The 7th AICS INTERNATIONAL SYMPOSIUM

Emerging numerical techniques for exascale and post-Moore era

Feb 23-24, 2017

Venue: Integrated Research Center of Kobe University, Convention Hall

7-1-48 Minatojima-minami-machi, Chuo-ku, Kobe, Japan

The 7th AICS International Symposium will focus on computing and computational science in the exascale and post-Moore era, and explore the directions and possibilities from the perspective of numerical algorithms.

Specifically, researchers are being invited from around the world to discuss the algorithms and approaches that will be necessary to fill the gaps between the speed of computing and bandwidths of memory and communication in exascale systems.

Such algorithms and approaches include explicit methods as reduced speed of sound, temporal blocking to reduce required memory bandwidth, and use of the Fast Multipole Method (FMM) to reduce required communication bandwidth.

Organized by
RIKEN Advanced Institute for Computational Science

Supported by
Priority issues (9 issues) on Post-K computer
RIKEN Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS)
Research Organization for Information Science and Technology (RIST)

This symposium is subsidized by
FOCUS Establishing Supercomputing Center of Excellence

http://www.aics.riken.jp/AICS-Symposium/2017/index.php

RIKEN AICS

Program

■ Thursday, February 23 (Day 1)

09:30-10:00	**Opening Session** (Convention Hall, 2nd Floor)
	Opening Address: RIKEN AICS Overview Akira Ukawa (AICS)
10:00-10:30	**Break**
10:30-11:10	**Keynote talk**
	"Multi-Hybrid Platform for Next Generation Computational Science" Taisuke Boku (University of Tsukuba)
11:10-12:00	**Session Ⅰa: Explicit methods and their variations**
11:10-11:40	Invited talk : "Modeling the coupled solar convection zone/photosphere/corona system" Matthias Rempel (NCAR)
11:40-12:00	Contributed talk: "Variable Inertia Method: an explicit method for mantle convection simulations" Takayuki Saitoh (Tokyo Institute of Technology)
12:00-13:30	**Lunch and Coffee Break**
13:30-14:00	**Session Ⅰb: Explicit methods and their variations** (continued)
	Invited talk : "Solar global convection and dynamo with reduced speed of sound technique" Hideyuki Hotta (Chiba University)
14:00-15:00	**SessionⅡ: Temporal blocking and related techniques**
14:00-14:30	Invited talk : "Making sense of temporally blocked stencil performance via analytic model" Georg Hager (RRZE)
14:30-15:00	Invited talk : "Automatic Generation of Stencil Codes from Mathematical Specifications" Takayuki Muranushi (AICS)
15:00-15:30	**Break**
15:30-16:30	**SessionⅢa: FMM**
15:30-16:00	Invited talk : "Energy Conservation of Fast Multipole Methods in Classical Molecular Dynamics Simulations " Rio Yokota (Tokyo Institute of Technology)
16:00-16:30	Invited talk : "Optimizing FMM in GROMACS" Berk Hess (KTH)
16:30–18:00	**Poster Session** (Entrance Hall, 1st Floor)
18:00–19:30	**Banquet** (Lounge, 4th Floor)

■ Friday, February 24 (Day 2)

9:40-10:20	**Keynote talk** (Convention Hall, 2nd Floor)
	"The Sunway TaihuLight Supercomputer: Updates on System and Application" Lin Gan (Tsinghua University)
10:20-10:50	**Break**
10:50-12:30	**SessionⅢb: FMM**
10:50-11:20	Invited talk : "Improving Performance and Scaling of Molecular Dynamics Simulation with the Multilevel Summation Method" David J. Hardy (UIUC)
11:20-11:50	Invited talk : "Exascale N-body algorithms for data analysis and simulation" George Biros (U.T.Austin)
11:50-12:10	Contributed talk: "Is fast multipole method fast? Another multipole method in FFT convolution over uniform mesh" Keigo Nitadori (AICS)
12:10-12:30	Contributed talk : "Towards Exascale in High-Order Computational Fluid Dynamics" Niclas Jansson (KTH)
12:30-14:30	**Lunch and Coffee Break**
13:30-14:15	**AICS and K computer Tour** (AICS Entrance Hall and Visitor Hall, 6th Floor)
14:30-15:30	**Session Ⅳ: New Problems and New approaches** (Convention Hall, 2nd Floor)
14:30-15:10	Invited talk : "Large-Scale Deep Learning: Achievements and Challenges" Takuya Akiba (Preferred Networks)
15:10-15:30	Contributed talk : "Parallel Implementation of Shift-invert Rational Krylov method" Yuka Hashimoto (Keio University)
15:30-16:00	**Break**
16:00-17:00	**Closing Panel Discussion** (Convention Hall, 2nd Floor)

23 Feb (Thu), 10:30-11:10

Multi-Hybrid Platform for Next Generation Computational Science

Taisuke Boku[*1]
[*1] University of Tsukuba

Abstract

For so called "Post-Moore" era, we should prepare various aspects of effort on processor speed, semiconductor utilization, power consumption, memory speed and capacity, etc. For the transistor utilization and performance per power consumption, we have been achieved great success with high performance accelerating devices such as GPU additionally to main CPU. However, such a simple solution based on large degree of SIMD type parallel execution on data parallel manner is not perfect for large variety of applications. Moreover, the issue on interconnection latency and throughput becomes clearly the serious bottleneck on the system scalability and performance. Another trend of accelerating device research focuses on FPGA for HPC although the programmability and applicability for real application are still unclean. In this talk, I will focus on the possibility to apply FPGA for problems partially not suitable for GPU nor CPU in real computational sciences in next generation. There are several challenges on FPGA utilization including programming, performance/power, I/O bottleneck by weak interface, etc. The talk will provide some idea and hints to solve these problems with several examples which we are working for now. Finally, the perspective on the next generation's multi-hybrid system is mentioned.

23 Feb (Thu), 11:10-11:40

Modeling the coupled solar convection zone/photosphere/corona system

Matthias Rempel[*1], Hideyuki Hotta
[*1] National Center for Atmospheric Research

Abstract

The solar convection zone and atmosphere are highly stratified systems. From the base of the convection zone into the solar photosphere the density drops by about a factor of 10^6, from the photosphere into the corona by about a factor of 10^8. As a consequence there is a wide spread in characteristic time scales that makes numerical modeling very challenging and time consuming. While in the photosphere the flow velocity (v), speed of sound (cs) and Alfven velocity (va) are comparable, this is not the case in the deep convection zone where v is on the order of 10-100 m/s, whereas cs reaches 200 km/s. In the corona v and cs are on the order of a few 100 km/s, but va can exceed values of 100,000 km/s. While either implicit treatment or filtering of fast processes are widely used, I will present in this talk methods that artificially limit the fastest characteristic speeds. The advantage of this approach is that the system of equations remains hyperbolic and the required modification to the codes are minor. Since the numerical method remains fully explicit good scaling to large processor counts is almost trivial. I will discuss in particular the Reduced Speed of Sound Technique (RSST) that is applicable to low Mach number flows in the deep convection zone. I will present a few applications ranging from convection to dynamo simulations and present a recent scaling experiment that demonstrates scaling of this approach to the full size of the K-Computer. The second approach I will discuss is the "Boris correction" (semi-relativistic MHD with an artificially reduced speed of light) that has been widely used in simulations of magnetospheric dynamics. Applied to solar physics, this method enables efficient simulations of the solar corona by limiting the Alfven velocity. Also here the method required only the the addition of computationally inexpensive correction terms that do not change the hyperbolic and explicit character of the numerical approach. I will present recent examples including a comprehensive simulation of a solar flare that is triggered by photospheric flux emergence.

23 Feb (Thu), 11:40-12:00

Variable Inertia Method: an explicit method for mantle convection simulations

Takayuki Saitoh[*1], Kosuke Takeyama, Natsuki Hosono, Junichiro Makino, Satoko Yamamoto, Takaaki Takeda, Daisuke Namekata, Takayuki Muranushi
[*1] Tokyo Institute of Technology

Abstract

3D mantle simulation is one of the challenging topics for the current numerical studies, since the mantle has extreme conditions such as high Prandtl number (Pr$\sim 10^{24}$), low Mach number (M$\sim 10^{-16}$), and Low Reynolds number (Re$\sim 10^{-20}$). In order to handle this extreme material, almost all simulations adopt implicit methods by ignoring the inertia term and by adopting the (extended) Boussinesq and incompressible approximations. However, this requires frequent global matrix inversions, resulting in low performance in the present-day massive parallel supercomputers. Instead of using the conventional formulation, we developed a new explicit scheme for mantle convection. Since non-dimensional numbers of the mantle are extremely high or low and a flow's property does not change as long as the changes of non-dimensional numbers are sufficiently small, we can change the non-dimensional numbers by introducing several parameters so that we can use larger time steps. This is based on the reduced sound speed method (Rampel 2005) which is used to achieve a large simulation of the Sun (e.g., Hotta et al. 2012, 2014). To obtain larger time steps, we multiply a factor to the inertia term. In addition to this, we introduce two factors to alter the viscosity and thermal diffusion terms in order to keep the original Rayleigh number of the mantle. We call this method the variable inertia method (VIM; Takeyama, Saitoh, & Makino 2017). According to our numerical experiences, as long as the modified non-dimensional numbers are being reasonable ranges (Pr > 10, M < 1, and Re < 1), we obtain the almost identical results with the conventional formulation. We believe that this explicit method is a clue to reveal the whole thermal history of the Earth. In this presentation, as well as the introduction of the formulation of VIM, we will explain the recent effort to implement a parallel version of VIM using the framework for developing particle simulator (Iwasawa et al. 2016).

