, Germany*
[2] *RIKEN AICS, Kobe, Japan*

Abstract

Shallow water models with porosity are a family of shallow water equations-based models designed to compute floods at a coarse resolution for computational efficiency purposes. Topographic data is indirectly accounted for by means of porosity terms. The integral porosity shallow water model is a specific type of this model family, and defines two types of porosity terms: (1) storage porosity inside the mesh cell; (2) conveyance porosity at the cell edge. This presentation will give an overview of the mathematical model concept, capability and limitation of the model. Recent developments such as the dual integral porosity shallow water model and depth-dependent porosity terms will be presented. The possibility and feasibility of combining the integral porosity shallow water model with high-performance computing will be discussed.

Biography

Ilhan özgen is a visiting scientist at RIKEN AICS. From 2012, he worked as a research associate at Technische Universität Berlin, Germany, and received his Ph.D. degree with distinction in 2017. His main research interests are the development of numerical methods for hyperbolic systems with discontinuous coefficients or source terms, and high-performance scientific computing for the solution of hyperbolic partial differential equations.

Connection between numerical weather simulation and hydrological simulation for disaster prevention and mitigation

Satoru OISHI[1]
[1] *RIKEN AICS, Kobe, Japan*

Abstract

A numerical weather simulation uses discrete grids which have scale of 100m to 1000m or larger. The shape of the grids is polyhedra and most of the case they are rectangular parallelepipeds which does not take the boundary of the basin into account because there is no boundary in the atmosphere. However, hydrological simulation completes in one basin. Water current on the slope is from higher topography to lower, river water runs from upper stream to lower river mouth. Natural water transportation does not cross the basin. The problem of unmatched grid between atmospheric simulation and hydrological simulation will be solved by higher resolution simulation of atmosphere with considering effect of shape edge of the boundary of the basin. Then we can have an optimized evacuation strategy by using the result of multi basin hydrological simulation.

Extreme rain and flood can be accompanied by storm surge and sediment related hazards. The sediment related hazards include slope failure and mud water. Storm surge prevents river water from releasing itself to the ocean. Even tide embankment protects cities from intrusion of storm surge. River water from upper river basin could not be stored in the river mouth followed by river water flooding, bank collapse and large area inundation. The sediment related hazards destroy houses, roads and paddy fields. Moreover, water running through a river deposits sediment on river bed and it reduces discharge capacity of the river then it raises risk of large area inundation.

Unlike continental rivers flooding, Japanese rivers are running on steep slope. It means flow current has strong energy to destroy urban areas which consists of many kinds of infrastructure like roads, railways, internet lines, electricity cables, gas pipes and domestic water supplies. The recovery of the area takes much time than disaster of continental rivers flooding.

For better adaptation strategy to face the disaster starts from the knowledge of all stakeholders in the river basin. Every stakeholder shares which kinds of disasters are expected, how much does it give damage to our society for better preparation against future extreme disasters.

Biography

Satoru OISHI is a research unit leader at RIKEN AICS from 2017, a professor of Kobe University from 2009 and the director of the research center for urban safety and security of Kobe University from 2016 for two years. He received his Doctor of Engineering degree from Graduate School of Engineering, Kyoto University. He worked in Disaster Prevention Research Institute of Kyoto University as an assistant professor from 1993 to 2000 and an associate professor of University of Yamanashi from 2000 to 2009. His main research interests are application of meteorology to engineering especially disaster mitigation, radar hydrology, high-performance computing for simulation of extreme hazard and risk communications.

How does the regional climate assessment make more advance?

Hirofumi TOMITA[1], Sachiho A. ADACHI[1], Seiya NISHIZAWA[1], Ryuji YOSHIDA[1,2], Tsuyoshi YAMAURA[1], Kazuto ANDO[1], Hisashi YASHIRO[1], Yoshiyuki KAJIKAWA[1,2]
[1] *RIKEN Advanced Institute for Computational Science*
[2] *Research Center for Urban Safety and Security, KOBE University*

Abstract

To take one step further into the assessment of regional climate change, we should clarify what is dominant factor on such change, how the change does occur in the regional scale, to what extent the assessment results are reliable. Although regional climate models have been used for the future projection of regional climate, the above questions are still debated. So far, many studies for the regional climate assessment were proposed. Nishizawa et al. (2018) summarized them and discussed about the boundary condition of regional climate model in a phase space of mean-state and perturbation, defined as large-scale state averaged during a certain long time and as variability from the mean state, respectively. One of the well-known conventional methods is the pseudo-global warming method. However, this method has a problem; it considers only the influence of changes in mean state, i.e., the influence of changes in perturbation on the regional climate projection is negligible. Recently, our research group proposed a novel assessment method, which quantifies the contribution from the changes in mean-state and the changes in perturbation and their interaction (Adachi et al. 2017, Nishizawa et al. 2018). The idea is as follows. We extract two mean-states and two perturbations from the present and future projections obtained by a general circulation model. Considering the combination by exchanging the two mean states and the two perturbations, we can construct four boundary conditions for regional climate model: two boundary conditions are used for the so-called the direct downscaling runs and the other two are used for artificial downscaling runs. By adding the latter runs, we extract each of sensitivities against the changes in mean-state, changes in perturbation, and the non-linear effect at the simultaneous changes. If we take the method as an abstraction, it would be applicable to extract the uncertainties of regional climate assessment, including future scenario, model bias, and so on. In this talk, we will introduce our conceptual method and discuss how the regional climate assessment can be more-advanced using our method form the viewpoint of disaster prevention.

References

[1] Adachi, S. A, S. Nishizawa, R. Yoshida, T. Yamaura, K. Ando, H. Yashiro, Y. Kajikawa, and H. Tomita (2017): Contributions of changes in climatology and perturbation and the resulting nonlinearity to regional climate change", Nature Communications, doi: 10.1038/s41467-017-02360-z

[2] Nishizawa S., S. A. Adachi, Y. Kajikawa, T. Yamaura, K. Ando, R. Yoshida, H. Yashiro, H. Tomita (2018): Decomposition of the large-scale atmospheric state driving downscaling: A perspective on dynamical downscaling for regional climate study, Progress in Earth and Planetary Science, in press,

Acknowledgement

This study was supported by the Foundation for Computational Science (FOCUS) establishing Supercomputing Center of Excellence and JST CREST (Grant Number JPMJCR1312), Japan. We thank Drs. I. Takayabu, R. Mizuta, and O. Arakawa for providing the MRI-AGCM3.2S data. The results of this study were obtained using the computational resources of the K computer at RIKEN/AICS.

Molecular dynamics simulations emerging as platform for the structural biology of large biomolecular assemblies

Yibing Shan[1]
[1] *D. E. Shaw Research*

Abstract

Despite the elegance of the basic concept underlying molecular dynamics (MD) simulations, for a couple of decades it had been limited to the small timescales of jiggles and wiggles of macro-biomolecules, which is largely irrelevant to their biological functions. An interdisciplinary approach led to the development of special purpose Anton supercomputer that increases the timescale by 2-3 orders of magnitude, allowing MD simulations of the micromachines in action. Moreover, the speed-up makes simulations of assembling of macro-biomolecules into biological functional units possible. Consequently, MD simulations emerges as an increasingly credible platform for structural biology of large biomolecular complexes and as a promising tool for early stage drug discovery for challenging drug targets. To demonstrate the new potentials afforded by the special-purpose supercomputer, we will report de novo modeling of full-length JAK2 kinase and Ras oligomer as examples.

