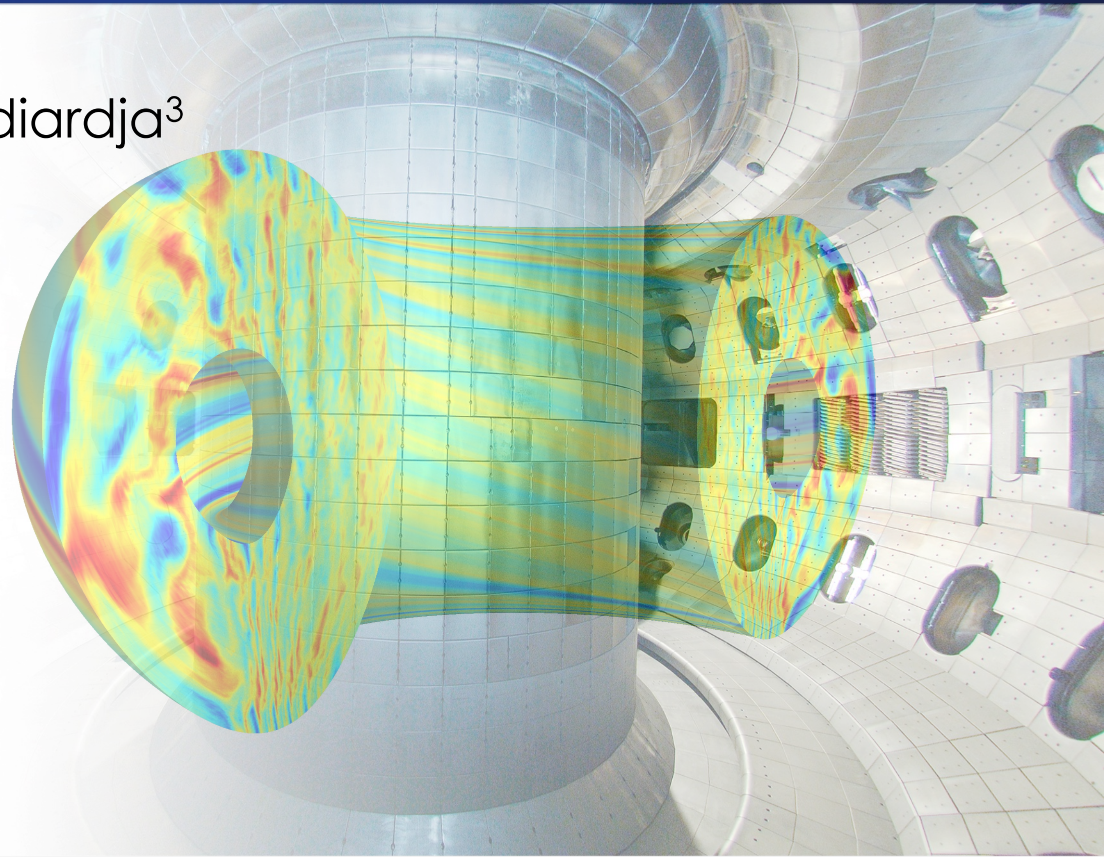
Discovering New Clues to Improving Confinement in Fusion Burning Plasmas

By **E.A. Belli**[1], J. Candy[1],
I. Sfiligoi[2], J. Garcia[4], R. Budiardja[3]

[1] **General Atomics**
[2] UCSD San Diego Supercomputing Center
[3] CEA
[4] ORNL NCCS

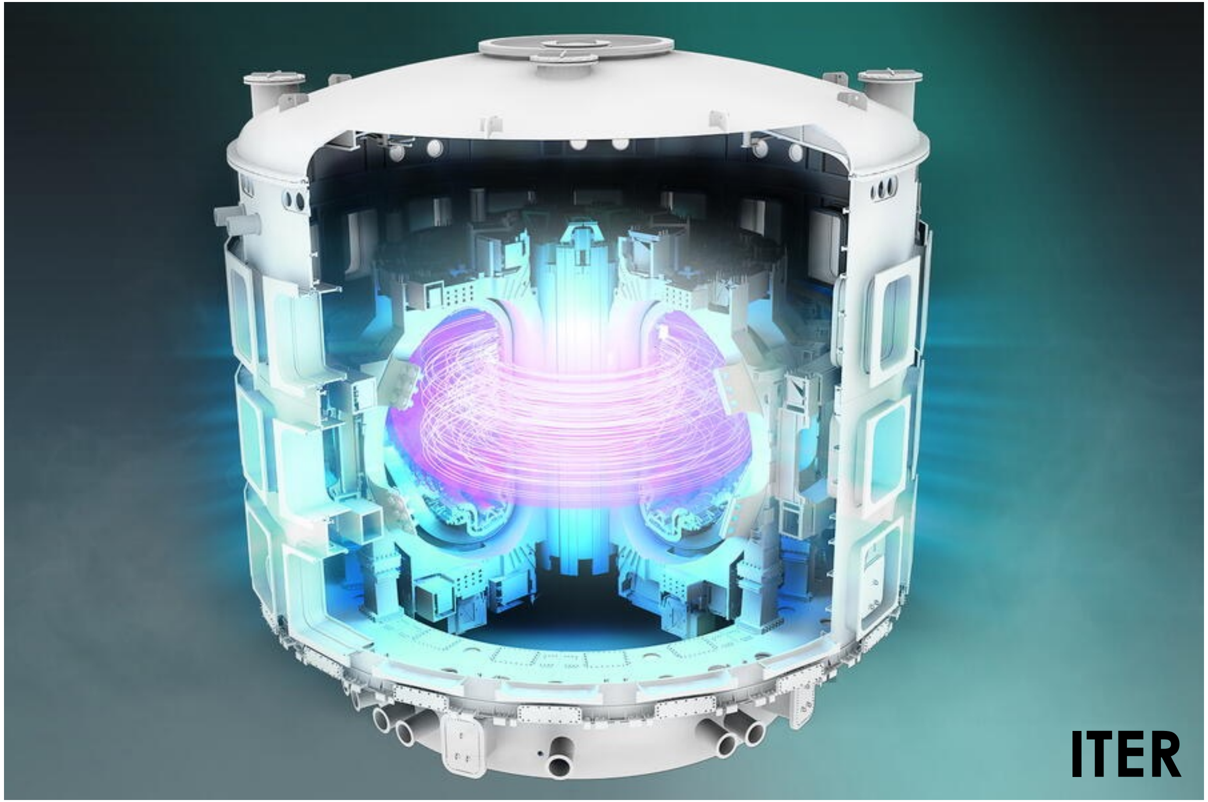


Presented at the
OLCF User Group Meeting

Aug 2025

Supported by U.S. DOE Grants DE-FC02-06ER54873 and DE-SC0024425

Magnetic confinement fusion holds a promising solution as a nearly limitless source of energy for the future.



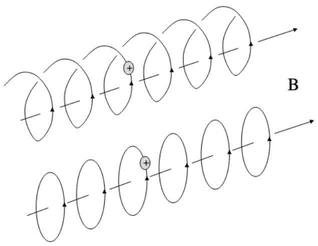
GENERAL ATOMICS

A quantitatively accurate simulation capability for plasma turbulence is essential to optimize and develop scenarios for ITER and FPPs.

- Confinement is limited by **slow particle and energy losses** due to **turbulence** driven by unstable plasma waves.

- **CGYRO** is an **Eulerian-based gyrokinetic code** that computes plasma turbulent transport.
  - Extrapolate from experiments & understand turbulence regimes expected in conditions unique to burning plasmas
  - GK eqn derived from **6D Fokker-Planck equation**

$$\frac{\partial F}{\partial t} + \vec{\mathrm{v}} \cdot \nabla_x F + \frac{q}{m}\left(\vec{E} + \frac{\vec{\mathrm{v}}\times\vec{B}}{c}\right) \cdot \nabla_{\mathrm{v}} F = \left(\frac{\partial F_a}{\partial t}\right)_{coll}$$



  - **Gyro-average** reduces dimensionality: tracks motion of "gyrocenters" (rings of charge) along B
  - **EM fields are dynamically coupled with the distribution function** (GK Maxwell eqns)

- **The Science**:
  - Critical need for simulations of D-T burning plasmas

- **The Computation**:
  - Challenges of tokamak edge “pedestal” turbulence simulations
  - CGYRO: A scalable, GPU-optimized (spectral) GK solver for multiscale turbulence