23 Feb (Thu), 13:30-14:00

Solar global convection and dynamo with reduced speed of sound technique

Hideyuki Hotta[*1], Matthias Rempel, Takaaki Yokoyama
[*1] Chiba University

Abstract

Thermal convection and turbulent dynamo are essential keys to understand the 11-year cycle of the solar activity. Since the speed of sound in the solar convection zone is much faster than the convection velocity and the Alfven velocity, which is related to the magnetic field, the time step is severely restricted by the CFL condition. The anelastic approximation where the speed of sound is assumed infinite and the propagation of the sound wave is not solved, is frequently used. The anelastic approximation, however, also has problems especially in parallel computing. The infinite speed of sound transports information in a CPU to all CPUs every time step. This requires frequent global communications. We newly adopt the reduced speed of sound technique (RSST) to increase spatial resolution with using massive supercomputers such as K. The RSST reduces the speed of sound with modifying the perturbation density and no global communication is required. With the RSST, we nicely use the entire system of the K-computer. We have revealed important small-scale physics about the differential rotation and the solar dynamo. In the presentation, the results of our high-resolution calculations are reviewed.

23 Feb (Thu), 14:00-14:30

Making sense of temporally blocked stencil performance via analytic model

Georg Hager[*1]
[*1] Erlangen Regional Computing Center (RRZE)

Abstract

Many techniques have been devised to improve the performance of stencil algorithms on cache-based multicore CPUs. The main goal is to decouple from the scarce resource of main memory bandwidth, but this is just where the real challenges begin: How can the equally scarce cache space be used most effectively? What is the next bottleneck beyond memory bandwidth? Does it make sense to block for higher-level caches? What is the role of low-level code quality? What is a good parallelization strategy? Some of these questions can be answered by auto-tuning techniques, but others require deeper analysis with the help of analytic performance models. Such models enable us to pinpoint relevant performance issues and sometimes lead to surprising insights. The talk gives an introduction to useful analytic performance models of streaming kernels and how they can by applied to temporally blocked stencil algorithms. Using relevant corner cases we demonstrate how far these models can take us and where they stop being purely predictive. Beyond this point one can still use the principles behind the models to learn more about the bottlenecks or shortcomings of running code by constructing the models not from first principles but from performance counter measurements. We call this approach ``phenomenological modeling." Backed by these concepts we present an analysis and comparison among state-of-the-art stencil frameworks.

23 Feb (Thu), 14:30-15:00

Automatic Generation of Stencil Codes from Mathematical Specifications

Takayuki Muranushi[*1], Hideyuki Hotta, Junichiro Makino, Seiya Nishizawa, Hirofumi Tomita, Keigo Nitadori, Masaki Iwasawa, Natsuki Hosono, Yutaka Maruyama, Hikaru Inoue, Hisashi Yashiro, Yoshifumi Nakamura

[*1] RIKEN Advanced Institute for Computational Science

Abstract

We have designed and demonstrated programming language Formura, that can generate highly-optimized stencil codes from the succinct specifications of the applications. Stencil computation is a type of simulation algorithm, where the simulated state is represented as multidimensional arrays, and the value of an array cell at the next timestep is defined as functions of the values of the cell's neighbor at the previous timestep. Stencil computanion is used in a wide range of applications such as weather simulations, earthquake simulations, metallurgy, and image processing. However, writing computer code to perform such simulations on supercomputers takes a lot of effort both in writing and the optimization of the codes. Particularly, such applications can be extremely memory-intensive. On the other hand, modern processors tend to have limited memory bandwidth. As a consequence, the lack of memory bondwitdth between processor speed and memory bandwidth has recently become a major headache for supercomputing applications. The syntax of formura allows intuitive expression over array elements while ensuring the program is a stencil computation. The syntax and the type system ensures that Formura programs are always stencil computations. Therefore, Formura compiler can succesfully apply state-of-the-art stencil optimization techniques such as temporal blocking to the programs it generates. We have demonstrated the usefulness of Formura by implementing a xpreliminary below-ground biology simulation. Optimized C-code are generated from 672 bytes of Formura program. The simulation was executed on the full nodes of the K computer, with 1.184 Pflops, 11.62% floating-point-instruction efficiency, and 31.26% memory throughput efficiency. In this presentation, we would like te give the overall introduction to the Formura lanugage, the internals of the language-processing system, the code generation system and the optimization system. We would also like to discuss the future directions of Formura.

23 Feb (Thu), 15:30-16:00

Energy Conservation of Fast Multipole Methods in Classical Molecular Dynamics Simulations

Rio Yokota[*1]
[*1] Tokyo Institute of Technology

Abstract

Fast Multipole Methods (FMM) have a rare combination of $O(N)$ computational complexity, high arithmetic intensity, and excellent scalability. These properties make it an interesting alternative to the FFT-based particle-mesh Ewald methods for large scale MD calculations. However, FMM has previously been reported to suffer from energy conservation issues. This talk will cover this issue of energy conservation of FMM in MD simulations.

23 Feb (Thu), 16:00-16:30

Optimizing FMM in GROMACS

Berk Hess[*1]
[*1] KTH Royal Institute of Technology

Abstract

Commonly used particle mesh electrostatics methods employ a spherical pair interaction cut-off, which fits well with the spherical cut-off for the Van der Waals interactions. However, on modern hardware with SIMD instruction or GPUs which perform up to 16 operations at once, a spherical interaction volume is inconvenient. The cell-cell pair interactions used in FMM are a much better fit for modern hardware. We have linked the ExaFMM package to GROMACS, where the short-range particle pair interactions are computed by GROMACS, together with the Van der Waals interactions. Here we show how we can take advantage of the more regular interaction volume by using a staggered FMM cell grid and infrequent cell pair updates. We also show how FFM regularization, which improves energy conservation, fits into this setup.

24 Feb (Fri), 9:40-10:20

The Sunway TaihuLight Supercomputer: Updates on System and Application

Lin Gan[*1]
[*1] Tsinghua University

Abstract

This talk focuses on the latest updates on applications that are using the Sunway TaihuLight supercomputer. I will briefly introduce the architectures of both the homegrown many-core processor and the 125 PetaFlops supercomputing system. Applications from earth system, green energy, big data and life science will then be presented to show the inspiring results.

24 Feb (Fri), 10:50-11:20

Improving Performance and Scaling of Molecular Dynamics Simulation with the Multilevel Summation Method

David J. Hardy[*1]
[*1] University of Illinois Urbana-Champaign

Abstract

Molecular dynamics (MD) simulation has for decades been a valuable computational approach for investigating biomolecules. The emergence of exascale computing provides the opportunity for a commensurate increase in simulated system sizes to enable the study of macromolecular assemblies comprised of billions of atoms. The rate-limiting part of MD is the calculation of the non-bonded Coulomb forces, which must be computed billions of times when simulating microsecond timescales. The most commonly used fast Coulomb solver, the particle-mesh Ewald (PME) method, causes a significant bottleneck to parallel scalability due to calculation of the FFT. This talk presents an alternative approach with better scaling, the multilevel summation method (MSM). MSM imposes a splitting on the $1/r$ interaction kernel, calculating the short-range part exactly and interpolating the remaining long-range parts from a hierarchy of nested grids, where interpolation between grid levels produces an algorithm with operation count that scales linearly in the number of atoms. The methodology is quite flexible, permitting use for simulations that employ non-periodic, semi-periodic, or periodic boundary conditions, and allowing generalization to other interaction kernels, most notably the long-range dispersion forces. The use of B-spline interpolation is shown to improve the accuracy of the approximation by an order of magnitude over previous formulations. The parallelization of MSM in the MD program NAMD makes use of a combined domain decomposition and force decomposition approach to provide scaling to large numbers of processors, together with vector processing employed to greatly enhance the performance of the localized convolution calculations.

