jh210049-MDH

Developing data driven analysis methods for extreme scale numerical simulations

Yuuichi Asahi（Japan Atomic Energy Agency）

Abstract

We aim to establish an in-situ data analysis method for large scale fluid simulation data and develop deep learning based surrogate models to predict fluid simulation results. Firstly, we have developed an in-situ data processing approach, which loosely couples the MPI application and python scripts. It has been shown that this approach is simple and efficient which offers the speedup of 2.7 compared to post hoc data processing. Secondly, we have developed a deep learning model for predicting multi-resolution steady flow fields. The deep learning model can give reasonably accurate predictions of simulation results with orders of magnitude faster compared to simulations. Finally, we have presented the optimization strategies applied on a kinetic plasma simulation code that makes use of OpenACC/OpenMP directives and Kokkos performance portable framework to run across multiple hardware platforms.

1. Basic Information

(1) Collaborating JHPCN Centers

Tokyo, Tokyo-Tech, Nagoya

(2) Research Areas

- ☑ Very large-scale numerical computation
- ☑ Very large-scale data processing
- ☐ Very large capacity network technology
- ☐ Very large-scale information systems

(3) Roles of Project Members

Project representative Yuuichi Asahi works to develop ML/DL model to extract features from fluid simulation data. Shinya Maeyama performs the local plasma turbulence simulations. Julien Bigot works on the in-situ data analysis of GYSELA. Xavier Garbet works for theoretical analysis of non-local transport processes. Virginie Grandgirard gives the advice for the large scale plasma turbulence simulation. Takashi Shimokawabe gives advice for large scale simulation and deep learning models. Keisuke Fujii contributes on data-driven analysis. Naoyuki Onodera gives the advice for the large scale LBM simulation. Yuta Hasegawa gives the optimization on GPUs. Prof. Watanabe comments on characteristics of local transport processes. Yasuhiro Idomura gives advice for the large scale simulations. Prof. Katagiri supports the optimization on Flow, particularly the intranode parallelization. Prof. Aoki gives advices about the usage of TSUBAME3.0.

2. Purpose and Significance of Research

In this project, we aim to develop an in-situ data analysis method for large scale fluid simulation data and to establish an interpretable deep learning model for feature extraction and data reduction methods.

The simulation data size is getting larger and larger as the computational capability is improved on the way to exa-scale supercomputing. Accordingly, the efficient managements of large simulation data are essential in post-processing. We will establish the in-situ machine learning methodology to avoid saving and keeping the large simulation data to the storage.

Deep learning (DL) techniques can be used to predict the computational fluid dynamics (CFD) simulation results. Unfortunately, most of the deep learning models are not straightforward to interpret. In this work, we will develop a deep learning model to extract features from fluid simulation data and offer some meanings to the extracted data through the comparison with the established approaches.

3. Significance as JHPCN Joint Research Project

There are two major difficulties in this project: the technical difficulty to handle the large data and the difficulty to give physically reasonable interpretation. In FY2020, we have developed an incremental PCA for the time series of 5D fusion turbulence data (larger than 10 TB) based on Dask. Though the computational cost of the forementioned method is acceptable, the storage cost is hardly acceptable. It is thus needed to develop the in-situ version of this analysis to avoid saving the large amount of data. The French group has already succeeded to read the simulation data on memory from a python script through PDI library (developed by Dr. Bigot). By coupling their scripts with ours, we can apply the proposed method in the in-situ way.

Collaborating with the French group is also important for better understanding of the extracted features by our methods. Dr. Garbet is well-known expert in the field of plasma turbulence. Under the collaboration with the University of Tokyo, we will sophisticate our deep learning models to predict the fluid simulation results. Since both our members and Dr. Shimokawabe have rich experiences on GPU computing, we can share techniques to perform large amount of simulations to create datasets.

A deep collaboration based on JHPCN is essential for achieving physics and HPC objectives. In FY2021, we newly added a HPC and data science specialist from the Univ. Tokyo to develop the deep learning model for the neutral flow field prediction. Since we need GPU resources for deep learning, GPU environments TSUBAME3.0 and BDEC offered by JHPCN framework are essential for this work.

4. Outline of Research Achievements up to FY2020

This is a new project.

5. Details of FY2021 Research Achievements

<u>Scalable data analysis</u>

In FY2021, we have demonstrated a new data processing approach to combine the simplicity of post hoc approaches and the performance on in-situ by coupling MPI simulation codes with Dask and PDI [https://github.com/pdidev]. By loosely coupling the MPI-app and Dask scripts though PDI, we can directly transfer the data from the MPI-app to Dask scripts, bypassing files. We have shown that it is both simple to use and offers performance that already outperform post hoc expectations by a factor of 2.7. This work has been presented in international conferences [1, 7].