Biography

Professor Yibing Shan is a senior scientist of D. E. Shaw Research in New York City and a visiting chair professor of Beijing Computational Science Research Center. Trained as a physicist, his current research focuses on computational structural biology studies of functional assemblies of macro-biomolecules and early-stage drug discovery based on long-timescale molecular dynamics simulations combined with experimental means.

Development of large-scale ab initio quantum chemical code based on massively-parallel sparse-matrix library

Keisuke Sawada[1], William Dawson[1], and Nakajima Takahito[1]
[1] *Computational Molecular Science Research Team, RIKEN Advanced Institute for Computational Science*

Abstract

The computational prediction for the chemical reactivity and electronic properties in large-scale molecules composed of several thousands to ten thousands of atoms such as supramolecules and biomolecules plays an important role in drug discovery and nanomaterial design. Ab initio quantum chemical calculation represented by Hartree-Fock and density functional theory (DFT) methods is a powerful tool to achieve the accurate prediction for such properties. In these methods, the eigenvalue problem for the Fock matrix is solved to obtain the density matrix. In standard quantum chemical codes, the arrays of the Fock and density matrices are not separated per CPU core, that is, all CPU cores have full-size matrices. Since these array sizes increase with the square of the molecular size, we often suffer from the problem of over memory in large molecular systems. Therefore, the development of the memory-divided code is required to realize the large-scale molecular calculation.

In this study, we develop the memory-divided quantum chemical code. The Fock and density matrices are treated as the sparse matrix and we use the massively-parallel sparse-matrix library NTPoly. Our code enables to calculate large molecular system consisting of more than 3,500 atoms on the K computer, whereas the possible molecular size treated in using the non memory-divided code is about 1,500 atoms at a maximum. Moreover, by exploiting some fast and efficient calculation methods in quantum chemistry, we find that our code shows a good scaling with respect to used CPU cores.

Application of DFT calculations for predicting the structural and electronic properties of materials

Mohammad Khazaei[1] and Seiji Yunoki[1]
[1] *Computational Materials Science Research Team, RIKEN Advanced Institute for Computational Science (AICS), Kobe, Hyogo 650-0047, Japan*

Abstract

Among the different methods, density functional theory (DFT) has proven to be a reliable method to predict various physical and chemical phenomena on the atomic scale. Hence, we have used DFT calculations to examine many properties and applications of materials. Here, we present some examples of our calculations to investigate electron field emission simulation, scanning tunneling microscopy simulations of surfaces, structural phase transition of materials at high pressure, exfoliation possibility of layered materials, and electronic properties of materials.

Poster Session

(Day 1)

February 7

P-01

An Approach from computational chemistry to the interpretation of the rotation and vibration spectra of NO3 molecule

Umpei Nagashima[1] and Tsuneo Hirano
[1] *FOCUS*

Abstract

NO3 : Ro-vibrational Spectroscopy
NIST, WebBook. Data up to 2013 (Fujimori et al. 2013)

Long-standing debate on 1492 cm-1 band
1492 cm-1 n3 Hirota (1985), IR
or n1+n4 Stanton (2007), Calc. EOMIP-CC
or n3+n4 Kawaguchi et al (2011) , FTIR
n4 = 365.48 cm-1, but n3 not detected
NO3 electronic ground-state structure is mostly considered to be D3.

CASSCF, MR-SDCI + Q: Multi-reference calc.
Global minimum: C2v structure
D3h structure: 2nd-order TS.

P-02

First principle calculation of thermodynamic properties of Mg-IV-N2

Atchara Punya Jaroenjittichai[1], Chaiyawat Kaewmeechai and Yongyut Laosiritaworn
[1] *Chiang Mai university, Thailand*

Abstract

Recently, electronic properties of Mg-IV-N2 compounds have been investigated. This is as they have potential properties for utilization in optoelectronic devices. However, some devices may operate in temperature variation environments. Sometimes the variation can go extremely and beyond the materials tolerance. This could then bring fatigue or unexpected results to the devices if thermal related effect is not in careful consideration. As a result, not only electronic properties but also thermal properties of the materials should be profoundly considered in electronic device design. Therefore, in this work, we investigated phonon and other associated properties of the Mg-IV-N2 compounds, i.e Helmholtz free energy, internal energy , entropy, specific heat capacity at constant volume, high frequency and static dielectric constants. The calculations were carried out by using the norm-conserving pseudopotential plane-wave method under the framework of the density functional perturbation theory (DFPT) and the local-density approximation (LDA).

P-03

Flexibility of Quantum Computer for Plasma Simulation to Characterize the Plasma Parameters

Shankar Bhattarai[1]
[1] *Patan Multiple Campus, Tribhuvan University Kathmandu Nepal*

Abstract

Computer simulations are carried out to understand the consequences of fundamental physical laws, design and predict new experiments and interpret experiments can provides information which is difficult to measure. There are various was to describe a plasma; Klimontovich formalism, Closed Vlasov-Maxwell equation and Closure problem. Particle Simulations take the Klimontovich approach, it simulate a plasma system by following a number of particle trajectories. It is difficult to sample a 6D phase space by cells but PIC codes use particles to efficiently sample the phase space distribution. PIC codes as a first-principle kinetic model for plasma physics. It efficiently tracks particle trajectories to sample particle phase space and structure is scalable to massively parallel computer systems. But plasma turbulence is a big problem in nuclear fusion reactors. Fusion reactors could be made a lot smaller if plasma turbulence could be controlled. A reactor with completely controlled plasma would be truck sized vs. warehouse sized. Current supercomputers take a lot of computing time to solve the models for plasma turbulence. This research paper explores while thinking beyond the K computer era, can a true quantum computer solve the models and equations behind plasma turbulence significantly better, faster, than their silicone counterparts.

P-04

The variants of Arnoldi method for solving nonlinear eigenvalue problem

Hideaki Nagasaka[1] and Takashi Nodera[1]
[1] *Keio University*

Abstract

Arnoldi method is one of the schemes for solving nonlinear eigenvalue problems. We consider to solve a nonlinear eigenvalue problem $M(\lambda)x = 0$. In other words, we will find eigenvalues and eigenvectors of a matrix whose elements depend on its eigenvalue. Before we solve its problem, we do use Taylor expansion of $M(\lambda)$. And $M(\lambda)$ has a structure where high order terms of its Taylor expansion have a low-rank structure. Nowadays, there are "finite Arnoldi method" and "a rank exploiting infinite Arnoldi method". The former is modified version of Arnoldi method, and the latter is improved version of finite Arnoldi version. In this talk, we will propose two procedures. One is restart. The restart reduces computation time. Second is to get more than one Krylov vectors in one iteration. This procedure reduces number of iterations and computation time. At last, results of our numerical experiments show that our approach is effective and useful.