24 Feb (Fri), 11:20-11:50

Exascale N-body algorithms for data analysis and simulation

George Biros[*1]
[*1] The University of Texas at Austin

Abstract

N-body algorithms are ubiquitous in science and engineering and form the core methods for many production codes in industrial, academic, and government labs. The find application in both computational physics and machine learning. Tree-based methods typically require irregular memory access patterns that result in reduced off- and on-node performance. Although significant progress has been made in improving off-node performance, on-node performance remains an open problem. This is especially true for production tree-based codes that have multi-stage computations involving data reshuffles and multiple computational kernels. This on-node utilization wall—a chronic problem since the early nineties—not only remains unresolved but has become much more acute with the emergence of deeper memory hierarchies and manycore and heterogeneous architectures. In this talk, I will outline the computational kernels used in N-body methods and I will describe the challenges in scaling them efficiently.

24 Feb (Fri), 11:50-12:10

Is fast multipole method fast? Another multipole method in FFT convolution over uniform mesh

Keigo Nitadori[*1],
[*1] RIKEN Advanced Institute for Computational Science

Abstract

The naming of 'fast' multipole method is presumably derived from its $O(N)$ nature established on the hierarchical octree structure (for the three-dimensional case). Distant interactions of the long-range force are evaluated through larger size cells, and the total computational cost is dominated by a constant number of M2L (multipole-to-local) transformations between the smallest size cells. Hence, it it $O(N)$. In this talk, we discuss on a different approach to efficiently evaluate the all-to-all multipole interactions between the cells. From the octree structure, it retrogresses to mere a uniform mesh structure. To speed up the all-to-all M2L transformations, the convolution theorem with FFT (fast Fourier transform) is applied. As a result, one smallest size cell requires only one M2L translation which can be 189 or 875 in the original octree case. In exchange for the reduces number of M2L translations, the presented approach scales $O(N \log N)$, however, we believe that in a practical molecular dynamics simulation, this approach is faster than the $O(N)$ method.

24 Feb (Fri), 12:10-12:30

Towards Exascale in High-Order Computational Fluid Dynamics

Niclas Jansson[*1]
[*1] KTH Royal Institute of Technology

Abstract

The complex nature of turbulent fluid flows implies that the computational resources needed to accurately model problems of industrial and academic relevance is virtually unbounded. Computational Fluid Dynamics (CFD) is therefore a natural driver for exascale computing and has the potential for substantial societal impact, like reduced energy consumption, alternative sources of energy, improved health care, and improved climate models. Extreme-scale CFD poses several cross disciplinary challenges e.g. algorithmic issues in scalable solver design, handling of extreme sized data with compression and in-situ analysis, resilience and energy awareness in both hardware and algorithm design. The wide range of topics makes exascale CFD relevant to a wider HPC audience, extending outside the traditional fluid dynamics community. The main goal of the EU funded Horizon 2020 project ExaFLOW is to address key algorithmic challenges in CFD in order to facilitate simulations at exascale, e.g. accurate and scalable solvers, data reduction methods and strategies to ensure fault tolerance and resilience. We present our recent progress in the algorithmic development e.g. enhanced accuracy through adaptive error control, improved solver scalability via hybrid continuous and discontinuous methods, optimized preconditioners and efficient low latency communication kernels. Our work is implemented in several open source CFD codes (Nek5000, Nektar++, OpenSBLI), and performance is evaluated for both academic and industrial use cases.

24 Feb (Fri), 14:30-15:10

Large-Scale Deep Learning: Achievements and Challenges

Takuya Akiba[*1]
[*1] Preferred Networks, Inc.

Abstract

Deep learning has broken through previous state-of-the-art machine learning methods and is playing key roles in many fields such as image recognition, speech recognition, and natural language processing. In this talk, we first review the foundation of deep learning and then discuss the challenges and recent advances for larger scale deep learning.

24 Feb (Fri), 15:10-15:30

Parallel Implementation of Shift-invert Rational Krylov method

Yuka Hashimoto[*1], Takashi Nodera
[*1] Keio University

Abstract

Evolution equations have been used to solve problems in various fields of science and technology. For example, the heat equation, the convection diffusion equation, and the wave equation in fluid mechanics and building physics, ordinarily differential equations of matrix form are derived after the spatial discretization. In order to get the solution of the ODE, the computation of matrix function and vector products $\phi_k(\mathrm{A})\mathrm{v}$ needs to be implemented repeatedly, where ϕ_k is the function defined recursively about k, A is a large matrix, and v is a vector. Krylov subspace methods such as Shift-invert Arnoldi method and Rational Krylov method are viable alternatives for computing these large matrix functions. However, these methods need the shifts γ_j for transforming A into $(\gamma_j I - A)^{-1}$, and the choice of appropriate shifts is difficult. We have proposed Shift-invert Rational Krylov method to deal with this problem. Shift-invert Rational Krylov method uses the very simple shifts $\gamma_j = \mathrm{N} - \mathrm{j}$, where $\mathrm{N} \in \mathbb{N}$ and j is the number of iterations. This choice of shifts realizes first convergence. Moreover, the computation is efficient. Shift-invert Rational Krylov algorithm is suitable for parallel implementation theoretically, because we can treat $(\gamma_j I - A)^{-1}$ parallely for j. However, the computation often becomes unstable in practice due to the appearance of ill-conditioned matrices. One of the approach avoiding the unstable computation is restarting the algorithm with some information which have already been obtained. Thus, we propose the restarted version of Shift-invert Rational Krylov method for stable parallel computation. We show that the condition of matrices appear in Shift-invert Rational Krylov method does not get worse with this restated algorithm. In addition, the Restarted Shift-invert Rational Krylov method also has following two advantages. First, we can discard the old information used before restarting, so we save the memory. The information used before restarting is compressed into one vector called restarting vector and we restart the algorithm with this vector. The cost of computing restarting vector is cheap. Second, the shifts are completely determined. In Shift-invert Rational Krylov

method without restarting, N in γ_j is not completely determined automatically. The restarted version gives the way choosing N. As a result, we can realize the efficient parallel computation related to evolution equations with Restarted Shift-invert Rational Krylov method. Numerical experiments which show the effectiveness of our proposed method are also presented.

List of Poster Presentations

P-01	Hiroshi Murakami (Tokyo Metropolitan University): Filter Diagonalization Method for a Real Symmetric-Definite Generalized Eigenproblem by Using a Filter which Is a Polynomial of a Resolvent
P-02	Shunji Kotsuki (AICS): Model Parameter Estimation Using Ensemble Data Assimilation: A Case with the Nonhydrostatic Icosahedral Atmospheric Model NICAM and the Global Satellite Mapping of Precipitation Data
P-03	Yasumitsu Maejima (AICS): Impacts of the surface observations on predicting torrential rainfalls on September 9, 2015 around Tochigi and Ibaraki prefectures
P-04	Akihisa Yamakawa (Tokyo Institute of Technology): A parallel implementation of the Particle-Particle Particle-Tree scheme for simulating planetesimal systems
P-05	Koji Terasaki (AICS): Two-year analysis experiments with NICAM-LETKF
P-06	Atsushi Okazaki (AICS): Data assimilation in paleoclimate
P-07	Keiichi Kondo (AICS): Assimilating satellite radiances without vertical localization using the Local Ensemble Transform Kalman Filter with up to 1280 ensemble members
P-08	Ashok Sharma (MIET University of Jammu India): Efficient storage of Data in cloud
P-09	Shaligram Prajapat (Devi Ahilya University Indore): AVK based symmetric Cryptosystems and Cryptic Mining
P-10	Shigenori Otsuka (AICS): Applying data assimilation to three-dimensional precipitation nowcasting with phased-array weather radar
P-11	William David Fletcher Dawson (AICS): Large Scale Matrix Polynomial Computation for Linear Scaling Quantum Chemistry