<u>Development of AI-based surrogate models</u>

In FY2021, we have developed a convolutional neural network (CNN) model, AMR-Net, to predict multi-resolution steady flow fields (u and v) from the object shapes (signed distance function). Signed distance functions (SDFs)

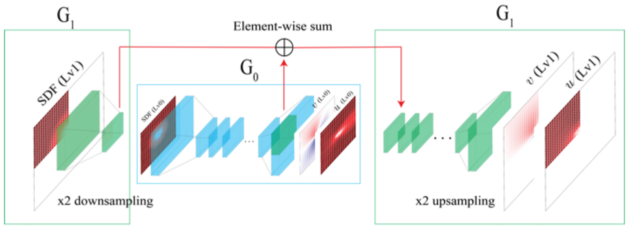

Fig. 1 The network architecture of AMR-Net. Instead of using high-resolution inputs, we use low-resolution global (un-patched) data and high-resolution local (patched) data to predict multi-resolution flow fields.

represent both simple and complex object shapes in a universal way. By extending the image-to-image translation model pix2pixHD, our model can predict high resolution flow fields from the set of patched signed distance functions. The network architecture is shown in Fig. 1. By combining the low-resolution global and high-resolution local (patched) data in the feature space, the model can predict high resolution local flow fields while keeping the global consistency (the global structure can be obtained from the low-resolution network).

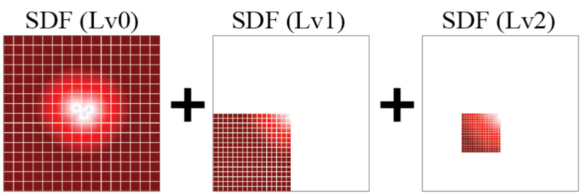

Fig. 2 Input data structure of AMR-Net.

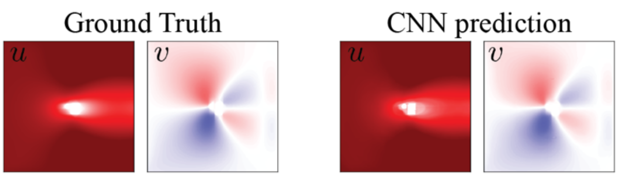

Fig. 3 The CNN prediction of global flow fields (u and v) from multi-resolution patched SDFs.

As shown in Fig. 2, the inputs of the network are multi-resolution signed distance functions (SDFs) at Lv0 (unpatched, low resolution), Lv1 (patched middle resolution) and Lv2 (patched high-resolution).

By patching the high-resolution data, our model uses roughly the one third of the memory used by pix2pixHD. The accuracy of our model is almost the same as the U-Net model (a conventional CNN model for image-to-image translation tasks) using the unpatched high-resolution data (see Fig. 3). This model gives the reasonably accurate predictions of simulation results with significant speedups.

This work has been presented in an international conference [3]. The model and dataset are publicly available [10].

<u>Researches on HPC and plasma physics</u>

For high performance computing, we have developed optimization strategies applied on a kinetic plasma simulation code that makes use of OpenACC/OpenMP directives and Kokkos performance portable framework to run across multiple CPUs and GPUs. We have evaluated the impacts of optimizations on multiple hardware platforms: Intel Xeon Skylake, Fujitsu Arm A64FX, and Nvidia P100, V100 and A100 GPUs (Fig. 4).

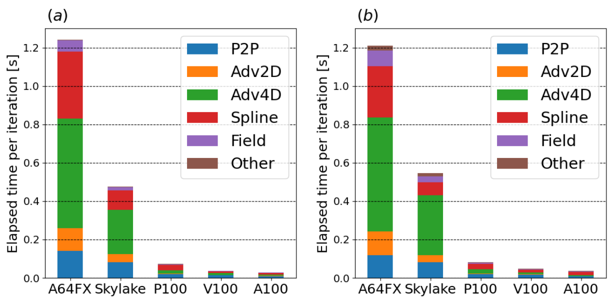

Fig. 4 The elapsed time of 6 parts of the mini-app including "P2P (P2P communications)", "Adv2D (2D advection)", "Adv4D (4D advection)", "Spline (Spline construction)", "Field (Poisson solver)", and "Other (other parts)" with (a) OpenACC/OpenMP and (b) Kokkos. 16 MPI processes are mapped in each case. 4 A64FX, 8 Skylake, 16 P100, 16 V100, and 16 A100 GPUs are used.

With vectorization and cache tuning, the OpenACC/OpenMP version achieved speedups of x1.07 to x1.39 on these architectures. The Kokkos version in turn achieved speedups of x1.00 to x1.33 with Memory Layout and execution policy tuning.

Since the impact of optimizations under multiple combinations of kernels, devices and parallel implementations is demonstrated, this work provides a widely available approach to accelerate codes keeping performance portability. This work has been presented in the supercomputing conference [2, 9].

The research achievements related to plasma physics have been reported in international conferences [4, 5, 6, 8].

