GPU-Acceleration of Plasma Turbulence Simulations for Fusion Energy

by

J. Candy\textsuperscript{1}, I. Sfiligoi\textsuperscript{1}, E. Belli\textsuperscript{1}, N. Howard\textsuperscript{2}, C. Holland\textsuperscript{3}

\textsuperscript{1}General Atomics, San Diego, CA
\textsuperscript{2}MIT, Cambridge, MA
\textsuperscript{3}UCSD, San Diego CA

Presented at

SC18/NVIDIA
Dallas, TX
11-16 Nov 2018
Background and Motivation

1. **General Atomics** (GA) is a private contractor in San Diego.
Background and Motivation

1. **General Atomics** (GA) is a private contractor in San Diego
2. The GA **Magnetic Fusion** division does DOE-funded research
Background and Motivation

1. **General Atomics** (GA) is a private contractor in San Diego
2. The GA **Magnetic Fusion** division does DOE-funded research
3. Hosts **DIII-D National Fusion Facility**

![Image of DIII-D National Fusion Facility control room and tokamak](image_url)
Background and Motivation

1. **General Atomics** (GA) is a private contractor in San Diego
2. The GA **Magnetic Fusion** division does DOE-funded research
3. Hosts **DIII-D National Fusion Facility**
4. **THIS TALK**: GPU-based plasma turbulence simulation using **gyrokinetic model**
Important locations for CGYRO

**Source code**

github.com/gafusion/gacode

**DOI**

www.osti.gov/doecode/biblio/20298

**User Documentation**

gafusion.github.io/doc

**Documentary Video (for GYRO)**

www.youtube.com/watch?v=RLI6QW2x4Lg
ITER Facility (35 nations) under construction in France
GOAL: Simulate turbulent plasma in core (magenta) region
Why such a large facility?
Tokamak confinement improves with LARGE PLASMA VOLUME
Plasma theory in closed fieldline region well-understood
Helical field perfectly confines plasma (almost)
There is a small amount of radial energy/particle loss

- Collisions (1970s): $\Gamma_{\text{collision}}$
- Turbulence (1980s): $\Gamma_{\text{turbulence}}$
- Both exhibit **gyroBohm scaling**

\[ \text{flux} \quad \Gamma \sim v(\rho/a)^2 \]
\[ \text{confinement time} \quad \tau = \frac{a}{\Gamma} \sim \frac{a^3}{v\rho^2} \]

- $a = \text{torus radius}$
- $\rho = \text{particle orbit size}$
- $v = \text{particle velocity}$
CGYRO computes the turbulent flux
DIII-D Tokamak at General Atomics in San Diego, CA
CGYRO computes the turbulent flux
DIII-D Tokamak at General Atomics in San Diego, CA
CGYRO fully ported to GPU

NCCS TITAN (Oak Ridge, TN) – K20x
CGYRO fully ported to GPU

NCCS TITAN (Oak Ridge, TN) – K20x
CGYRO fully ported to GPU

CSCS PIZ DAINT (Lugano, Switzerland) – P100
CGYRO fully ported to GPU

CSCS PIZ DAINT (Lugano, Switzerland) – P100
CGYRO fully ported to GPU

General Atomics Power9 (San Diego, CA) – V100
CGYRO fully ported to GPU

General Atomics Power9 (San Diego, CA) – **V100**
CGYRO fully ported to GPU

General Atomics Power9 (San Diego, CA) – **V100**
History of Energy Research at GA

General Atomics – June 25th, 1959
Gyrokinetic equation for plasma species $a$

Typically: $a = (\text{deuterium, carbon, electron})$

$$\frac{\partial \tilde{h}_a}{\partial \tau} - i \Omega_s X \tilde{h}_a - i (\Omega_\theta + \Omega_\xi + \Omega_d) \tilde{H}_a - i \Omega_* \tilde{\Psi}_a + \Omega_{NL}(\tilde{h}_a, \tilde{\Psi}_a) = C_a$$

Symbol definitions

particles

$$\tilde{H}_a = \tilde{h}_a + \frac{z_a T_e}{T_a} \tilde{\Psi}_a$$
Gyrokinetic equation for plasma species $a$

Typically: $a = \text{(deuterium, carbon, electron)}$

$$\frac{\partial \tilde{h}_a}{\partial \tau} - i \Omega_s X \tilde{h}_a - i \left( \Omega_\theta + \Omega_\xi + \Omega_d \right) \tilde{H}_a - i \Omega_* \tilde{\Psi}_a + \Omega_{NL} (\tilde{h}_a, \tilde{\Psi}_a) = C_a$$