P-12	Michio Katouda (AICS): Massively parallel and multi-GPU implementation of RI-MP2 energy calculations
P-13	Hazuki Arakida (AICS): Extending data assimilation with MODIS LAI observations and the dynamic global vegetation model SEIB-DGVM to multiple locations
P-14	Guo-Yuan Lien (AICS): Implicit thinning and localization of dense observation data in the LETKF: A case of phased array weather radar
P-15	Takumi Honda (AICS): Assimilating All-Sky Himawari-8 Satellite Infrared Radiances: A Case of Heavy Rainfalls and Floods
P-16	Takateru Yamagishi (RIST): Accelerating a Non-Hydrostatic Ocean Model with Lagrangian Particle Tracking using a GPU
P-17	Takaharu Mori (RIKEN Theoretical Molecular Science Laboratory): Molecular mechanisms underlying proton transfer through SecDF
P-18	Yusuke Hirota (AICS): Development of Banded Eigenvalue Solvers for Shared Memory Parallel Computers
P-19	Takehiro Yonehara (AICS): Parallel computation in solving Liouville-von Neumann equation intended for a study on electron dynamics including spin-orbit, laser-matter and nonadiabatic couplings
P-20	Akie Mayumi (Japan Atomic Energy Agency): Left-preconditioned communication avoiding CG solver for multiphase CFD code JUPITER
P-21	Masahiro Nakao (AICS): Productivity and Performance of the XcalableACC language on Acceralated Clusters
P-22	Taeka Awazu (AICS): Forecast Verification by Pattern Recognition with Integrated Precipitation Areas
P-23	Hiroya Suno (AICS): Eigenspectrum calculation of the O(a)-improved Wilson-Dirac operator in lattice QCD using the Sakurai-Sugiura method
P-24	Nobuaki Ohno (University of Hyogo): Vectorization of Volume Rendering Methods

P-25	Satoko Yamamoto (Tokyo Institute of Technology): A Formulation of Consistent Particle Hydrodynamics in Strong Form
P-26	Toshiki Teramura (AICS): Non-Gaussianity in the ensemble Kalman filter
P-27	Tsuyoshi Yamaura (AICS): Optimum numerical calculation with mixed precision floating point number for a regional shallow-water model
P-28	Akiyoshi Kuroda (AICS): Improvement of the Usage of Computing Resources and Reduction of the Total Energy for Calculation on the K computer
P-29	Hao Zhang (AICS): Optimizing On-Chip Networks for FPGA-based Neural Network Accelerator
P-30	Masaki Iwasawa (AICS): Framework for Developing Particle Simulators (FDPS)
P-31	Jaewoon Jung (AICS): Multiple program/multiple data molecular dynamics method with multiple time step integrator for large biological systems
P-32	Chigusa Kobayashi (AICS): Molecular dynamics simulations for reaction with large conformational changes in biological system
P-33	Takashi Shimokawabe (Tokyo Institute of Technology): Performance Evaluation of Wind Simulation Based on a GPU-computing Framework to Realize Large-scale Stencil Computations Beyond Device Memory Capacity
P-34	Shinichi Mineo (RIST): Call for proposals of General Trial Use projects for K computer

P-01

Filter Diagonalization Method for a Real Symmetric-Definite Generalized Eigenproblem by Using a Filter which Is a Polynomial of a Resolvent

Hiroshi Murakami[*1]

[*1] Tokyo Metropolitan University

Abstract

For a real symmetric-definite generalized eigenproblem $A\mathbf{v} = \lambda B\mathbf{v}$ $(B > 0)$, we solve eigenpairs whose eigenvalues are in a real interval $[a, b]$ by the filter diagonalization method. In our current study, the filter is the real part of a polynomial of a resolvent: $\mathcal{F} = Re\sum_{k=1}^{n} \gamma_k \mathcal{R}(\rho)^k$. Here $\mathcal{R}(\rho) \equiv (A - \rho B)^{-1}B$ is the resolvent with an imaginary shift ρ, and γ_k are coefficients. In our experiments, the (half) degree n is 15 or 20. By tuning the shift ρ and coefficients γ_k well, the filter passes those eigenvectors well whose eigenvalues are in a neighbor of $[a, b]$, but strongly reduces those ones whose eigenvalues are separated from the interval. We apply the filter to a set of sufficiently many B-orthonormal random vectors $\mathbf{x}^{(\ell)}$ to obtain another set $\mathbf{y}^{(\ell)}$. From both sets of vectors and considering properties of the filter, we construct a basis which approximately spans an invariant subspace whose eigenvalues are in a neighbor of $[a, b]$. An application of the Rayleigh-Ritz procedure to the basis gives approximations of all required eigenpairs. For banded problems, experiments showed this approach worked well.

P-02

Model Parameter Estimation Using Ensemble Data Assimilation: A Case with the Nonhydrostatic Icosahedral Atmospheric Model NICAM and the Global Satellite Mapping of Precipitation Data

Shunji Kotsuki[*1], Koji Terasaki, Hisashi Yashiro, Hirofumi Tomita, Masaki Satoh, Takemasa Miyoshi

[*1] RIKEN Advanced Institute for Computational Science

Abstract

This study aims to improve precipitation forecasts from numerical weather prediction (NWP) models through effective use of satellite-derived precipitation data. Kotsuki et al. (2016, JGR-A) successfully improved the precipitation forecasts by assimilating the Japan Aerospace eXploration Agency (JAXA)'s Global Satellite Mapping of Precipitation (GSMaP) data into the Nonhydrostatic Icosahedral Atmospheric Model (NICAM) at 112-km horizontal resolution. Kotsuki et al. mitigated the non-Gaussianity of the precipitation variables by the Gaussian transform method for observed and forecasted precipitation using the previous 30-day precipitation data. This study extends the previous study by Kotsuki et al. and explores an online estimation of model parameters using ensemble data assimilation. We choose two globally-uniform parameters, one is the cloud-to-rain auto-conversion parameter of the Berry's scheme for large scale condensation and the other is the relative humidity threshold of the Arakawa-Schubert cumulus parameterization scheme. We perform the online-estimation of the two model parameters with an ensemble transform Kalman filter by assimilating the GSMaP precipitation data. The estimated parameters improve the analyzed and forecasted mixing ratio in the lower troposphere. Therefore, the parameter estimation would be a useful technique to improve the NWP models and their forecasts. This presentation will include the most recent progress up to the time of the symposium.

P-03

Impacts of the surface observations on predicting torrential rainfalls on September 9, 2015 around Tochigi and Ibaraki prefectures

Yasumitsu Maejima[*1], Takemasa Miyoshi
[*1] RIKEN Advanced Institute for Computational Science

Abstract

To investigate the impact of the surface observations on a severe rainfall event occurred on September 9, 2015 around Tochigi and Ibaraki prefectures, we perform a series of data assimilation (DA) experiments using the Local Ensemble Transform Kalman Filter (LETKF) with the SCALE regional NWP model (Nishizawa et al. 2015). In this event, an active rainband was maintained for an extended period and caused torrential rainfalls over 500 mm/day with catastrophic flooding. Two DA experiments were performed: the control experiment (CTRL) at 4-km resolution with only conventional observations (NCEP PREPBUFR), and the other with additional surface observation data with every minute (TEST). CTRL showed generally similar rainfall patterns, although the intensity was smaller, and rainfall area was shifted westward compared to the JMA analyzed precipitation based on the radar and gauge networks. By contrast, TEST showed stronger rainfall intensity, better matching with the JMA analyzed precipitation. Surface DA contributed to improve the moisture field in the lower layer, leading to intensified rainfall amount. The results suggest that the surface DA have a potential to improve the forecast accuracy for severe rainfall events.

P-04

A parallel implementation of the Particle-Particle Particle-Tree scheme for simulating planetesimal systems

Akihisa Yamakawa[*1], Junichiro Makino, Takayuki Saitoh, Junko D. Kominami
[*1] Tokyo Institute of Technology

Abstract

N-body simulation is widely used to investigate formation processes of terrestrial planets through the planetesimal accretion. Since it requires long time integration with high accuracy and large computational cost [$O(N^2)$], the number of particles used in N-body simulations of planet formation is limited. In order to achieve high-resolution simulations of planet formation, we developed a parallel code adopting the Particle-Particle Particle-Tree(P^3T) scheme (Oshino et al. 2011). In P^3T, the mutual gravitational force is split into two parts; short-range and long-range interactions. The short-range interaction is evaluated by the direct summation [the computational cost is $O(N^2)$] and integrated with the fourth order Hermite scheme. The long-range interaction is computed by the Barnes-Hut tree algorithm [the computational cost is $O(N\log N)$] and integrated with the leapfrog integrator. For short-range part with parallel computation, we have developed a new algorithm which is based on the connected graph of neighbor particles and keeps a good parallel performance to a few hundred CPUs. To accelerate the calculation of the tree part with parallel computation, we adopted Framework for Developing Particle Simulator (Iwasawa et al. 2016). In this presentation, we explain details of our implementation. Then, we show that our code has a good performance and accuracy and is feasible for large scale N-body simulations.

