Advances and Optimizations of Gyrokinetic Turbulence Code GKV towards Exa-scale Computing

M. Nunami\textsuperscript{1,2}, M. Nakata\textsuperscript{1,2}, T.-H. watanabe\textsuperscript{3}, S. Maeyama\textsuperscript{3}, K. Nishioka\textsuperscript{3}, A. Ishizawa\textsuperscript{4}, and Y. Idomura\textsuperscript{5}

\textsuperscript{1}National Institute for Fusion Science, Japan
\textsuperscript{2}The Graduate University for Advanced Studies, Japan
\textsuperscript{3}Nagoya University, Japan,
\textsuperscript{4}Kyoto University, Japan
\textsuperscript{5}Japan Atomic Energy Agency, Japan
Acknowledgements

We greatly appreciate collaborators,

• R. Kanno, S. Matsuoka, K. Tanaka, Y. Suzuki, K. Nagaoka, H. Takahashi (NIFS)
• N. Kasuya (Kyushu U.)
• Y. Kasai (Fujitsu Ltd.)
• M. Honda, E. Narita, M. Yoshida, Y. Asahi (QST)
• J. Velasco, A. Alonso (CIEMAT)
• P. Xanthopoulos, G. Plunk (IPP, Greifswald)
• J. Proll (TUe)
• D. Mikkelsen (PPPL)

This work is supported by the MEXT, Grant for Post-K priority issue No. 6: Accelerated Development of Innovative Clean Energy Systems.

The simulations have been done by using Plasma Simulator (NIFS), Helios & JFRS-1 (IFERC-CSC), and K-computer (Riken R-CCS).
Outline

- Introduction & Gyrokinetic Vlasov code: GKV
- Recent studies from GKV simulations
  - Multi-scale turbulence simulations
  - Isotope effects and multi-species particle transport
- Optimizations in GKV
  - Overlap techniques between compt. & comm., and Segmented MPI-procs. mapping
  - More optimization toward Exa-scale computing
- For further speed up of GKV
  - Implicit solvers for collision operator
  - Moment extract approach
- Summary
Outline

- **Introduction & Gyrokinetic Vlasov code: GKV**
  - Recent studies from GKV simulations
    - Multi-scale turbulence simulations
    - Isotope effects and multi-species particle transport
  - Optimizations in GKV
    - Overlap techniques between compt. & comm., and Segmented MPI-procs. mapping
    - More optimization toward Exa-scale computing
  - For further speed up of GKV
    - Implicit solvers for collision operator
    - Moment extract approach
- Summary
Introduction

- Transport of magnetic confined plasmas
  - Collisions and orbits of particle motions etc. makes neoclassical transport.
  - Turbulence driven by microinstabilities causes anomalous transport.
  - Particle and heat transport observed in experiments are much larger than predictions by collisional transport theory (NC transport).
  - As a first-principle framework, 
    **Gyrokinetic model is powerful tool**
    for qualitative and quantitative analyses
    for the transport phenomena.

Gyrokinetic eq. (ES case)
Introduction

**Gyrokinetic simulations for multi-scales, multi-species, and stellarators**

- Multi-scales from ion to electron scale sims:
  - Need quite higher resolution in perp. space.
- Multi-species turbulent transport:
  - Higher comp. costs & complicated collision term.
- 3D complicated field structure:
  - Higher resolution for config. than tokamaks (~ 100 X [Tokamak cases]).
- This is the computational challenge.
- Developments of HPC & optimizations of the code makes us to perform above gyrokinetic simulations.
- We can perform gyrokinetic simulations to evaluate the anomalous contributions to the plasma transport.
GKV code: GyroKinetic Vlasov code

- Local flux-tube 5D gyrokinetic simulation code
  - \( \delta f_s \) model for multi-species with fixed-background \( F_{s0} \) \( f_s = F_{s0} + \delta f_s \)
  - GKV solves GK eq. & Poisson/Ampere eqs. for \( \delta f \) in local flux-tube domain,
  - Eulerian (Continuum CFD) solver.
    - Spectral (FFT) in 2D (kx,ky)-space perp. to B-field
    - Finite-difference in 3D (z, v||, \( \mu \))-space
  - Parallelization with multi-dim. domain decomposition in (ky, z, v||, \( \mu \), s).
  - Hybrid parallelization with MPI/OpenMP.
  - GKV is a free software (under GNU General Public License).
HPC applications of GKV

Physical capability of GKV is extended along with HPC.

Multi-species & multi-scale turbulence simulations from ion to electron scales


Ion-scale turbulence simulations in complex LHD plasma


Ion-scale turbulence simulations in simple Tokamak plasma

Outline

- Introduction & Gyrokinetic Vlasov code: GKV
- Recent studies from GKV simulations
  - Multi-scale turbulence simulations
  - Isotope effects and multi-species particle transport
- Optimizations in GKV
  - Overlap techniques between compt. & comm., and Segmented MPI-procs. mapping
  - More optimization toward Exa-scale computing
- For further speed up of GKV
  - Implicit solvers for collision operator
  - Moment extract approach
- Summary
Multi-scale plasma turbulence simulations

- Wide-range instabilities from ion to electron scale.
- Electromagnetic (finite-\(\beta\)) effects stabilize ion modes.
- Multi-scale interactions between ion-scale turbulence and electron-scale turbulence are observed.
- Contrary to the conventional scale separation assumption, multi-scale interactions change the heat transport.
- Physical mechanism of multi-scale interactions is revealed via nonlinear triad interaction analysis.

