Tera-Scale Computers in Controlled Fusion Research: - The Potential for Breakthroughs in our Understanding of Complex Plasma Phenomena

Paul Bonoli, MIT PSFC

With Contributions from:
Darin Ernst and John Wright, MIT PSFC
Martin Greenwald, MIT PSFC

IAP Lecture: January 11, 2005
What is a “Tera-scale” Computer?

- A cluster of processing units (cpu’s) performing arithmetic simultaneously (in parallel) and resulting in about $10^{12}$ (Tera) floating point operations per second (flops/s).
  - “Tera-Scale” means “Tera-flop – Tflop”

- Maximum theoretical flop-rate ($R_{\text{flop}}$) is:
  $$R_{\text{flop}} \approx (0.5 – 2.0) \ R_{\text{cpu}} \times N_{\text{cpu}}$$
  where:
  - $R_{\text{cpu}}$ is the processor clock speed (Hz)
  - $N_{\text{cpu}}$ is the total number of processors
What is a “Tera-scale” Computer?

• Original CRAY1 “supercomputer” was capable of 133 Mflop/s!
  – $R_{cpu} = 80 \text{ MHz}$
  – Contained a single CPU capable of vector arithmetic (array manipulation in a single clock cycle).

• Today, CPU’s are very fast - and inexpensive!
  $R_{cpu} \approx 0.3 – 2.6 \text{ GHz}$
  $N_{cpu} = 1$ (a desktop computer)
  $R_{flop} \approx 0.5 – 1.0 \text{ Gflop/s}$
Beowulf architecture is an economical alternative that is part way to Tera-flop.

PSFC has two Beowulf clusters: MARSHALL & UNITY

←MARSHALL

\[ R_{\text{cpu}} = 1.66 \, \text{GHz} \]
\[ N_{\text{cpu}} = 50 \]
\[ R_{\text{flop}} \approx 50 \, \text{Gflop/s} \]
Tera-scale Computer at NERSC - Seaborg

\[ R_{\text{cpu}} = 375 \text{ MHz} \quad N_{\text{cpu}} = 6080 \quad R_{\text{flop}} \approx 9 \text{ Tflop/s} \]
Fast CPUs and Massive Number of Compute Nodes Present a Challenge for Communication Technology

• Computing clusters require fast interconnect speeds for compute nodes to efficiently share information.

• **Need large bandwidth and low latency:**
  – Ethernet
  – Myrinet

• **Ancillary benefit of massive number of compute nodes is greatly increased machine memory to handle larger problems.**
Beowulf and Tera-Scale Computers Can be Used to Understand Complex Phenomena in Fusion Plasmas

- Phenomena in fusion plasmas are inherently multi-scale (time and space) and nonlinear in nature:
  - Micro-turbulence and transport
  - Wave – particle interaction
  - MHD behavior

- Massively parallel platforms make it possible to simulate nonlinear and multi-scale phenomena.
In Year 2001, OFES made a Long-Range Commitment to Fund Initiatives that Would take Advantage of Tera-Scale Computers

• Program is called the Scientific Discovery Through Advanced Computing Initiative - “Sci-Dac”
  – Created an opportunity for researchers in the computer science and controlled fusion communities to work together.

• This talk will discuss “breakthroughs” from three of these Initiatives:
  – Center for Extended Magnetohydrodynamic Modelling (CEMM)
  – Center for Simulation of Wave-Plasma Interactions
  – Computational Center for the Study of Plasma Microturbulence
Multiple Time and Spatial Scales

Relevant timescales for a burning plasma experiment

- ELECTRON TRANSIT
- SAWTOOTH CRASH
- ENERGY CONFINEMENT
- CURRENT DIFFUSION
- TURBULENCE
- ISLAND GROWTH

(a) RF codes
(b) Micro-turbulence codes
(c) Extended-MHD codes
(d) Transport Codes
Fundamental Challenges to Fusion Simulation

• Extreme range of time scales – wall equilibration/electron cyclotron $O(10^{14})$
• Extreme range of spatial scales – machine radius/electron gyroradius $O(10^4)$
• Extreme anisotropy – mean free path parallel to magnetic field/perpendicular $O(10^{10})$
• Non-linearity – turbulence
• Sensitivity to geometric details
• High dimensionality – basic object of plasma is 7D $\rightarrow f(x, v, t)$, described by non-linear Boltzmann equation

$$\frac{\partial f}{\partial t} + v \cdot \nabla f + \frac{q}{m} [E + v \times B] \cdot \nabla_v f = C(f)$$