Symbol definitions

particles

$$\tilde{H}_a = \tilde{h}_a + \frac{z_a T_e}{T_a} \tilde{\Psi}_a$$

fields

$$\tilde{\Psi}_a = J_0(\gamma_a) \left( \delta \phi - \frac{v_\parallel}{c} \delta A_\parallel \right) + \frac{v_\perp^2}{\Omega_{ca} \gamma_a} \frac{J_1(\gamma_a)}{\gamma_a} \delta B_\parallel$$
Electromagnetic GK-Maxwell Equations

Coupling to fields is a MAJOR complication!

\[
\begin{align*}
\left(k^2 \lambda_D^2 + \sum_a z_a^2 \frac{T_e}{T_a} \int d^3\nu \frac{f_{0a}}{n_e} \right) \delta \tilde{\Phi} &= \sum_a z_a \int d^3\nu \frac{f_{0a}}{n_e} J_0(\gamma_a) \tilde{H}_a \\
\frac{2}{\beta_{e,\text{unit}}} k^2 \rho_s^2 \delta \tilde{A}_\parallel &= \sum_a z_a \int d^3\nu \frac{f_{0a} \nu_\parallel}{n_e c_s} J_0(\gamma_a) \tilde{H}_a \\
- \frac{2}{\beta_{e,\text{unit}}} \frac{B}{B_{\text{unit}}} \delta \tilde{B}_\parallel &= \sum_a \int d^3\nu \frac{f_{0a} m_a v^2_\perp}{n_e T_e} J_1(\gamma_a) \gamma_a \tilde{H}_a
\end{align*}
\]
Gyrokinetic equation for plasma species $a$

Typically, deuterium, some carbon, and electrons

\[
\frac{\partial \tilde{h}_a}{\partial \tau} - i \Omega_s X \tilde{h}_a - i (\Omega_\theta + \Omega_\xi + \Omega_d) \tilde{H}_a - i \Omega_* \tilde{\Psi}_a + \Omega_{NL}(\tilde{h}_a, \tilde{\Psi}_a) = C_a
\]

\textbf{E} \times \textbf{B} \text{ flow}

\[
-i \Omega_s = -i \frac{k_\theta L_a}{2\pi c_s} \gamma_E
\]

$\textit{acc parallel loop}$
Gyrokinetic equation for plasma species \( a \)

Typically, deuterium, some carbon, and electrons

\[
\frac{\partial \tilde{h}_a}{\partial \tau} - i \Omega_s X \tilde{h}_a - i \left( \Omega_\theta + \Omega_\xi + \Omega_d \right) \tilde{H}_a - i \Omega_* \tilde{\Psi}_a + \Omega_{NL}(\tilde{h}_a, \tilde{\Psi}_a) = C_a
\]

Streaming

\[-i \Omega_\theta = \frac{v_\parallel}{w_s} \frac{\partial}{\partial \theta}\]

$acc$ parallel loop
Gyrokinetic equation for plasma species $a$

Typically, deuterium, some carbon, and electrons

\[
\frac{\partial \tilde{h}_a}{\partial \tau} - i \Omega_s X \tilde{h}_a - i \left( \Omega_\theta + \Omega_\xi + \Omega_d \right) \tilde{H}_a - i \Omega_* \tilde{\Psi}_a + \Omega_{NL}(\tilde{h}_a, \tilde{\Psi}_a) = \mathcal{C}_a
\]

Trapping

\[
-i \Omega_\xi = - \frac{v_{ta}}{w_s} \frac{u_a}{\sqrt{2}} \left( 1 - \xi^2 \right) \frac{\partial \ln B}{\partial \theta} \frac{\partial}{\partial \xi} \left( 1 - \xi^2 \right) \frac{\partial}{\partial \xi}
\]

\[
- \frac{1}{2u_a} \frac{\partial \lambda_a}{\partial \theta} \left[ \frac{v_{||}}{w_s} \frac{\partial}{\partial u_a} + \frac{\sqrt{2}v_{ta}}{w_s} \left( 1 - \xi^2 \right) \frac{\partial}{\partial \xi} \right]
\]

Fold into collision operator
Gyrokinetic equation for plasma species $a$

Typically, deuterium, some carbon, and electrons

\[
\frac{\partial \tilde{h}_a}{\partial \tau} - i \Omega_s \times \tilde{h}_a - i \left( \Omega_\theta + \Omega_\xi + \Omega_d \right) \tilde{H}_a - i \Omega_\ast \tilde{\Psi}_a + \Omega_{\text{NL}}(\tilde{h}_a, \tilde{\Psi}_a) = C_a
\]

