High-Performance Computing in Computational Astrophysics

Robert Fisher
University of Massachusetts Dartmouth

UMass Amherst HPC in CSE : September 28, 2012
In collaboration with

University of Massachusetts Dartmouth

Suoqing Ji
Kevin Jumper

<table>
<thead>
<tr>
<th>University of Chicago</th>
<th>University of Alabama</th>
<th>SUNY Stony Brook</th>
<th>Polytechnic University of Catalonia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donald Lamb</td>
<td>Dean Townsley</td>
<td>Alan Calder</td>
<td>Pablo Lorén-Aguilar Enricó García-Berro</td>
</tr>
<tr>
<td>Jim Truran</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>George Jordan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carlo Graziani</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Monday, October 8, 12
With Special Thanks To

<table>
<thead>
<tr>
<th>University of Chicago</th>
<th>Argonne National Laboratory</th>
<th>Lawrence Berkeley National Laboratory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anshu Dubey</td>
<td>Ray Bair</td>
<td>Katie Antypas</td>
</tr>
<tr>
<td>Brad Gallagher</td>
<td>Susan Coghlan</td>
<td></td>
</tr>
<tr>
<td>Lynn Reid</td>
<td>Randy Hudson</td>
<td></td>
</tr>
<tr>
<td>Paul Rich</td>
<td>John Norris</td>
<td></td>
</tr>
<tr>
<td>Dan Sheeler</td>
<td>Mike Papka</td>
<td></td>
</tr>
<tr>
<td>Klaus Weide</td>
<td>Katherine Riley</td>
<td></td>
</tr>
</tbody>
</table>
Outline

I. Modeling Turbulence in Astrophysical Simulations

II. Type Ia Supernovae

III. Asymptotically-Large Simulations
I. Modeling Turbulence in Astrophysical Simulations
Post-Millenial Computational Astrophysics

- Computation in astrophysics serves as the third pillar of science, alongside observation and theory.

- Remarkably successful 3-D simulations of cosmological structure, galaxy formation, star formation, stellar explosions, and many other systems can now be carried out.
### Post-Millenial Computational Astrophysics

**Table:**

<table>
<thead>
<tr>
<th></th>
<th>M_{\text{vir}} \times 10^{12} M_{\odot}</th>
<th>V_{\odot} \text{ [km/s]}</th>
<th>M^* \times 10^{10} M_{\odot}</th>
<th>f_b</th>
<th>B/D</th>
<th>R_d \text{ [kpc]}</th>
<th>M_l</th>
<th>SFR \text{ [M}_\odot \text{ yr}^{-1}]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eris</td>
<td>0.79</td>
<td>206</td>
<td>3.9</td>
<td>0.12</td>
<td>0.35</td>
<td>2.5</td>
<td>-21.7</td>
<td>1.1</td>
</tr>
<tr>
<td>MW</td>
<td>1 ± 0.2</td>
<td>221 ± 18</td>
<td>4.9 - 5.5</td>
<td>?</td>
<td>0.33</td>
<td>2.3 ± 0.6</td>
<td>?</td>
<td>0.68 - 1.45</td>
</tr>
</tbody>
</table>

*Guedes et al, 2011*
Weakly-Compressible Hydrodynamic Turbulence

The Universal Nature of Turbulent Flows

• Turbulence at high Reynolds is *universal* - the inertial scaling laws of a homogeneous, isotropic turbulent velocity field are independent of the driving.

• This is one of the deepest lessons of Kolmogorov (1941).
Parameter spaces sampled by simulation, experiment, and geophysical / astrophysical observation do not generally overlap.

\[ Re_\lambda \simeq 10^2 - 10^3 \]

\[ Re_\lambda \simeq 10^2 - 10^5 \]

\[ Re_\lambda \simeq 10^4 - 10^8 \]

The parameter space divide is one of the principal challenges that both simulation and scaled experiment face in understanding physical processes in nature.
Large-scale homogeneous, isotropic compressible fully-developed turbulence:

- $1856^3$ base grid size
- $256^3$ Lagrangian tracer particles
- 3D turbulent RMS Mach number = 0.3 (1D = .17) in steady-state
- $Re_\lambda \sim 600$
- Roughly one week wall clock on 65,536 processors in CO mode
We will use this overlap between simulation and experiment to ask the question:

How well can we simulate fully-developed turbulence?
Fundamental Equations

- Fundamental equations solved are the conservative form of the Euler equations, with stochastic forcing (Eswaran & Pope, 1988). The system is closed assuming an ideal equation of state with $\gamma = 7/5$:

\[
\begin{align*}
\frac{\partial \rho}{\partial t} + \nabla \cdot (\mathbf{v} \rho) &= 0 \\
\frac{\partial \rho \mathbf{v}}{\partial t} + \nabla \cdot (\mathbf{v} \mathbf{v} \rho) &= -\nabla P + \mathbf{F} \\
\frac{\partial \rho E}{\partial t} + \nabla \cdot [\mathbf{v} (\rho E + P)] &= \mathbf{F} \cdot \mathbf{v} \\
\end{align*}
\]

\[
P = (\gamma - 1) \rho U
\]

\[
E = U + \frac{1}{2} \rho \mathbf{v}^2
\]
Visualization of Lagrangian Tracers
Hierarchy of Fidelity in Turbulence Modeling

- **Direct Numerical Simulation (DNS)**
  - Resolves Kolmogorov scale
    \[ n \simeq 2 - 4\Delta x \]

- **Large Eddy Simulation (LES)**
  - Introduces a subgrid model below the filter scale \( \lambda_{SGS} \)

- **Reynolds-Averaged Navier Stokes (RANS)**

Modeled in RANS

Simulated in DNS

Modeled in LES
Hierarchy of Fidelity in Turbulence Modeling

- Implicit Large Eddy Simulation (ILES)
  - Numerical solution to Euler equations
  - Introduces an effective subgrid model and an effective viscosity through numerical dissipation

\[ \eta \approx \Delta x \]

\[ k^{-5/3} \]

- Modeled in RANS
- Simulated in DNS
- Modeled in ILES
Universality of Lagrangian Structure of Turbulence

\[ S_p(\tau) = \langle |v(t + \tau) - v(t)|^p \rangle \propto \tau^{\zeta_p} \] 

(Arneodo et al, 2008)
Astrophysical Simulation Modeling

• Current turbulent modeling succeeds because the velocity field is universal

• Turbulence modeling may pose significant challenges to future astrophysical studies of coupled multifluid, multiphysics processes:
  • Turbulent Combustion (SNe Ia)
  • Turbulent Mixing (Planet Formation, Convective Mixing in Stars, etc.)
Buoyancy-Driven Turbulent Nuclear Combustion

- Ongoing work targets the issue of turbulent nuclear combustion
- Simulations resolve the Gibson scale and the flame-polishing scale
- Adaptive-mesh refinement calculations using FLASH3 up to full scale of ANL BG/P Intrepid, $\sim 10^5$ cores and $10^5$ grids
Buoyancy-Driven Turbulent Nuclear Combustion

(Bair et al, 2011)
Summary Part I

• ILES models of turbulent flows, commonly used in astrophysics and geophysics, can capture the detailed properties of homogeneous, isotropic turbulence with great accuracy.

• Inhomogeneous, anistropic, or multiphysics turbulent flows present challenges to simplified turbulence modeling.
II. Type Ia Supernovae
What are SNe Ia?

• No one knows.

Two leading models.

Single Degenerate

Double Degenerate
Why Are SNe Ia Fundamentally Possible?

• Two remarkable coincidences make the thermonuclear explosion of C/O white dwarfs fundamentally possible:

• 1) The maximum mass a white dwarf may attain is close to one solar mass.

\[ M_{\text{ch}} \sim (\frac{hc}{G})^{3/2} \frac{1}{m_p^2} \sim \left( \frac{m_{\text{Planck}}}{m_p} \right)^2 m_{\text{Planck}} \]
Why Are SNe Ia Fundamentally Possible?

The total nuclear energy release of a C/O white dwarf exceeds its gravitational binding energy.

\[ E_{\text{nuc}} = \ldots \approx 2 \times 10^{51} \text{ ergs} \]

\[ E_{\text{bind}} \approx 3 \times 10^{51} \text{ ergs} \]
Our most recent work overturns both of these fundamental assumptions about SNe Ia.
Why are Type Ia Supernovae Important?

- It is possible to fit most known Type Ia supernovae to a universal light curve using one free parameter specified by width of the curve. (Phillips, 1993)

- By calibrating the light curve to nearby supernovae, the Phillips relation allows astronomers to use Type Ia events as standard candles of cosmological distances.