Isotope effects on turbulent transport

- Ion mass and collisionality dependence of the mixing length diffusivity,

\[ \chi_{\text{turb}} \sim \frac{\gamma}{k^2} = \frac{\widetilde{\gamma}_s}{k^2_s} \sqrt{A_s} \chi^{\text{GB}}(H) \]

- For ITG: almost no isotope-dep. in \( \widetilde{\gamma}_s \) \( \Rightarrow \tau_{ii}^{-1}/\omega_{ti} \propto A_i/A_c \)
- For TEM: reduction in \( \widetilde{\gamma}_s \) due to the isotope-dep. in \( \tau_{ei}^{-1}/\omega_{ti} \propto (A_i/A_c)^{1/2} \)

- Linear GK analysis predicts the improved confinement for TEM cases in a certain \( \nu_{ei}^* \)-regime (\( \nu_{ei}^* > 0.04 \)), beyond the Gyro-Bohm scaling.

- In TEM turb. sims. in LHD H- and D-plasmas, in addition to linear stabilization, transport reduction resulting from enhanced ZFs is identified in D-plasma.
Particle transport in multi-species plasmas

- LHD high-$T_i$ (> 5keV) discharge heated by NBI
- Hollow impurity density “impurity hole” (Yoshinuma, NF2009)
  \[ \Rightarrow \text{Advantage to avoid impurity accumulation} \]

There exists multiple particle species (e, H, He, C).
- Particle balances with neoclassical & turbulent fluxes;
  \[ \frac{\partial n_s(\rho)}{\partial t} + \frac{1}{V'} \frac{\partial}{\partial \rho} V' \left( \Gamma^{(\text{neo})}_s + \Gamma^{(\text{trb})}_s \right) = S^{(\text{aux})}_s \]
- Particle transport has various types of dependences on grad-T & grad-n.
Outline

- Introduction & Gyrokinetic Vlasov code: GKV
- Recent studies from GKV simulations
  - Multi-scale turbulence simulations
  - Isotope effects and multi-species particle transport
- Optimizations in GKV
  - Overlap techniques between compt. & comm., and Segmented MPI-procs. mapping
  - More optimization toward Exa-scale computing
- For further speed up of GKV
  - Implicit solvers for collision operator
  - Moment extract approach
- Summary
Optimizations in GKV

- Some communications exist in multi-dimensional CFD in GKV.
  - Spectral methods in \((x, y)\) ⇔ data transpose communication
  - Finite difference in \((z, v, \mu)\) ⇔ 1-to-1 communication
  - Integration over \((v, \mu, s)\) ⇔ reduction communication

- Communications should be masked and reduced.

Overlap techniques between compt. & comm.

- Communication thread enables overlaps for All-to-All.
- Overlap techniques are applied for
  - Spectral in \((x, y)\); Pipelined overlaps
  - FD in \((z, v, \mu)\);
  - Overlaps for MPI/OpenMP hybrid parallelization

Segmented MPI-procs. mapping of 5D problem on 3D torus network

- In K computer, 3D torus network is available.
- To reduce the costs of communication time, MPI mapping of 5D problem is optimized.

1. Arrange rank_xy:
   Data transpose is performed in a segment.

2. Arrange rank_z, _v, _m:
   Point-to-point communications are performed between adjacent segments.

3. Arrange rank_s:
   Reduction is performed in a cross section.


Performances on K computer

- Mapping + Overlaps
- Mapping
- No optimization

Number of cores [x10^3] vs. Speed up
Strong scaling toward million cores

- The optimized GKV code enables to perform GK simulations with high efficiencies and performances.
- Excellent strong scaling has been realized up to ~600k cores
- Achieved parallelization rate ~99.99994%
- Computation performance ~780 TFLOPS (Flops/Peak: 8.3 - 10.8%)

Problem size on K: 
\((n_{x}, n_{v}, n_{z}, n_{v}, n_{\mu}, n_{s}) = (1024, 1024, 96, 96, 32, 2)\)

Parallelization on K: 
\((N_{xy}, N_{z}, N_{v}, N_{\mu}, N_{s}, N_{\text{threads}}) = (8-64, 12, 12, 4, 2, 8)\)
Issues in comm.-compt. overlap in many-core system

- In many-core system as Exa-scale computers, the dynamic scheduling overhead in thread parallelizations are crucial.
- Comm.-Comp. overlap performance will degrade.