P-05

FPGA vs GPU implementation of GAMESS computational chemistry kernel

Withana Kankanamalage Umayanganie Klaassen[1] and Shirley Moore
[1] *University of Texas at El Paso*

Abstract

Heterogeneous computing with FPGAs (field-programmable gate arrays) is a potential area of interest for next generation supercomputers for solving exascale problems. The amount of additional effort invested into integrating FPGA device programming into existing workflows, for accelerating compute intensive kernels, is significant. Writing scientific code to run on FPGAs using HDLs (Hardware Description Languages) is complex. The Intel OpenCL SDK converts OpenCL code to FPGA executable code. The OpenARC compiler provides easier FPGA programmability by translating OpenACC directive based code into OpenCL optimized for FPGAs. Directive based programming models like OpenACC hide low level language complexities. We port a GAMESS computational chemistry kernel containing the Hartree-Fock procedure to FPGA and GPU enabled machines using OpenARC and evaluate the performance results. We insert OpenACC directives and translate to cuda code using OpenARC for Nvidia GPUs. The GAMESS-SIMINT Hartree-Fock quantum chemistry method is used both to compute molecular properties and as a starting point for higher accuracy, more computationally demanding methods. The computational bottleneck of the Hartree-Fock procedure is construction of the Fock matrix, which requires computation of many electron repulsion integrals (ERIs). The SIMINT integral package is a highly vectorized, high performance implementation of the Obara-Saika ERI evaluation method.

P-06

Case Study of a Flow Control Mechanism for Inter-FPGA Communication

Antoniette Pangilinan Mondigo[1] and Kentaro Sano
[1] *Tohoku University*

Abstract

In stream computing applications, the use of multiple FPGAs provide high performance and scalable solutions. To facilitate higher throughput, inter-FPGA communication should utilize a low latency, high-speed serial communications protocol such as Intel SerialLite III, which supports high-bandwidth applications. However, the numerous data transfers between the FPGAs may cause a receiving end to be temporarily unable to accept incoming data streams. In this case, flow control is necessary to avoid data loss. Unfortunately, the flow control feature in SerialLite III for Intel Arria 10 FPGAs is not available. With this, we propose a credit-based flow control mechanism for Arria 10 FPGAs, to allow the transmitter to be aware of the status of its corresponding receiver and decide the next transmission cycle. In this work, we evaluate the performance and resource utilization of the FPGAs with the proposed flow control. For verification, we implement an actual numerical simulation in a 1D ring topology and present the results.

P-07

Ensemble Forecast Sensitivity to Observations Verified with Multiple References

Shunji Kotsuki[1], Kenta Kurosawa, and Takemasa Miyoshi
[1] *RIKEN AICS*

Abstract

It is important to evaluate the impact of assimilated observations on forecast skill in numerical weather prediction (NWP). Several methods have been proposed to estimate the impact of observations with adjoint-based or ensemble-based sensitivity analysis. Most studies evaluate the observational impact using their own analyses as a verification reference. This study investigates the sensitivity of the impact estimates to the choice of the verification reference. We implemented the ensemble-based forecast sensitivity to observation (EFSO) method proposed by Kalnay et al. (2012) with a global atmospheric data assimilation system NICAM-LETKF (Terasaki et al. 2015 and follow-on studies), which comprises the Nonhydrostatic ICosahedral Atmospheric Model (NICAM) and Local Ensemble Transform Kalman Filter (LETKF). We evaluate the impact of observations with the moist total energy norm verified against the NICAM-LETKF ' s own analysis and the ERA Interim reanalysis. In addition, we implemented an observation-based verification metric proposed by Sommer and Weissmann (2016) and Cardinali (2017).

The results suggest that the observational impact be overestimated in 6-h forecasts if the NICAM-LETKF analysis is used for the verification reference. However, no overestimation is observed if we use the ERA-Interim reanalysis and radiosonde observations for the reference. The results imply that the impact of observations at the analysis time would persist in the analysis 6-h later. In the observation-based verification metric, each type of observations mainly contributes to the improvement of the observed variable. For instance, assimilation of the AMSU-A radiances significantly improves the first guess of the AMSU-A radiances at the next assimilation cycle, but this is not necessarily true to other observations. This presentation will include the most recent progress up to the time of the symposium.

CANCELLED

P-08

Dam Operation Optimization by Machine Learning

Marimo Ohhigashi[1], Shunji Kotsuki, and Takemasa Miyoshi
[1] *RIKEN AICS*

Abstract

Hydraulic power generation is an important renewable energy resource. Current operations of hydroelectric dams mainly depend on operators' decision based on their experience. We may improve dam operations for more efficient hydraulic power generation. This study aims to optimize dam operations using machine learning methods. We develop three machines (Machine $1-3$). Machine 1 improves precipitation forecast from available data including numerical weather prediction (NWP). The Japan Meteorological Agency (JMA) Mesoscale Model (MSM)' s operational precipitation forecasts are improved by supervised learning with observed radar data. Machine 2 estimates river inflow into dams using the precipitation forecasts from Machine 1. We use the observed river inflow records to teach Machine 2. Machine 3 aims to maximize the total power generation by reinforcement learning. We concatenate these three machines and apply them for Sai River basin in Nagano Prefecture with 11 hydraulic dams. This poster presents our ongoing effort on developing objective dam operations.

P-09

Accelerating assimilation development for new observing systems using EFSO

Guo-Yuan Lien[1], Daisuke Hotta[2], Eugenia Kalnay[3],
Takemasa Miyoshi[1], and Tse-Chun Chen[3]
[1] *RIKEN AICS*
[2] *JMA/MRI*
[3] *University of Maryland*

Abstract

To successfully assimilate data from a new observing system, it is necessary to develop appropriate data selection strategies, assimilating only the generally useful data. This development work is usually done by trial-and-error using Observing System Experiments (OSEs), which are very time- and resource-consuming. We propose a new, efficient methodology to accelerate the development using the Ensemble Forecast Sensitivity to Observations (EFSO). First, non-cycled assimilation of the new observation data is conducted to compute EFSO diagnostics for each observation among a large sample. Second, the average EFSO conditionally sampled in terms of various factors is computed. Third, potential data selection criteria are designed based on the non-cycled EFSO statistics, and tested in cycled OSEs to verify the actual assimilation impact. We demonstrate the usefulness of this method with the assimilation of satellite precipitation data. It is shown that the EFSO based method can efficiently suggest data selection criteria that significantly improve the assimilation results.

P-10

Towards a coastal ocean forecasting system with data assimilation

Kohei Takatama[1], Takemasa Miyoshi[1], Jun Kikuchi[2], and Kengo Ito[2]
[1] *RIKEN AICS*
[2] *RIKEN CSRS*

Abstract

Coastal ocean is closely related to human activities, but its prediction is generally difficult because of its complicated nature where the ocean, atmosphere, land and river interact with each other. In this study we focus on Tokyo Bay, one of the most important inland seas in Japan. To forecast the ocean currents and the biochemical fields in the bay, we are now developing a coupled system consisting of a regional ocean model known as ROMS, a regional atmospheric model known as SCALE and a river run-off model. We will also implement a data assimilation system based on the Ensemble Kalman Filter for the regional ocean-atmosphere coupled system to use station observations in the bay. In this poster presentation, we will show an overview of the system design and a limitation of an ocean-only system.

P-11

Impacts of dense surface observations on predicting a torrential rainfall event on September 9 and 10, 2015 in the east Japan area

Yasumitsu Maejima[1], Guo-Yuan Lien, and Takemasa Miyoshi
[1] *RIKEN AICS*

Abstract

To investigate the impact of dense surface observations on a torrential rainfall event occurred on September 9 and 10, 2015 in the east Japan area, we perform a series of data assimilation (DA) experiments using the Local Ensemble Transform Kalman Filter (LETKF) with the SCALE regional numerical weather prediction model. In this event, an active rainband was maintained for an extended period, associated with two typhoons (T1508; Etau and T1510; Krovanh), and caused torrential rainfalls over 600 mm/2days with catastrophic flooding of Kinu river.