- SNe Ia play a key role in the discovery and the characterization of dark energy (Riess et al, 1998; Perlmutter et al, 1999)
Cracking the SNe Ia progenitor problem is crucial to an improved understanding of dark energy.
Type Ia Supernovae Mechanisms

Flame Ignition
Nomoto, Thielemann, Yokoi (1984)

Pure Deflagration
Niemeyer, Hillebrandt, Woosley (1996)

Deflagration to Detonation Transition
Khokhlov (1991)

Gravitationally-Confined Detonation
Plewa, Calder, Lamb (2004)
Fundamental Equations

- Equations solved are Euler equations of hydrodynamics coupled to Poisson’s equation for self-gravity and an advection-diffusion reaction model of combustion front:

\[ \frac{\partial \rho}{\partial t} + \nabla \cdot (\mathbf{v} \rho) = 0 \]

\[ \frac{\partial \rho \mathbf{v}}{\partial t} + \nabla \cdot (\mathbf{v} \mathbf{v} \rho) = -\nabla P - \rho \nabla \Phi \]

\[ \frac{\partial \rho E}{\partial t} + \nabla \cdot [\mathbf{v} (\rho E + P)] = \rho \mathbf{v} \cdot \nabla \Phi + \rho \epsilon_{\text{nuc}} \]

\[ E = U + \frac{1}{2} \rho v^2 \]

\[ \nabla^2 \Phi = 4\pi G \rho \]

\[ \frac{\partial \phi}{\partial t} + \mathbf{v} \cdot \nabla \phi = \kappa \nabla^2 \phi + \frac{1}{\tau} R(\phi) \]

\[ R(\phi) = \frac{f}{4} (\phi - \epsilon_0) (1 - \phi + \epsilon_1) \]
First Successful 3-D Simulation of a Type Ia Detonation

- Background is a cold white dwarf model in initial equilibrium with initial mass 1.38 Msun.

- Nuclear bubble is ignited within a spherical region slightly offset from the center of the white dwarf.

- Supercomputer simulation “marches” this condition forward in time in 3D, using full equations describing the flame, hydrodynamics, and self-gravity.
The Center for Astrophysical Thermonuclear Flashes

Simulation of the Deflagration and Detonation Phases of a Type Ia Supernovae

30 initial bubbles in 100 km radius.
Ignition occurs 80 km from the center of the star.
Hot material is shown in color and stellar surface in green.

This work was supported in part at the University of Chicago by the DOE NNSA ASC ASAP and by the NSF. This work also used computational resources at LBNL NERSC awarded under the INCITE program, which is supported by the DOE Office of Science.
“Failed” Supernovae

Jordan et al, 2012b
Summary of Part II

• Simulations of SNe Ia are providing crucial insights into the explosion mechanism.

• It is very likely that there is more than one channel active, and HPC simulations will help shed light on this mystery.
III. Asymptotically-Large Simulations
A Brief History of Supercomputing

Earth Simulator

<table>
<thead>
<tr>
<th>Year Introduced</th>
<th>Peak Speed (flops)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1940</td>
<td>1E+16</td>
</tr>
<tr>
<td>1950</td>
<td>1E+14</td>
</tr>
<tr>
<td>1960</td>
<td>1E+12</td>
</tr>
<tr>
<td>1970</td>
<td>1E+10</td>
</tr>
<tr>
<td>1980</td>
<td>1E+8</td>
</tr>
<tr>
<td>1990</td>
<td>1E+6</td>
</tr>
<tr>
<td>2000</td>
<td>1E+4</td>
</tr>
<tr>
<td>2010</td>
<td>1E+2</td>
</tr>
</tbody>
</table>

Doubling time = 1.5 yr.
Blue Gene Series

- **BG/L, 2004**
  - 2 Cores/node
  - 700 MHz/core, 512 MB/core
- **BG/P, 2007**
  - 4 Cores/Node
  - 850 MHz/core, 1 GB/core
- **BG/Q, 2012**
  - 18 Cores/node
  - 1.6 GHz/Core, 1 GB per user core
Theory of Ideal, Asymptotically-Large, Explicit Simulations

- Ideal -
  - Perfect load balance
  - Perfect scalability
  - Infinite memory bandwidth (no memory wall!)
  - Neglect cost and power considerations