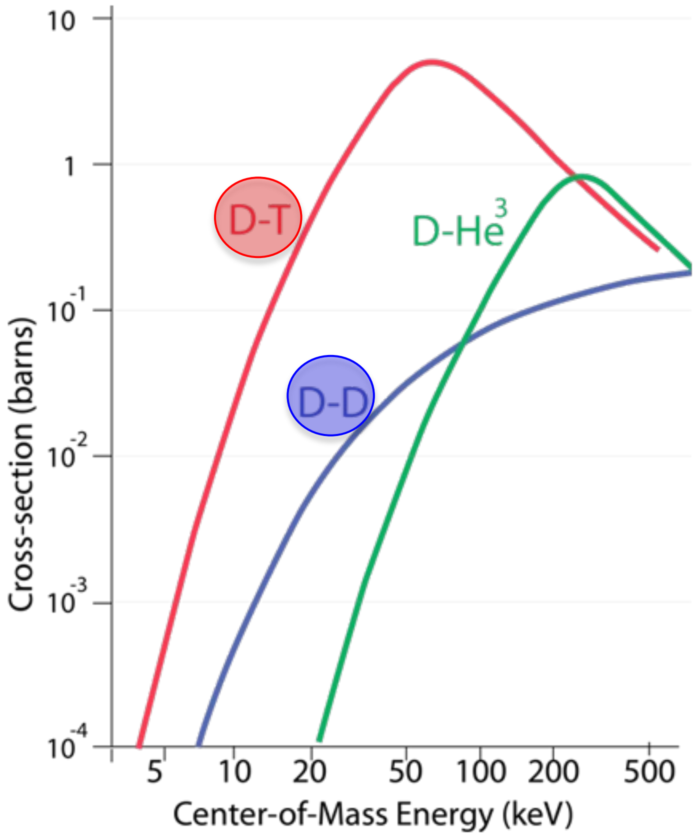
- **The Discoveries**:
  - Improving energy confinement in D-T fusion plasmas

- **The Science**:
  - Critical need for simulations of D-T burning plasmas

- **The Computation**:
  - Challenges of tokamak edge “pedestal” turbulence simulations
  - CGYRO: A scalable, GPU-optimized (spectral) GK solver for multiscale turbulence

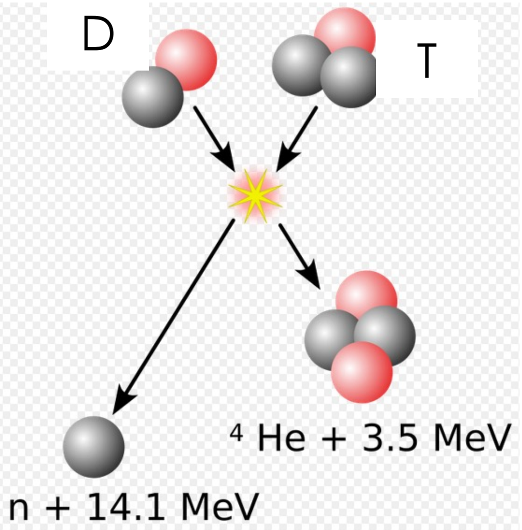
- **The Discoveries**:
  - Improving energy confinement in D-T fusion plasmas

Understanding scaling of confinement time with hydrogenic isotope is important in moving toward reactor-relevant D-T plasmas.

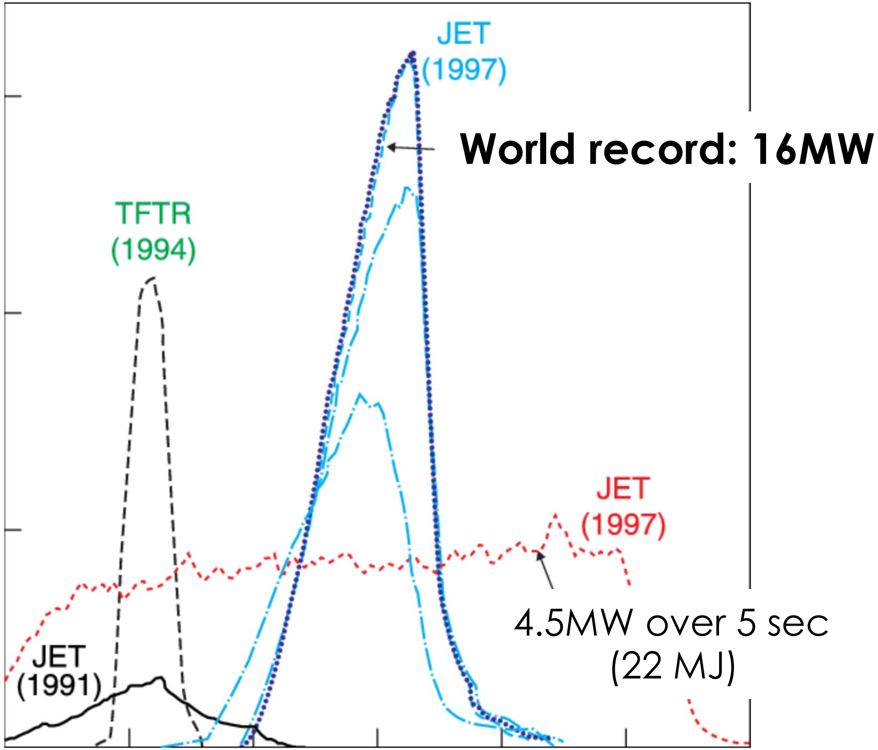


ITER Operational Phases:

- 1st phase: **H**/He
- **2nd phase: D**
- **3rd phase: 50:50 D-T**



Understanding scaling of confinement time with hydrogenic isotope is important in moving toward reactor-relevant D-T plasmas.



Fusion Power vs. time

ITER Operational Phases:

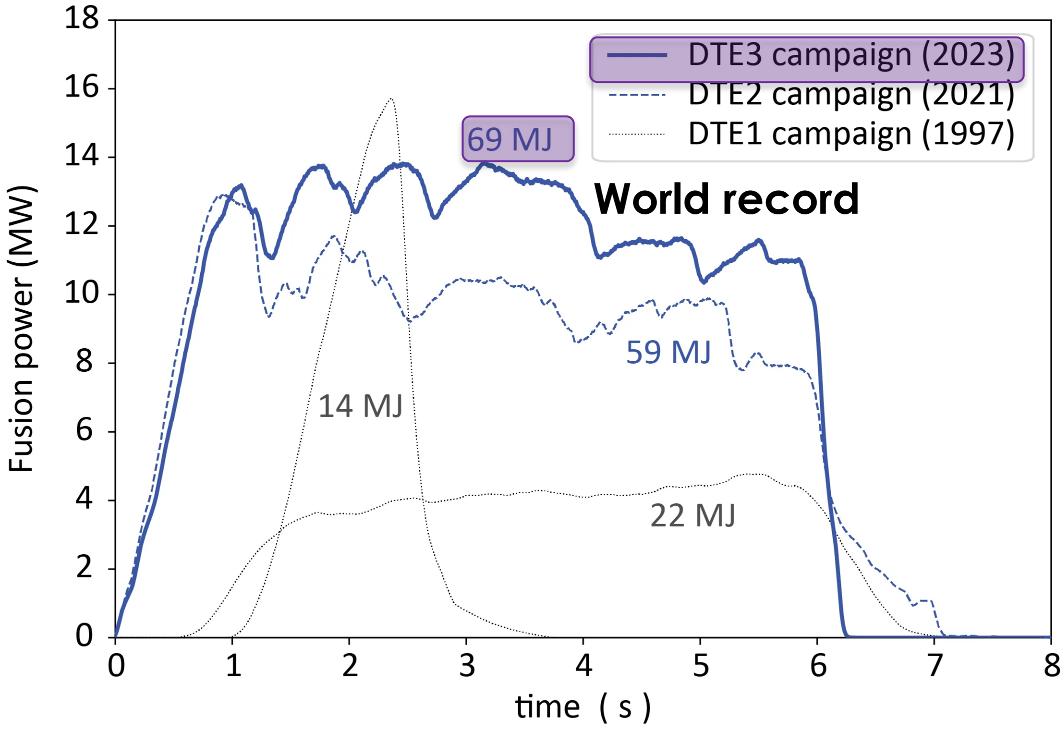
- **1st phase: H**/He
- **2nd phase: D**
- **3rd phase: 50:50 D-T**

D-T Tokamak Experiments:

- JET (90:10) 1991
- **TFTR 1993-1997**
- **JET DTE1 1997**

GENERAL ATOMICS

Modeling is playing an essential role in planning the ramp up stages in ITER to reactor-level D-T.



ITER Operational Phases:

- **1st phase: H**/He
- **2nd phase: D**
- **3rd phase: 50:50 D-T**

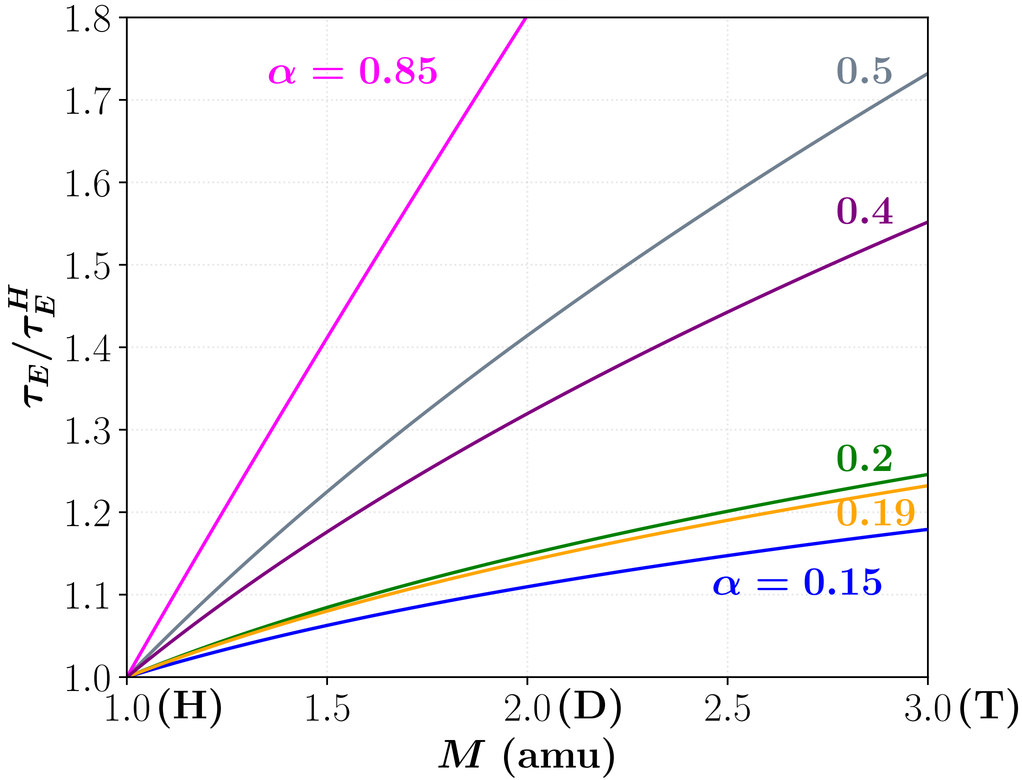
D-T Tokamak Experiments:

- JET (90:10) 1991
- TFTR 1993-1997
- JET DTE1 1997
- **JET DTE2/3 2021-2023**

GENERAL ATOMICS

The Mystery of the "Isotope Effect"

$\tau_E \sim M^{\alpha}$



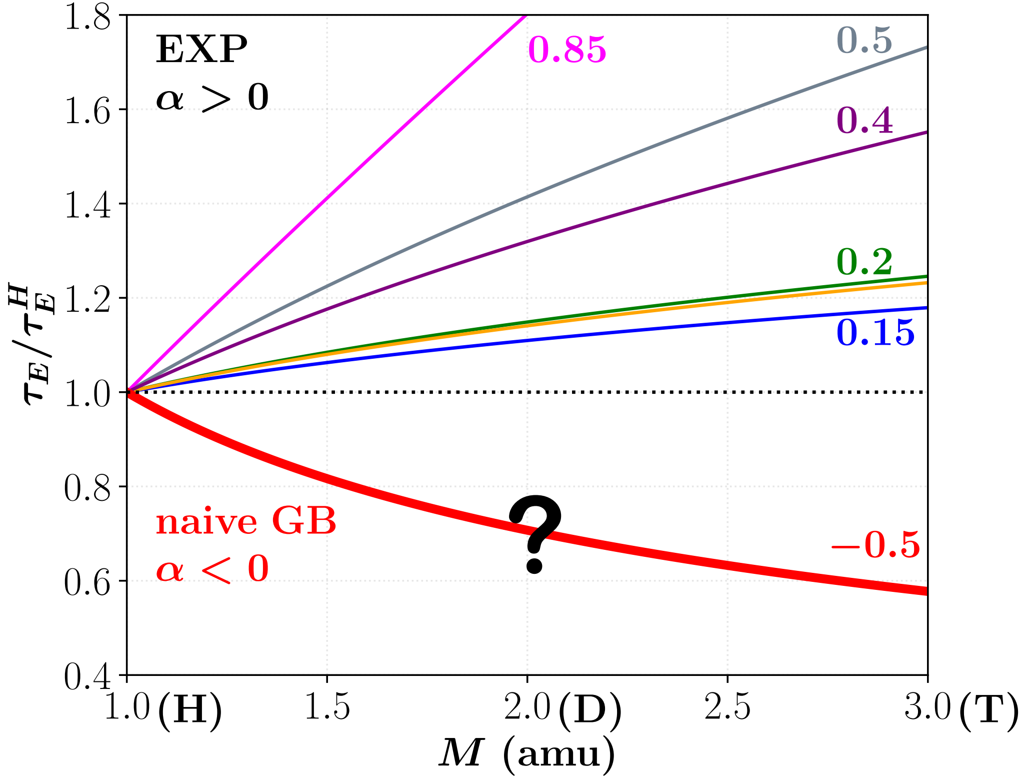
Experiment

$$\tau_E^H < \tau_E^D < \tau_E^{DT}$$

- TFTR supershot
- ASDEX ohmic
- JET-ILW H-mode
- ITER97-L
- ITERH-98P(y, 2)
- JET-ILW L-mode

The Mystery of the "Isotope Effect"

$$\tau_E \sim M^{\alpha}$$



Experiment

$$\tau_E^H < \tau_E^D < \tau_E^{DT}$$

Theory: Naive GB Scaling

$$\chi_i \sim \frac{\Delta x^2}{\Delta t} \sim \frac{\rho_i^2}{(a/v_{ti})} \sim \sqrt{m_i}$$

$$\tau_E \sim a^2/\chi_i$$

$$\tau_E^H > \tau_E^D > \tau_E^{DT}$$

GENERAL ATOMICS

Unraveling the mystery of the isotope effect with CGYRO: Non-adiabatic electron physics leads to “reversal” from core to edge

Tokamak core

$$\chi_i = c_0 \chi_{GBi} = c_0 \chi_{GB} \sqrt{\boldsymbol{m_i}}$$

$$\chi_H < \chi_D < \chi_{DT} \qquad \boldsymbol{\tau_E^H > \tau_E^D > \tau_E^{DT}}$$

⬇

Tokamak edge

$$\chi_i = \tilde{c}_0 \left(\frac{\boldsymbol{m_e}}{\boldsymbol{m_i}} \right) \chi_{GBi}$$

$$\chi_H > \chi_D > \chi_{DT} \qquad \boldsymbol{\tau_E^H < \tau_E^D < \tau_E^{DT}}$$

- **The Science**:
  - Critical need for simulations of D-T burning plasmas

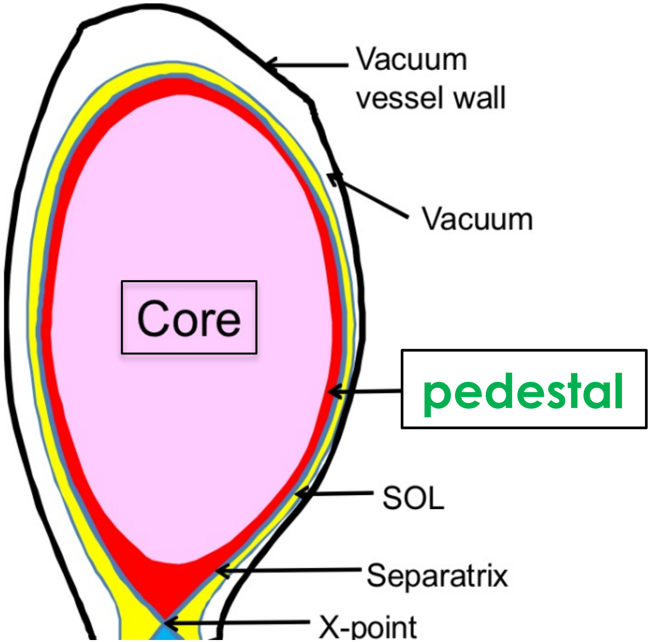
- **The Computation**:
  - Challenges of tokamak edge “pedestal” turbulence simulations
  - CGYRO: A scalable, GPU-optimized (spectral) GK solver for multiscale turbulence

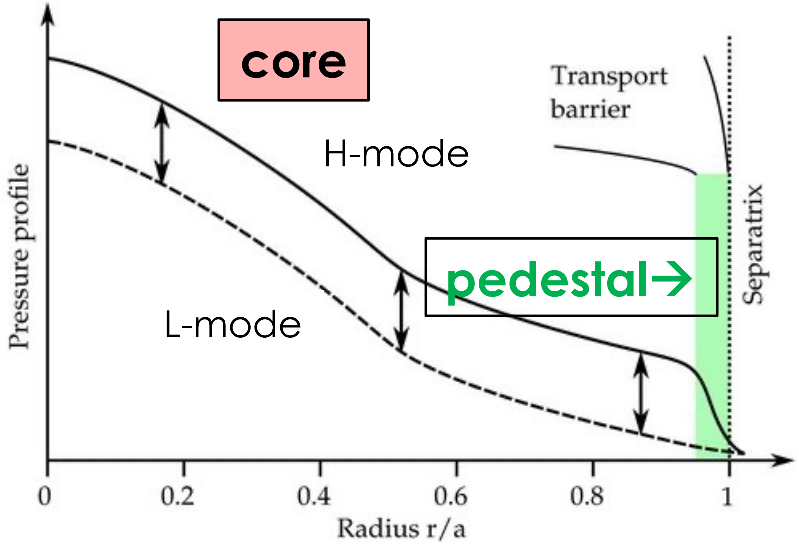
- **The Discoveries**:
  - Improving energy confinement in D-T fusion plasmas

GENERAL ATOMICS

The plasma edge/pedestal is known to play a key role in determining global energy confinement, but is difficult to simulate.

- High-confinement (H-mode) is characterized by steep pressure gradients
  → "**pedestal**" structure at the edge
- Understanding turbulence in the pedestal can help develop operating regimes for **optimal confinement and fusion performance**.



