HPC Challenges in Plasma Physics

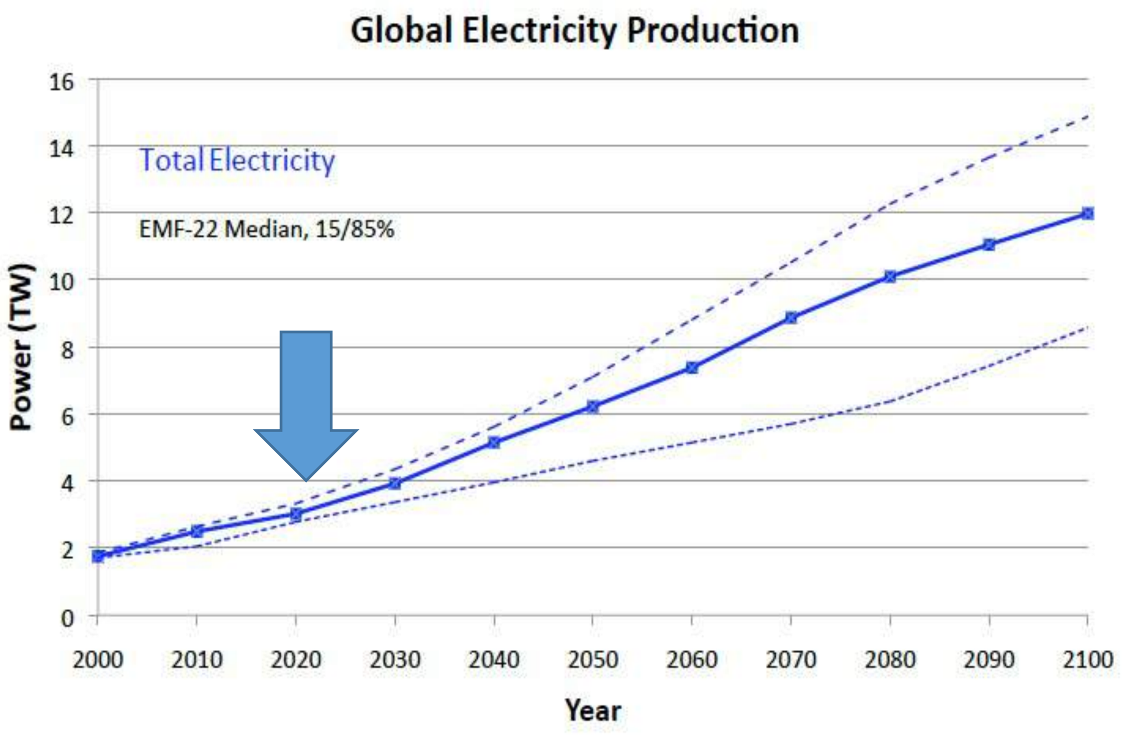
Frank Jenko

Max Planck Institute for Plasma Physics, Garching
Technical University of Munich
University of Texas at Austin

IHPCSS 2021 – July 26, 2021

Global electricity needs will keep increasing

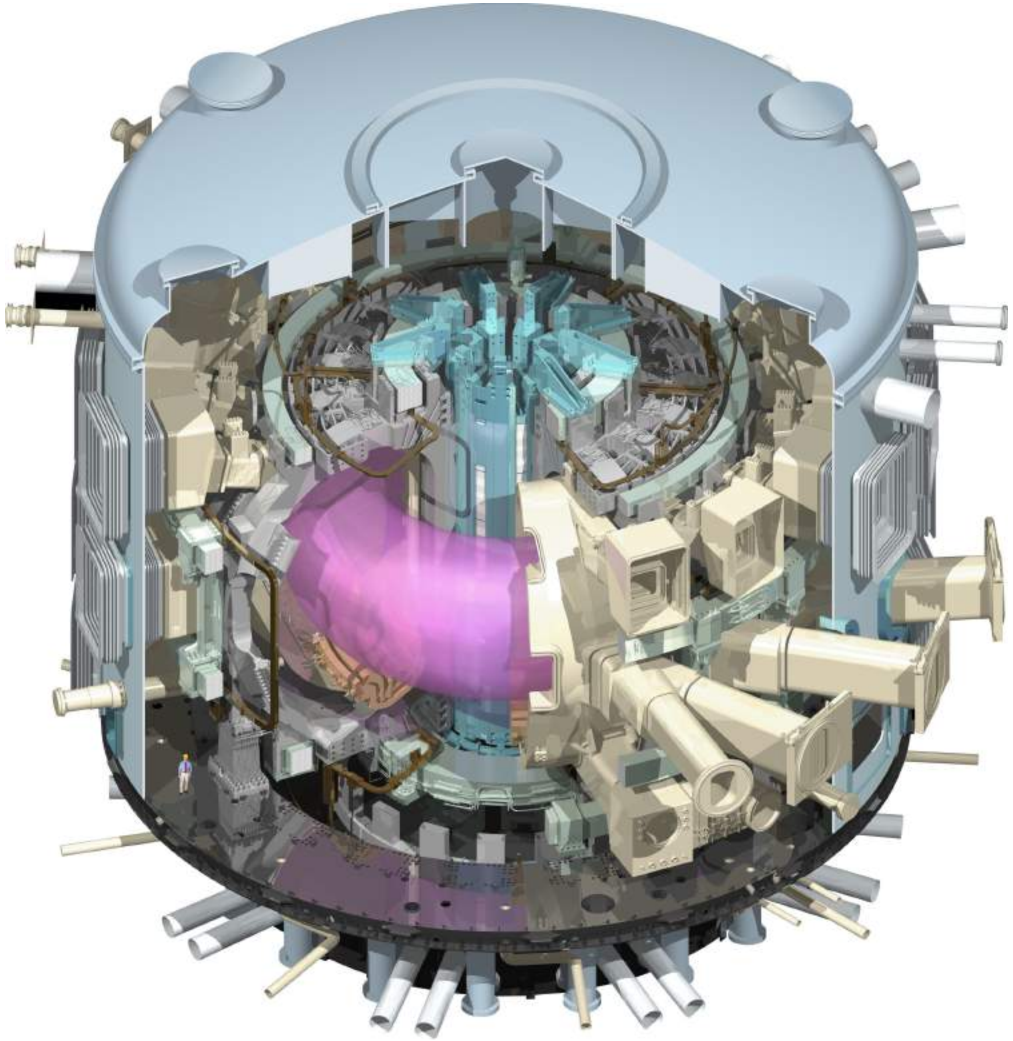
Energy Modeling Forum 22
100 models from 15 research groups (Clarke 2009)

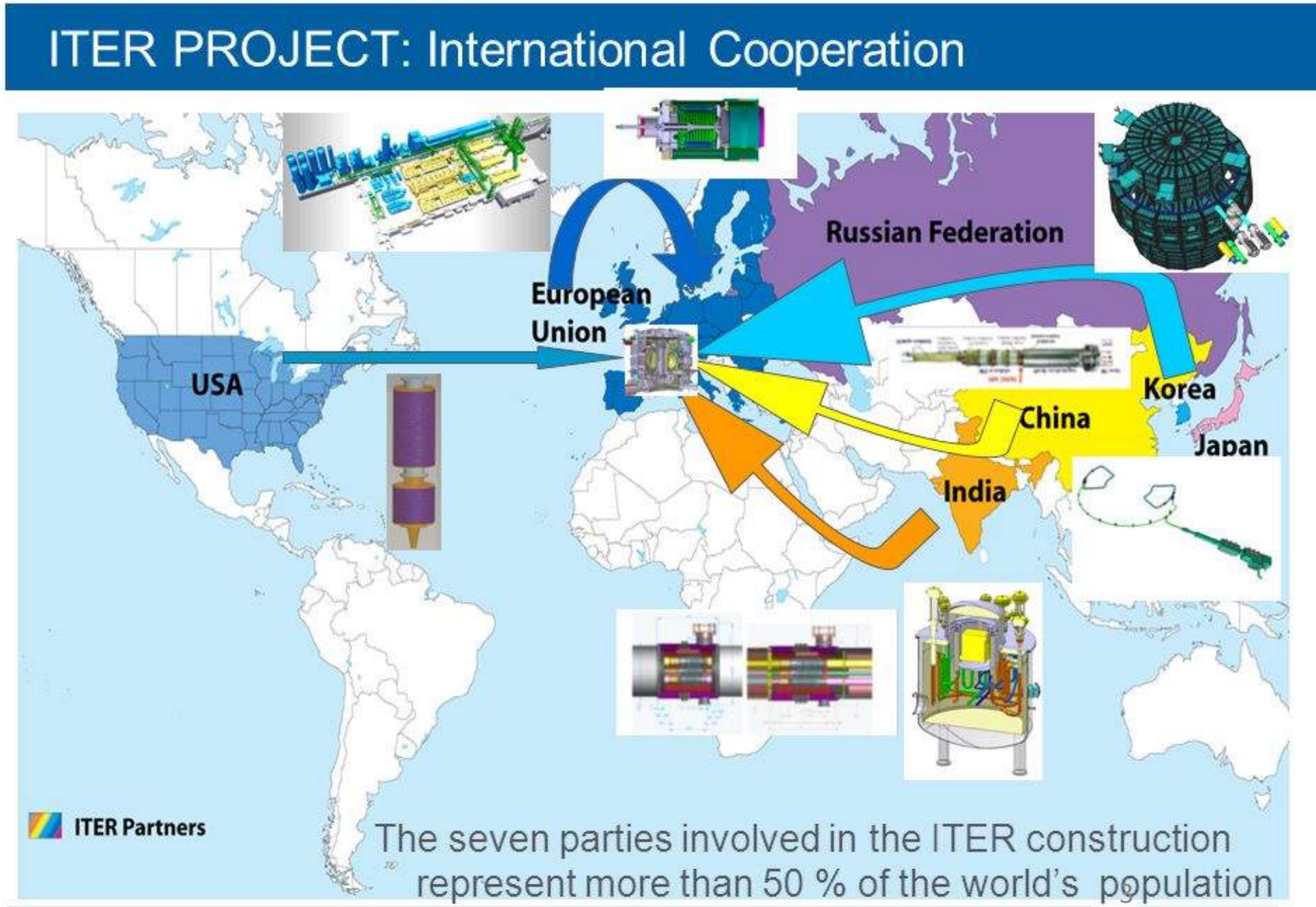


The international ITER project

Goal: 500 MW of fusion power

www.iter.org





ITER construction site in Southern France

From trial-and-error to predict-first:
The role of High Performance Computing

At the forefront of supercomputing since the 70's

NERSC HISTORY

Powering Scientific Discovery Since 1974

Contact: Jon Bashor, jbashor@lbl.gov, +1 510 486 5849

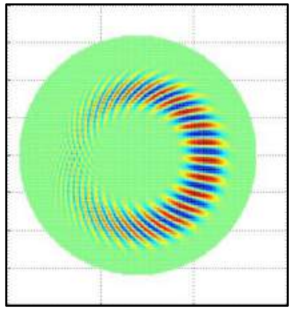
The oil crisis of 1973 did more than create long lines at the gas pumps — it jumpstarted a supercomputing revolution.