- Convection
- Convection in velocity space
- Collisional relaxation toward Maxwellian in velocity space
Microturbulence codes describe the small scale fluctuations that presently dominate transport of matter and heat in fusion plasmas, $\tau \sim 10^{-7} - 10^{-4}$ sec

Gyrokinetic equation – 5D

\[
\frac{\partial \tilde{h}_a}{\partial t} + \left( v_{\chi a} + v_{da} + v_{||} \hat{b} \right) \cdot \nabla \tilde{h}_a = -v_{\chi a} \cdot \nabla f_{0a} - q_a \frac{\partial f_{0a}}{\partial W} \frac{\partial \tilde{\chi}}{\partial t} + \text{collisions + sources/sinks,}
\]

Direct solution of PDE - sparse linear matrix solve - **GS2, GYRO**

Particle-in-cell (PIC) - scatter/gather operations – **GTC**

Significant activity to compare codes and approaches.
PIC Methods Have a High Degree of Scalability

3D Gyrokinetic Turbulence Code (GTC)
Scalable on Massively Parallel Computers

Y-axis: the number of particles which move 1 step in 1 second
PIC Simulation (GTC) Showing the Effect of Plasma Flow on an ITG Mode
Contours of Turbulent Electrostatic Potential (W. Lee, PPPL)
Nonlinear, global simulation (GYRO) of ion turbulence in a shaped tokamak (DIII-D).
Radial annulus is 130 ion gyro-radii.

GYRO gives superlinear scaling up to 1024 processors on FIXED problem size.

J. Candy, GA
RF and Fokker Planck codes describe the interaction of injected radio waves with plasma particles

$$\nabla \times \nabla \times E + \frac{\omega^2}{c^2} E = J_p \circ E + J_{ant} : + \text{boundary conditions}$$

• Solution method via:
  – Full-wave EM field solutions - matrix solve
  – Ray tracing - integrate ray equations (ODE’s)

• Multiple spatial scales result from conversion of long wavelength modes to short wavelength:

• Injected waves may be used for heating, current drive, or plasma control.
Parallel Computing has made it possible to accurately simulate the signal detected by a laser as it scatters from mode converted ICRF waves in a tokamak (Alcator C-Mod). “Synthetic Diagnostic”
Comparison Between Theory and Experiment is Remarkable

S. Wukitch, Y. Lin, A. Parisot, M. Porkolab, J. Wright, 2005
Three waves – IBW, ICW, FW – are resolved at the MC layer.

The 2-3 wavelengths of IBW seen in simulations have been verified by PCI measurements on Alcator C-Mod.

The intermediate wavelength ICW travel towards the low field side above and below the midplane.

Poloidal field allows coupling to ICW (Perkins 1977)
Field solvers employing different conductivity models give very similar results.

TORIC at $240 N_r \times 255 N_m$

AORSA at $230 N_x \times 230 N_y$
All orders spectral technique has been extended to minority heating in 3D –AORSA3D

- Preliminary calculation for Fast Wave minority heating on LHD stellarator – 5% minority H in $^4$He
- 16×50×50 modes in $\phi$, x, y (10 independent solutions - one per field period, technique of H. Weitzner)

Fast wave heating in LHD Stellarator

- Gigantic, dense linear system $\rightarrow$ NERSC Seaborg, 1600 processor IBM SP
  8 hr processor time at ~1.7 teraflops

Memory = 750Mb/processor = 1,200 Gbytes
Self-consistency: Need to iterate physics modules

- High power waves can drive the distribution far from Maxwellian
- Significant non-Maxwellian components can be produced by neutral injection or fusion alpha particles

Non-Maxwellian distributions can:
- Affect local damping rate and wavelength
- Modify heating and current drive profiles

Calculation of wave fields self-consistently with the plasma distribution requires closed loop coupling of four significant physics models
Full-wave solvers (AORSA) have been coupled directly to nonthermal particle distributions

HHFW heating on NSTX with D beam injection:
Comparison of CQL3D distribution with 2 temperature Maxwellian model