Non-blocking comm. by assistant cores on SMaC architecture

- In Japanese FLAGSHIP2020 project, Post-K computer is being developed towards Exa-scale computing, and Scalable Many Core (SMaC) architecture with Assistant cores (ACs) will be applied.
- In Fujitsu FX100, ACs (2cores with 32comp. Cores/node) are useful not only for reducing OS Jitter, but also for performance improvement.
  - Optimized Comm.-Comp. overlap w/o Master thread comm.
  - Fully non-blocking ISend/IRecv, IAllreduce, IAlltoAll.
  - Static or chunk-size-optimized Dynamic scheduling (More efficient calc. to mask overheads)
Numerical performance is compared among 3 types of the overlap:
- **MS(D)**: Master thread comm. with Dynamic sched. (chunk size=1, As-Is)
- **AC(D)**: Assistant core comm. with Dynamic sched. (chunk size=1)
- **AC(S)**: Assistant core comm. with Static sched.

On FX100:
- 4x12x18 nodes
- ~10 billion grid points
- 432 MPI proc. with 16 SMP

Comm. on MS is successfully masked by AC:
\[\frac{16}{15} = 6.25\%\text{ improve.}\] of comp. dynamic sched.

Static scheduling leads to not only reduction in OMP overhead, but also improved load/store and cache performance (incl. prefetch) with the continuous memory access for larger chunk size:
>15\% improve.
Outline

- Introduction & Gyrokinetic Vlasov code: GKV
- Recent studies from GKV simulations
  - Multi-scale turbulence simulations
  - Isotope effects and multi-species particle transport
- Optimizations in GKV
  - Overlap techniques between compt. & comm., and Segmented MPI_procs. mapping
  - More optimization toward Exa-scale computing
- For further speed up of GKV
  - Implicit solvers for collision operator
  - Moment extract approach
- Summary
For further speed up

- Our next target is **Post-K** computer around 2020 (Designed for Exa-Flops, Many-core processors, 100k - 1M nodes)
- Computation / communication cost ratio becomes severe.
- We should explore physically / computationally efficient time integration;
  - Implicit solver for collision operator
  - Moment extract approach

Implicit collision operator

- Velocity-dependent collision frequency ($\nu \propto 1/v^3$) restricts CFL.
- Since collision is an integro-differential operator over $(v_\parallel, \mu, s)$,
  - Data transpose by MPI_alltoall
    $$f \left( n_x, n_y, n_z, n_v, n_\mu, n_s \right) = f \left( n_v, n_\mu, n_s, \frac{n_z}{P_z}, \frac{n_x}{P_x}, \frac{n_y}{P_y} \right)$$
  - Iterative implicit solver for $f(v_\parallel, \mu, s)$, independent to $(x, y, z)$
  - Transpose back again by MPI_alltoall
- Arithmetic intensity and computational performance are enhanced. $\Rightarrow$ Promising for manycore processor.
In Vlasov simulations, some restrictions exist due to the condition, $C \downarrow w = \omega \Delta t < 1$.

As a novel semi-implicit solver for time integrations in GKV, operator splitting scheme is applied in the momentum extract (ME) approach.

**Operator splitting (Strang splitting)**

\[
\frac{\partial h}{\partial t} = f(h) + g(h)
\Rightarrow h(t + \Delta t) = e^{\Delta t(f+g)} h(t)
= e^{\Delta t g/2} e^{\Delta t f} e^{\Delta t g/2} h(t) + \mathcal{O}(\Delta t^3)
\]

In a test for toroidal ITG mode with
\[
C_v = v_{\text{lim}} \Delta t / \Delta z = 0.47, C_w = v_{\text{pmax}} \Delta t / \Delta z = 0.40,
\]
precise calculation are successfully performed compared with conventional semi-implicit method (A-SIRK).

For lower wavenumbers, the width of time step $\Delta t$ can be extended to $1.7 \times$ RKG limit.
Outline

- Introduction & Gyrokinetic Vlasov code: GKV
- Recent studies from GKV simulations
  - Multi-scale turbulence simulations
  - Isotope effects and multi-species particle transport
- Optimizations in GKV
  - Overlap techniques between compt. & comm., and Segmented MPI-procs. mapping
  - More optimization toward Exa-scale computing
- For further speed up of GKV
  - Implicit solvers for collision operator
  - Moment extract approach

Summary
Summary

Gyrokinetic simulations and GKV code

- Gyrokinetic simulation is most powerful tool for qualitative/quantitative analyses for turbulent plasma transport phenomena.
- Owing to recent HPC and developments, analyses for multi-scale turbulence, isotope effects and multi-species particle transport become possible by means of gyrokinetic code, GKV.

Optimizations and advances in GKV

- **Optimization**
  - For efficient computations, communications should be masked and reduced.
  - Overlap techniques for compt. & comm., and optimized mapping of 5D problem strong cause excellent strong scaling toward million cores.
  - In Comm.-Comp. overlap by AC, all the OpenMP threads concentrate their computations, and the usage of Static scheduling leads to ~22.5% improvement of performance.

- **For further speed up**
  - Implicit solver for collision operator leads to arithmetic intensity and enhancements of computational performance.
  - Semi-implicit solver for time integrations, operator splitting scheme is applied in momentum extract.
Thank you very much!