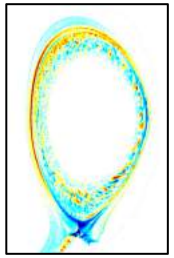
The quest for alternative energy sources led to increased funding for the Department of Energy's Magnetic Fusion Energy program, and simulating the behavior of plasma in a fusion reactor required a computer center dedicated to this purpose. Founded in 1974 at Lawrence Livermore National Laboratory, the Controlled Thermonuclear Research Computer Center was the first unclassified supercomputer center and was the model for those that followed.

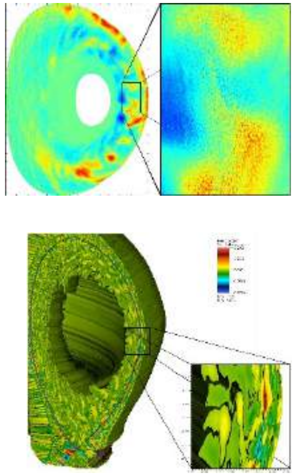
Over the years the center's name was changed to the National Magnetic Fusion Energy Computer Center and later the National Energy Research Supercomputer Center (NERSC). In 1983 NERSC's role was expanded beyond the fusion program, and it began providing general computing services to all of the programs funded by the DOE Office of Energy Research (now the Office of Science). The current name was adopted in 1996 when NERSC relocated to Lawrence Berkeley National Laboratory and merged with Berkeley Lab's Computing Sciences program. The name change — from "Supercomputer Center" to "Scientific Computing Center" — signaled a new philosophy, one of making scientific computing more productive, not just providing supercomputer cycles.

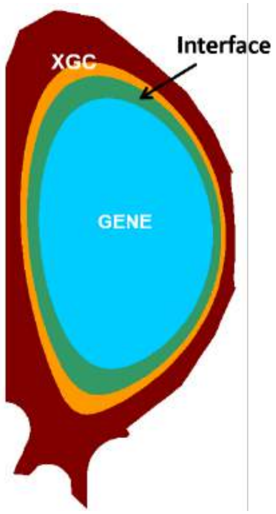
From highly idealized models to virtual fusion systems

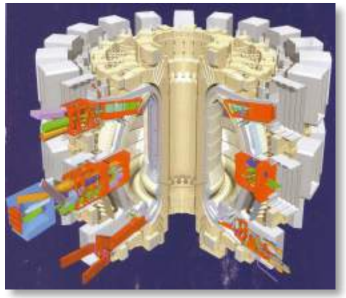
Increasing fidelity & modeling capability with increasing computing power

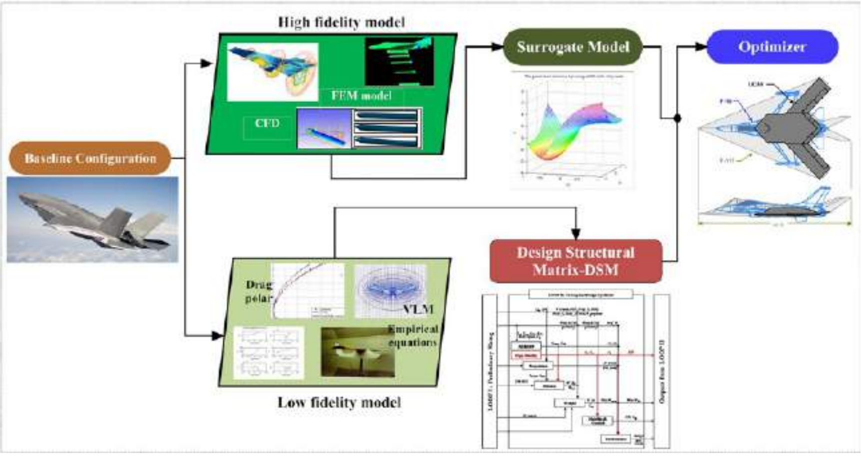




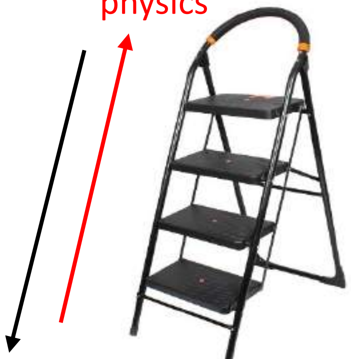








more complete physics



faster

Multi-fidelity approach:

- HiFi models for reliable extrapolation/prediction
- LoFi models (based on HiFi models) for high-throughput computing & real-time applications (incl. control)

Both are needed – together

A high-fidelity model for determining turbulent transport (i.e., the energy confinement time): The GENE code

Fluid models don't work – use (gyro-)kinetics!

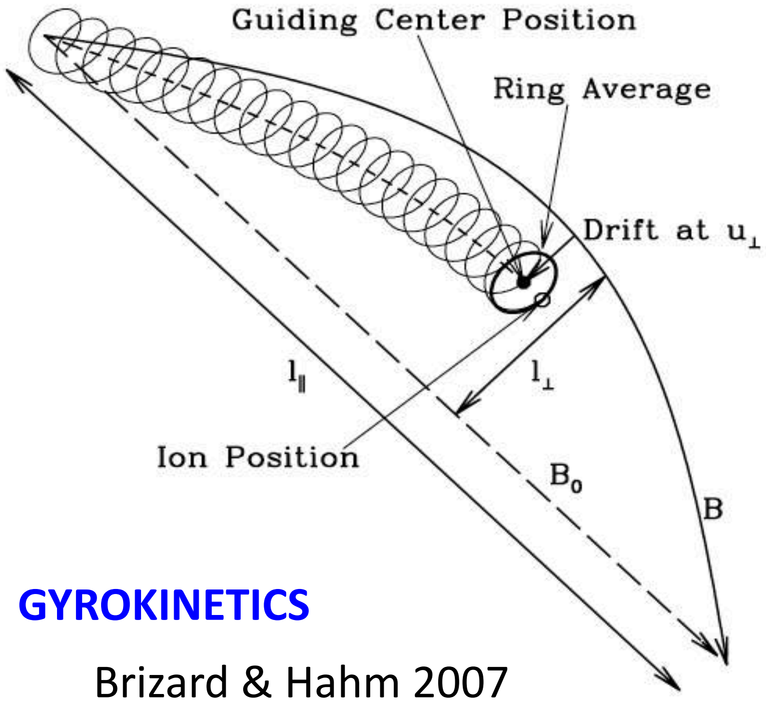
Hot and/or dilute plasmas are only weakly collisional: 6D Vlasov-Maxwell equations

$$\frac{\partial f_\alpha}{\partial t} + \mathbf{v}\cdot\nabla f_\alpha + \frac{q_\alpha}{m_\alpha}\left[\mathbf{E} + \frac{\mathbf{v}\times\mathbf{B}}{c}\right]\cdot\nabla_v f_\alpha = 0$$

α = particle species

$$f_\alpha = f_\alpha(\mathbf{x}, \mathbf{v}, t)$$

...from the Liouville equation via the BBGKY hierarchy



GYROKINETICS

Brizard & Hahm 2007

Strong background magnetic field:

Eliminate fast gyromotion; consider slow dynamics of **guiding centers**

$$f = f(\mathbf{X}, v_\parallel, \mu; t)$$

$$\boxed{\frac{\partial f}{\partial t} + \dot{\mathbf{X}}\cdot\frac{\partial f}{\partial \mathbf{X}} + \dot{v}_\parallel\frac{\partial f}{\partial v_\parallel} = 0}$$

Reduction of effort by ~12 orders of magnitude (elimination of irrelevant spatio-temporal scales & reduction from 6D to 5D)