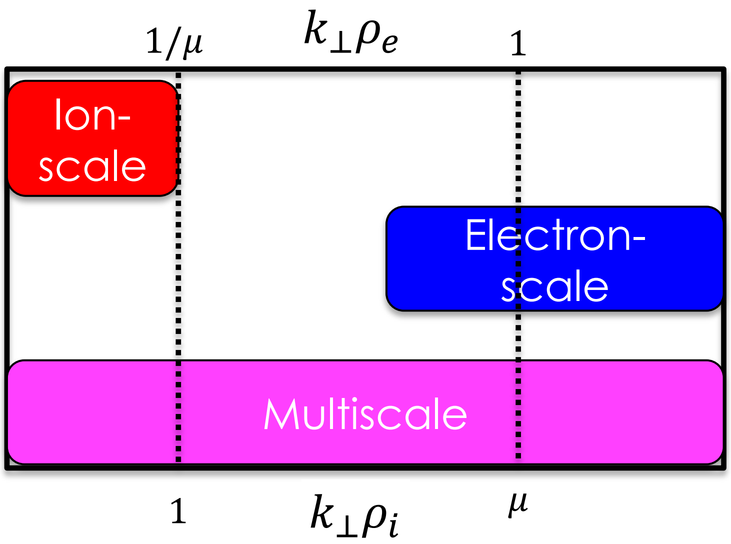


GENERAL ATOMICS

In comparison to the core, GK simulations in the pedestal are far more challenging due to the multiscale nature of turbulence.

Steep gradients drive multiple instabilities across broad range of spatial scales

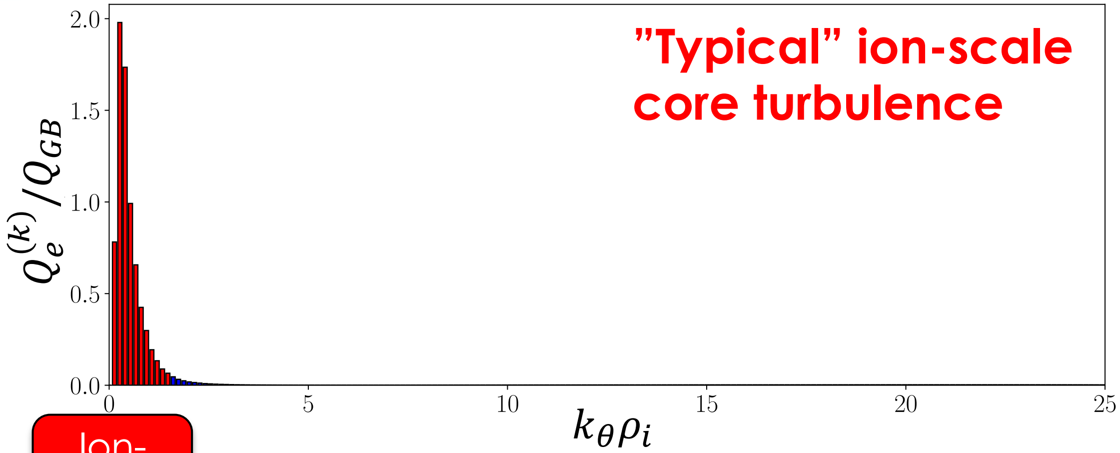
- **Ion-scales ($k_\perp \rho_i < 1$)**
- **Electron-scales ($k_\perp \rho_e \sim 1$)**
- Typically, only **single-scale** GK simulations are done
- **Multiscale** needed for H-mode pedestal turbulence
- **Complex nonlinear cross-scale mode coupling** requires extremely fine mesh in real space

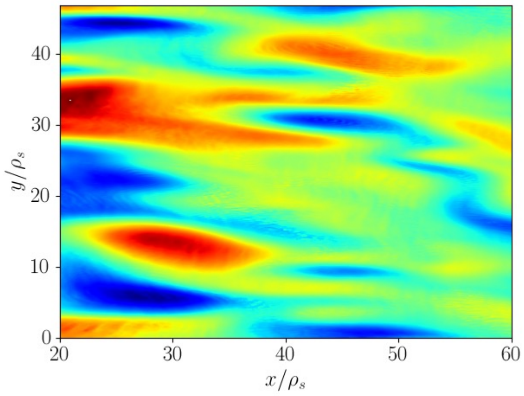


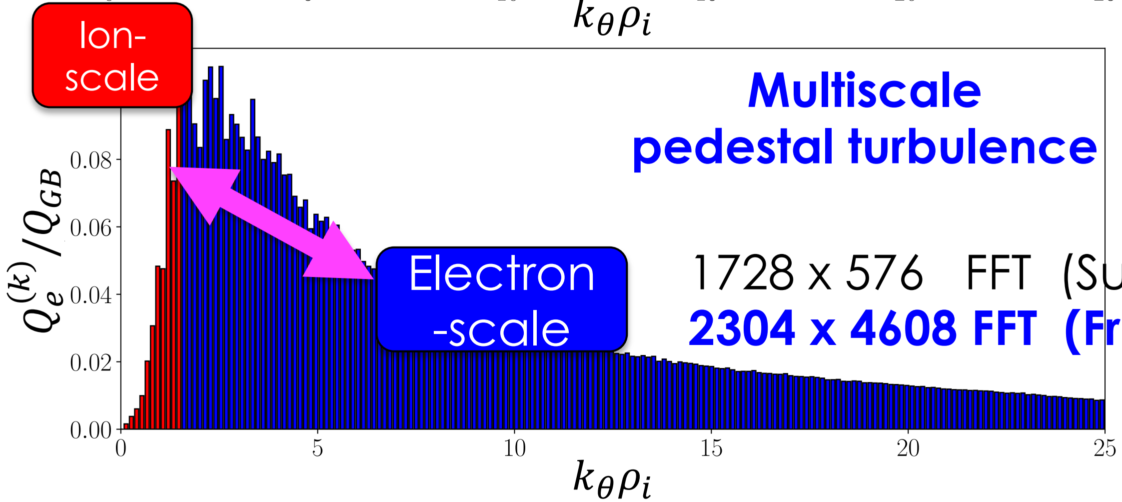
$$\mu = \sqrt{\frac{m_i}{m_e}} \sim 60$$

GENERAL ATOMICS

Multiscale GK pedestal simulations require leadership-scale computing resources and highly optimized solvers.

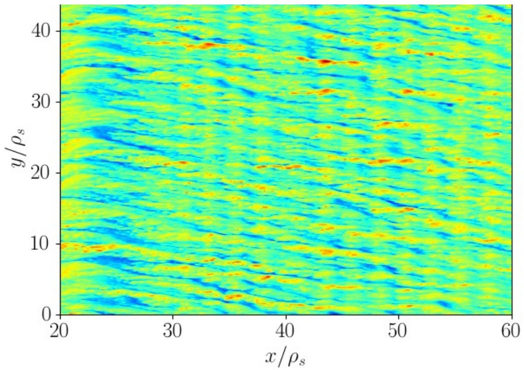






1728 x 576 FFT (Summit)

2304 x 4608 FFT (Frontier)



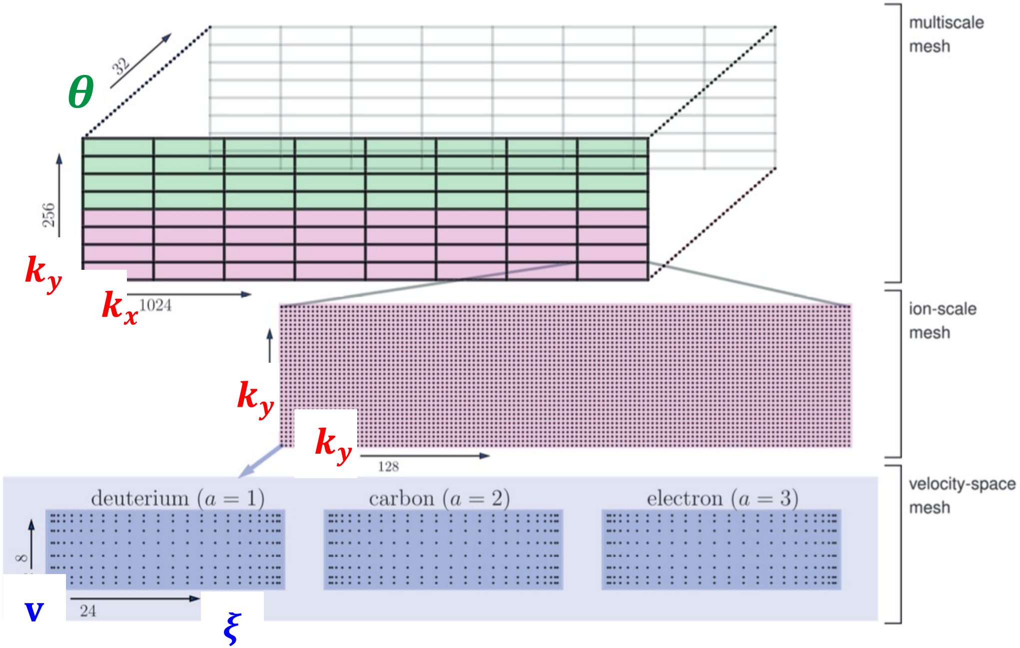
Belli, PPCF, 2023/2024

Advances in supercomputing have accelerated multiscale simulation.