P-05

Two-year analysis experiments with NICAM-LETKF

Koji Terasaki[*1], Shunji Kotsuki, Takemasa Miyoshi
[*1] RIKEN Advanced Institute for Computational Science

Abstract

We applied the Local Ensemble Transform Kalman Filter (LETKF) to the Non-hydrostatic ICosahedral Atmospheric Model (NICAM). Observation operators to assimilate the conventional observations, satellite-borne Atmospheric Microwave Sounder Unit-A (AMSU-A), and the Global Satellite Mapping of Precipitation (GSMaP) data were developed. The purpose of this study is to verify the long-term stability of the NICAM-LETKF system. We performed experiments to assimilate all observations for two years and two months from June 2014 to July 2016. The first experiment was not successful. We found that the NICAM-LETKF system became unstable due to an extreme outlier of the 100-member ensemble. Therefore, we applied the relaxation to prior spread (RTPS) instead of the default setting of an adaptive multiplicative inflation method, and found that the NICAM-LETKF system was stable for more than two years. The analyzed atmospheric fields were largely improved by assimilating the AMSU-A radiances. The humidity bias is also improved by assimilating the GSMaP data while the NICAM is known to have a dry bias, especially over land.

P-06

Data assimilation in paleoclimate

Atsushi Okazaki[*1], Kei Yoshimura, Takemasa Miyoshi
[*1] RIKEN Advanced Institute for Computational Science

Abstract

Knowledge of past climate conditions is crucial to understand the climate system. Recently, data assimilation (DA) has been applied to reconstruct paleoclimate, in which proxy data serves as observations. Here, proxies are not the direct record of climate variables such as temperature, winds, and pressure, but natural records representing climate such as tree-ring width and isotopic composition in ice sheets. DA has long been used for forecasting the weather and is a well-established method. However, the DA algorithms used for weather forecasts cannot be directly applied to paleoclimate due to the different temporal resolution, spatial extent, and type of information contained within the observation data. The temporal resolution and spatial distribution of proxy data are significantly lower (seasonal at best) and sparser than the present-day observations used for weather forecasts. Therefore, DA applied to paleoclimate is only loosely linked to the methods used in the more mature field of weather forecasting. Several DA methods have been proposed for paleoclimate reconstruction, and paleoclimate studies using DA have successfully determined the mechanisms behind the past climate changes. In this presentation, we review the previously proposed methods and develop a new method in which a proxy is directly assimilated for the first time.

P-07

Assimilating satellite radiances without vertical localization using the Local Ensemble Transform Kalman Filter with up to 1280 ensemble members

Keiichi Kondo[*1], Koji Terasaki, Takemasa Miyoshi
[*1] RIKEN Advanced Institute for Computational Science

Abstract

Covariance localization plays an essential role in the ensemble Kalman filter (EnKF) to avoid bad influence of spurious covariance from sampling errors when the ensemble size is limited. In our previous study, we performed 10240-member ensemble data assimilation experiments with the global atmospheric model NICAM (Nonhydrostatic Icosahedral Atmospheric Model) to investigate horizontal and vertical error correlations. As a result, we found that roughly 1000 ensemble members would be large enough to avoid vertical covariance localization for satellite radiance data. In this study, we perform the Local Ensemble Transform Kalman Filter (LETKF) experiments with NICAM using the ensemble sizes from 20 to 1280. We compare the results with and without vertical localization for satellite radiance data.

P-08

Efficient storage of Data in cloud

Ashok Sharma[*1], RamJeevan Singh Thakur
[*1] MIET University of Jammu India

Abstract

In order to analyse the performance of symmetric key based cryptic Algorithms on different parameters of data files in cloud Figure 3.1 presents a framework. The tool developed on this framework would be termed as Cloud Crypter tool provides actual statistics generated during encryption or decryption in four different categories and provides results in the Graphical and Tabular form for better analysis. Following Four categories of analysis with graphical form are supported by Tool. Category I: Analysis of Encryption of Data files of variable sizes with fixed key. Category II: Analysis of Decryption of Data files of variable sizes with fixed key. Category III: Analysis of Encryption of Data files of fixed size with Variable key sizes. Category IV: Analysis of Decryption of Data files of fixed size with Variable key sizes. These four categories will be used to analyze the behaviour of constant files (Text files and Image file) i.e. having fixed content with respect to time and for dynamic contents (Audio and Video files). The behaviour and outcome of the analysis result will be able to draw useful inference and optimize the cryptic algorithm on the basis of need and constraint of the client.

P-09

AVK based symmetric Cryptosystems and Cryptic Mining

Shaligram Prajapat[*1], Ashok Sharma
[*1] Devi Ahilya University Indore

Abstract

Automatic Variable Key based enciphering schemes are claimed to be energy efficient for communication among IOT based devices and hand held devices for information exchange. This chapter presents parametric versions of symmetric key based cryptic algorithms. In parametric versions emphasis is given on generation and usage of key based on parameter only. The common method of key construction was based on numeric keys. In this chapter the key construction process has been extended for generation of alphanumeric keys and domain of parameter selection is chosen based on personnel information. The reason for this is the success of cryptic model depends on freedom given to user for the parameter selection and variation in the parameters according to comfort instead of using series, recurrence relation or location information. Subsequent section investigates parametric model in the light of Association Rule Mining for Automatic variable key based symmetric cryptosystem. Useful inferences and results from testing of Cryptic association rule mining supports the view of auditing of cryptic algorithm and identifies power of using a large number of parameters for secure information exchange.

P-10

Applying data assimilation to three-dimensional precipitation nowcasting with phased-array weather radar

Shigenori Otsuka[*1], Takemasa Miyoshi
[*1] RIKEN Advanced Institute for Computational Science

Abstract

The phased-array weather radar (PAWR) is a new generation rapid-scanning weather radar. The PAWR at Osaka University can perform three-dimensional volume scans at about 100 elevation angles within 60 km every 30 seconds. To take advantage of the dense and frequent observations, Otsuka et al. (2016b) performed three-dimensional precipitation nowcasting experiments with space-time extrapolation of PAWR data, and demonstrated that the three-dimensional space-time extrapolation outperformed the conventional two-dimensional space-time extrapolation. However, erroneous motion vectors sometimes degrade the forecast performance. Here, data assimilation is employed to improve the motion vector field. Otsuka et al. (2016a) implemented the Local Ensemble Transform Kalman Filter (LETKF) with the two-dimensional space-time extrapolation system, and applied it to the Global Satellite Mapping of Precipitation (GSMaP) data. In this study, we extend the nowcasting system with the LETKF to the three-dimensional motion vector field, and applied it to the PAWR precipitation nowcasting. A case study on an isolated convective system showed that the three-dimensional space-time extrapolation with the LETKF outperformed that without data assimilation in terms of precipitation threat scores.

References:

Otsuka, S., S. Kotsuki, and T. Miyoshi, 2016a: Wea. Forecasting, 31, 1409-1416.
Otsuka, S., G. Tuerhong, R. Kikuchi, Y. Kitano, Y. Taniguchi, J. Ruiz, S. Satoh, T. Ushio, and T. Miyoshi, 2016b: Wea. Forecasting, 31, 329-340.

P-11

Large Scale Matrix Polynomial Computation for Linear Scaling Quantum Chemistry

William David Fletcher Dawson[*1], Takahito Nakjima
[*1] RIKEN Advanced Institute for Computational Science

Abstract

Today, large scale calculations based on ab initio quantum chemistry methods are routinely performed to study systems such as photovoltaic cells and novel catalysts. These calculations, an alternative to experimental techniques, can provide atom level insight into the origin of material properties. Quantum chemistry calculations involve first constructing a discrete representation of the linear operator describing the relevant physics, and then computing some matrix function of that operator. Traditionally, this is done by computing the eigendecomposition of that matrix, an approach that has been aided by the development of large scale eigenvalue libraries. However, the cubic scaling cost associated with computing the eigendecomposition has become a bottleneck as researchers try and investigate larger and larger systems. In the case where the operator can be represented as a sparse matrix, however, there exist methods for computing matrix functions in linear time. In particular, methods based on matrix polynomials have been shown to efficiently approximation the matrix functions used in quantum chemistry. Recently, our group has begun development of a highly parallel library for computing polynomials of sparse matrices on the K computer. This library will be able to accelerate quantum chemistry calculations across a number of different codes. In this talk, I will present the parallel algorithms at the basis of this library, and our current progress and insights.

P-12

Massively parallel and multi-GPU implementation of RI-MP2 energy calculations

Michio Katouda[*1], Akira Naruse, Yukihiko Hirano, Takahito Nakajima
[*1] RIKEN Advanced Institute for Computational Science

Abstract

We performed the multi-GPU massively parallel implementation of resolution-of-identity second order Møller-Plesset perturbation (RI-MP2) energy calculation suitable for calculations of large molecules on CPU/GPU hybrid supercomputers. In this presentation, we report the overview of implementation and the results of performance evaluation of the implementation using up to 1,349 nodes and 4,047 GPUs of TSUBAME2.5 supercomputer. The GPU computation speeds up considerably (4.1-6.6 times) the RI-MP2 calculations. Parallel scalability of present GPU implementation is good with the number of nodes. 514.7 TFLOPs of the measured peak performance is attained for the GPU job of RI-MP2/cc-pVTZ calculation of nanographene dimer $(C_{96}H_{24})_2$ using 1,349 nodes and 4,047 GPUs of TSUBAME 2.5, which is much higher than that of CPU jobs (87.5 TFLOPs). We also present application of the inter-molecular interaction analysis of nano carbon molecular assemblies such as nanographenes.