Additional gain from using field-aligned coordinates

The gyrokinetic code GENE

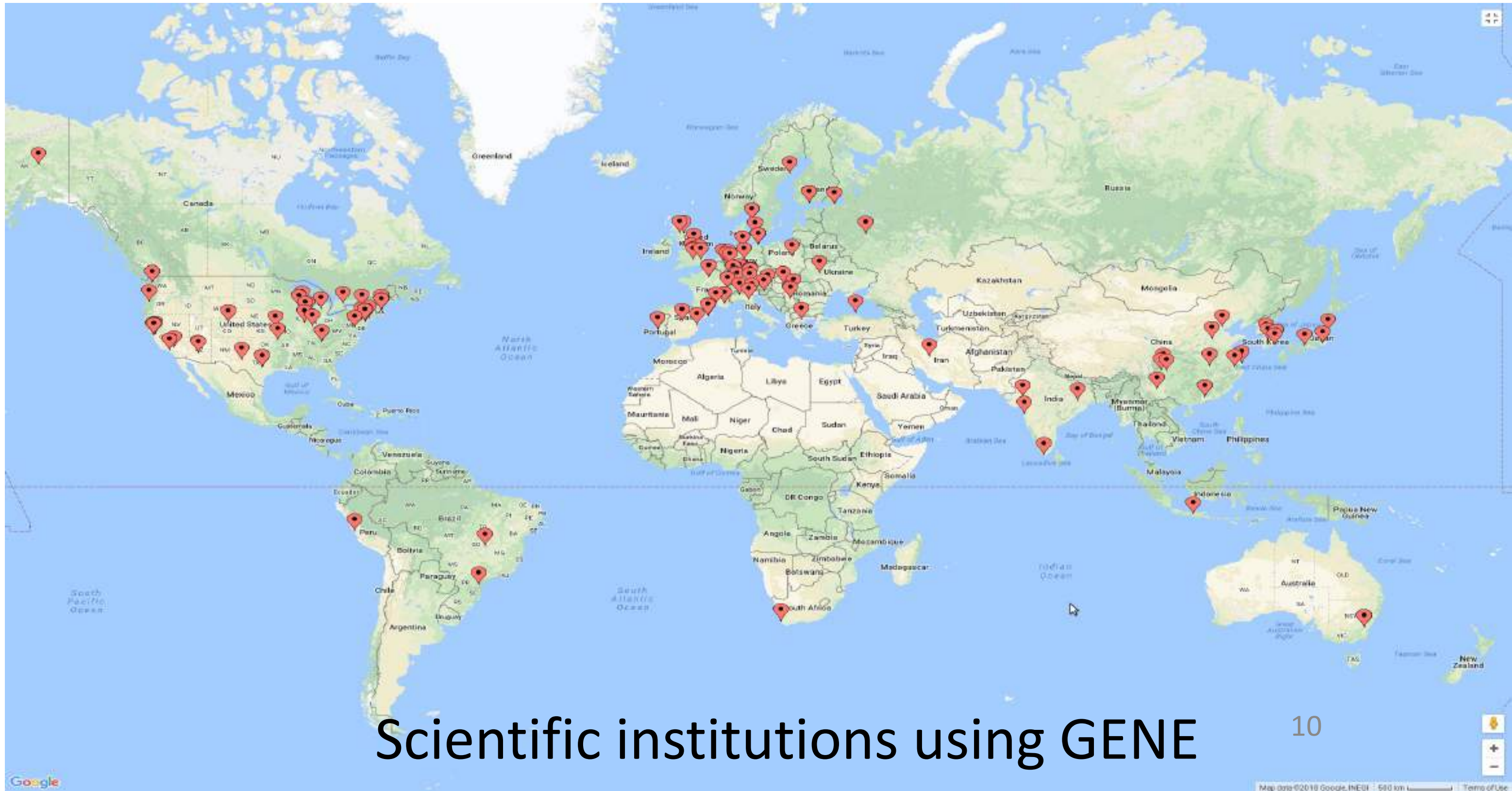
- First GENE publication: Jenko et al., Physics of Plasmas 2000 (>600 citations – WoS)
- More than 100,000 lines of source code, plus 200,000 lines for pre-/post-processing
- Based on numerical methods from **Computational Fluid Dynamics**

- Open source policy

- World-wide user base:
  genecode.org
  support@genecode.org

Part of an ecosystem of codes for fusion research

Also used in astrophysics



Scientific institutions using GENE

Global Gyrokinetic Simulation of Turbulence in

ASDEX Upgrade

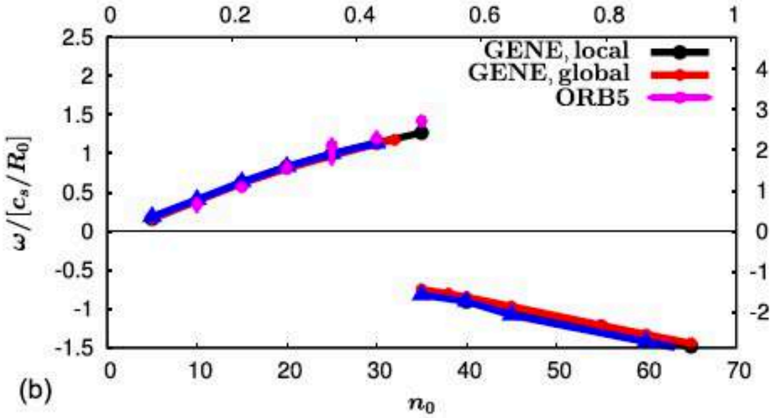
GENE

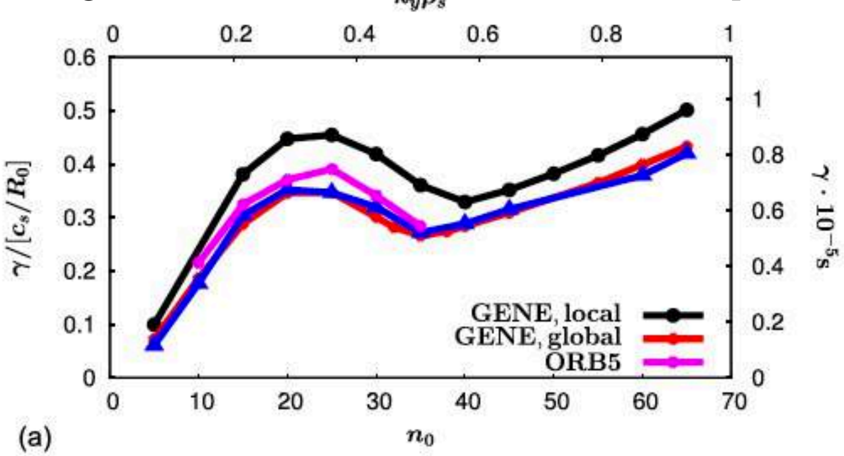
gene.rzg.mpg.de

Verification and validation

Code verification: Are we solving the equations right?

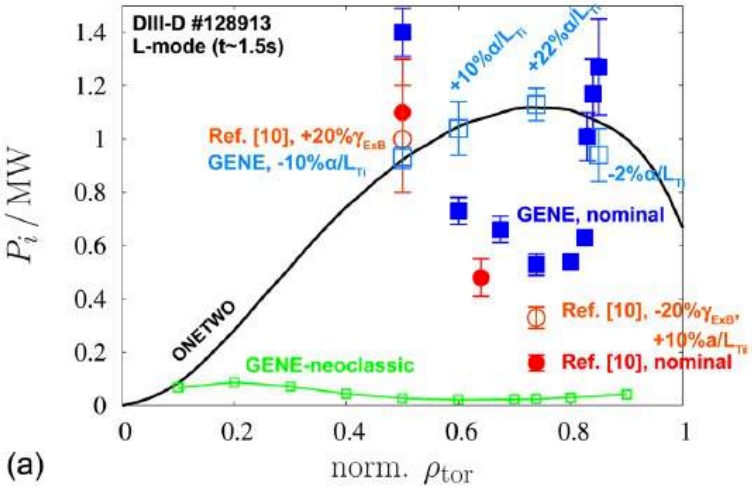
First benchmark of global electromagnetic GK codes [Görler PoP 2016]

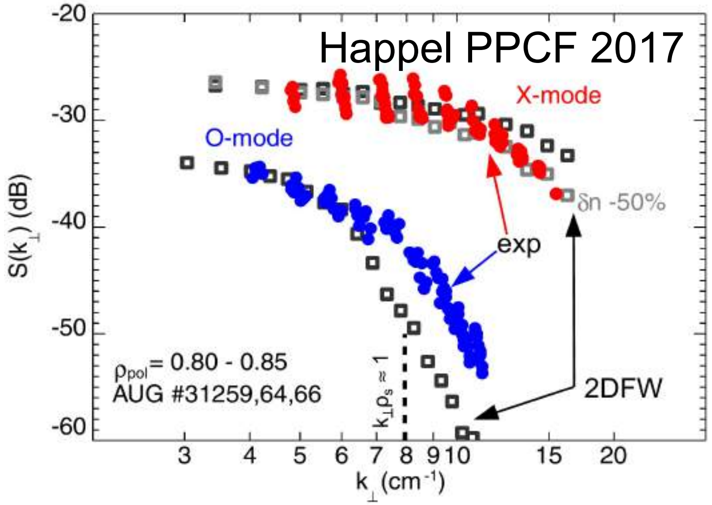




Code validation: Are we solving the right equations?