- Significant differences are seen compared to two temperature Maxwellian model:
- Non-Maxwellian presently requires about 13× more CPU time
Parallel Computing Has Made it Possible to do Full-Wave Simulations of Lower Hybrid Waves ($\lambda < 1\,\text{mm}$)

- Eliminates need to do ray tracing.
- Injected LH waves convert between fast and slow modes.
- Loci of edge reflections and mode conversions in the full wave field pattern forms a “ring” of accessible LH waves.
- Simulation required 8064 MPP’s on Beowulf cluster (7 day utilizing 48 nodes)
LH wave Absorption is found to be different in full-wave and ray tracing

- All wave power is absorbed on the electrons by Landau damping.
- Caustic formation is apparent.
- There is also a suggestion of resonance cones in the full wave field patterns.
MHD codes describe gross plasma motion in a fluid model with kinetic and non-ideal effects, $\tau \sim 10^{-6} – 10^{-2}$ sec

\[
\rho \left( \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) \mathbf{v} = -\nabla p - \nabla \cdot \Pi + \mathbf{J} \times \mathbf{B}
\]

\[
\left( \frac{\partial}{\partial t} + \mathbf{v}_\alpha \cdot \nabla \right) p_\alpha = -\gamma p_\alpha \mathbf{v}_\alpha \cdot \nabla \mathbf{v}_\alpha + (\gamma - 1) Q_\alpha - \nabla \cdot \mathbf{q}_\alpha
\]

\[
\mathbf{E} = -\mathbf{V} \times \mathbf{B} + \eta \mathbf{J} + \frac{1}{\epsilon_0 \omega_p^2} \left[ \frac{\partial \mathbf{J}}{\partial t} + \nabla \cdot (\mathbf{J} \mathbf{V} + \mathbf{V} \mathbf{J}) + \sum_{\alpha} \frac{q_\alpha}{m_\alpha} (\nabla p_\alpha + \nabla \cdot \Pi_\alpha) \right]
\]

\[
\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E} \; , \; \nabla \cdot \mathbf{B} = 0
\]

- Computation presently dominated by sparse linear solves, maybe PIC in future
  NIMROD and M3D

Snapshot of a 3D calculation of a reconnection event in a low-aspect ratio tokamak. Two iso-pressure surfaces are shown
Calculation of sawtooth phenomena in a NSTX - Hot inner region interchanges with colder outer region via magnetic reconnection – M3D
NIMROD Simulation of DIII-D Disruption (Wall Loading)

S. Krueger, 2004
NIMROD Simulation of DIII-D Disruption
Nonlinear simulations find faster than exponential growth rate as expected from theory
Number one priority for DOE science is International Thermonuclear Experimental Reactor – ITER

ITER (≈$5B scale device)

Fusion experimental devices throughout the world already range in scale from < $50M to ~ $1B

We require massive computations to:

• Design experiments on fusion devices
• Interpret and understand experimental results from devices
• Reliably design and evaluate large new machines
• Invent and evaluate new, higher performing fusion concepts
A project is being designed whose ultimate (~ 15yr) objective is to predict reliably the behavior of plasma discharges in a toroidal magnetic fusion device on all relevant time and space scales → Fusion Simulation Project (FSP)
Initial Phase of the FSP will Consist of Focused Integration Initiatives

- **RF / MHD Simulation**
  - Study control of sawteeth or tearing modes using RF current drive
- **Edge Plasma Simulation**
- Initially “whole device” simulations might consist of:
  - Detailed MHD or transport physics module
  - Simple source models for heating, current drive, etc.
Summary

• Use of massively parallel computers has made disparate spatial and time scales of fusion plasmas possible to simulate:
  – Most simulations were not possible 5 years ago
    • MHD (sawtooth reconnection and disruption)
    • Transport (ion turbulence and turbulent suppression)
    • Wave – particle (ICRF mode conversion and LH full-wave)

• Numerical algorithms used in physics modules (PIC and matrix solve) have excellent scalability – well beyond present processor numbers.
Summary

• Whole device simulations for existing fusion devices will probably require \(10^2 - 10^3\) Tflop/s capability
  – **Burning plasma simulations will likely require > 10^3** Tflops/s.
  – **Device size (1-2 m) becomes a real “killer” in simulating fusion reactors.**
  – **New parallel – vector architectures (CRAY X1) may be a necessity.**