

Optimizations of a spectral/finite difference gyrokinetic code for improved strong scaling toward million cores

Shinya Maeyama
Japan Atomic Energy Agency
2-166 Omotedate, Obuchi,
Rokkasho, Aomori, Japan
maeyama.shinya@jaea.go.jp

Tomohiko Watanabe
National Institute for
Fusion Science
322-6 Oroshi-cho,
Toki, Gifu, Japan
watanabe.tomohiko@nifs.ac.jp

Yasuhiro Idomura
Japan Atomic Energy Agency
5-1-5 Kashiwanoha,
Kashiwa, Chiba, Japan
idomura.yasuhiro@jaea.go.jp

Motoki Nakata
Japan Atomic Energy Agency
2-166 Omotedate, Obuchi,
Rokkasho, Aomori, Japan
nakata.motoki@jaea.go.jp

Masanori Nunami
National Institute for
Fusion Science
322-6 Oroshi-cho,
Toki, Gifu, Japan
nunami.masanori@nifs.ac.jp

Akihiro Ishizawa
National Institute for
Fusion Science
322-6 Oroshi-cho,
Toki, Gifu, Japan
ishizawa@nifs.ac.jp

ABSTRACT

Plasma turbulence is governed by multi-scale physics, extending over spatio-temporal scales of electrons and ions, whose simulations based on the gyrokinetic theory require huge computations. In this work, we present optimizations of the gyrokinetic Eulerian code GKV, which performs five-dimensional (5D) computational fluid dynamics calculations by means of spectral and finite difference methods, on massively parallel platforms. First, a segmented rank mapping on the 3D torus network utilizes the bi-section bandwidth and reduces the collisions of the messages for simultaneous point-to-point communications, and therefore, reduces communication cost. Second, computation-communication overlaps with pipelining methods effectively mask the communication cost. Thanks to the optimizations, GKV achieves excellent strong scaling up to $\sim 600k$ cores with high parallel efficiency $\sim 99.99994\%$ on the K computer (K), and enables us to address the first multi-scale plasma turbulence simulations employing the real ion-to-electron mass ratio.

Categories and Subject Descriptors

[**Computing methodologies**]: Massively parallel and high-performance simulations; [**Applied computing**]: Physics

General Terms

Algorithms, Performance

Keywords

Plasma turbulence, Parallel performance, Three-dimensional torus network, Computation-communication overlap

1. INTRODUCTION

Turbulence in magnetic fusion plasma is anisotropic due to strong magnetic fields. Its perpendicular scales are in the order of the gyroradii of charged particles, which are different for electrons and ions by a factor of the square root of the mass ratio. Therefore, plasma turbulence is inherently a multi-scale phenomenon, including electron- and ion-scale physics, whose simulations require huge computations. To carry out the multi-scale plasma turbulence simulation using the real ion-to-electron mass ratio, we have extended the gyrokinetic Eulerian code GKV [1, 2]. Here, we present the optimization of inter-node communications and computation-communication overlap techniques on massively parallel platforms such as K and BX900 (Nehalem-EP cluster), and report their impacts on the strong scaling.

2. THE GKV CODE

The GKV code is based on the gyrokinetic theory and solves time evolution of distribution functions of ions and electrons and associated electromagnetic potentials. From numerical viewpoints, the code performs computational fluid dynamics (CFD) calculations in the 5D space (x, y, z, v, μ) for each species (s). The procedures mainly consist of three parts: spectral calculations in the x and y [using the parallelized fast Fourier transforms (FFTs) and the $3/2$ de-aliasing rule], finite difference calculations in the z , v and μ , and a field solver (including integrations over v , μ and s). Then, the time integration is performed by means of the fourth-order Runge-Kutta-Gill method. The computations are parallelized by using the OpenMP/MPI hybrid parallelization. Since multi-dimensional domain decomposition is applied for x (or y), z , v , μ and s , the required MPI communications are data transpose for the parallel FFTs in x and y , point-to-point communications in z , v and μ , and reduction communications over v , μ and s . The most time consuming calculation is the parallel FFTs, where data transpose often degrades the scalability. Since the multi-scale turbulence simulations require high resolutions in the coordinates perpendicular to the magnetic field, x and y , we need to improve the strong scaling by overcoming this severe bottleneck.

3. OPTIMIZATIONS

Our main target is K, which consists of 82,944 nodes (2 GHz, 16 GFlops/core, Memory-BW 8 GB/s/core, 8 cores/node) connected by Torus fusion (Tofu) interconnect (6D mesh/torus topology, Interconnect-BW 5 GB/s $\times$ 4, Multi-way simultaneous communications: 4 sends + 4 receives). We developed the following optimizations to improve the strong scaling.