Can gyrokinetics describe outer-core L-mode plasmas? [Görler PoP 2014]

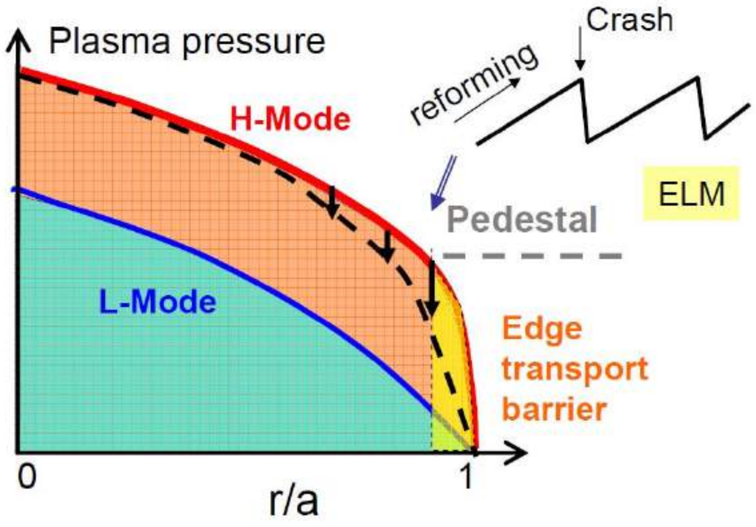


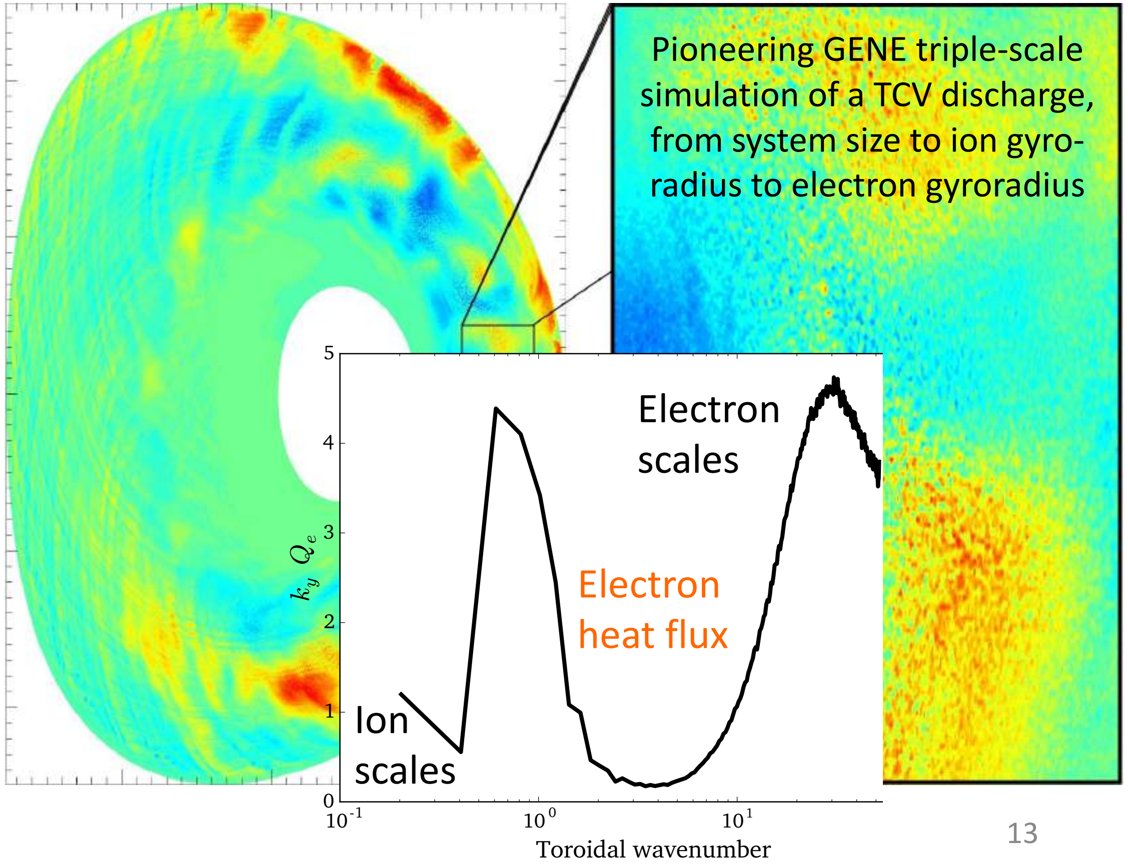


Science highlights (just two examples)

Prediction of relevant contributions to turbulent transport at very small – hiterto neglected – spatio-temporal scales (experimentally confirmed)

Indications that the physics of the "pedestal region" in ITER may differ from that in present-day devices