P-13

Extending data assimilation with MODIS LAI observations and the dynamic global vegetation model SEIB-DGVM to multiple locations

Hazuki Arakida[*1], Shunji Kotsuki, Shigenori Otsuka, Yohei Sawada, Takemasa Miyoshi
[*1] RIKEN Advanced Institute for Computational Science

Abstract

In the previous study, Arakida et al. developed a data assimilation system based on a particle filter approach with a dynamical global vegetation model known as the SEIB-DGVM (Spatially Explicit Individual-Based Dynamic Global Vegetation Model), and assimilated the MODIS LAI (Leaf Area Index) observations successfully. In this study, we extend the previous study to different locations and estimate the state variables including carbon flux, water flux, heat flux, vegetation structure, and parameters related to the phenology of the deciduous needle leaved tree and grass. The results showed that the DA system performed well at multiple locations.

P-14

Implicit thinning and localization of dense observation data in the LETKF: A case of phased array weather radar

Guo-Yuan Lien[*1], Takemasa Miyoshi
[*1] RIKEN Advanced Institute for Computational Science

Abstract

Observation data from advanced remote-sensing platforms are getting bigger and bigger. Past studies have shown that, to effectively assimilate dense observations, a proper thinning or superobing method to reduce the data density is usually necessary. In general, these techniques have been employed to deal with various factors such as observation error correlations, representativeness errors, and computational costs. However, they also unavoidably decrease the resolution of data, which is contradictory to the pursuit of high-resolution observing systems and numerical models. We point out that, when using an ensemble data assimilation method, another important, but likely neglected reason to thin the data is to stay in the range that all observations can be effectively assimilated by the limited ensemble size. This issue has been usually addressed by covariance localization methods, but probably not in an optimal way. Recently, the local ensemble transform Kalman filter (LETKF) systems at European Centre for Medium-Range Weather Forecasts (ECMWF) and Deutscher Wetterdienst (DWD) have adopted an "implicit localization" method that significantly reduces the assimilated observation numbers while preserving high-resolution information, by selecting N nearest neighbors of observations from the analyzed grid point. We demonstrate the usefulness of this method on the assimilation of very dense phased array weather radar data, and explain it as an ideal combination of thinning and localization.

P-15

Assimilating All-Sky Himawari-8 Satellite Infrared Radiances: A Case of Heavy Rainfalls and Floods

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

[*1] RIKEN Advanced Institute for Computational Science

Abstract

To predict heavy rainfalls and floods, it is important to get better initial conditions with accurate moisture transport via data assimilation. To do so, infrared (IR) radiance observations by geostationary satellites can give useful information in a wide area because some IR bands are sensitive to moisture. In particular, the new Japanese geostationary satellite "Himawari-8" can provide high-spatiotemporal resolution observations with many bands. The present study aims to assimilate all-sky IR radiance observations by Himawari-8 and investigate its impact on the analyses and forecasts of a heavy rainfall event in Japan. The results show that northward moisture transport over the ocean south of Japan is enhanced due to Himawari-8 data. The improved analyses give much better precipitation forecasts compared to the control experiment without Himawari-8 IR observations. The improved precipitation forecasts are essential for more accurate river model forecasts.

P-16

Accelerating a Non-Hydrostatic Ocean Model with Lagrangian Particle Tracking using a GPU

Takateru Yamagishi[*1], Yoshimasa Matsumura

[*1] Research Organization for Information Science and Technology

Abstract

Lagrangian modeling of particles in the ocean is an effective tool for understanding and reproducing several significant multiphase flows. This method can track several tracers, such as nutrients or isotopes, and helps study ocean physics and dynamics. Matsumura and Ohshima (2015) formulated an explicit expression of the particles by the application of Lagrangian particle tracking to a non-hydrostatic ocean model on a CPU and performed detail studies on frazil ice formation. The demand for conducting studies on numerous and various types of particles has been growing, and GPUs are expected to meet this demand because of their massive cores and their architecture, which is highly specialized for massive parallel computing. This study aims to accelerate the non-hydrostatic ocean model with particle tracking using NVIDIA GPU and PGI CUDA Fortran and demonstrates some experimental cases and techniques that would be applicable to other studies in numerical ocean modeling on GPUs. In this implementation to a GPU, we have revised the algorithm for particle tracking and optimized the kernels for ocean dynamics calculation. The sorting of all particles at every time step was introduced to coalesce the access to the GPU global memory, and a texture cache was assigned to the ocean current velocity array to accelerate particle tracking. Thread-level parallelism was exploited with the help of additional calculations on adjacent grids, and registers were effectively used for the ocean dynamics kernels. When comparing its execution on a Fujitsu SPARC64 IXfx to that on the NVIDIA K20C, the GPU-implemented model was three times faster. This model successfully reproduced the nonlinear distribution of the particles in the ocean.

P-17

Molecular mechanisms underlying proton transfer through SecDF

Takaharu Mori[*1], Tomoya Tsukazaki, Yuji Sugita
[*1] RIKEN Theoretical Molecular Science Laboratory

Abstract

About 30% of proteins are secreted across membranes or integrated into membranes. The Sec translocon plays a central role in these protein translocations. It consists of protein complexes: SecA ATPase, protein channel SecYEG, and membrane chaperone SecDF. Recently, it has been revealed that SecDF undergoes large conformational change by the proton-motive force to enhance protein export. However, the relationship between conformational change and proton transport has not been understood well. To elucidate their molecular mechanisms, we carried out molecular dynamics simulations of SecDF. We found that dynamics of the conserved residues in the transmembrane region is important for water channel formation. We discuss detailed mechanisms for the function of SecDF.

P-18

Development of Banded Eigenvalue Solvers for Shared Memory Parallel Computers

Yusuke Hirota[*1], Toshiyuki Imamura
[*1] RIKEN Advanced Institute for Computational Science

Abstract

The problem to find the pair $(\lambda, \mathbf{x})$ to satisfy the equation $A\mathbf{x} = \lambda B\mathbf{x}$, where A is a real symmetric banded matrix and B is a real symmetric positive definite banded matrix, is called banded generalized eigenvalue problem (banded GEP). If B is an identity matrix, the problem is called banded standard eigenvalue problem (banded SEP). The solvers for banded SEP/GEP (banded solvers) are important for many computational science applications, since the banded solvers are used as subsolvers of dense SEP/GEP solvers (dense solvers) which are important building blocks of the applications. Especially on highly parallel shared memory computers, the non-tridiagonal banded (e.g. pentadiagonal) SEP/GEP solvers are useful to reduce the total execution time of the dense solvers. However, the conventional banded solvers are not very efficient on the highly parallel shared memory computers, such as on Intel Xeon Phi Knights Corner. We have developed solvers based on the divide-and-conquer algorithm for banded SEP/GEP for the computational environment. In the poster, we will present our recent results on the theoretical analysis to the divide-and-conquer algorithm, and techniques for the acceleration of the implementation of the divide-and-conquer algorithm.

P-19

Parallel computation in solving Liouville-von Neumann equation intended for a study on electron dynamics including spinorbit, laser-matter and nonadiabatic couplings

Takehiro Yonehara[*1], Takahito Nakajima
[*1] RIKEN Advanced Institute for Computational Science

Abstract

Excited electrons in molecular aggregation systems play important roles in many types of solar energy conversions from radiation field to chemical and material functionality. For studying this in a microscopic level with molecular resolution, efficient theoretical computation schemes for describing electron dynamics in molecular systems are highly demanded. The combination of the well-known Liouville-von Neumann equation and molecular orbital theories provides a natural description of the dynamics of electronic systems involved with spin-orbit, laser-matter and nonadiabatic electron-nuclear couplings associated with molecular motions in realistic situations. In this poster presentation, we show that the MPI and openMP parallel computations have possibilities to provide a high efficiency in a computation of this types of dynamics.