3.1 Segmented mapping on 3D torus networks

The Tofu interconnect provides a 3D torus network as a user view. To minimize the communication cost, we design the segmented rank mapping on the 3D network as follows. First, a segment of the nodes, which is involved in a local 3D box, is defined so that data transpose in x and y is performed only in a segment. Second, the segments are set up so that point-to-point communications in z , v and μ are performed between adjacent segments. This constructs a segment group for each species. Finally, the segment groups are arranged so that reduction communications over v , μ and s are performed through a cross section of the 3D network.

The segmented rank mapping has advantages for reducing the communication cost. First, the mapping fully utilizes the bisection bandwidth and minimizes the cost of the data transpose. On K, additionally, high-performance and highly-scalable MPI_AlltoAll is available when the communicating processes are arranged in such a box shape. Second, the mapping minimizes the distance of the nodes for point-to-point communications and avoids collisions of the message during the simultaneous point-to-point communications. We carry out communications in z and v simultaneously (two sends and two receives in each of z and v), which takes full advantage of the multi-way channels on the Tofu interconnect. Finally, the mapping also minimizes the reduction communications for the field solver. As a result, the communication cost are significantly reduced.

3.2 Computation-communication overlaps

To mask the communication cost, we implement the computation-communication overlaps for the spectral calculations. Taking advantage of independence of the parallel FFTs in the μ direction, the procedures are pipelined in the μ loop, and the data transpose and spectral computations with adjacent μ indexes are overlapped by using a communication thread, which is implemented on the MASTER thread. To improve the strong scaling, it is important that all computations of the spectral calculations (not only FFTs but also real space calculations, buffer copies, and data expansion/truncation for the de-aliasing) are overlapped.

In the similar way, the point-to-point communications are also overlapped with the finite difference calculations. While the communications in the μ direction are simply overlapped with the independent x and y space computations, the overlaps of the communications in z and v are realized by using a communication thread and pipelining in the μ loop. Although the split of the stencil calculations into two parts (i.e., the edge part which refers the buffer data from neighboring domains and the inner part which can be computed without the buffer data) has often been used for the overlaps of finite difference calculations, we note that this approach is not sufficient in our case. This is because pursuing the excellent strong scaling, where the employed array size becomes

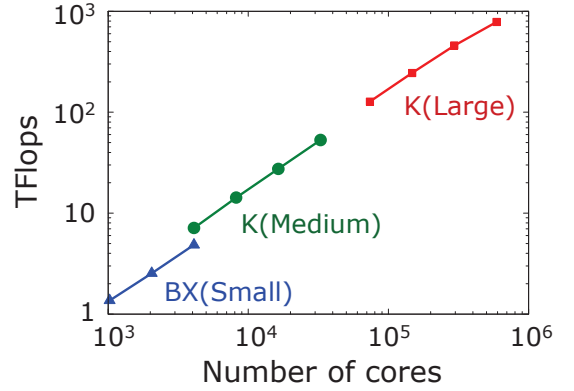


Figure 1: Scalings of GKV on BX900 and K. [Problem size: $(n_x, n_y, n_z, n_v, n_\mu, n_s) = (256, 256, 32, 32, 32, 2)$ for BX900(Small), $(1024, 1024, 32, 32, 32, 1)$ for K(Middle), $(1024, 1024, 96, 96, 32, 2)$ for K(Large).]

close to the buffer size as the number of the parallelization increases. Thus, the inner part of the computations becomes too small to mask the communication cost.

3.3 Impacts on the strong scaling

The impacts of the optimizations are examined via the strong scaling test on BX900 and K. Figure 1 demonstrates that the GKV code shows good strong scaling up to ~ 600 k cores. The parallel efficiency estimated from the Amdahl's law is $\sim 99.99994\%$, which is improved by an order of magnitude compared with the previous results ($\sim 99.9998\%$ in Ref. [2]). This is because the implementation of the species parallelization increases the number of available cores, and the presented optimizations improve scalability by reducing and masking the communication cost.

4. CONCLUSIONS

The GKV code, which performs 5D CFD calculations by using spectral and finite difference methods, is highly optimized for K. The code shows the excellent strong scaling and demonstrates its applicability and efficiency toward million cores. It strongly accelerates the calculations and enables multi-scale plasma turbulence simulations using the real ion-to-electron mass ratio.

5. ACKNOWLEDGMENTS

This work is supported by the HPCI Strategic Programs Field No. 4, the G8 Research Council Initiative (NuFuSE), and the JAEA-NIFS collaboration program.

6. REFERENCES

- [1] T.-H. Watanabe, and H. Sugama. Velocity-space structures of distribution function in toroidal ion temperature gradient turbulence. *Nucl. Fusion*, 46(1):24–32, January 2006.
- [2] S. Maeyama, T.-H. Watanabe, Y. Idomura, M. Nakata, M. Nunami, and A. Ishizawa. Computation-communication overlap techniques for parallel spectral calculations in gyrokinetic vlasov simulations. *Plasma Fusion Res.*, in press.