	Physics	System	Core hrs used	System Rpeak (PF)
GYRO 2007	Reduced m_i/m_e	OLCF-1 Phoenix (2004)		0.018 Terascale
		OLCF-2 Jaguar (2005-10)		0.025→2.3 Petascale
GYRO 2015	First full mass (core)	NERSC Edison	125M CPU hrs	2.5
		OLCF-3 (GPU) Titan (2012)		27
CGYRO 2022/23	First full mass (pedestal)	OLCF-4 (GPU) Summit (2018)	250k GPU node hrs	200
CGYRO 2024/25	EM full mass (pedestal)	OLCF-5 (GPU) Frontier (2022)	100k GPU node hrs	1685 Exascale

CGYRO: A scalable, GPU-optimized (spectral) GK solver for multiscale turbulence

CGYRO implements **highly efficient spectral/pseudo-spectral** numerical schemes optimized for multiscale.



x	Radial	spectral
y	Binormal	spectral
θ	Poloidal	Finite diff

ξ	Pitch angle	pseudospectral
v	Velocity	psuedospectral

Candy, JCP, 2016

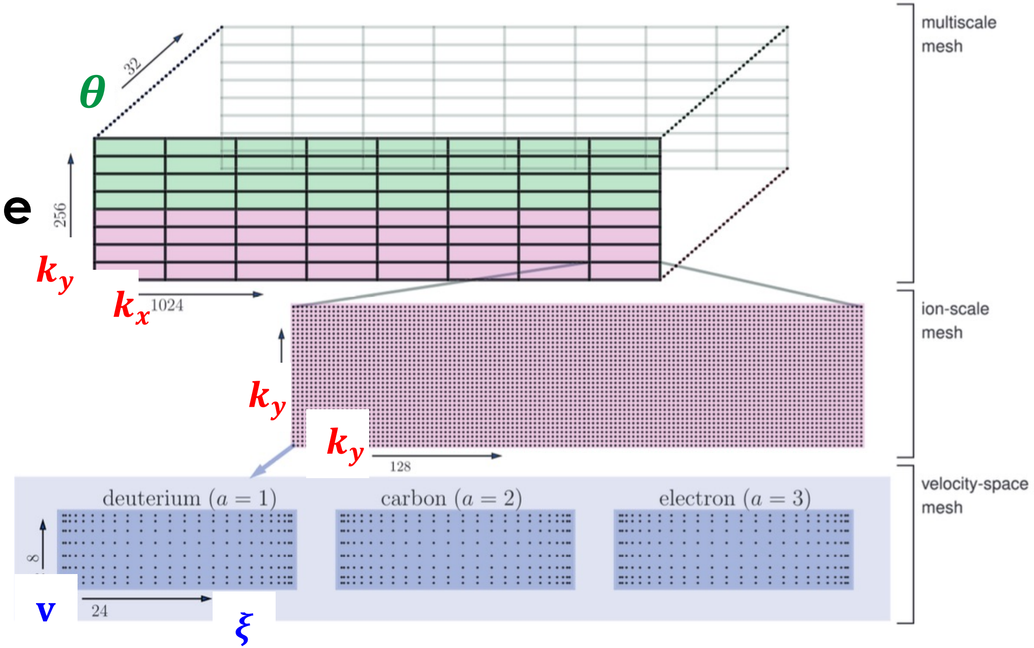
GENERAL ATOMICS

CGYRO: A scalable, GPU-optimized (spectral) GK solver for multiscale turbulence

Main numerical difficulties of scalable multiscale GK simulation:

- **High dimensionality** (6D grid): 3D spatial + 2D velocity + 1D species
  - Allows for massive parallelism but memory intensive

x	**Radial**	**spectral**
y	**Binormal**	**spectral**
θ	**Poloidal**	**Finite diff**



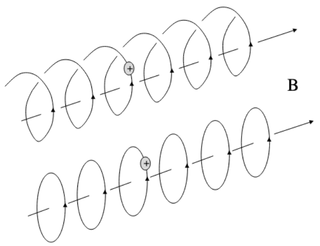
ξ	**Pitch angle**	**pseudospectral**
v	**Velocity**	**psuedospectral**

CGYRO: A scalable, GPU-optimized (spectral) GK solver for multiscale turbulence

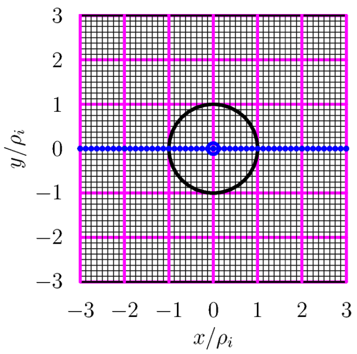
$$\frac{\partial H_a}{\partial \tau} - \frac{e_a}{T_a}\frac{\partial \phi}{\partial \tau} + L(H_a, \phi\,) + NL(H_a, \phi\,) + C_{ab}(H_a, H_b) = 0$$

Main numerical difficulties of scalable multiscale GK simulation:

- **Gyro-averaging**
  - To evolve position & velocity of gyrocenters, need gyro-avg transformation for EM fields and charge density
  - **Fully spectral in (x, y)** provides most efficient & accurate evaluation



gyrocenters



$J_0(k_\perp \rho)$

GENERAL ATOMICS

CGYRO: A scalable, GPU-optimized (spectral) GK solver for multiscale turbulence

$$\frac{\partial H_a}{\partial \tau} - \frac{e_a}{T_a}\frac{\partial \phi}{\partial \tau} + L(H_a, \phi) + \boldsymbol{NL(H_a, \phi)} + C_{ab}(H_a, H_b) = 0$$

Main numerical difficulties of scalable multiscale GK simulation:

- **Nonlinearity** (mode-mode coupling)
  - Convolution over 2D wavenumber space
  - Evaluation of nonlinear term on GPUs (**cuFFT/rocFFT**) ensures maximum performance, scalability, and portability.

$$NL:\ \frac{c}{B}\sum_{\vec{k}'_\perp + \vec{k}''_\perp = \vec{k}_\perp}\left[\vec{b}\cdot\left(\vec{k}'_\perp \times \vec{k}''_\perp\right)\right]\phi\left(\vec{k}'_\perp\right) H_a\left(\vec{k}''_\perp\right)$$

CGYRO: A scalable, GPU-optimized (spectral) GK solver for multiscale turbulence

$$\frac{\partial H_a}{\partial \tau} - \frac{e_a}{T_a}\frac{\partial \boldsymbol{\phi}}{\partial \tau} + L(H_a, \boldsymbol{\phi}) + NL(H_a, \boldsymbol{\phi}) + C_{ab}(H_a, H_b) = 0$$

Main numerical difficulties of scalable multiscale GK simulation:

- **Field solve** (integro-differential)
  - EM fields dynamically coupled with the distribution functions (GK Maxwell eqns)

$$\phi \sim \sum_a \int d^3\mathrm{v}\, f(H_a)$$

GENERAL ATOMICS

Why Eulerian? PIC vs. Eulerian GK Solvers

- **PIC/Lagrangian**:
  - First nonlinear GK codes were PIC; **Easier to implement**
  - Easier to implement **complex geometry**, particularly for edge
  - Can be subject to numerical noise
  - **Field solve is difficult**: Particles move across gridpoints, then must be "deposited" onto the field mesh using a distance weighting, an intrinsically diffusive process (a particle diffuses to differing meshpts)
- **Eulerian**:
  - Can be **hand-tuned** in each dimension
  - **Exact field-distribution coupling**: Treats fields and distributions on the same grid

CGYRO has the first pseudo-spectral implementation of the collision operator in a GK code.

$$\frac{\partial H_a}{\partial \tau} - \frac{e_a}{T_a}\frac{\partial \phi}{\partial \tau} + L(H_a,\phi) + NL(H_a,\phi) + C_{ab}(H_a,H_b) = 0$$

$$H_a(x, y, \theta, \xi, \mathrm{v})$$

ξ	Pitch angle	pseudospectral
v	Velocity	psuedospectral

- **Multi-species collisions**
  - 2D diffusion
  - Requires **implicit** time advance
    - **Trade memory intensity for lower compute**
      - Compute matrix once per sim
      - Accounts for > 10x size of all memory buffers combined
      - Complex matrix-vector multiply fast on GPUs

$$\begin{bmatrix} H_1^+ \\ H_2^+ \\ \vdots \\ H_{Na}^+ \end{bmatrix} = \mathbb{M} \begin{bmatrix} H_1^- \\ H_2^- \\ \vdots \\ H_{Na}^- \end{bmatrix}$$