P-20

Left-preconditioned communication avoiding CG solver for multiphase CFD code JUPITER

Akie Mayumi[*1], Yasuhiro Idomura, Susumu Yamada, Takuya Ina, Toshiyuki Imamura
[*1] Japan Atomic Energy Agency

Abstract

The left-preconditioned communication avoiding conjugate gradient (LP-CA-CG) method is applied to the pressure Poisson equation in the multiphase CFD code JUPITER. The arithmetic intensity of the LP-CA-CG method is analyzed, and is dramatically improved by loop splitting for inner product operations and for three term recurrence operations. Two LP-CA-CG solvers with block Jacobi preconditioning and with underlap preconditioning, in which point Jacobi preconditioning approximation is applied to boundary regions, are developed. The former is developed based on a hybrid CA approach, in which CA is applied only to global collective communications for inner product operations. The latter is a full CA approach, in which CA is applied also to local point-to-point communications in sparse matrix-vector (SpMV) operations and preconditioning. CA-SpMV requires additional computation for overlapping regions. CA-preconditioning is enabled by underlap preconditioning, which approximates preconditioning for overlapping regions by point Jacobi preconditioning. It is shown that on the K computer, the former hybrid approach is faster, because the performance of local point-to-point communications scales well, and the convergence property becomes worse with underlap preconditioning. The LP-CA-CG solver shows good strong scaling up to 30,000 nodes, where the LP-CA-CG solver becomes 2x faster than the original CG solver by reducing the cost of global collective communications by 69 percent.

P-21

Productivity and Performance of the XcalableACC language on Acceralated Clusters

Masahiro Nakao[*1], Hitoshi Murai, Hidetoshi Iwashita, Akihiro Tabuchi, Taisuke Boku, Mitsuhisa Sato

[*1] RIKEN Advanced Institute for Computational Science

Abstract

We introduce XcalableACC (XACC) language, which is a hybrid model of the XcalableMP (XMP) Partitioned Global Address Space (PGAS) language and OpenACC. XACC defines directives that enable programmers to mix XMP and OpenACC directives in order to develop applications that can use acceralated clusters with ease. In this poster, we evaluate the productivity and performance of XACC through implementations of several benchmarks, for example Lattice QCD. The results show that thank to the productivity improvements, the source lines of codes in XACC requires much less than those in MPI+CUDA or MPI+OpenACC, which are commonly used as a typical programming model. Moreover, performance results of XACC are almost the same as those of MPI+CUDA or MPI+OpenACC.

P-22

Forecast Verification by Pattern Recognition with Integrated Precipitation Areas

Taeka Awazu[*1], Shigenori Otsuka, Takemasa Miyoshi
[*1] RIKEN Advanced Institute for Computational Science

Abstract

The human eye recognizes many patches of rainfall areas as a single precipitation system. However, precipitation forecast verification methods usually recognize individual patches. To mimic the human recognition, it is desirable to combine the individual patches into a single precipitation system. Hence, we propose a verification method using the integrated rainfall areas. In addition, the proposed method evaluates the location error and shape of the rainfall areas. Grid-based verification methods such as the Threat Score (TS) and Root Mean Squared Error (RMSE) are commonly used but have difficulties to evaluate the location error and the shape. Thus, the proposed method evaluates the integrated rainfall areas using the shape features, precipitation rate and the distance between the areas of the observation and the forecast. This study used the forecast data of space-time extrapolation in Global Satellite Mapping of Precipitation (Otsuka et al. 2015). The evaluation of the forecast data was compared between the proposed method and the traditional methods. The results showed that with longer lead time the TS and RMSE change only slightly, while the proposed score gets worse linearly with time. This is more consistent with human recognition.

P-23

Eigenspectrum calculation of the O(a)-improved Wilson-Dirac operator in lattice QCD using the Sakurai-Sugiura method

Hiroya Suno[*1], Y. Nakamura, K.-I. Ishikawa, Y. Kuramashi, Y. Futamura, A. Imakura, T. Sakurai

[*1] RIKEN Advanced Institute for Computational Science

Abstract

We have developed a computer code to find eigenvalues and eigenvectors of non-Hermitian sparse matrices arising in lattice quantum chromodynamics (lattice QCD). The Sakurai-Sugiura (SS) method is employed here, which is based on a contour integral, allowing us to obtain desired eigenvalues located inside a given contour of the complex plane. We apply the method here to calculating several low-lying eigenvalues of the non-Hermitian O(a)-improved Wilson-Dirac operator D. Evaluation of the low-lying eigenvalues is crucial since they determine the sign of its determinant detD, important quantity in lattice QCD. We are particularly interested in such cases as finding the lowest eigenvalues to be equal or close to zero in the complex plane. Our implementation is tested for the Wilson-Dirac operator in free case, for which the eigenvalues are analytically known. We also carry out several numerical experiments using different sets of gauge field configurations obtained in quenched approximation as well as in full QCD simulation almost at the physical point. Various lattice sizes are considered from $8^3 \times 16$ to 96^4, amounting to the matrix order from 98,304 to 1,019,215,872.

P-24

Vectorization of Volume Rendering Methods

Nobuaki Ohno[*1]
[*1] University of Hyogo

Abstract

Current computer simulations on high performance computers produce very large scale data. They are too large to analyze on PCs. Thus, in-situ visualization, which simulation and visualization are carried out simultaneously on the same computer, is drawing attention. For in-situ visualization on vector computer such as the Earth Simulator, vectorization of visualization methods is indispensable. In this poster, vectorization of volume rendering methods and its performance evaluations are presented. The particle-based volume rendering and ray casting methods have been vectorized and its performances have been examines on the Earth Simulator.

P-25

A Formulation of Consistent Particle Hydrodynamics in Strong Form

Satoko Yamamoto[*1], Junichiro Makino
[*1] Tokyo Institute of Technology

Abstract

In fluid dynamical simulations in astrophysics, large deformations are common and surface tracking is sometimes necessary. Smoothed Particle Hydrodynamics (SPH) method has been used in many of such simulations. Recently, however, it has been shown that SPH cannot handle contact discontinuities or free surfaces accurately. There are several reasons for this problem. The first one is that SPH requires that the density is continuous and differentiable. The second one is that SPH does not have the consistency, and thus the accuracy is zeroth order in space. In addition, we cannot express accurate boundary conditions with SPH. In this presentation, we propose a novel, high-order scheme for particle-based hydrodynamics of compressible fluid. Our method is based on kernel-weighted high-order fitting polynomial for intensive variables. With this approach, we can construct a scheme which solves all of the three problems described above. We have applied our method to many test problems and obtained excellent result. Our method is not conservative, since particles do not have mass or energy, but only their densities. However, because of the Lagrangian nature of our scheme, the violation of the conservation laws turned out to be small. We name this method Consistent Particle Hydrodynamics in Strong Form (CPHSF).

P-26

Non-Gaussianity in the ensemble Kalman filter

Toshiki Teramura[*1], Takemasa Miyoshi
[*1] RIKEN Advanced Institute for Computational Science

Abstract

Non-Gaussianity is an important problem in the ensemble Kalman filter (EnKF) since it leads to suboptimal filtering. We evaluate how the non-Gaussian distribution emerges due to the non-linear effect of the equation of motion using the Lorenz three-variable model. Non-Gaussianity is evaluated by a bias of EnKF relative to a merging particle filter (MPF) using a large particle size (K=1000). In this low dimensional model, MPF with 1000 particles is an accurate non-Gaussian filter. Since the bias disappears when the prior distribution is completely Gaussian, this bias represents non-Gaussianity of the distribution. This bias is explained by the square root of the trace of the prior covariance, which we call simply the standard deviation. If the system is linear, there is no correlation between the bias and the standard deviation except for the sampling noise. The experimental results reveal a power low between the bias and standard deviation, and its exponent is between 1.2 and 1.7. Hence, the non-Gaussian bias can be estimated from the prior covariance, which is also able to be estimated by the linear analysis. The state dependence of non-Gaussianity and extension to more realistic problems are also discussed.

P-27

Optimum numerical calculation with mixed precision floating point number for a regional shallow-water model

Tsuyoshi Yamaura[*1], Hirofumi Tomita
[*1] RIKEN Advanced Institute for Computational Science

Abstract

We investigate the impact of numerical error of floating point numbers (FPNs) to the equilibrium and instability condition experiments, using a regional shallow-water model with mixed precision FPN by a software emulator. To express the numerical errors due to mixed precision FPN quantitatively, we define the reproducibility index that is the mean ratio of root-mean-square-error to standard deviation of prognostic variables in the shallow-water model. The results of experiments correspond to the theory of the reproducibility index, suggesting that higher spatial resolution requires larger size of significand bit width. In the case of the shallow-water model, preparing a reference value, which is made from horizontal mean of a variable before time integration, is effective in reducing loss of significand digits. Reducing accuracy of FPN before making the reference value of geopotential may induce large loss of significand digits, while that of velocity is relatively small contribution to the loss. A careful summation algorithm for a large amount of grids can avoid loss of trailing digits that induces low accuracy of the reference value. Following the above results, we construct an optimum shallow-water model that uses single precision FPN to dynamics module. The optimum model can obtain the results with slight numerical errors, compared with the shallow-water model fully using double precision FPN. In contrast, execution time of the optimum model is comparable to that of the shallow-water model fully using single precision FPN.