6. Progress during FY2021 and Future Prospects

Scalable data analysis

French group has demonstrated a new data processing approach to combine the simplicity of post hoc approaches and the performance on in-situ by coupling MPI simulation codes with Dask and PDI. By loosely coupling the MPI-app and Dask scripts though PDI, we can directly transfer the data from the MPI-app to Dask scripts, bypassing files.

In parallel, we have also applied the incremental PCA on the 3D potential data from running GYSELA simulation through hdf5 files. In FY2022, we will couple GYSELA and Dask scripts through PDI to perform the incremental PCA on the 3D-2V (3D space and 2D velocity space) distribution function data from running GYSELA simulation.

Development of AI-based surrogate models

We have successfully developed a surrogate model for CFD simulations. The model can give the reasonably accurate predictions of simulation results with significant speedups.

In FY2022, we will develop AI-based models which partially replace the direct simulations while keeping the physics. We will investigate the emerging approaches like physics informed neural network (PINN) or probabilistic models which remains as a future task.

Researches on HPC

In FY2021, we have developed optimization strategies applied on a kinetic plasma simulation code that makes use of MPI + OpenACC/OpenMP directives and MPI + Kokkos performance portable framework to run across multiple CPUs and GPUs. Some test implementations of the mini-apps with OpenMP4.5 and stdpar (C++ parallel algorithm) have also been made.

In FY2022, we will investigate the performance of multiple applications on AMD GPUs. Firstly, we are planning to port the mini-apps to AMD GPUs with Kokkos and OpenMP4.5 to support performance portability. Secondly, we will also port a production simulation code in CUDA to AMD GPUs with HIP [https://github.com/ROCm-Developer-Tools/HIP].

In parallel, we will explore the performance of mini-apps over CPUs and GPUs implemented with the language standard parallelism (stdpar) and the language standard high dimensional array support (mdspan).

7. List of Publications and Presentations

Please note that items in status of "to be submitted/presented" and "submitted" cannot be included.

(1) Journal Papers (Refereed)

(2) Proceedings of International Conferences (Refereed)

[1] A. Gueroudji (+), J. Bigot (+), and B. Raffin (+), "DEISA Dask-Enabled In Situ Analytics", Proceeding of the 28th IEEE International Conference on High Performance Computing, Data, and Analytics, December 17th 2021.

[2] Y. Asahi, G. Latu (+), J. Bigot (+), and V. Grandgirard (+), "Optimization strategy for a performance portable Vlasov code", in 2021 IEEE/ACM International Workshop on Performance, Portability and Productivity in HPC (P3HPC), November 2021, pp. 79-91, doi: 10.1109/P3HPC54578.2021.00011.

[3] Y. Asahi, S. Hatayama, T. Shimokawabe, N. Onodera, Y. Hasegawa and Y. Idomura, "AMR-Net: Convolutional Neural Networks for Multi-resolution Steady Flow Prediction", AI4S, IEEE CLUSTER, September 2021.

(3) International conference Papers (Non-refereed)

[4] S. Maeyama, T.-H. Watanabe, M. Nakata, M. Nunami, Y. Asahi, and A. Ishizawa, "Gyrokinetic Simulations of Cross-Scale Interactions between Electron Temperature Gradient and Trapped Electron Modes on the Supercomputer Fugaku", The 30th International Toki Conference, November 2021 (Invited).

[5] Y. Asahi, K. Fujii, S. Maeyama and Y. Idomura, "Phase-Space Pattern Extraction from 5D gyrokinetic simulation data", The 30th International Toki Conference, November 2021 (Invited).

[6] S. Maeyama, "Exploring multiscale turbulent interactions in high electron temperature burning plasma", 5th Asia-Pacific Conference on Plasma Physics, September 2021 (Invited).

[7] A. Gueroudji (+), J. Bigot (+), and B. Raffin (+), "Preliminary Experiments in Coupling in situ Dask analytics with MPI Simulations", HPCDA, ISC High Performance 2021, July 2nd 2021.

[8] V. Grandgirard (+), Y. Asahi, J. Bigot (+), et al, "How to prepare the GYSELA gyrokinetic code to future exascale Edge-Core Simulations", PASC, July 6th 2021.

(4) Presentations at domestic conference (Non-refereed)

(5) Published library and relating data

[9] Y. Asahi, G. Latu (+), J. Bigot (+), and V. Grandgirard (+), "MPI+Kokkos, MPI+OpenACC, MPI + OpenMP4.5, and MPI + stdpar implementation of 2D+2V Vlasov Poisson code", https://github.com/yasahi-hpc/vlp4d_mpi

[10] Y. Asahi, "Multi-resolution steady flow prediction by Convolutional Neural Networks in PyTorch (Multi-resolution steady flow dataset also available)" https://github.com/yasahi-hpc/AMRNet

(6) Other (patents, press releases, books and so on)