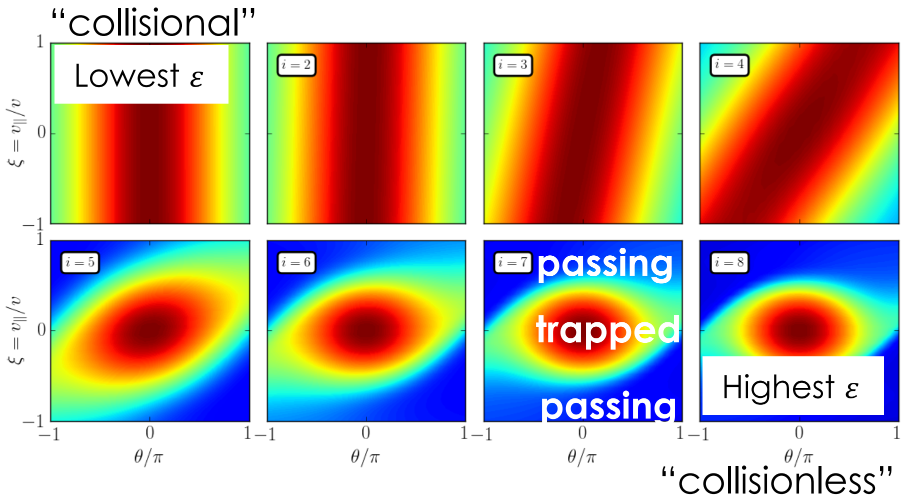
$\mathrm{Rank}(\mathbb{M}) = N_\xi N_v N_a$

The GK solve exhibits highly-collisional behavior at the lowest energies, transitioning to collisionless behavior at high energies.

$$\frac{\partial H_a}{\partial \tau} - \frac{e_a}{T_a}\frac{\partial \phi}{\partial \tau} + L(H_a, \phi\,) + NL(H_a, \phi\,) + \boldsymbol{C_{ab}(H_a, H_b)} = 0$$

Main numerical difficulties of scalable multiscale GK simulation:

- **Multi-species collisions**
  - spans equivalent of factor of 10^5 in effective collision frequency



GENERAL ATOMICS

CGYRO uses a spatial discretization & array distribution scheme that targets scalability on next-generation HPC systems.

Operator splitting for time integration

$$\frac{\partial h_a}{\partial \tau} + A(H_a, \Psi_a) + B(H_a, \Psi_a) = 0$$

Collisionless +nonlinear step:

$$\frac{\partial h_a}{\partial \tau} + A(H_a, \Psi_a) = 0 \qquad H_a(x, y, \theta, \xi, \mathrm{v})$$

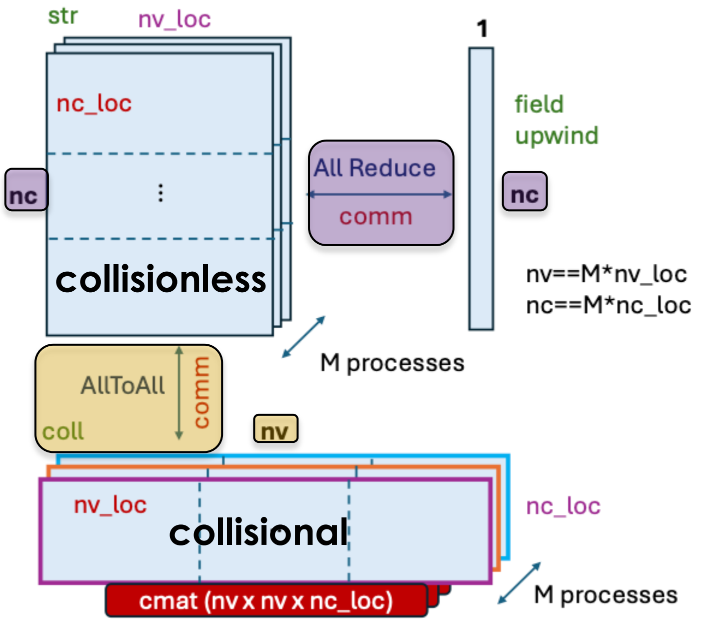
Collisional :

$$\frac{\partial h_a}{\partial \tau} + B(H_a, \Psi_a) = 0 \qquad H_a(x, y, \theta, \xi, \mathrm{v})$$

All compute kernels are ported to GPUs (OpenACC or OpenMP GPU-offloading)

Transferring data between array layouts requires frequent transpose operations, which can be communication heavy.

Kernel	Data dependence	Dominant operation
Collisionless	nc k_x, θ $[k_y]_1, [nv]_2$	Loop (lin)
Nonlinear	nt k_x, k_y $[\theta, [nv]_2]_1$	FFT
Collisional	nv $[k_y]_1, [nc]_2$	Mat-vec

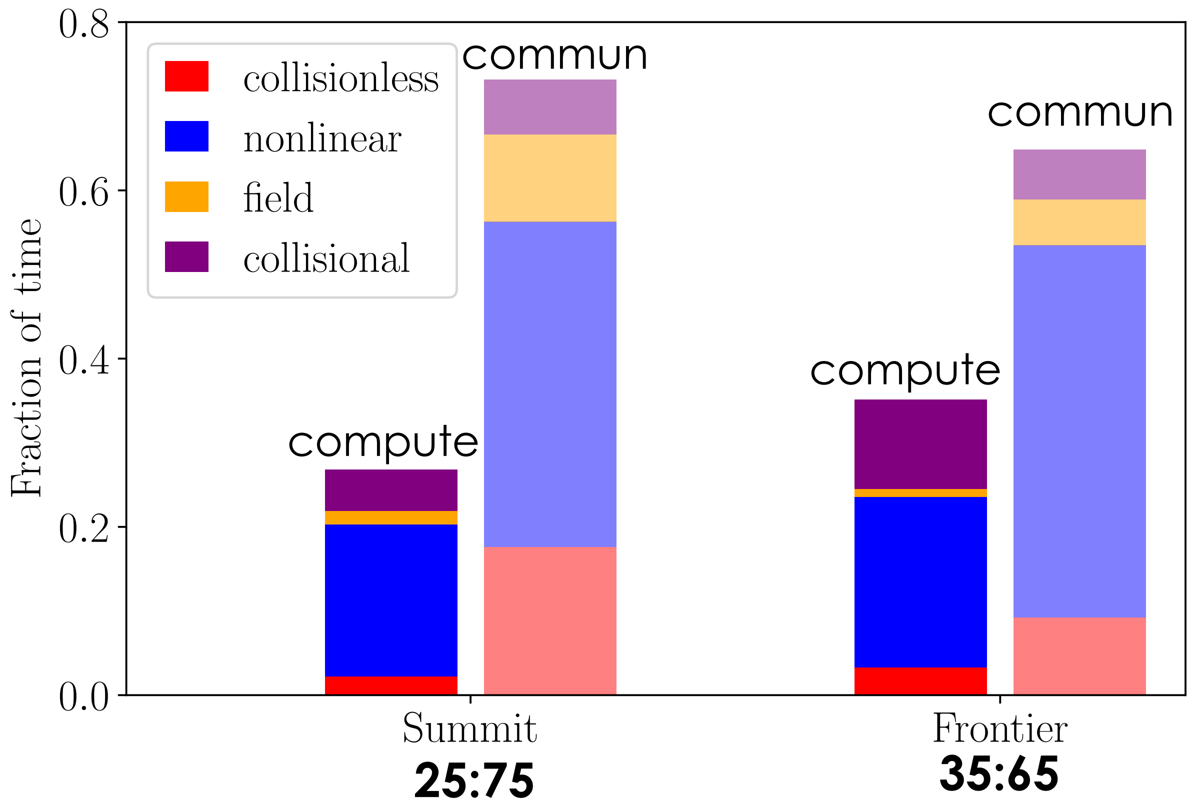


- Critical use of **GPU-aware MPI** minimizes cost of memory movement
  →30-40% reduction in comm timing

Sfiligoi, PEARC, 2023; ICPP, 2025

GENERAL ATOMICS

CGYRO is communication heavy, but CGYRO's large multiscale mesh can scale to large number of CPUs/GPUs.

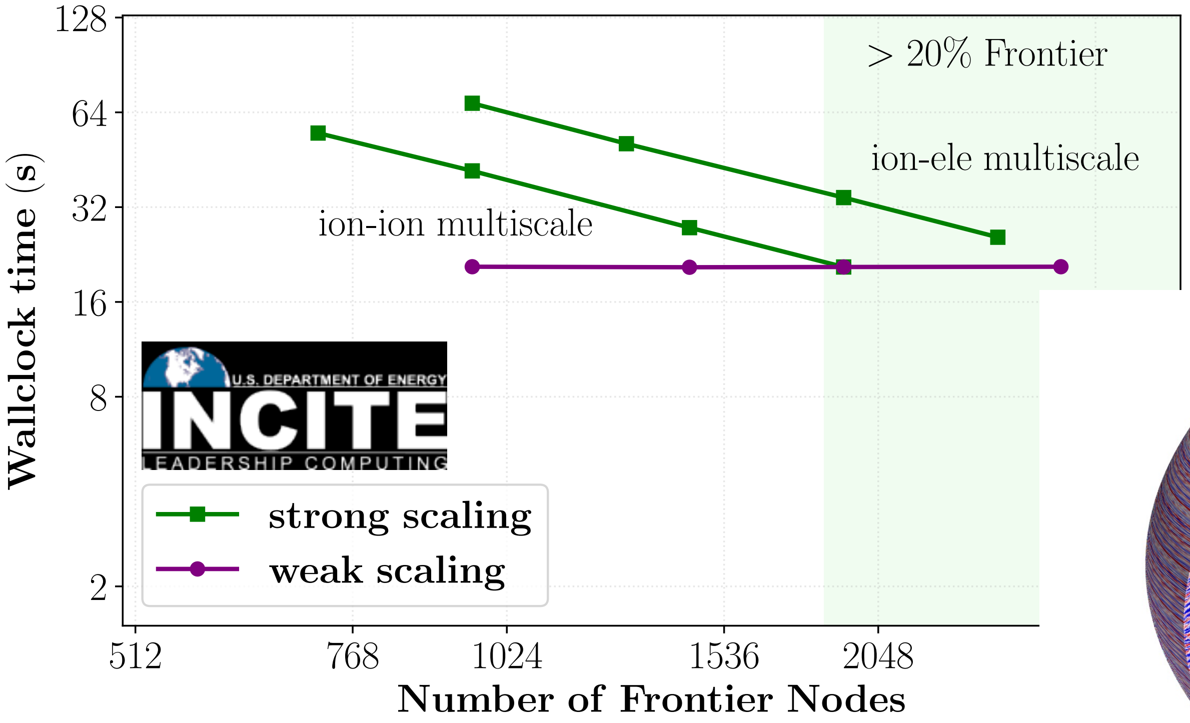


Reflects **high absolute performance of GPU** compute, rather than poor interconnect