P-28

Improvement of the Usage of Computing Resources and Reduction of the Total Energy for Calculation on the K computer

Akiyoshi Kuroda[*1], Shunsuke Inoue, Kentaro Koyama, Ryuichi Sekizawa, Kazuo Minami

[*1] RIKEN Advanced Institute for Computational Science

Abstract

The effective use of computing resources is important for the operation of the K computer. Improving the job performance is necessary to improve the usage of computing resources and reduce the total energy for calculation. We constructed the system to collect information concerning jobs for reducing the total energy for calculation and usage of computing resources. This system transparently collects the information and does not require additional tasks from the users. It can automatically collect the FLOPS, the memory throughput, the time for internode imbalances, and the amount of internode communications for all the jobs without using a profiler. A job that requires large computational resources occupies a greater proportion of the total system resources. We analyzed collected information using a weighted average of the compute node time product. Then we found that the huge sleep time occupied the many computational time due to load imbalances including communication in these jobs. The large sleep time is the cause of the waste the usage of computing resources. Job performance tuning is important to reduce the sleep time, and the total energy for calculation can also be reduced as well as the usage of computing resources. We can illustrate the relation between the job performance and power consumption. The power consumption is a two-storied structure: basic part (standby power) and increments part (active power). Even though the power consumption is increased by job tuning, the total energy for calculation will be reduced by decreasing the dominant standby power. To improve the usage of computing resources and reduce the total energy for calculation by tuning the large scale jobs running on the K computer, we continue to analyze the information collected by this system.

P-29

Optimizing On-Chip Networks for FPGA-based Neural Network Accelerator

Hao Zhang[*1], Makoto Taiji
[*1] RIKEN Advanced Institute for Computational Science

Abstract

Deep Neural Network (DNN), a human brain inspired machine learning algorithm, has recently been shown to provide much better performance than other machine learning algorithm. It has been applied to fields like computer vision, speech recognition, nature language process and bioinformatics where it has been shown to produce state-of-art results on various tasks. However, a conventional general CPU or GPU cannot meet the performance and power requirement when the data is huge and DNN scale is big such as drug discovery. The target of this research is to improve the performance and energy efficiency for the deep neural network accelerator. I would like to propose a mapping algorithm to map the neurons to on-chip networks by using computing the minimum hops. This mapping algorithm can be applied to different cases. We also employed bufferless Network-on-Chip(NoC) architecture to make the chip area and energy more efficient.

P-30

Framework for Developing Particle Simulators (FDPS)

Masaki Iwasawa[*1], Daisuke Namekata, Ataru Tanikawa, Natsuki Hosono, Keigo Nitadori, Takayuki Muranushi, Jun Makino
[*1] RIKEN Advanced Institute for Computational Science

Abstract

Particle-based simulations are widely used in the fields of science and engineering. Examples include gravitational N-body simulations, Molecular dynamics (MD) simulations, Smoothed Particle Hydrodynamics (SPH) simulations, Discrete Element Method (DEM) simulations, etc. To develop an efficient program for particle-based simulation for large-scale parallel computers is not easy, and to some extent the efforts of many researchers have been spent on the programming and tuning. However, the algorithms of particle-based simulations are largely similar. For example, some tree structure is used for interaction calculation and the neighbor search, and domain decomposition is used for parallelization on distributed-memory parallel computers. Thus we have developed a framework which helps the researchers to develop efficient programs for particle-based simulation on large parallel machines, which we call Framework for Developing of Particle Simulators, or FDPS. FDPS is a C++ template library. A user of FDPS need to provide the particle class and the function to evaluate particle-particle interactions. FDPS automatically decomposes domain, scatters particles and calculate interactions. The user program then integrates the orbits (and other physical quantities if necessary) of particles using the calculated interaction. Since the interaction calculation and relocation of particles are done in the FDPS part, there would be no need for the user to implement these complicated procedures, which we believe greatly improves the productivity of the researchers. The current version of FDPS (Ver.3) has Fortran interface. Thus a user also can develop simulations codes in Fortran. In this presentation, we introduce concept and implemantaion of FDPS. We also discuss some application programs using FDPS.

P-31

Multiple program/multiple data molecular dynamics method with multiple time step integrator for large biological systems

Jaewoon Jung[*1], Yuji Sugita
[*1] RIKEN Advanced Institute for Computational Science

Abstract

Parallelization of molecular dynamics (MD) simulation is essential for investigating conformational dynamics of large biological systems, such as ribosomes, viruses, and multiple proteins in cellular environments. To improve efficiency in the parallel computation, we have to reduce the amount of data transfer between processors by introducing domain decomposition schemes. Also, it is important to optimize the computational balance between real-space non-bonded interactions and reciprocal-space interactions for long-range electrostatic interactions. Here, we introduce a novel parallelization scheme for large-scale MD simulations on massively parallel supercomputers consisting of only CPUs. We make use of a multiple program/multiple data approach for separating the real-space and reciprocal-space computations on different processors. We also utilize the r-RESPA multiple time step integrator on the framework of the multiple program/multiple data approach in an efficient way: when the reciprocal-space computations are skipped in r-RESPA, processors assigned for them are utilized for half of the real-space computations. The new scheme allows us to use twice as many as processors that are available in the conventional single program approach. The best performances of all-atom MD simulations for 1 million (STMV), 8.5 million (8_STMV), and 28.8 million (27_STMV) atom systems on K computer are 65, 36, and 24 ns/day, respectively. The multiple program/multiple data scheme can accelerate 23.4, 10.2, and 9.2 ns/day from the maximum performance of single-program approach for STMV, 8_STMV, and 27_STMV systems, respectively, which correspond to 57%, 39%, and 60% speed up. This suggests significant speedups by increasing the number of processors without losing parallel computational efficiency.

P-32

Molecular dynamics simulations for reaction with large conformational changes in biological system

Chigusa Kobayashi[*1], Yasuhiro Matsunaga, Jaewoon Jung, Yuji Sugita
[*1] RIKEN Advanced Institute for Computational Science

Abstract

Molecular dynamics (MD) simulation of biomolecules has been widely used to elucidate relation between its functions and conformational dynamics. It is, however, difficult to simulate reaction process coupled with large conformational changes of biomolecules because of the slow time-scale. To overcome the difficulty, we have developed an MD program, GENESIS to perform the simulation efficiently on K computer and PC clusters. In particular, we have developed a structure-based coarse-grained (CG) model describing large conformational changes and introduced enhanced sampling methods using the CG model. We will discuss the schemes and the applications to reactions in a representative ion pump.

P-33

Performance Evaluation of Wind Simulation Based on a GPU-computing Framework to Realize Large-scale Stencil Computations Beyond Device Memory Capacity

Takashi Shimokawabe[*1], Takashi Shimokawabe, Toshio Endo, Naoyuki Onodera, Takayuki Aoki

[*1] Tokyo Institute of Technology

Abstract

Grid-based physical simulations are important applications in the field of high-performance computing. We are developing a GPU-computing framework that makes it easy to describe stencil calculations using multiple GPUs and develop large-scale stencil applications with high productivity. This framework is implemented in the C ++ language and CUDA with its template techniques. In the case of the stencil calculations on a GPU, the size of the computational domain is limited in the device memory capacity of the GPU. We are currently developing a mechanism to realize the computation that exceeds the device memory capacity of the GPU and introducing it into this framework. To realize this, we exploit the HHRT library, which provides a mechanism of automatic swap -out/in the data between the host memory and device memory, with a temporal blocking technique in this framework. The temporal blocking technique is capable of improving locality of memory access, which can suppress performance degradation due to frequent memory swapping between the host memory and the device memory. In this poster, a lattice Boltzmann method for wind simulations is developed by using this framework. The performance of these simulations running on multiple GPUs of the TSUBAME 2.5 supercomputer at Tokyo Institute of Technology is evaluated.

P-34

Call for proposals of General Trial Use projects for K computer

Shinichi Mineo[*1],
[*1] Research Organization for Information Science and Technology

Abstract

The High Performance Computing Infrastructure (HPCI) is a platform realized by connecting the K computer and other major Japanese supercomputers via high speed networks. Research Organization for Information Science and Technology (RIST) calls for proposals of the projects using the K computer and the other HPCI computers. After screening processes, the computational resources will be awarded for researchers in academia and also in industry. With regard to the use of the K computer, anybody from around the world is eligible to apply for by following the instructions found on the HPCI Portal Site: https://www.hpci-office.jp/folders/english.We will introduce permanent call for proposals of General Trial Use projects for K computer and related information.