Two DA experiments were performed: the control experiment (CTRL) at 4-km resolution with only hourly conventional observations (NCEP PREPBUFR), and the other with additional hourly dense surface observation data (TEST). CTRL showed general agreement with the observed weather patterns, although the track of T1508 was shifted westward. The heavy rainfall area was shifted to the west as well compared to the JMA analyzed precipitation based on the radar and gauge networks. By contrast, TEST showed stronger rainfall intensity, better matching with the observed precipitation likely due to an improvement of the track of T1508. The dense surface DA contributed to improve the moisture field in the lower layer, leading to an intensified rainfall amount. The results suggest that the dense surface DA have a potential to improve the forecast accuracy for severe rainfall events.

P-12

Data assimilation using shape features of rainfall areas

Taeka Awazu[1], Shigenori Otsuka[1], and Takemasa Miyoshi[1]
[1] *RIKEN AICS*

Abstract

Many verification methods have been proposed to evaluate similarity between two images. In the case of meteorology, pixel-by-pixel verification methods are commonly used to evaluate weather forecast. These methods have a limitation to evaluate the structure of an image. Therefore, this study aims to propose a new verification method that can evaluate the rainfall structure for weather forecast. Using the new verification method, we develop a method to assimilate the features of rainfall image structure in numerical weather prediction.

First, we split the entire rainfall image into small sections as in the human recognition. Next, we identify matched patterns between the observed rainfalls and forecasted ones. The image structure features are computed for each matched pair. The features represent slope, shape pattern, area, and rain intensity. We test the method with the JAXA ' s GSMaP NRT data (GSMaP: Global Satellite Mapping of Precipitation, NRT: Near-Real-Time). Moreover, we assimilate the features using the SPEEDY-LETKF system (SPEEDY: Simplified Parameterizations, Primitive Equation Dynamics, LETKF: Local Ensemble Transform Kalman Filter).

P-13

Assimilation of Himawari-8 All-Sky Radiances Every 10 Minutes: A Case of the September 2015 Kanto-Tohoku rainfall

Takumi Honda[1], Shunji Kotsuki, Guo-Yuan Lien, Yasumitsu Maejima, Kozo Okamoto, and Takemasa Miyoshi
[1] *RIKEN AICS*

Abstract

Severe rainfalls and their flood risk have attracted considerable attention from society because of their profound impacts including loss of life. To mitigate disasters associated with severe rainfalls, it is essential to obtain accurate precipitation forecasts in terms of intensity, location, and timing. To do so, data assimilation plays an essential role to provide better initial conditions. In particular, geostationary satellites are among the most important data sources because they can frequently observe a cloud system with broad coverage. In 2015, Japan Meteorological Agency (JMA) started full operations of a new generation satellite Himawari-8, which is capable of every-10-minute full disk observation and enables to refresh precipitation and flood predictions as frequently as every 10 minutes. This study aims to demonstrate the advantage of frequent updates of precipitation and flood risk predictions by assimilating all-sky Himawari-8 infrared (IR) radiances. We use an advanced regional data assimilation system known as the SCALE-LETKF, composed of a regional numerical weather prediction (NWP) model (SCALE-RM) and the Local Ensemble Transform Kalman Filter (LETKF). We focus on a major disaster case in Japan known as September 2015 Kanto-Tohoku heavy rainfall in which a meridional precipitation band associated with a tropical cyclone induced a record-breaking rainfall and eventually caused a collapse of a Kinu River levee. By assimilating a moisture sensitive IR band (band 9, 6.9 ?m) of Himawari-8 every 10 minutes into a 6-km mesh SCALE-LETKF, the heavy precipitation forecasts are greatly improved. We run a rainfall-runoff model using the improved precipitation forecasts and predict flood risk with longer lead times.

P-14

Multi-GPU Implementation of a Non-Hydrostatic Numerical Ocean Model with GPUDirect RDMA Transfer

Takateru Yamagishi[1], Matsumura Yoshimasa, and Hiroyasu Hasumi
[1] *Research Organization for Information Science and Technology*

Abstract

The new Pascal GPU architecture and GPUDirect technology are expected to accelerate numerical ocean models that involve systematic memory accesses and calculations and require frequent data transfers between computational nodes. We have implemented our “ kinaco ” numerical ocean model on Tokyo University ’s Reedbush supercomputer, which incorporates the latest NVIDIA P100 GPUs and GPUDirect technology. This study presents our introductory results and discusses how we have implemented and optimized our model for future detailed experiments with numerous cells.

While the optimization of our model is still at a preliminary stage, some functions are already highly accelerated? the velocity diagnosis functions, for example, are about seven times faster than on Intel Xeon E5-2695v4 CPUs. Our model also exhibits good scalability up to 16 processors. For exchanging halo data comprising three-dimensional values, the performance of inter-node data transfers using a CUDA-aware MPI library with GPUDirect is similar to that of transfers using a typical MPI library with CPUs.

Multigrid preconditioner performance, which contributes to stable convergence when using numerous computational nodes, is significantly worse for the Poisson/Helmholtz solver than when using CPUs, especially for coarse grids with fewer than 1024 (= 32 $\times$ 32) cells. As the first step in determining the cause of this problem, we extracted the inter-node data transfer function for coarse grids and evaluated its performance separately. We found that the inter-node communication latency was approximately same for both CPUs and GPUs, i.e., approximately 10 μ s with the help of GPUDirect ’ s high performance. However, the functions for packing/unpacking halo data to/from one-dimensional data performed poorly on GPUs because CPUs can take advantage of their L1 cache in this case while no cache tuning was applied to the multigrid preconditioner functions for GPUs. This indicates that it is necessary to effectively utilize the read-only data cache, registers, and shared memory with such a small number of cells, and we believe that this also applies to other preconditioner functions, such as matrix?vector multiplication.

In future work, we will focus on enhancing data locality on coarse grids for the multigrid preconditioner by explicitly utilizing the read-only data cache, registers, and shared memory. In such cases, CUDA ’ s intrinsic warp shuffle function could also be effective for matrix?vector multiplication. For other functions, such as tracer advection?diffusion, we will consider explicit register utilization to enhance data locality and increase instruction-level parallelism. We will also overlap kernel computation with inter-node data transfers.

P-15

Optimization of electromagnetic particle simulation code for distributed-memory parallel computer system

Hiroaki Ohtani[1], Youhei Miyake, Hiroshi Nakashima, Ritoku Horiuchi, and Shunsuke Usami
[1] *National Institute for Fusion Science*

Abstract

The sizes of simulations become larger and larger as the capabilities of computers are improved. One of important issues is how to develop an efficient simulation code. We have been developing a three-dimensional electromagnetic particle-in-cell simulation code PASMO [1] in order to investigate magnetic reconnection from the microscopic viewpoint. For performing the code on a distributed memory and multi-processor computer system with a distributed parallel algorithm, we decomposed three-dimensionally the simulation domain under the distributed parallel processing by MPI, and distributed the information of particles under thread parallel processing by OpenMP and automatic parallelization. We also introduced the charge conservation scheme [2,3] in order to exclude the global calculation, such as Poisson solver with FFT. On the other hand, in the simulation of magnetic reconnection, the number density of plasma is much higher in the central plasma sheet than in the lobe region. If the simulation domain is decomposed uniformly, a load balance problem appears in the parallelization. In order to avoid the load unbalance, we introduce a dynamic load balancing library, OhHelp [4]. Further, we optimize the algorithm on pusher and gather processes for performing on the Fujitsu FX100. In this paper, we report the performance of the optimized PASMO code.