Drift motion

\[
-i \Omega_d = a \frac{\nu_{ta}}{c_s} \mathbf{b} \times \left[ u_a^2 (1 + \xi^2) \frac{\nabla B}{B} + u_a^2 \xi^2 \frac{8\pi}{B^2} (\nabla p)_{\text{eff}} \right] \cdot i \mathbf{k}_\perp \rho_a
\]

\[
+ M_a \frac{2av_\parallel}{c_s R_0} \mathbf{b} \times \left( \frac{R}{\partial \psi} \frac{\partial R}{\partial \theta} \nabla \varphi - \frac{B_t}{B} \nabla R \right) \cdot i \mathbf{k}_\perp \rho_a
\]

\[
+ \frac{a}{c_s} \mathbf{b} \times \left( -\frac{\nu_{ta}}{T_a} \mathbf{F}_c + \frac{c}{B} \nabla \Phi_\ast \right) \cdot i \mathbf{k}_\perp \rho_a
\]

Fold into streaming (diagonal)
Gyrokinetic equation for plasma species $a$

Typically, deuterium, some carbon, and electrons

$$\frac{\partial \tilde{h}_a}{\partial \tau} - i \Omega_s X \tilde{h}_a - i (\Omega_\theta + \Omega_\xi + \Omega_d) \tilde{H}_a - i \Omega_\star \tilde{\Psi}_a + \Omega_{NL}(\tilde{h}_a, \tilde{\Psi}_a) = C_a$$

Gradient drive

$$-i \Omega_\star = \left[ \frac{a}{L_{na}} + \frac{a}{L_{Ta}} \left( u_a^2 - \frac{3}{2} \right) + \gamma_p v || \frac{a}{v_{ta}} \frac{R B_i}{R_0 B} \right] ik_\theta \rho_s$$

$$+ \left\{ \frac{a}{L_{Ta}} \left[ \frac{z_a e}{T_a} \Phi_\star - \frac{M_a^2}{2 R_0^2} \left( R^2 - R(\theta_0)^2 \right) \right] \right\} ik_\theta \rho_s$$

$$+ M_a^2 \frac{a R(\theta_0)}{R_0^2} \frac{d R(\theta_0)}{d r} + M_a \gamma_p \frac{a}{v_{ta} R_0^2} \left( R^2 - R(\theta_0)^2 \right) \right\} ik_\theta \rho_s$$

Fold into streaming (diagonal)
Gyrokinetic equation for plasma species $a$

Typically, deuterium, some carbon, and electrons

\[
\frac{\partial \tilde{h}_a}{\partial \tau} - i \Omega_s X \tilde{h}_a - i (\Omega_\theta + \Omega_\xi + \Omega_d) \tilde{H}_a - i \Omega_s \tilde{\Psi}_a + \Omega_{NL}(\tilde{h}_a, \tilde{\Psi}_a) = C_a
\]

Nonlinearity

\[
\Omega_{NL}(\tilde{h}_a, \tilde{\Psi}_a) = \frac{a c_s}{\Omega_{cD}} \sum_{k'_\perp + k''_\perp = k_\perp} (b \cdot k'_\perp \times k''_\perp) \tilde{\Psi}_a(k'_\perp) \tilde{h}_a(k''_\perp)
\]
Gyrokinetic equation for plasma species $a$

Typically, deuterium, some carbon, and electrons

$$\frac{\partial \tilde{h}_a}{\partial \tau} - i \Omega_s X \tilde{h}_a - i \left( \Omega_\theta + \Omega_\xi + \Omega_d \right) \tilde{H}_a - i \Omega \tilde{\Psi}_a + \Omega_{NL}(\tilde{h}_a, \tilde{\Psi}_a) = \mathcal{C}_a$$

Cross-species collision operator

$$\mathcal{C}_a = \sum_b C_{ab}^L \left( \tilde{H}_a, \tilde{H}_b \right)$$

$$C_{ab}^L(\tilde{H}_a, \tilde{H}_b) = \frac{\nu_{ab}^D}{2} \frac{\partial}{\partial \xi} \left( 1 - \xi^2 \right) \tilde{H}_a + \frac{1}{\nu^2} \frac{\partial}{\partial \nu} \left[ \frac{\nu_{ab}^\parallel}{2} \left( \frac{\nu^4}{T_b} \frac{\partial \tilde{H}_a}{\partial \nu} + m_a \nu^5 \tilde{H}_a \right) \right]$$

$$-\tilde{H}_a k^2_a \rho_a^2 \frac{\nu^2}{4 \nu_{ta}^2} \left[ \nu_{ab}^D (1 + \xi^2) + \nu_{ab}^\parallel (1 - \xi^2) \right] + R_{mom}(\tilde{H}_b) + R_{ene}(\tilde{H}_b)$$

