Data Management and Analysis Challenges Associated with Validating Multi-Scale Gyrokinetic Simulation of Experimental Discharges

N.T. Howard, ORISE Postdoctoral Fellow
nthoward@psfc.mit.edu

Primary Panel: F ; Oral Presentation: No

Background

Steady increases in computing and the increased utilization of MPI and OpenMP have, for the first time, allowed for full physics, gyrokinetic simulation that simultaneously captures all of the spatio-temporal scales believed to be relevant for electrostatic plasma turbulence in the core of fusion devices. However, with the grid resolutions required to model all of the relevant physical phenomena, additional challenges related to data management and analysis have arisen. As simulation data is output on increasingly finer computational grids, file sizes can become prohibitively large and direct comparisons with experiment can even become infeasible. In this paper, challenges in data management and analysis arising from attempts to validate full physics, multi-scale gyrokinetic simulation against experimental measurements are discussed. However, it should be noted that these issues are applicable to any high-resolution simulations such as those performed in the wide range of fields in plasma physics.

Issues Associated with Multi-Scale Simulation Output and Comparison With Experiment

A. Output file size

The simplest issue associated with multi-scale gyrokinetic simulation (or possible any other high resolution simulations) is related to the size of the output files. If computational grids are made extremely fine to resolve the phenomenon of interest, and the corresponding outputs are generated on these same grids (as a general rule this is true), the resulting file sizes may limit the ability of the user to transfer, copy, and even open the outputs in some cases. For example, the high-resolution multi-scale simulations performed output complex data arrays associated with the fluctuating quantities (ϕ, n, T, etc.) of with dimensions of approximately 3 x 1800 x 342 x 4800. These dimensions represent the total number of kinetic species, radial grid points, toroidal grid points, and time points respectively. Such large data arrays can result in file sizes approaching 180 gb for a single fluctuating field, even when data is output at only a single poloidal location. With multiple fluctuating fields, total simulation sizes can easily exceed a 0.5-1.0 terabytes, making copying and transfer of the files difficult and time consuming. The memory requirements of devices capable of opening and manipulating such output files are also beyond typical desktops. Therefore, systems with a large amount of available memory are required to perform analysis and are not always readily available. Techniques to work around these limitations have been developed but more permanent solutions are required, as simulations will inevitably continue to push to higher and higher resolutions.

B. Synthetic Diagnostics

In recent years, comparison of gyrokinetic simulation with experiment has pushed beyond derived quantities such as heat and particle fluxes and has begun to make quantitative comparison with the output fluctuations themselves. These comparisons are made using synthetic diagnostics designed to reproduce the response of an experimental diagnostic on simulation data, allowing for direct comparison between simulation and experiment. Such comparisons have been made in a variety of fields of plasma simulation and are extremely valuable as they provide a more stringent test of the model fidelity. However, as most measurements provide information from specific regions or line integrals through the plasma, simulation outputs are generally required with finite radial and poloidal extent, further expanding the size of output files and making them prohibitively large for any significant manipulation. If the desired synthetic diagnostic comparisons are known *a priori,* it should be possible for a series of calculations to be implemented in the simulations themselves to aid in comparison with experiment (such as evaluation of a particular line integral, etc.). However, this obviously eliminates the possibility of additional comparisons after the simulation is complete. Solutions for aiding in synthetic diagnostic comparisons must be developed, as these comparisons provide perhaps the best opportunities to validate physics models and will become increasingly difficult as simulation resolutions increase.