Acknowledgement

This work was partially supported by the National Institute for Fusion Science (NIFS) Collaborative Research Program (NIFS16KNSS075, NIFS16KNXN335, and NIFS17KNTS046) and "Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures" in Japan.

References

[1] H.Ohtani and R.Horiuchi: Plasma and Fusion Research, Vol.4, 024 (2009).

[2] T.Zh.Esirkepov: Computer Physics Communications, Vol.135, pp. 144-153 (2001).

[3] H.Ohtani, R.Horiuchi, M.Nunami, S.Usami, N.Ohno: APS DPP Meeting Abstracts, PP8.130, (2014).

[4] H.Nakashima, Y.Miyake, H.Usui, and Y.Omura: Proc. Intl. Conf. Supercomputing, pp. 90-99 (2009).

P-16

Solidus and liquidus phases in Cu-rich area of Al-Cu alloys by Monte Carlo/Molecular dynamics (MC/MD) simulations

Sankar Kumar Deb Nath[1], Yasushi Shibuta, Munekazu Ohno,
Tomohiro Takaki, and Tetsuo Mohri
[1] *Tohoku Uniersity*

Abstract

We study the behaviors of solidus-liquidus curves in Cu rich area of the phase diagram of Al-Cu alloys using MC/MD simulations [1-2]. Using the solid-liquid interface coexisting system [3], the melting point of Cu is obtained applying embedded atom method (EAM) in MD simulation. Solid and liquid Cu are obtained heating Cu atoms at temperatures 300K and 2000K, respectively, using NVT (N; number of atoms, V; volume, T; temperature in K) by MD simulations. Then both the solid and liquid Cu are heated separately below their melting temperatures using NPT (N; number of atoms, P; pressure, T; temperature in K) and NVT canonical ensembles by MC/MD simulations to obtain a value of the average composition Al (xAl) by imposing a value of $\Delta\mu$, which changes some Cu atoms to Al [1-2], where $\Delta\mu$ is the difference in chemical potential between Al and Cu for a particular phase. Below the melting temperature of Cu, we obtain $\Delta\mu$ vs. xAl relationships for each phase. Mixing free energy (Δ Gmix) for each phase of Al-Cu alloys is determined integrating the relationship $\Delta\mu$ vs. xAl with respect to xAl at given temperatures. The grand canonical potential $\Omega = \Delta$ Gmix - $\Delta\mu$ .xAl vs. $\Delta\mu$ for each phase is plotted together to obtain the equilibrium solidus and liquidus compositions from the intersection point. The obtained solidus-liquidus curves in Cu-rich area are compared with those obtained by CALPHAD method to examine the present study [3].

References

[1] B. Sadigh, P. Erhart, A. Stukowski, A. Caro, E. Martinez, and L. Zepeda-Ruiz, Physical Review B 85, 184203(2012).

[2] S.J. Plimpton, J. Comput. Phys. 117, 1 (1995).

[3] S.K. Deb Nath, Y. Shibuta, M.Ohno, T.Takaki, and T.Mohri, ISIJ International 57, 1774 (2017).

P-17

Accounting for the observation error correlation in data assimilation

Koji Terasaki[1] and Takemasa Miyoshi
[1] *RIKEN AICS*

Abstract

It is known that some observations such as satellite radiances have the spatial and inter-channel error correlations. The current data assimilation systems in the operational centers neglect the observation error correlations. It is important to account for the observation error correlations to effectively draw the information from the observation " BigData. "

In this study, we developed a method to account for the observation error correlations in the local ensemble transform Kalman filter (LETKF: Hunt et al. 2007), and performed idealized experiments with the Lorenz-96 model (Lorenz et al. 1998) and the non-hydrostatic icosahedral atmospheric model (NICAM: Satoh et al. 2013). The condition number of the observation error covariance matrix (R matrix), or the ratio of the maximum and minimum eigenvalues, is essential for the stable performance of the LETKF. Reconditioning can make the R matrix well-conditioned by adding a small constant to all the diagonal terms of the R matrix. We examined that the reconditioning not only stabilizes the LETKF but also greatly improves the analysis accuracy by including the observation error correlations.

P-18

Development of VISMO

Nobuaki Ohno[1] and Hiroaki Ohtani
[1] *University of Hyogo*

Abstract

We have been developing an in-situ visualization tool VISMO for about 5 years. This tool is provided as a module of Fortran90, which makes it easy to combine simulation codes with VISMO. Usually, in-situ visualization tools output pictures as the results. Because of that, simulation researchers, who carry out in-situ visualization, cannot observe visualization objects such as Isosurface from various viewing points. We implemented a function to the tool to output visualized isosurfaces, color slices, arrows and stream lines as forms of point clouds that can be 3d-reconstructed. By this, visualization objects such as Isosurface can be observed from all viewing points. Furthermore, they can be displayed them even in the CAVE systems.

CANCELLED

P-19

Data assimilation experiments with MODIS LAI observation and the dynamical global vegetation model SEIB-DGVM at a deciduous broad-leaved forest in Japan

Hazuki Arakida[1], Hazuki Arakida[1], Sachiho A. Adachi[1], Shunji Kotsuki[1], Shigenori Otsuka[1], Hisashi Sato[2], and Takemasa Miyoshi[1]
[1] *RIKEN AICS*
[2] *JAMSTEC*

Abstract

In the previous study, Arakida et al. developed a data assimilation system with a dynamical global vegetation model known as the SEIB-DGVM (Spatially Explicit Individual-Based Dynamic Global Vegetation Model), and assimilated satellite-observed Leaf Area Index (LAI) successfully over eastern Siberia. In this study, we extend the experiment to a Japanese forest at the Takayama flux site and investigate the performance of the data assimilation system. We estimate the state variables including carbon flux, water flux, heat flux, vegetation structure, and model parameters related to the phenology of the deciduous broad-leaved tree and the undergrowth. The original data assimilation system developed for Siberia turned out not to be suitable for the Japanese forest because the LAI stays at the saturated level for an extended period in the summer in Japan, while in Siberia, the LAI starts to drop soon after it reaches to the maximum value. Therefore, a model parameter which controls the maximum LAI value in the summer was considered for additional parameter estimation. The results show that the DA system performs generally well in Japan with the additional parameter estimation.

CANCELLED

P-20

Consideration on Extending Silk Road Framework for Large-scale Distributed Collaborative Simulation

JIACHAO ZHANG[1], Yoshitaka Kumada, Shinji Fukuma, and Shin-ichiro MORI
[1] *University of Fukui*

Abstract

Interactive simulation allows scientific simulation with real-time interaction from the real world, which gives more dynamic, flexible, and ongoing analysis of phenomena. With the development of explosively emerging Internet of Things (IoT) devices, it 's expected to consider incorporating the possibility of interacting with these IoT sensing data in wide geographical distributions, which is important for real-time applications such as Cypher Physical System (CPS). However, for real-time interaction and high resolution accuracy on interested domains, the traditional centralized HPC center such as supercomputer cannot support both above requirements due to the existing of ultra-high network delay and the pitfalls of common batch processing; neither nor using a series of isolated regional servers due to the lack of mandatory boundary conditions for simulations since the simulated phenomena are continuous on space and temporal. In such background, we had proposed a framework named Silk Road realizing a sort of fog computing architecture to give both real-time interaction and high-resolution result. Its multi-scale bi-directional refinement model utilizes regional servers to perform high-resolution on limited domain of interest (DOI), and additionally utilize a central server to perform lower-resolution but global area simulation to couple each regional simulation in order to exchange necessary boundary conditions for each regional server.