$\text{acc parallel loop}$
Sonic Transport Fluxes

These are inputs to an independent TRANSPORT CODE

**Particle flux** \[ \Gamma_a = \sum_{k} \left\langle \int d^3v \, \tilde{H}_a^* c_1 \tilde{\Psi}_a \right\rangle \]

**Energy flux** \[ Q_a = \sum_{k} \left\langle \int d^3v \, \tilde{H}_a^* c_2 \tilde{\Psi}_a \right\rangle \]

**Momentum flux** \[ \Pi_a = \sum_{k} \left\langle \int d^3v \, \tilde{H}_a^* c_3 \tilde{\Psi}_a \right\rangle \]
What do we solve for
5-dimensional distribution for every plasma species

Six-dimensional array (mapped into internal 2D array in CGYRO)

\[ H_a(k_x, k_y, \theta, \xi, v, t) \]

The spatial coordinates are

\[ k_x \rightarrow \text{radial wavenumbers} \]
\[ k_y \rightarrow \text{binormal wavenumbers} \]
\[ \theta \rightarrow \text{field-line coordinate} \]

The velocity-space coordinates are

\[ \xi = \frac{v_\parallel}{v} \rightarrow \text{cosine of the pitch angle} \in [-1, 1] \]
\[ v \rightarrow \text{speed} \in [0, \infty] \].
Visual representation of computational mesh

deuterium ($a = 1$) carbon ($a = 2$) electron ($a = 3$)

velocity-space mesh
ion-scale mesh
multiscale mesh
CGYRO optimized for challenging multiscale turbulence
COMPLETE REDESIGN of world-renowned GYRO code
Simulation underway on Titan (NCCS)

4986 nodes = 4986 Tesla K20X GPUs
Recent aggressive GPU optimization
Significant progress at Boulder Hackathon (Summer 2018)

• Huge thanks to Craig Tierney and Brent Leback for guidance!
Recent aggressive GPU optimization
Significant progress at Boulder Hackathon (Summer 2018)

- Huge thanks to Craig Tierney and Brent Leback for guidance!
- CGYRO design anticipated aggressive thread/GPU utilization
  1. Huge nonlinear convolution (Poisson bracket) via FFT
  2. Large nested loops remain after MPI distribution
Recent aggressive GPU optimization
Significant progress at Boulder Hackathon (Summer 2018)

• Huge thanks to Craig Tierney and Brent Leback for guidance!
• CGYRO design anticipated aggressive thread/GPU utilization
  1. Huge nonlinear convolution (Poisson bracket) via FFT
  2. Large nested loops remain after MPI distribution
• Took full advantage of GPUs with minimal changes to code logic
  1. Existing FFTW code was ported directly to cuFFT
  2. Nested loops accelerated by OpenACC without restructuring or invasive changes
  3. Implemented GPU-aware MPI (utilizes GPUDirect and GPU-Infiniband RDMA)
## CGYRO kernels

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Data dependence</th>
<th>Dominant operation</th>
<th>GPU approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>str</td>
<td>$k_x, \theta, [k_y]_2, [\xi, v, a]_1$</td>
<td>loop</td>
<td>OpenACC</td>
</tr>
<tr>
<td>field</td>
<td>Same as str</td>
<td>loop</td>
<td>OpenACC</td>
</tr>
<tr>
<td>coll</td>
<td>$[k_x, \theta]_1, [k_y]_2, \xi, v, a$</td>
<td>mat-vec multiply</td>
<td>OpenACC</td>
</tr>
<tr>
<td>nl</td>
<td>$k_x, k_y, [\theta, [\xi, v, a]_1]_2$</td>
<td>FFT</td>
<td>cuFFT</td>
</tr>
</tbody>
</table>
Scaling: CGYRO n101
V100-GPU Performance improvement over time
Scaling: CGYRO n101 (individual kernels)

V100-GPU Performance improvement over time

Candy/SC18/Nov 2018
Scaling: CGYRO n101 (individual kernels)

V100-GPU Performance improvement over time
Scaling: CGYRO n101
GPU versus Skylake and KNL
Scaling: CGYRO n103 – much larger case
Skylake versus 3 different GPUs
Scaling: CGYRO n103 – much larger case

Skylake versus 3 different GPUs
Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, or those of the European Commission.