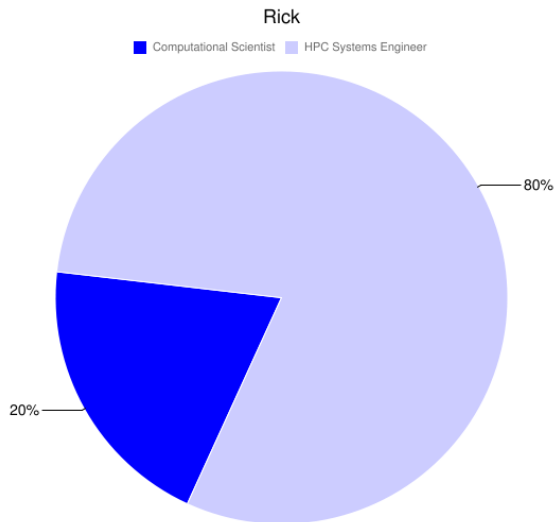# Computational Astrophysics with Enzo

Rick Wagner
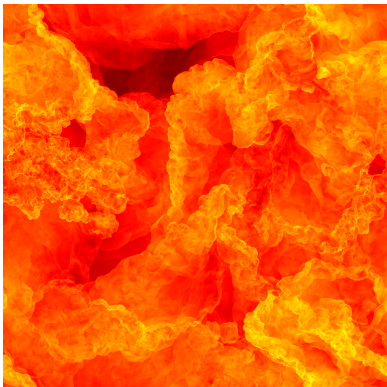
San Diego Supercomputer Center,
University of California, San Diego

CScADS / July 26, 2010
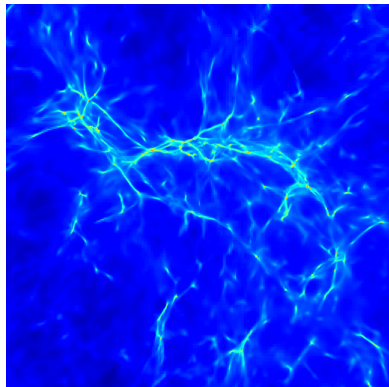
Duties & Responsibilities

Rick

■ Computational Scientist   ■ HPC Systems Engineer

80%

20%

Astrophysics: Turbulence & Cosmology
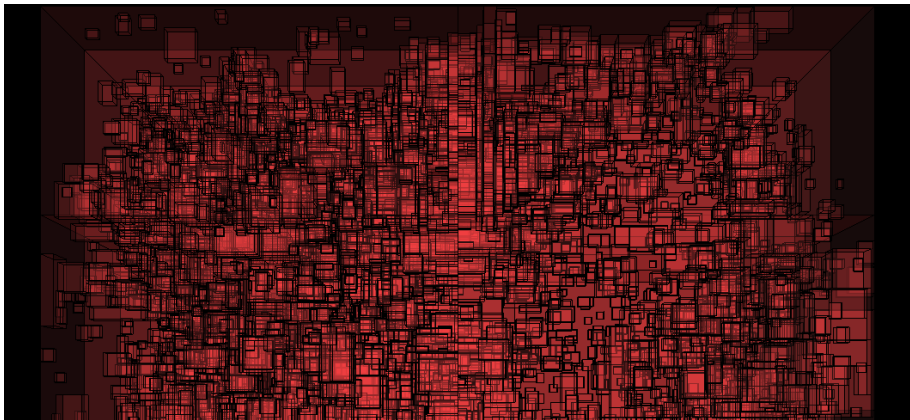

Driven isothermal turbulence in a 2048^3 uniform grid (Kritsuk, et al., 2007).

Halos at $z = 6$; 1024^3 root grid, 7 levels of refinement

SDSC UCSD

Our Tool: Enzo . . .

. . . an adaptive mesh refinement (AMR), grid-based hybrid code (hydro + N-Body).
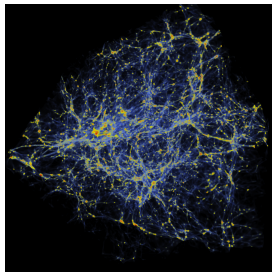

Grids in an AMR simulation.

Big Numbers: Blessing & Curse

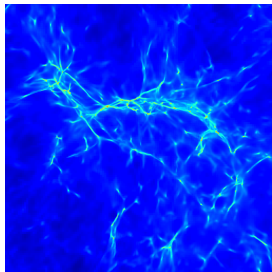*Yesterday's* simulations were big in many ways:

Raw data

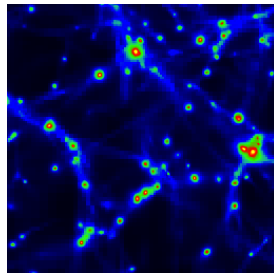4096^3 uniform grid:
> 3 TB per dataset

Data structures

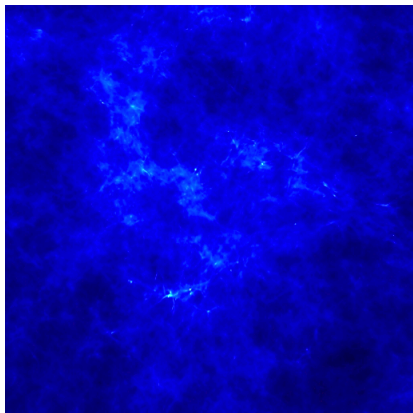1024^3 root grid, 7
levels of AMR:
$> 400,000$ grids

Interesting objects

512^3 Mpc/h AMR
simulation:
$> 50,000$ galaxy
clusters

Naturally, this isn't going to stop.

SDSC UCSD

The "Write Now, Analyze Later" Pattern Has Limits

512^2 section of a 6400^3 simulation.

- Full datasets from this simulation are ~ 30 TB
- 30 datasets (normally trivial) is almost a petabyte
- Unless a magical file system emerges, we need to reduce this at a much faster pace
- Tradeoff is the loss of going back with new questions

Things That We've Done

Subset the Data

Either

- Selective full outputs, or
- partial datasets with selected fields.

Issues:

- Still uses disk
- Limits the time or data resolution

Inline Analysis

I.e., when needed, stop the simulation, and do analysis

Issues:

- Blocks simulation
- Wasteful if analysis doesn't scale

My Hope: Co-Scheduled, Asynchronous, Analysis

Send In-Memory Data to Another Computational Resource

Issues:

1. Data transformation
   - Does Enzo create and send VisIt data structures?
   - Does VisIt learn to read Enzo data structures?
2. "Transportation" layer
   - MPI communicators?
   - TCP streams?
3. Network bandwidth
4. Co-scheduling is non-trivial, unless you own every resource

1 & 2 are what I'm hoping to work on this week.
But I'm happy to talk about 3 & 4.