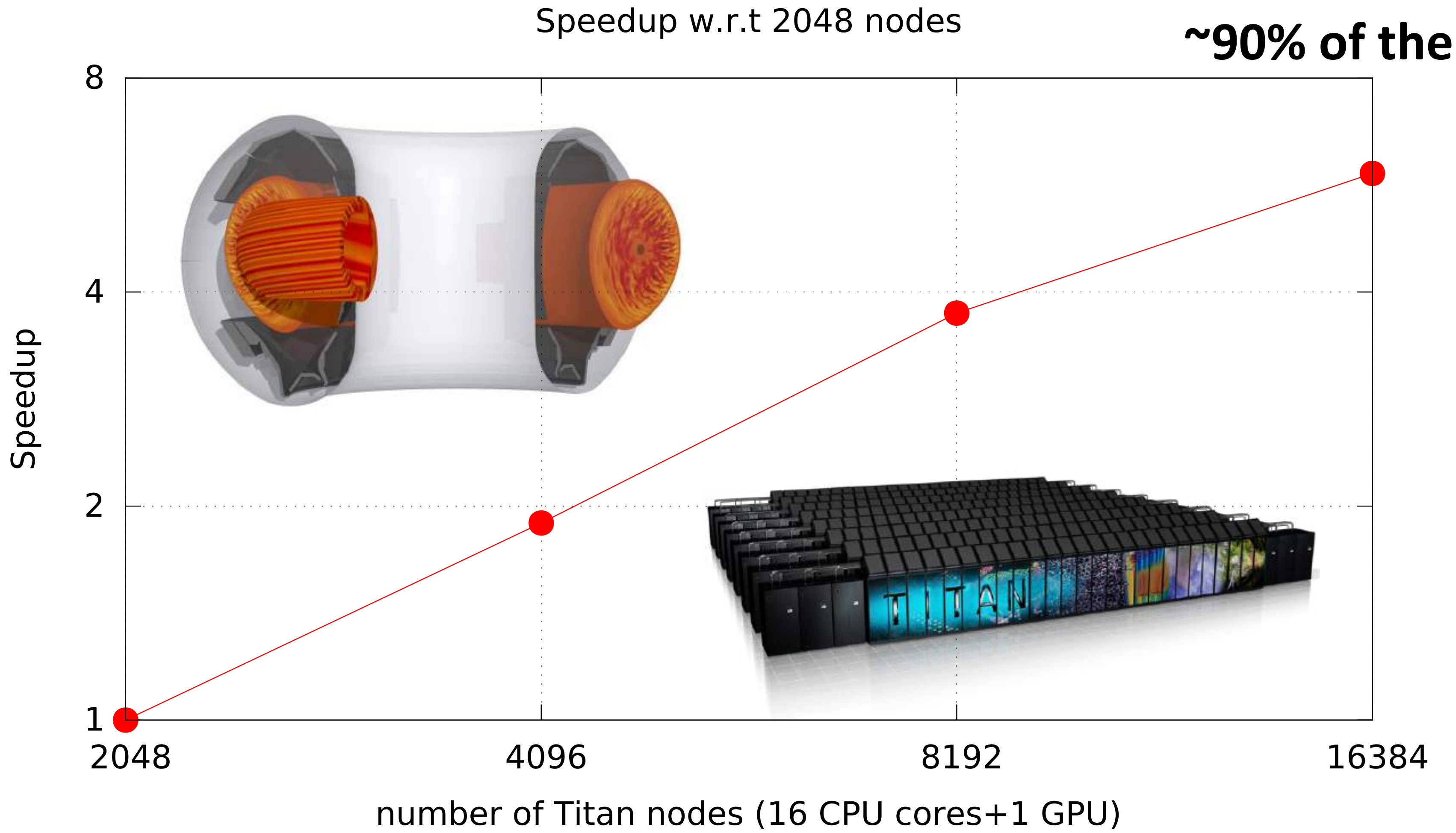


Strong scaling of GENE on Titan

MPCDF MAX PLANCK COMPUTING & DATA FACILITY

Tilman Dannert

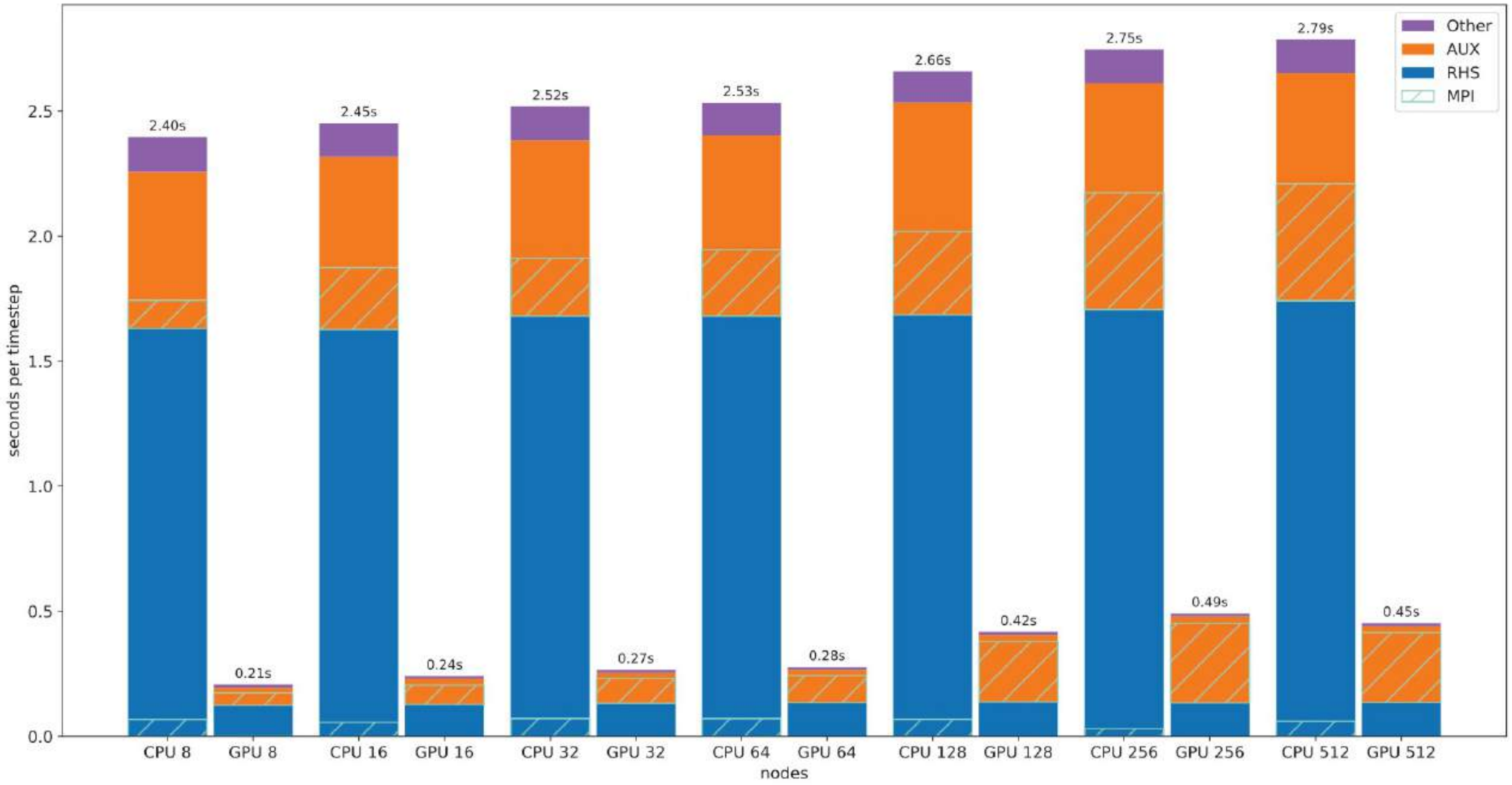
~90% of the machine



Optimizing large-scale codes (like GENE) on pre-exascale systems may take <u>person-years</u>

GENE performance on #2 supercomputer (Summit), using GPUs

Weak scaling from 8 to 512 nodes on Summit: GPU vs CPU
Per node: 2 x IBM Power9 and 6 x NVIDIA Volta V100



EXASCALE COMPUTING PROJECT

Speedups:

RHS = 12.3 – 13.2

Total (8 nodes) = 11.6

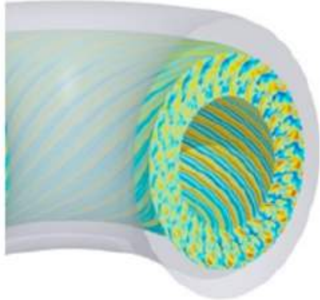
Total (512 nodes) = 6.2

K. Germaschewski et al., Phys. Plasmas 2021

Code development, maintenance, dissemination, application: GENE

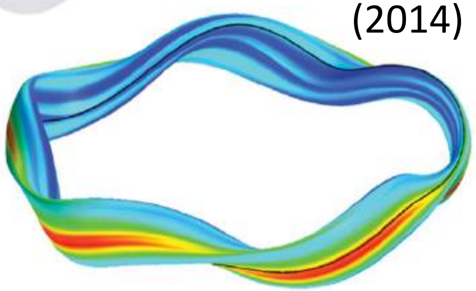
GENE family of codes (genecode.org)

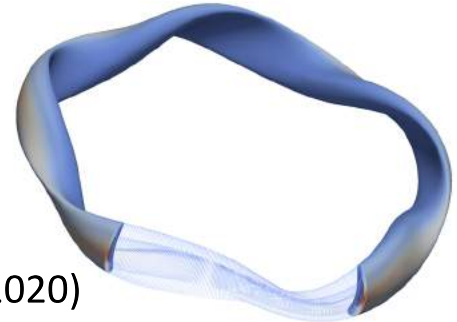
T flux tube (2000)
S flux tube (2002)



T global (2011)

S full surface (2014)

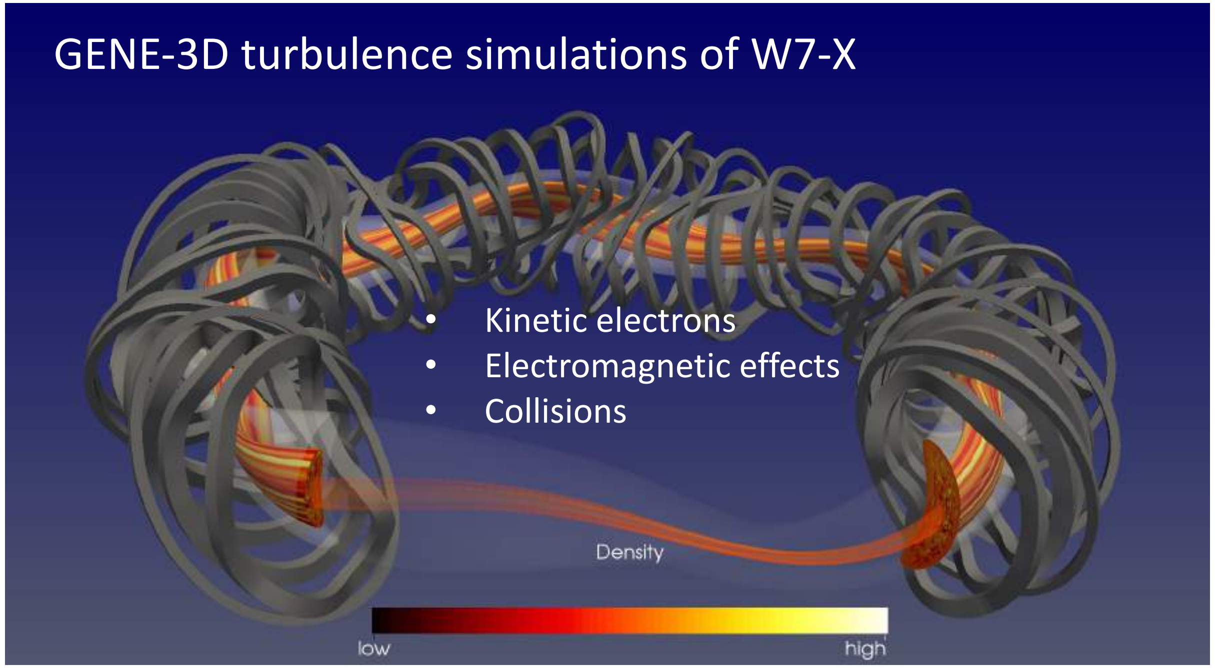




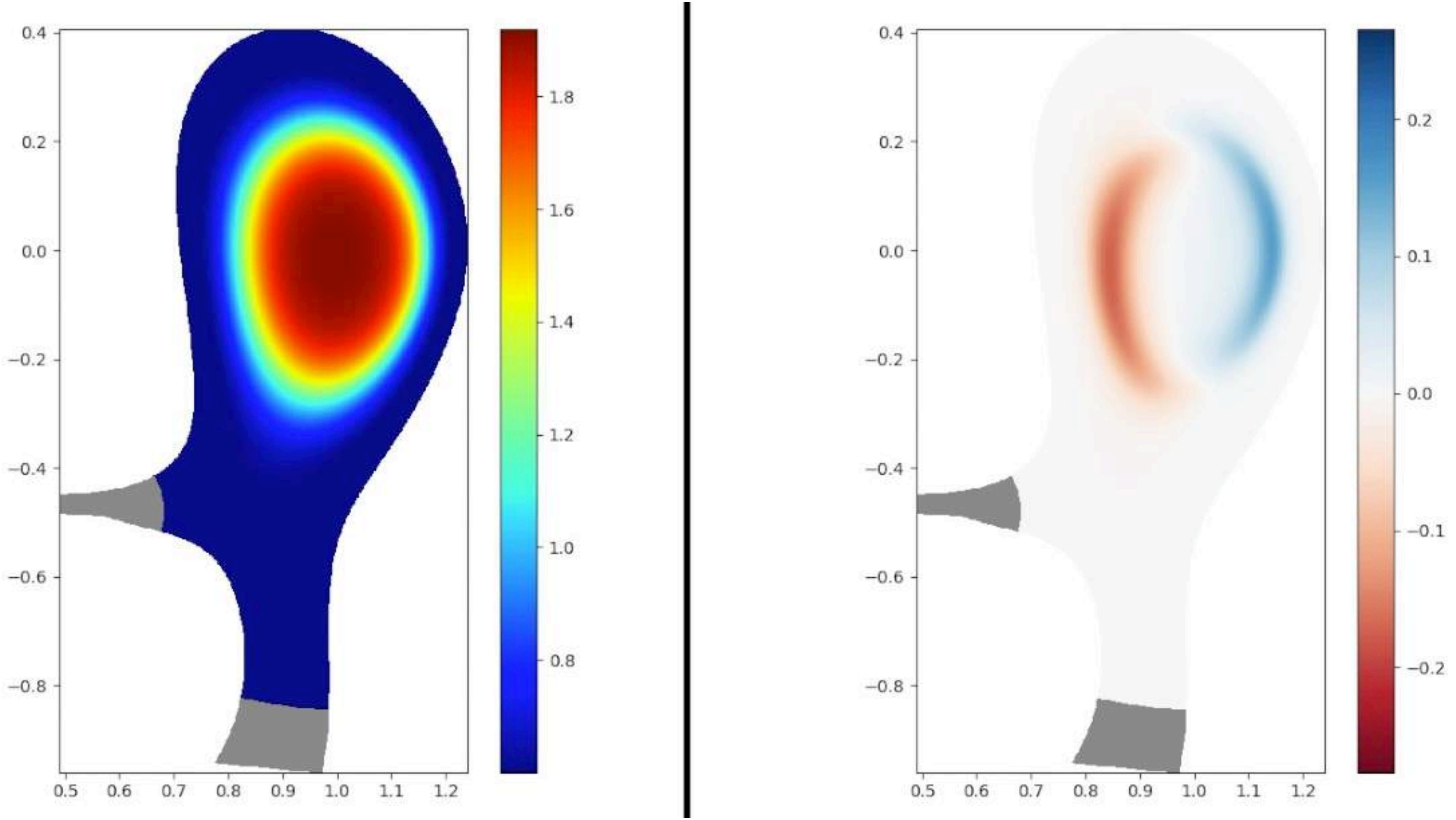
S global (2020)