However, the current framework performs computing task in a single MPI processor on either central server or regional server side, which cannot utilize the power of backend computing resources of a HPC center. It may not affect the effectiveness of framework when serves small-scale simulation, but computing may become performance bottleneck for real-time while handling large-scale simulation. So, an extending of current Silk Road Framework is considered to enable parallel computing on both central and regional server side. The plan is to utilize dynamic processors spawning, an advanced feature of MPI-2, to perform parallel computing on backend computing nodes. Typical domain decomposition will be considered as essential functions for parallelism, and input distribute and results merge will be also considered.

P-21

BONSAI (Benchtesting OpeN Software Autotuning Infrastructure)

Yaohung Mike Tsai[1], Matthew Bachstein, Piotr Luszczek, Jakub Kurzak,
Hartwig Anzt, Mark Gates, and Jack Dongarra
[1] *University of Tennessee, Knoxville*

Abstract

In this poster, we will introduce our auto-tuning framework: BONSAI.

The goal of the BONSAI (Benchtesting OpeN Software Autotuning Infrastructure) project is to develop a software infrastructure for using parallel hybrid systems at any scale to carry out large, concurrent autotuning sweeps in order to dramatically accelerate the optimization process of computational kernels for GPU accelerators and many-core coprocessors. We will demonstrate the key components of BONSAI, including LANAI for specifing the search space and pruning constraints, the process of compilation, benchmarking, analyzing, as well as tuning results we had for various kernels.

P-22

Event-driven Computing on HPC: Experiments with Scientific Applications

Hyungro Lee[1], and Geoffrey C. Fox
[1] *Indiana University*

Abstract

High performance computing requires to deal with diverse applications running on different environments, and HPC containers which encapsulate required software components and configurations in an image enable reproducibility of your workloads and increase portability across platforms. Serverless computing or Function-as-a-Service (FaaS) deploys a container to execute lines of codes without management of infrastructure which is more like a simple version of Platform-as-a-Service (PaaS) of cloud computing. Commercial cloud providers such as Amazon, Google, Microsoft and IBM offer serverless computing with various function runtimes along with open source implementations e.g. OpenWhisk, Fission and Kubeless but any of these platforms are suitable for HPC applications which require particular libraries and specific hardware e.g. GPUs and interconnects. We. however, this “ serverless ” model may find useful in processing workloads in certain circumstances, for example, a large number of small tasks that can be run in parallel for very short periods of time or a workflow consisting of small input and output messages in each component. We investigate existing serverless computing platforms to evaluate its possible use cases on HPC sites with regarding to 1) elasticity, 2) scalability, 3) portability, and 4) continuous deployment. This work conducts a study of building reproducible application environments on two different systems; HPC and Clouds with our comparison results for serverless platforms in regard to computational performance, latency, event handler capacity, and elasticity. Our results indicate that a function invocation spawns at least a thousand of concurrent computation containers in seconds and keep alive for another tasks until its deprovisioning whereas long-running HPC applications with a large size of libraries and binary executables are not yet

P-23

Assimilating dense precipitation radar data: a simulation study

Atsushi Okazaki[1], Takumi Honda, Shunji Kotsuki, and Takemasa Miyoshi
[1] *RIKEN AICS*

Abstract

Precipitation radar observations have been playing an important role in meteorology. Recently, a new radar system known as the phased array radar has been developed, scanning the three-dimensional structure of precipitation much faster and denser than the conventional parabolic type radar. The advancement can also be seen in the spaceborne precipitation radar. The GPM core satellite, the successor of the TRMM satellite, has the newly-developed dual-frequency precipitation radar (DPR). Precipitation radars keep advancing and tend to provide denser and more frequent observations. Therefore, it would be essential to develop a method to effectively use radar reflectivity data for numerical weather prediction through data assimilation. Previous studies showed some success in data assimilation of radar reflectivity for convective-scale and tropical cyclone analyses. Nevertheless, it is still difficult to build a general approach to data assimilation of radar reflectivity due to various factors such as the complex observation operator, strong nonlinearity and model errors in the moist physical processes. In this study, we aim to develop an effective data assimilation method which can fully exploit the radar reflectivity data. We perform an observing system simulation experiment, in which we assume that reflectivity data are available at all model grid points. As the first step, we focus on the case of Typhoon Soudelor (2015), the strongest typhoon in the West Pacific in 2015. In the presentation, we will report the impact of dense radar observations on the analyses and forecasts of Typhoon Soudelor.

P-24

RIKEN precipitation nowcasting systems

Shigenori Otsuka[1], Shunji Kotsuki[1], Marimo Ohhigashi[1], Yutaka Ishikawa[1], Takemasa Miyoshi[1], Shinsuke Satoh[2], and Tomoo Ushio[3]
[1] *RIKEN AICS*
[2] *NICT*
[3] *Tokyo Metropolitan U.*

Abstract

Data Assimilation Research Team is operating two real-time precipitation nowcasting systems for research purposes. In May 2017, RIKEN obtained a weather forecasting license from the Japan Meteorological Agency and started real-time predictions open to public. We are running the Phased-Array Weather Radar (PAWR) three-dimensional (3D) precipitation nowcasting system and the Global Satellite Mapping of Precipitation (GSMaP) RIKEN Nowcast (RNC) system.

The PAWR 3D nowcasting system transfers observation data from the radar site to RIKEN AICS every 30 seconds in real time using a newly-developed Just-In-Time Data-Transfer (JIT-DT) library. JIT-DT monitors new observation files in the radar system and transfers them to AICS. Three-dimensional motion vectors of radar echoes are computed and used for the 10-minute space-time extrapolation of the radar echoes. The PAWR 3D nowcasting covers a 40 km x 40 km area centered at Suita Campus, Osaka University, or NICT lab in Kobe city.

The GSMaP RNC system provides hourly-updated global precipitation nowcasts up to 12 hours. The GSMaP Near-Real-Time (NRT) product is provided by the Japan Aerospace eXploration Agency (JAXA) with a four-hour delay after the nominal time of satellite observations. GSMaP RNC computes motion vectors from consecutive two NRT images, and the Local Ensemble Transform Kalman Filter (LETKF) is used to obtain a full field of continuous motion vectors.

P-25

Automatized Algorithm Generation for Numerical Methods with Application to Limiter Design

Thomas Camminady[1], and Birte Schmidtmann
[1] *Karlsruhe Institute of Technology*

Abstract

In the past, developing numerical methods has gone a theoretical path. On the other side, modern approaches of data science and machine learning are able to make predictions without prior knowledge or theoretical foundation.

In this work we show that simple optimization techniques are capable of deriving parameters for numerical schemes which so far have only been derived through rigorous theoretical analysis.

This approach is not only limited to the finite volume framework, but can be generalized to multiple types of parameter estimation inside numerical methods.