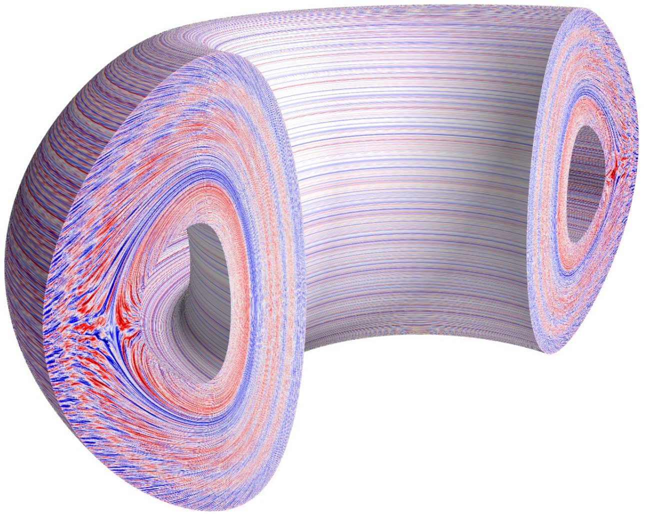
CGYRO capability-scale multiscale simulation:
1920 Frontier nodes
920 Summit nodes

GENERAL ATOMICS

CGYRO multiscale simulation is well-suited to capability simulation on accelerated systems like Frontier.



Requires >30 TB of GPU memory



Weak scaling: varying species

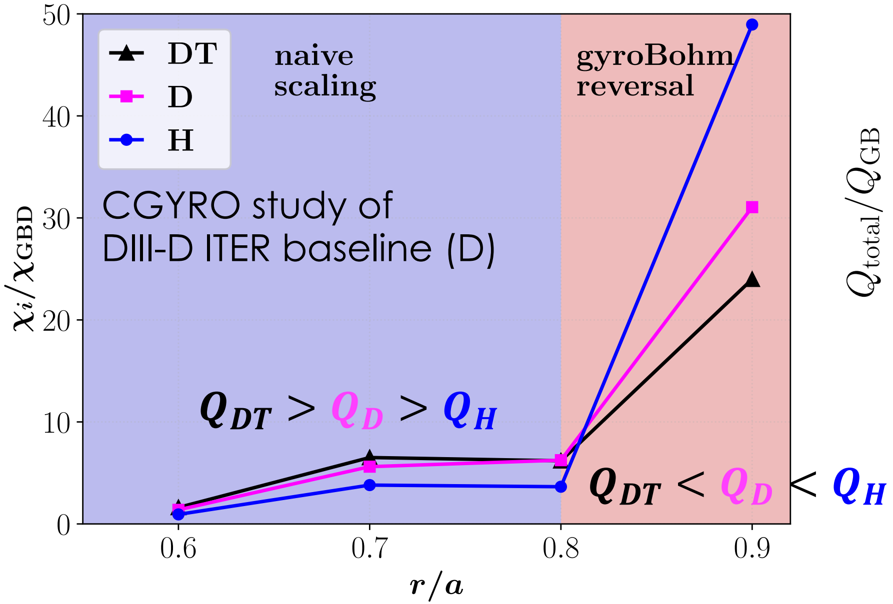
GENERAL ATOMICS

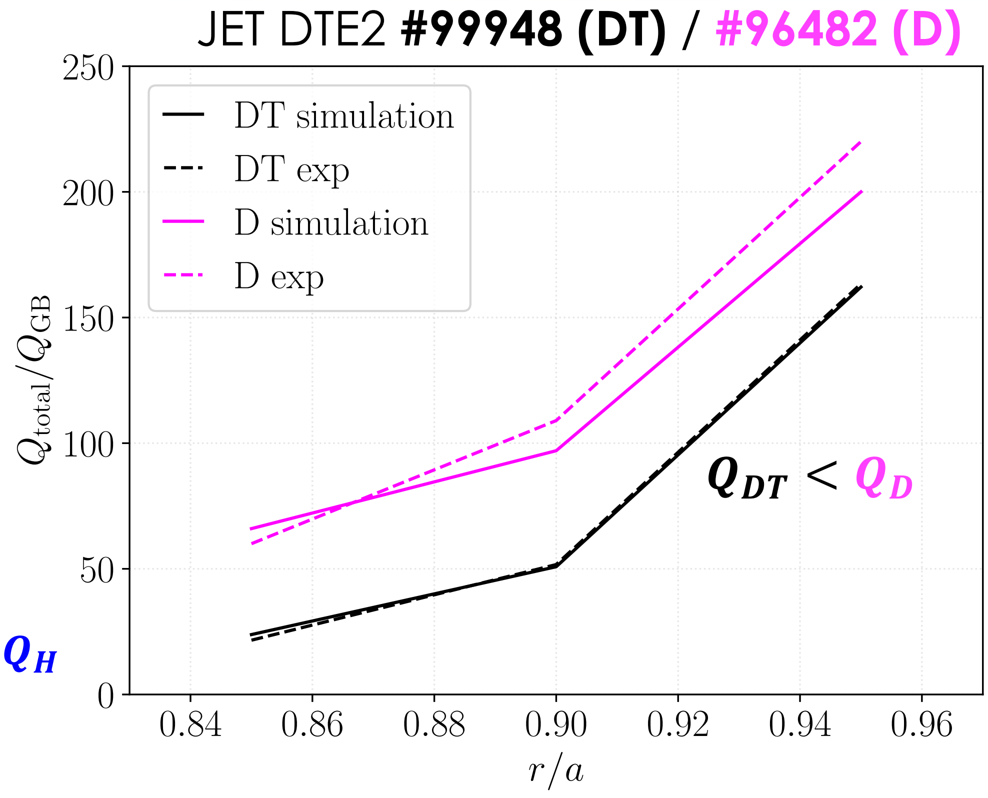
- **The Science**:
  - Critical need for simulations of D-T burning plasmas

- **The Computation**:
  - Challenges of tokamak edge "pedestal" turbulence simulations
  - CGYRO: A scalable, GPU-optimized (spectral) GK solver for multiscale turbulence

- **The Discoveries**:
  - Improving energy confinement in D-T fusion plasmas

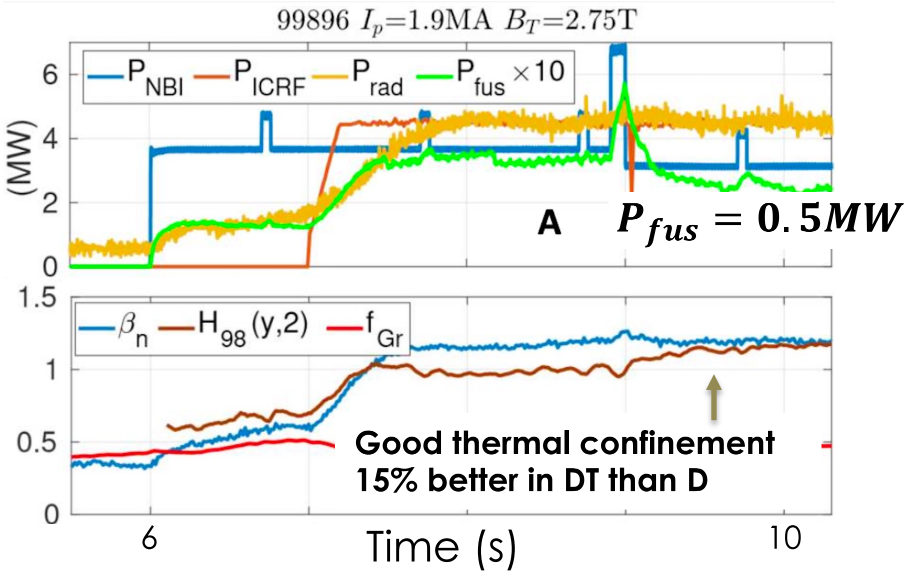
Isotope effect: CGYRO finds reversal of naive GB scaling in the pedestal, in agreement with better confinement in DT vs. D.

