State-of-the-art turbulence simulations for fusion and astrophysical plasmas with GENE

Max Planck Institute for Plasma Physics, Garching
University of Ulm

SuperMUC Review Workshop
Garching, 8-9 July 2014
Our plasma universe: More than 99% of the visible universe is in a plasma state

Turbulence is widely recognized as an important open problem in modern physics & astrophysics
EXCITING NEW OBSERVATIONS

*In situ* measurements of plasma turbulence at various scales!

Cluster mission
TURBULENCE LABORATORY

THE SOLAR WIND AS PLASMA TURBULENCE LABORATORY

Sahraoui et al., PRL 09

FIG. 3 (color online). High-pass filtered power spectra of the parallel (green line) and perpendicular (blue line) of the magnetic field spectrum. The black curve shows the power spectrum of the electric field in the same spacecraft. The low-frequency cutoff is due to reaching the noise level of the instrument.

The low-frequency part of the parallel magnetic field spectrum is shown in black, and the low-frequency part of the electric field spectrum is shown in green. The spectrum of the electric field is much flatter than that of the magnetic field, indicating that the electric field is less affected by the noise level.

The straight black lines are direct power law fits of the spectra. Vertical arrows are defined in the text.

The present observations suggest that the energy of the turbulence is only slightly damped at the proton gyroscale. The proton scale can indeed be explained solely by dispersive fluctuations measured by STAFF-SC during the time subproton gyroscales [1].

The energy spectrum above the proton scale is due to reaching the noise level of the instruments. The straight black lines are direct power law fits of the spectra. Vertical arrows are defined in the text.

The predictions of the kinetic Alfvén wave (KAW) turbulence as predicted by the GK theory are in striking agreement with these observations. KAW behavior down to electron scales where enhanced electron Landau damping, as it is shown below.

To confirm this scenario of KAW energy cascade and dissipation becomes evident. This can be explained by the large (proton) scales (up to 10 Hz). Here, we are observing dissipation becomes evident. This can be explained by the large (proton) scales (up to 10 Hz). Here, we are observing dissipation becomes evident. This can be explained by the large (proton) scales (up to 10 Hz).

The results shown in Fig. 5 of Sahraoui et al. [23] are in striking agreement with these observations. KAW behavior down to electron scales where enhanced electron Landau damping, as it is shown below.

The results shown in Fig. 5 of Sahraoui et al. [23] are in striking agreement with these observations. KAW behavior down to electron scales where enhanced electron Landau damping, as it is shown below.
Largest such simulation to date
WHY DO WE CARE?

Examples:
• heating of solar corona and wind
• radiation from our Galactic Center
Plasmas in fusion research: ITER

Idea: New source of CO$_2$ free energy for centuries to come

Magnetic confinement in a large tokamak

Goal: 500 MW of fusion power

The 7 ITER parties
The resources for fusion energy are practically unlimited

Deuterium in a bath tub full of water and Lithium in a used laptop battery suffice for a family over 50 years
Global Gyrokinetic Simulation of Turbulence in ASDEX Upgrade

gene.rzg.mpg.de
gene@ipp.mpg.de
Extreme computing with the GENE code
**Ab initio** microturbulence: NL gyrokinetics

Microturbulence in weakly collisional plasmas requires a kinetic description!

Vlasov-Maxwell equations

\[ \left[ \frac{\partial}{\partial t} + v \cdot \frac{\partial}{\partial x} + \frac{q}{m} \left( E + \frac{v}{c} \times B \right) \cdot \frac{\partial}{\partial v} \right] f(x, v, t) = 0 \]

**From kinetics (6D) to “gyrokinetics” (5D):**
For strongly magnetized plasmas, remove fast gyromotion, consider **guiding center dynamics**

\[ \frac{\partial f}{\partial t} + \dot{X} \cdot \frac{\partial f}{\partial X} + \dot{v}_\parallel \frac{\partial f}{\partial v_\parallel} = 0 \]

\[ f = f(X, v_\parallel, \mu; t) \]

\[ \dot{X} = v_\parallel b + \frac{B}{B^*} \left( \frac{v_\parallel}{B} \bar{B}_{1\perp} + \frac{c}{B^2} E_1 \times B + \frac{\mu}{m\Omega} b \times \nabla(B + \bar{B}_{1\parallel}) + \frac{v_\parallel^2}{\Omega} (\nabla \times b)_\perp \right) \]

motion along fluctuating field lines

ExB drift

\[ \dot{v}_\parallel = \frac{\dot{X}}{mv_\parallel} \cdot \left( e\bar{E}_1 - \mu \nabla(B + \bar{B}_{1\parallel}) \right) \]

Rigorously derived in the 1980s; enormous reduction of spatio-temporal scales

Howes et al. 06
Global GK runs for actual tokamaks

Example: GENE code

Today, there exists a variety of NL GK codes based on different numerical techniques.