P-26

Analysis of the Load of Local Meta Data Server Access from the Job I/O on the K computer

Akiyoshi Kuroda[1], S. Inoue, K. Koyama, R. Sekizawa, and K. Minami
[1] *RIKEN AICS*

Abstract

Recently, various problems have been reported in operations of the HPC system. The fluctuations of job execution times are one of the difficult problems that need to be investigated and solved as they cause complications. There are cases wherein the fluctuation of execution time is large. In such cases, jobs are not completed within their scheduled time. As a result of the analysis, it was found that there were cases wherein loads of metadata servers on the local I/O (LMDS) were saturated due to achieving its performance limit. In FY2016, we focused on I/O performance and analyzed the load of LMDS instructions by jobs.

In FY2015, we constructed the job performance database on the job analysis server [1]. In FY2016, the number of LMDS instructions of each job I/O was added to the database. We analyzed the LMDS loads due to job I/O using this database [2]. There are some heavy jobs for LMDS. It also contains the total and other jobs, excluding the three heavy jobs. This analysis shows that when one of job exceeded 25,000 [OPS], the other jobs could not access LMDS because of saturation. The number of MPI processes of jobs having a high load for LMDS was about 2,000, which was not so large. The jobs having a high load for LMDS often use the shared files instead of the local files on the local I/O file system, and then the number of LMDS instructions tends to increase in proportion to the number of MPI processes. Therefore, we investigated the number of MPI processes at which the load of LMDS instructions saturates. The number of the saturated MPI processes was calculated as follows: (1) Each job has its own number of LMDS instructions A and number of processes B. (2) The saturated MPI process is defined as $25,000 \div A \times B$. It can be observed that the number of jobs increases rapidly in the range from 300 to 1000 MPI processes. It was found that the load of LMDS was saturated easily even if the number of the MPI processes was small.

And further analysis illustrates that if the number of MPI process is doubled, the number of saturating jobs is increased by 6 times, the saturated time is increased by 14 times, and therefore the wasted node hour is increased by 28 times. It is predicted that the load saturation of the LMDS will occur more frequently with a significant increase of processes in the future. It is, therefore, important to monitor the load of the LMDS through jobs using shared files.

References

[1] A. Kuroda, S. Inoue, K. Koyama, R. Sekizawa and K. Minami: "Development of Job Analysis System for Power-saving and Operational Improvement on the K computer", IPSJ SIG Technical Reports (HPC), Otaru, JAPAN, Vol. 2016-HPC-156 No.7, pp.1-7 (2016.9.15-16).

[2] A. Kuroda, S. Inoue, R. Sekizawa and K. Minami: "Analysis of the Load of Local Meta Data Server Access from the Job I/O on the K computer", IPSJ SIG Technical Reports (HPC), Akita, JAPAN, Vol.2017-HPC-160, No.27, pp.1-6 (2017.7.26-28).

P-27

Parallel particle-based volume rendering with 234Compositor for large-scale unstructured volume data visualization

Kengo Hayashi[1], Yoshiaki Yamaoka[2], Naohisa Sakamoto[1], and Jorji Nonaka[3]
[1] *Graduate School of System Informatics Kobe University*
[2] *Kobe University*
[3] *RIKEN Advanced Institute for Computational Science*

Abstract

Numerical simulations using high performance computing (HPC) systems have generated a vast amount of distributed and complex data sets, such as the time-varying, unstructured volume data sets. The visualization of unstructured volume data, by using the traditional volume rendering technique, requires the time-consuming visibility sorting process. To minimize this visibility sorting problem, we focused on the PBVR (Particle-based Volume Rendering) method, which is a visibility sorting free approach. In this work, we propose a hybrid MPI/OpenMP parallel PBVR, which parallelizes the particle generation, rendering, and the final image composition process. For the latter stage, we utilized the 234Compositor for executing the parallel image composition. We verified the effectiveness by applying the parallel PBVR to a large-scale unstructured thermal fluid simulation result.

P-28

Performance evaluation of multi-GPU implementation of RI-MP2 method on Tesla P100 GPUs

Michio Katouda[1], Takahito NAKAJIMA[1]
[1] *RIKEN Advanced Institute for Computational Science*

Abstract

We performed the massively parallel implementation of resolution-of-identity second order Mller-Plesset perturbation (RI-MP2) energy calculation suitable for calculations of large molecules on K computer and CPU/GPU hybrid supercomputers. [1,2] Performance of the implementation was tested on the K computer (CPU clustering system) and TSUBAME 2.5 (CPU/GPU hybrid system) demonstrating high efficiency: The peak performance of 3.1 PFLOPS is attained using 80199 nodes of the K computer. The peak performance of the multi-node and multi-GPU implementation is 514 TFLOPS using 1349 nodes and 4047 GPUs of TSUBAME 2.5. Recently, we have ported the code for systems where Tesla P100 GPUs are installed. In this presentation, we report the overview of implementation and the results of performance evaluation. We also discuss the performance comparison between the results using Tesla P100 and those using Tesla K20X and K40.

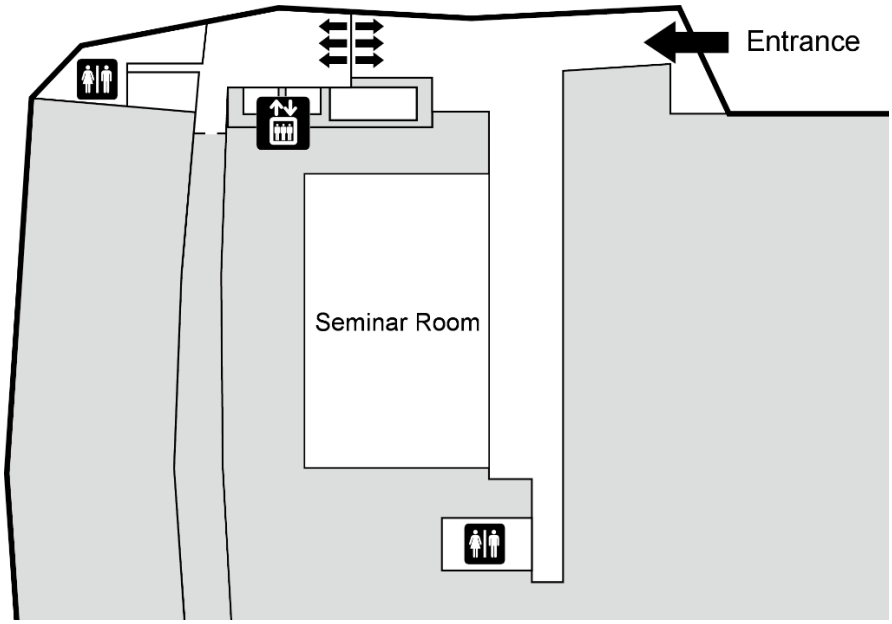
References

[1] M. Katouda and T. Nakajima, MPI/OpenMP Hybrid Parallel Algorithm of Resolution of Identity Second-Order Mller—Plesset Perturbation Calculation for Massively Parallel Multicore Supercomputers, J. Chem. Theory Comput., 2013, 9, 5373—5380.