- Explicit -
  - Timestep limited by CFL Condition
  - Idealized assumptions allow us to focus on deep limits to scalability and strategies to address these
Theory of Ideal, Asymptotically-Large, Explicit Simulations

- First consider scaling behavior of a serial, explicit, 3-D, uniform Eulerian code with $N^3$ cells:
  
  \[ \text{CPU Time} = CN^{4/3} \]
  
  \[ \text{Memory} = CN^{4/3} \]
  
  \[ \frac{\text{CPU Time}}{\text{Memory}} = \frac{4}{3} \]
Theory of Ideal, Asymptotically-Large, Explicit Simulations

- Given maximum memory and CPU time bounds, a serial simulation is memory-bound if

\[ \text{Max Memory} < \frac{\chi_{\text{mem}} C^{3/4}}{X_{\text{CPU}}} \left( \frac{\text{Max CPU Time}}{N_{\text{dyn}}} \right)^{3/4} \]

- In parallel, an ideal simulation has

\[ \text{Max Memory} = \frac{\text{Memory}}{\text{CPU} N_{\text{CPU}}} \]

\[ \text{Max CPU Time} = \text{Max Wall Clock} N_{\text{CPU}} \]
Theory of Ideal, Asymptotically-Large, Explicit Simulations

- The memory-boundedness criterion for a parallel simulation becomes

\[
\frac{\text{Memory}}{\text{CPU}} < \chi_{\text{mem}} \left( \frac{C \text{ Max Wall Clock}}{\chi_{\text{CPU}} N_{\text{dyn}}} \right)^3 \frac{1}{N_{\text{CPU}}} \right)^{1/4}
\]

- Scaling to typical values on a small cluster,

\[
\frac{\text{Memory}}{\text{CPU}} < 0.2 \text{ GB} \left[ \left( \frac{(N_{\text{state}}/10)(C/0.5) \text{ (Max Wall Clock/1wk)}}{\chi_{\text{CPU}}/10 \ \mu s)(N_{\text{dyn}}/10)} \right)^3 \frac{512}{N_{\text{CPU}}} \right]^{1/4}
\]

- Asymptotically-large, ideal, explicit simulations \( (N_{\text{CPU}} \to \infty) \) are always CPU-bound.
Theory of Ideal, Asymptotically-Large, Explicit Simulations

• Adaptive Mesh Refinement is a method to dynamically refine logically Cartesian meshes
Theory of Ideal, Asymptotically-Large, Explicit Simulations

• Consider an ideal AMR simulation with a total $N_{blocks}$ of $N_{grid}^3$ cells

$$\text{Wall Clock} = \left( \frac{\chi_{CPU} N_{\text{dy}}}{C N_{CPU}} \right) N^4$$

• Fixing the wall clock time barrier,

$$N \propto N_{CPU}^{1/4}$$
Theory of Ideal, Asymptotically-Large, Explicit Simulations

- The distribution of blocks over cores, fixing the wall clock time barrier and grind time,

\[
\frac{N_{\text{blocks}}}{N_{\text{CPU}}} = \frac{1}{N_{\text{grid}}^3} \left[ \frac{C(\text{Wall Clock})}{\chi_{\text{CPU}} N_{\text{dyn}}} \right]^{3/4} N_{\text{CPU}}^{-1/4}
\]

\[
\frac{N_{\text{blocks}}}{N_{\text{CPU}}} = 12 \left( \frac{32}{N_{\text{grid}}} \right)^3 \left[ \frac{(C/0.5)(\text{Wall Clock}/1 \text{ wk})}{(\chi_{\text{CPU}}/10 \mu s)(N_{\text{dyn}}/10)} \right]^{3/4} \left( \frac{N_{\text{CPU}}}{10^6} \right)^{-1/4}
\]
Theory of Ideal, Asymptotically-Large, Explicit Simulations

- As we go beyond the petascale, AMR simulations will face increasing tight load-balancing issues.
- Possible strategies:
  - Multithreading
  - Smaller block sizes
  - Improved load-balancing algorithms
  - Faster grind times through GPU or other technologies
Conclusions

• Continued success for computational astrophysics at scale will hinge upon our ability as a community to
  
• Think deeply about modeling of turbulence in ways not yet manifested in existing codes
  
• Think deeply about the ultimate limits to scalability and beginning to take long-term strategic directions to address these