Verification via numerous benchmarking activities.
Gyrokinetic code GENE (F. Jenko et al., 1999 – )

- Code is *publicly available and widely used* (http://gene.rzg.mpg.de)
- Part of the *Unified European Application Benchmark Suite (PRACE)*
- First PRACE Call: Ranked #1 out of 65 projects from all areas of science

On SuperMUC:
Up to ~16 kcores

Linear scaling from 65536 to 262144 cores
on BlueGene/P
Parallelization/optimization strategy:
- high-dimensional domain decomposition
- either pure MPI or mixed MPI/OpenMP paradigm
- optimal subroutines and processor layout determined during initialization phase (à la FFTW)
- time step is chosen in an optimal way

Thanks to F. Merz (now at IBM)
Some computational challenges

• GENE runs are compute intensive; large individual runs may require up to tens of millions of core-hours

• Large runs use many billion grid points and require many TB of short-term storage

• Many different HPC platforms are used in parallel

• Recently, GENE has been ported to GPGPU & MIC systems
GENE on GPGPU systems  

- we have *initially* achieved a speedup of 2 of the complete code with the previous hardware generation (Nehalem 4-core CPUs and Fermi M2090 GPUs)

- the speedup marginalizes on the current hardware generation (SandyBridge 8-core and Kepler K20x GPUs)

- we have achieved an in-depth understanding of the reasons:
  - the CPU version of GENE has improved during the GPU development
  - a roofline analysis shows: GENE is memory bound, and the data transfer limits the GPU speedup
  - possible improvements:
    - PCIe 3.0
    - move more computations to the GPU: transfer-to-compute ratio decreases
Applied math: Iterative eigenvalue solvers

Available methods:

- Power Iteration (also Inverse Power Iteration, Rayleigh Quotient Iteration)
- Subspace Iteration with Rayleigh-Ritz projection
- Arnoldi method
- Lanczos method
- Krylov-Schur (with and without Harmonic Extraction)
- Davidson methods (as of recently)

Recent result:

Using the preconditioned Jacobi-Davidson method is about twelve times faster as Krylov-Schur method with harmonic extraction and twice as fast as the preconditioned Krylov-Schur method.

Christoph Kowitz, TUM, Master’s Thesis
Sparse grid combination technique

With D. Pflüger (Stuttgart) & M. Griebel (Bonn) & H.-J. Bungartz (TUM) et al.

Cartesian grid
- Regular data structure
- Huge number of grid points for high-dimensional problems “curse of dimensionality”

Combination technique
- Good approximation of the Cartesian grid solution
- Smaller number of grid points
- Existing code (GENE) can be used more or less as is
- Applicable to other high-dimensional grid-based problems

<table>
<thead>
<tr>
<th>Resolution: 33 grid points per dimension</th>
<th>2D</th>
<th>5D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cartesian grid</td>
<td>1,089</td>
<td>39,135,393</td>
</tr>
<tr>
<td>Combination tech.</td>
<td>641</td>
<td>206,358</td>
</tr>
</tbody>
</table>

Resolution:
A new level of parallelism

**Dual parallelism**

- Independent grid setups from the combination technique + massively parallel GENE runs
- Run times of the instances tend to vary strongly

**Optimize the load balance**

- A simple *load-model* estimates the runtime required for each grid
- A *scheduler* creates an optimal load balancing to minimize idle cores
Spin-off: Algorithmic fault tolerance

Hardware failures ($10^5$-$7$ cores)
- The whole simulation has to be restarted from the last checkpoint file
- In the combination technique, only a single GENE instance would crash

Two ways to handle the failure
- The combination technique recovers an approximation
- Only a single GENE instance is rerun – which is much smaller than the full problem

Such techniques may be very useful on future exascale architectures
The future
Two key challenges

Virtual fusion devices

Space weather prediction

- Plasma turbulence: Where fascinating physics, extreme computing, and global challenges meet
- More information: gene.rzg.mpg.de