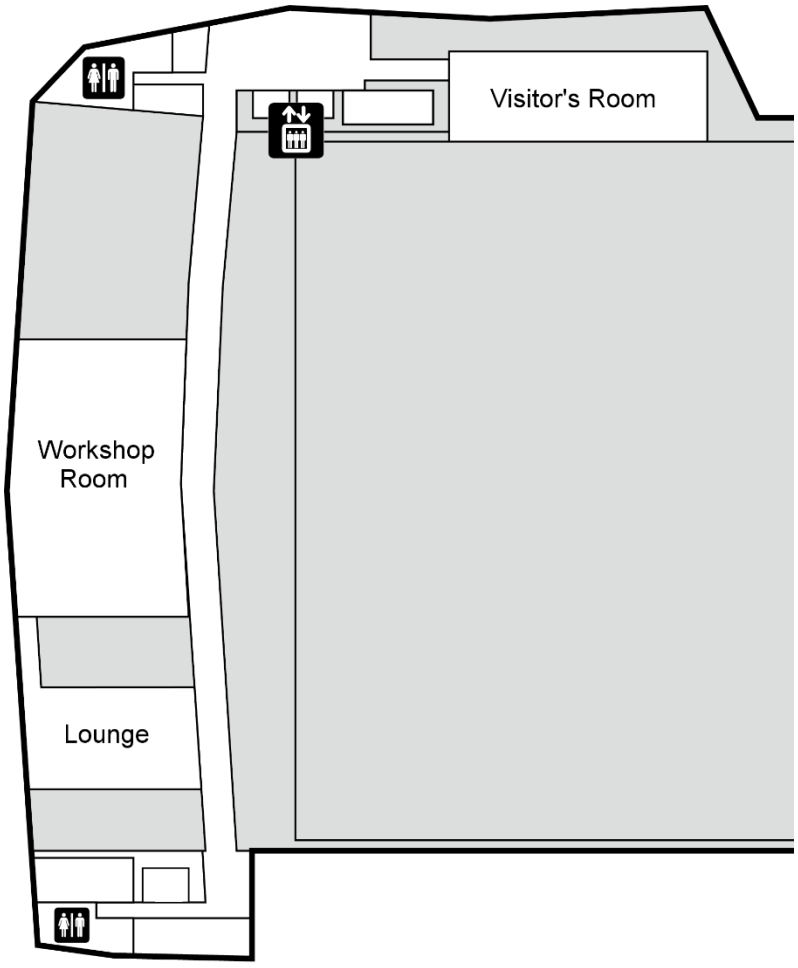
[2] M. Katouda, A. Naruse, Y. Hirano and T. Nakajima, Massively parallel algorithm and implementation of RI-MP2 energy calculation for peta-scale many-core supercomputers, J. Comput. Chem., 2016, 37, 2623—2633.

Floor Map

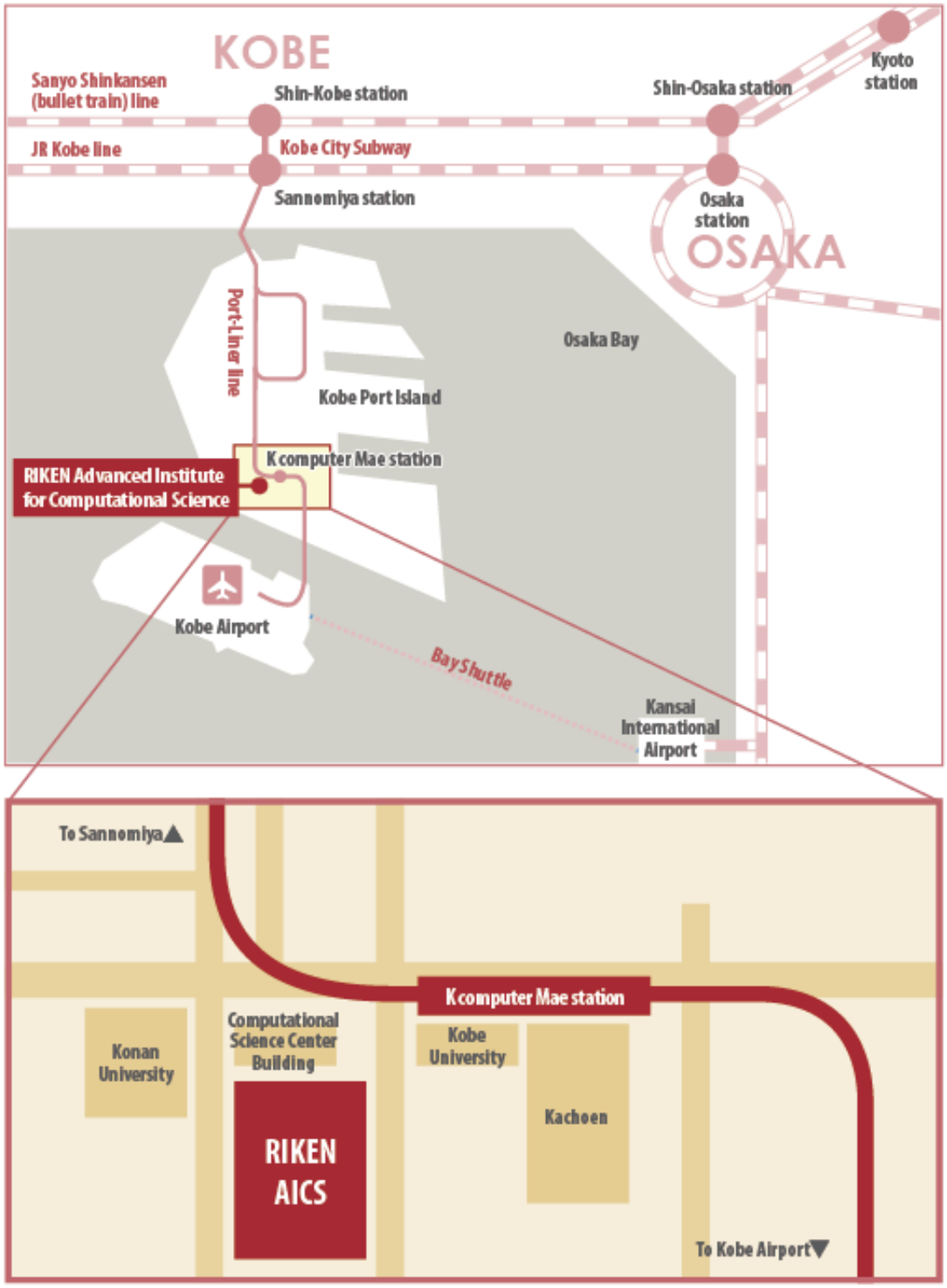
1st floor



6th floor



Area Map


KOBE
Sanyo Shinkansen (bullet train) line
Shin-Kobe station
Shin-Osaka station
Kyoto station
JR Kobe line
Kobe City Subway
Sannomiya station
Osaka station
OSAKA
Port-Liner line
Osaka Bay
Kobe Port Island
K computer Mae station
RIKEN Advanced Institute for Computational Science
Kobe Airport
Bay Shuttle
Kansai International Airport
To Sannomiya
K computer Mae station
Konan University
Computational Science Center Building
Kobe University
Kachoen
RIKEN AICS
To Kobe Airport

Editor & Publisher

RIKEN Advanced Institute for Computational Science
Minatojima-minami-machi 7-1-26, Chuo-ku, Kobe, 650-0047 JAPAN
E-mail: aics-sympo2018-reg@ml.riken.jp
Web: http://www.aics.riken.jp/

Published February, 2018