+ GENE-X
(later in this talk)

T = tokamak
S = stellarator



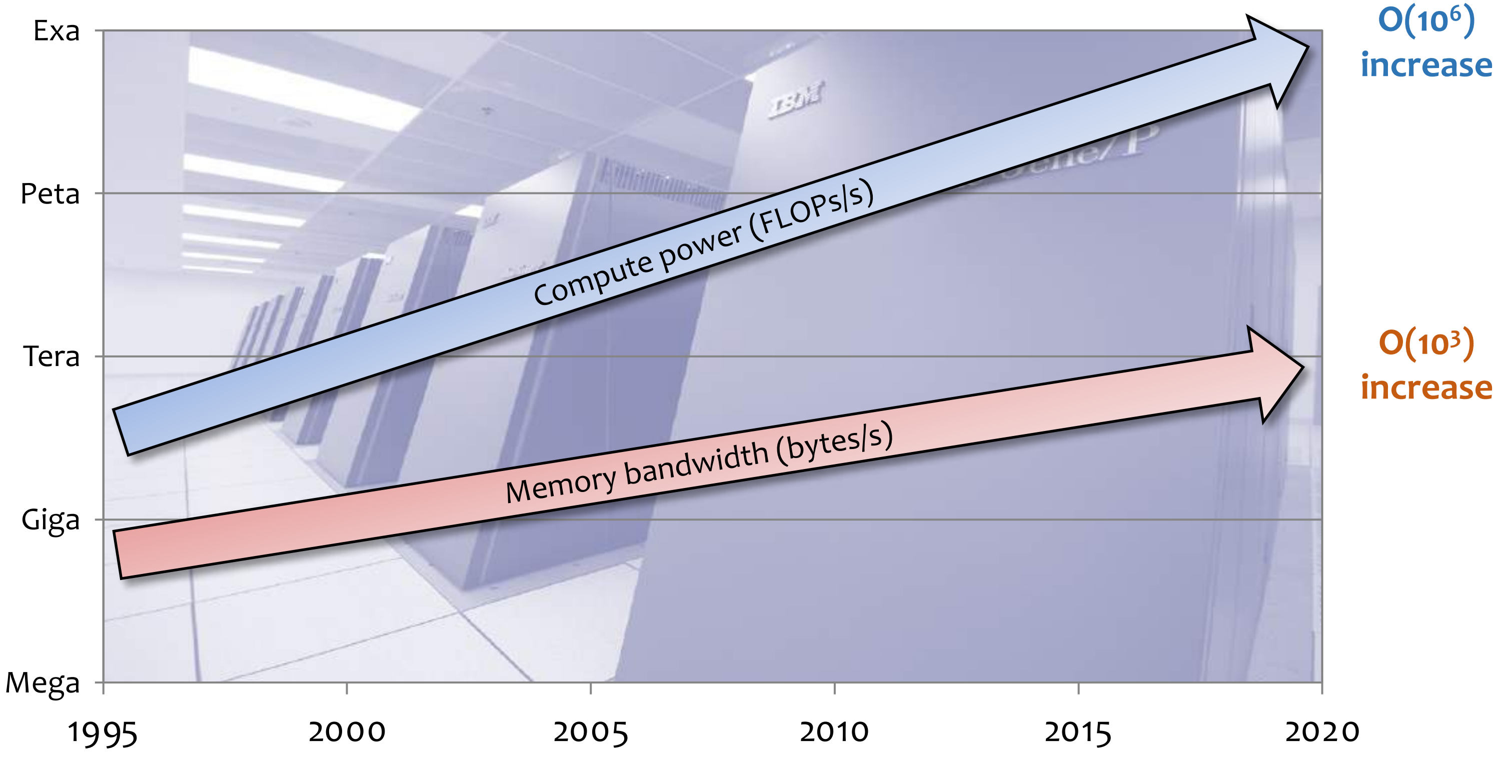
GENE-X: First whole-device simulations (D. Michels et al.)



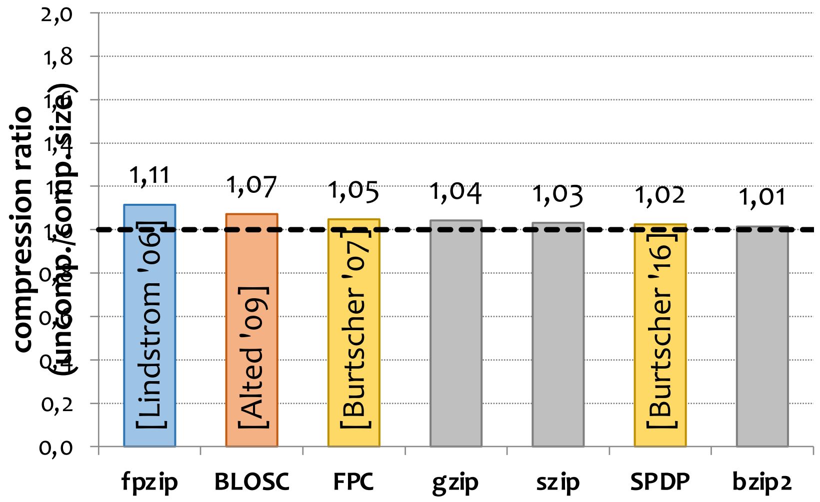
Lossy compression of scientific data in GENE simulations of plasma turbulence

Data outgrows compute, calls for data reduction techniques

P. Lindstrom, IPAM Workshop, UCLA (10/15/18)



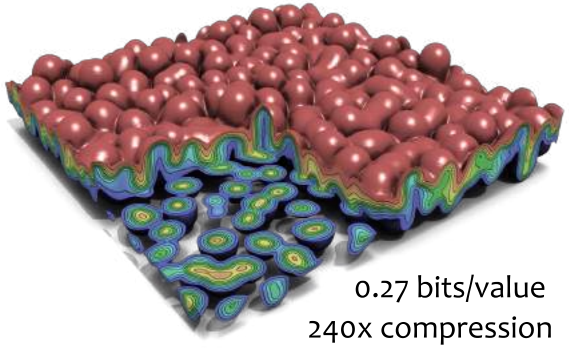
Lossless compression has very limited impact



Lossy compression enables greater reduction, but it is often met with skepticism by scientists

P. Lindstrom, IPAM Workshop, UCLA (10/15/18)

- Large improvements in compression are possible by allowing even small errors
  - Simulation often computes on meaningless bits
    - Round-off, truncation, iteration, model **errors abound**
    - Last few floating-point bits are effectively **random noise**



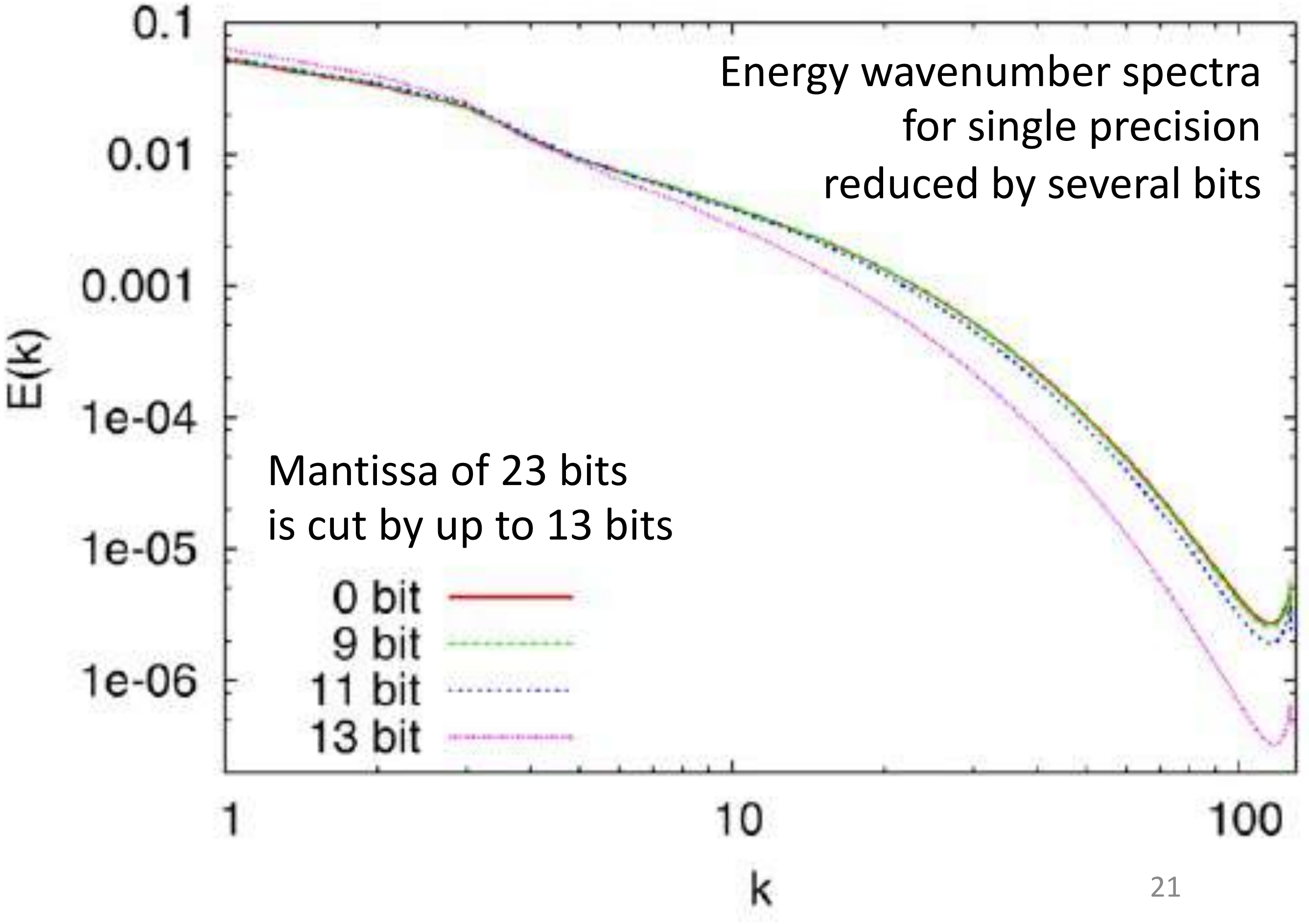
0.27 bits/value
240x compression

- Still, lossy compression often makes scientists nervous
  - Even though lossy **data reduction** is ubiquitous
    - **Decimation** in space and/or time (e.g. store every 100 time steps)
    - **Averaging** (hourly vs. daily vs. monthly averages)
    - **Truncation** to single precision (e.g. for history files)
  - State-of-the-art compressors support **error tolerances**