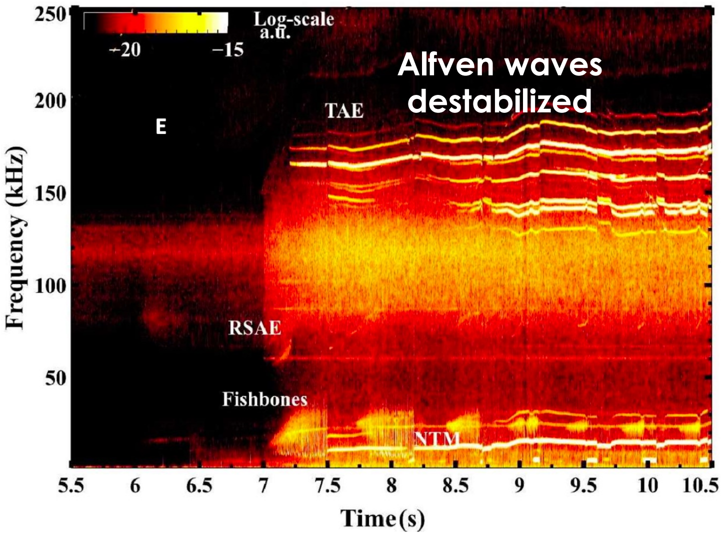




Belli, PRL, 2020; Kim, NF, 2023

JET DTE3 experiments suggest reactor-level operation with D-T & alphas may provide access to surprising new regimes with high confinement.

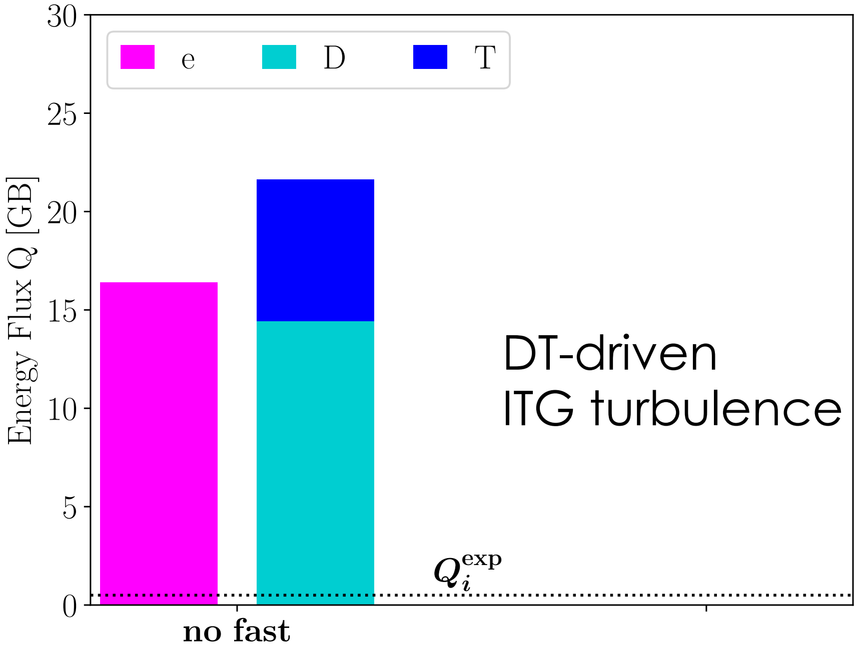




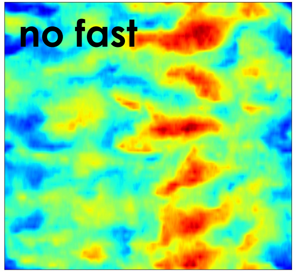
This is a unique and timely opportunity to assess predictability of D-T plasmas close to ITER conditions.

Garcia, Nature Comm, 2024

CGYRO simulations showed the importance of highly energetic ions to reduce turbulence and enhance confinement.

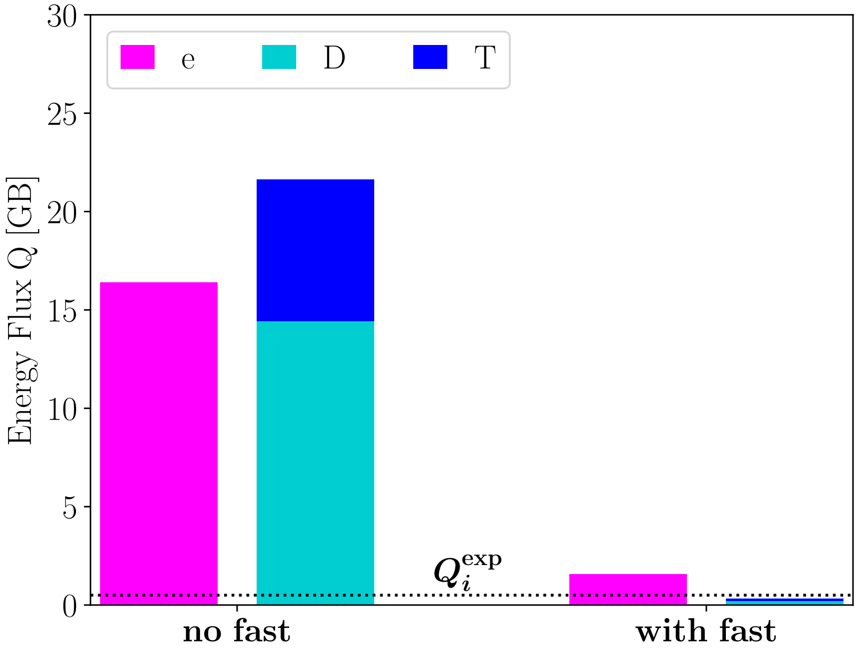


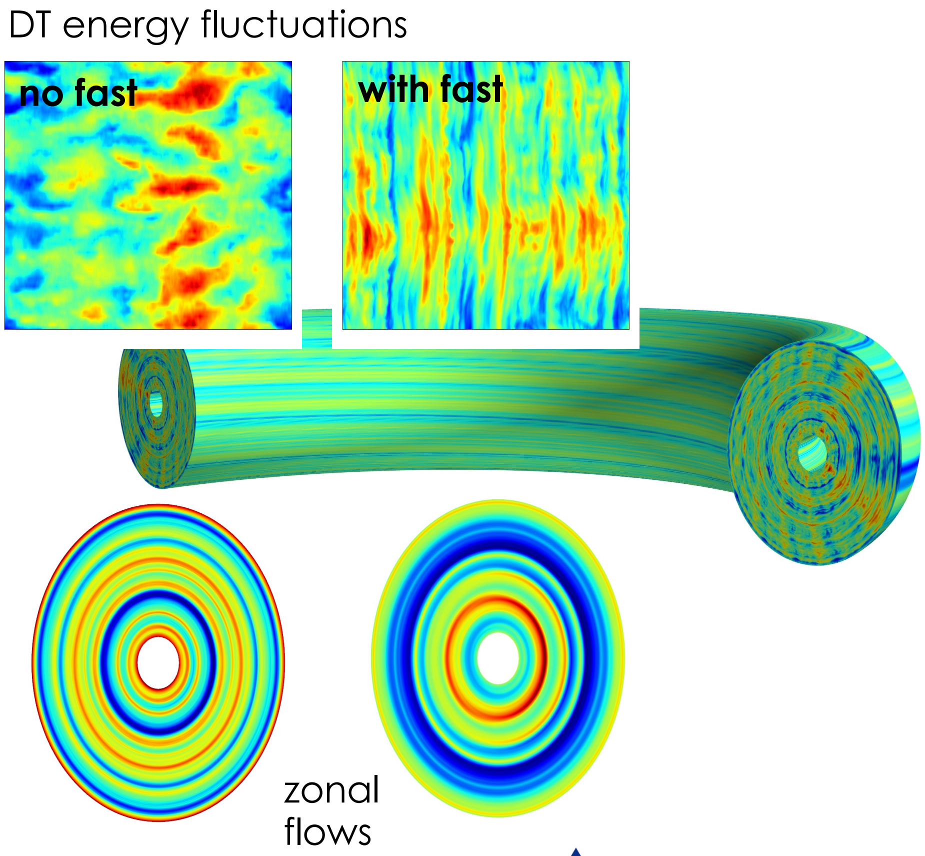
DT energy fluctuations



GENERAL ATOMICS

CGYRO simulations showed the importance of highly energetic ions to reduce turbulence and enhance confinement.





Simulations can provide a truly predictive capability for D-T burning plasmas with MeV alphas that are unique to fusion reactors.

Summary

- New experiments are providing a unique opportunity to assess **predictability of D-T plasmas w/ MeV alphas** in conditions expected in ITER.
- Burning fusion plasmas exhibit a broad (multiscale) spacetime spectrum of weak turbulence that requires high numerical resolution to simulate.
- **CGYRO is a scalable, GPU-optimized GK solver for multiscale turbulence**
  - Challenging due to high dim, nonlinearity, dynamic field coupling
  - Employs highly efficient **spectral/pseudo-spectral** numerical schemes
  - Nonlinear evaluation on GPUs (**cu/rocFFT**) ensures max performance and scalability
  - **GPU-aware MPI** critical for minimizing cost of memory movement
- **Using CGYRO sims on Frontier to understand turbulence regimes in D-T plasmas is essential to develop scenarios for next-generation FPPs with optimal fusion performance.**