Turbulence provides an interesting test case

- Turbulent dynamics is inherently nonlinear and high-dimensional
- It is characterized by a mix of disorder and order
- The detailed dynamics is chaotic, but its **statistical properties tend to be robust**
- It may be possible to trade accuracy for efficiency on the statistical level

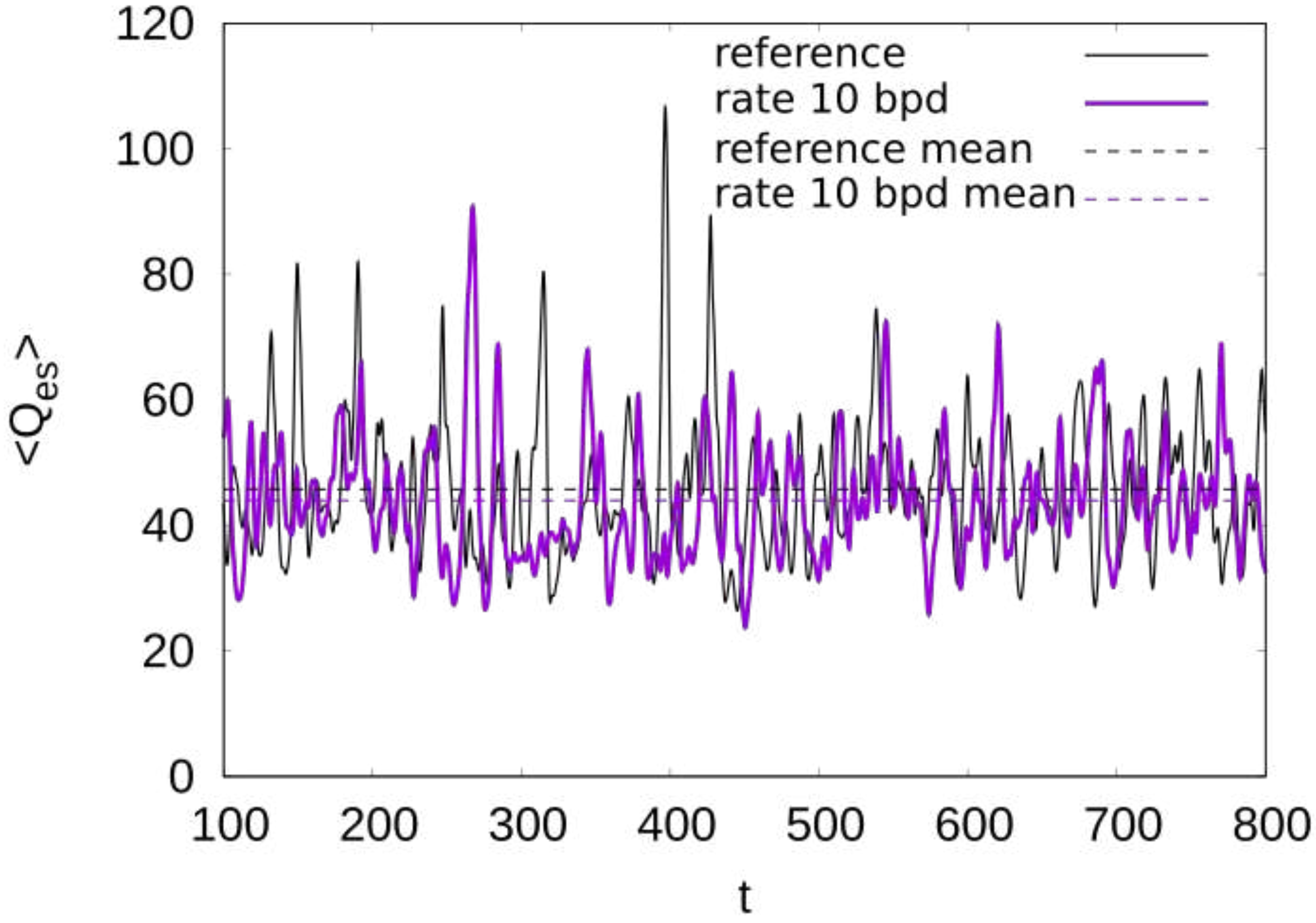
Homann+ CPC 2007



GENE-ZFP compression numerical experiment

- **Emulate simulations on compressed arrays** of grid-based data
- Steps in time loop:
  1. Compress and decompress 5D distribution function
  2. Compute time step
- Grid: $nx0 \times ny0 \times nz0 \times nv0 \times nw0 = 60 \times 128 \times 8 \times 32 \times 10$
- Tested ZFP modes (4D [$nx0 \times ny0 \times nz0 \times nv0$] compression):
  1. Fixed rate (fixed number of bits per floating point value)
  2. Fixed accuracy (fixed absolute error tolerance)
  3. Fixed precision (fixed number of uncompressed bits per value)

Heat flux as figure of merit

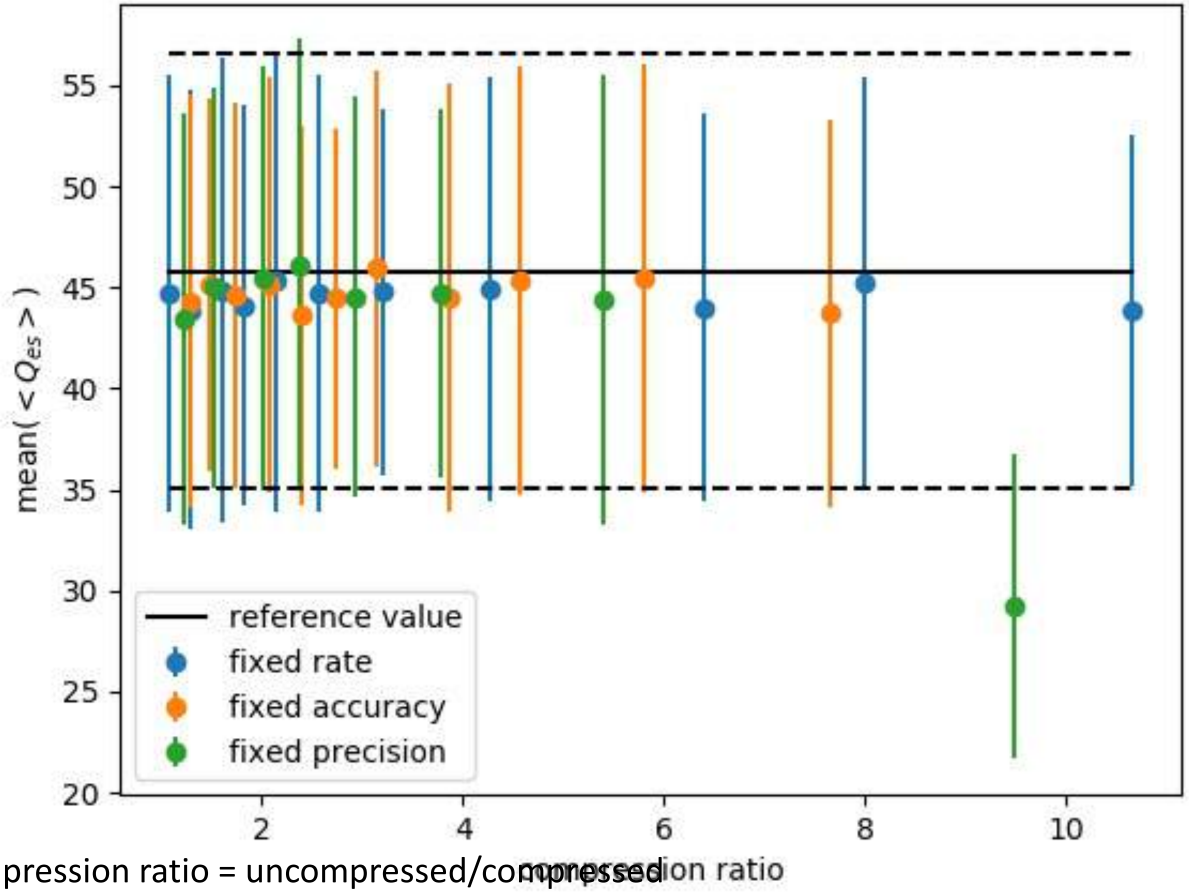


Heat flux in normalized units, averaged over space, as a function of time

Shown is a quasi-stationary saturated turbulent state

Plots for the reference and fixed rate (10 bits per double) compression runs

Heat flux for different compression modes and compression ratios



compression ratio = uncompressed/compressed size

Time-space averaged heat fluxes, with standard deviations:

- Fixed rate scan (bits per double): (unstable for 4) 6 , 8, 10 , 20, 25, 30, 35, 40, 50, 60
- Fixed accuracy scan (absolute error tolerance): (wrong results for 1e-2) 1e-3 … 1e-10, 1e-12, 1e-14, 1e-16
- Fixed precision scan (number of uncompressed bits per value): (unstable for 10) 15, 20, 25, 30, 35, 40, 50, 60

Lossy compression in GENE simulations of plasma turbulence

Denis Jarema, Peter Lindstrom & Frank Jenko

- **Goal:** Explore potential for data reduction in the GENE simulations; would help to reduce data motion and thus to increase code performance
- **Incentive:** We care mainly about *statistical* properties in a quasi-stationary saturated turbulent state, which may be quite robust (here: heat flux)
- **Results:** Compression ratios *up to ~10* with acceptable loss of quality
- **Conclusion:** Study points to *enormous potential* for data reduction in GENE

Combining computation with data analytics and machine learning for plasma physics

An important, timely topic of broad interest

"Science at extreme scales: Where big data meets large-scale computing"

Interdisciplinary Long Program @UCLA

September 12 - December 14, 2018

200+ participants, 50+ long-term participants

Speaker list includes:

- Yann LeCun (Director of AI Research @Facebook)
- Emmanuel Candes (Stanford University)
- Rajat Monga (Google)
- Matthias Troyer (Microsoft)
- James Sexton (IBM)
- Adrian Tate (Cray)
- Alan Lee (AMD)

Transformative Enabling Capabilities for fusion

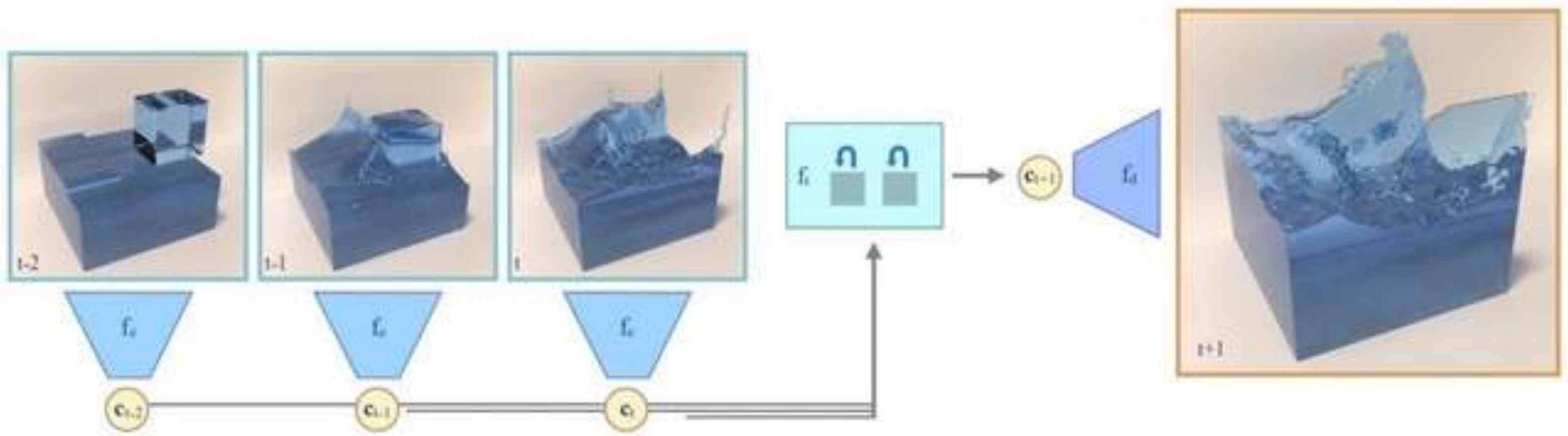
Advanced Algorithms – Advanced algorithms will transform our vision of feedback control for a power-producing fusion reactor. The vision will change from one of basic feasibility to the creation of intelligent systems, and perhaps even enabling operation at optimized operating points whose achievement and sustainment are impossible without high-performance feedback control. The area of advanced algorithms includes the related fields of mathematical control, machine learning, artificial intelligence, integrated data analysis, and other algorithm-based R&D. Given the pace of advances, control solutions that establish fusion reactor operation will become within reach, as will the discovery and refinement of physics principles embedded within the data from present experiments. This TEC offers tools and methods to support and accelerate the pace of physics understanding, leveraging both experimental and theoretical efforts. These tools are synergistic with advances in exascale and other high-performance computing capabilities that will enable improved physics understanding. Machine learning and mathematical control can also help to bridge gaps in knowledge when these exist, for example to enable effective control of fusion plasmas with imperfect understanding of the plasma state.

Integrating scientific knowledge into ANNs

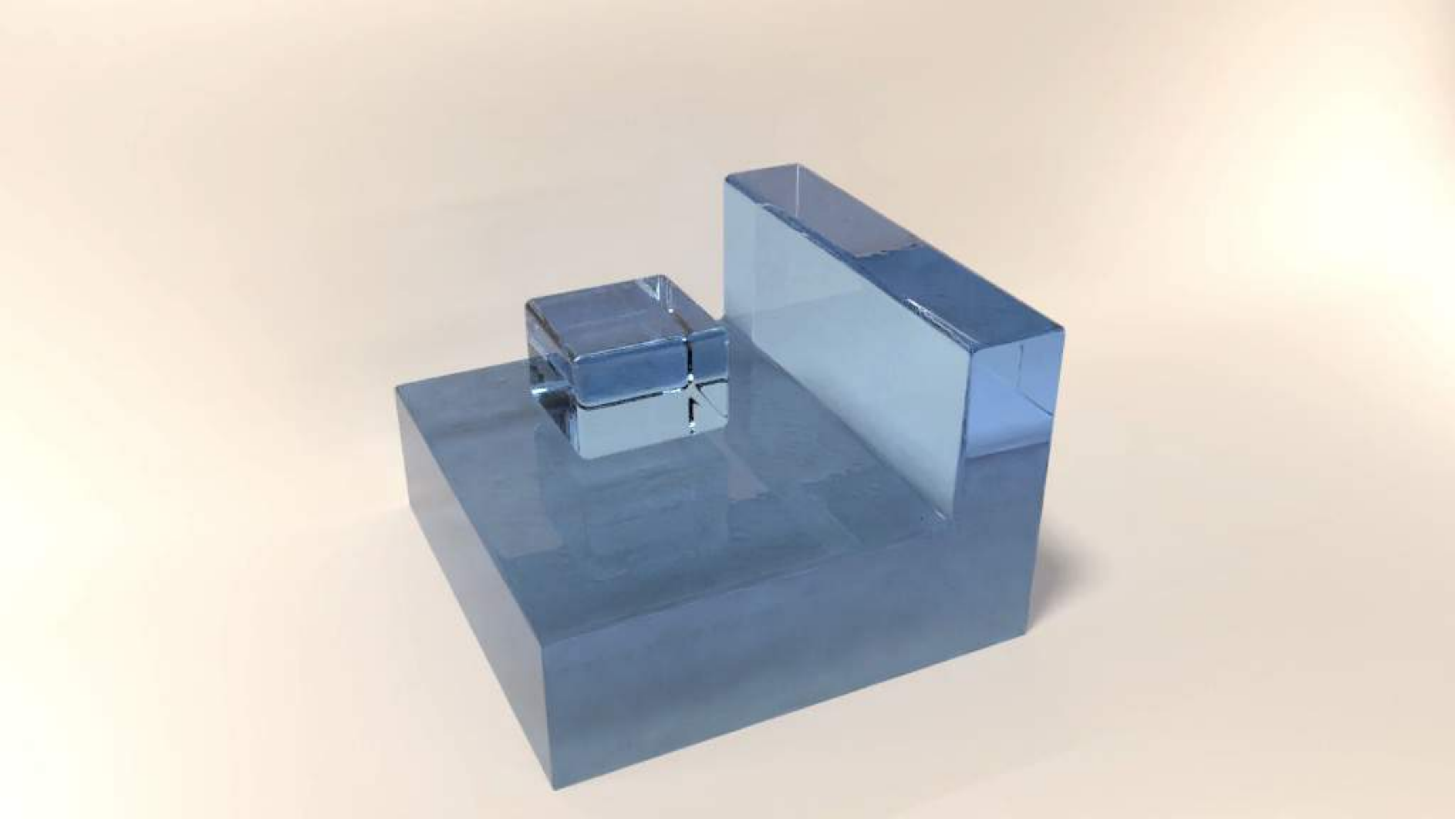
To maximize success, *Scientific ML* should not be physics-agnostic

- **Physics-guided design of ANNs**
- **Physics-guided learning of ANNs**
- **Combining data- and physics-based models**

Potential to improve accuracy, efficiency, interpretability, generalizability

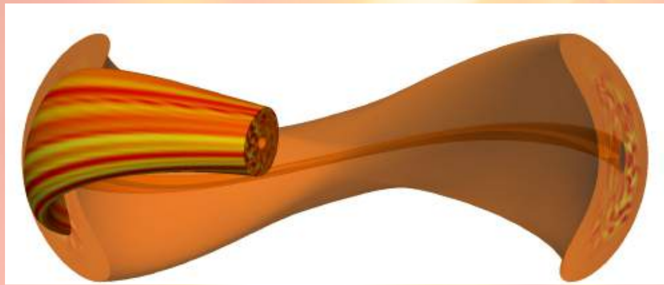


Innovative ideas (Nils Thuerey, TUM): Accelerating fluid simulations with Deep Learning



Conclusions

The big picture



Overarching goal: Contribute to the **gradual development of a validated predictive capability ("virtual fusion plasma"), helping to accelerate fusion energy research**

A beautiful example of how **fascinating science on some of the world's largest supercomputers contributes to solving grand challenges facing society**

The development and application of GENE illustrate the **fascinating challenges and opportunities at the interface of applied mathematics, computer science & physics**