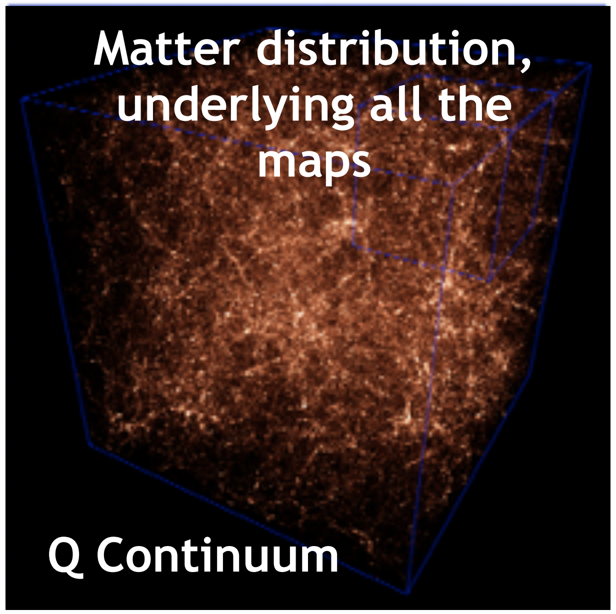
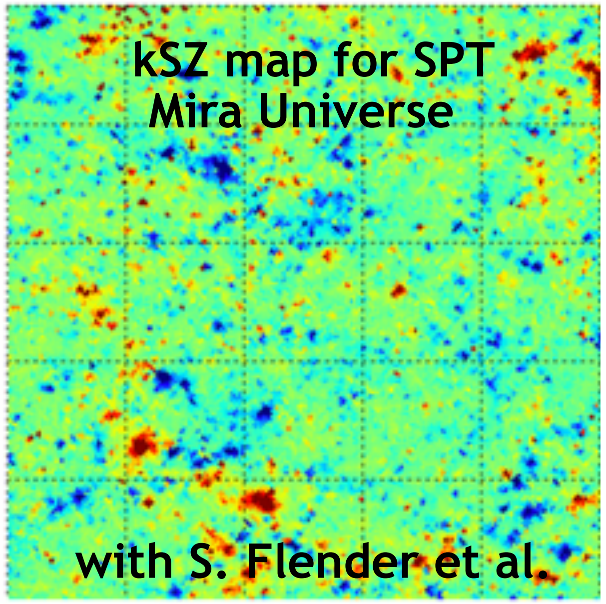
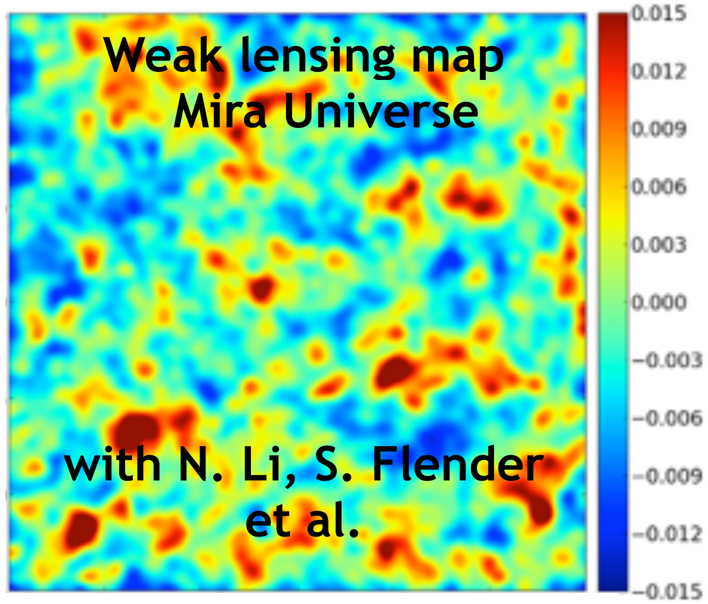
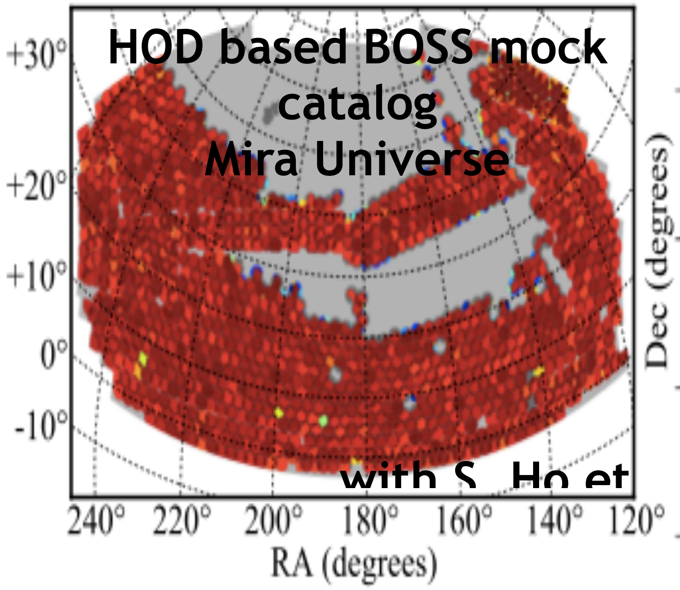
Cosmological Simulations: Under the Hood

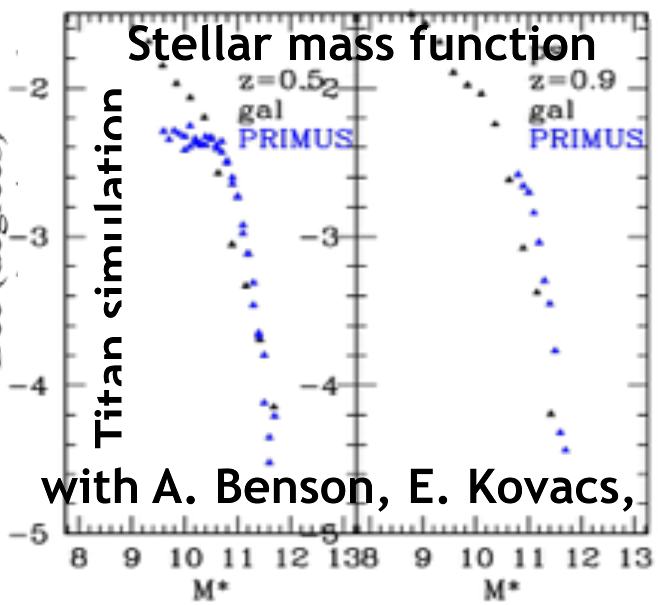
Salman Habib
Argonne National Laboratory
Kavli Institute for Cosmological Physics

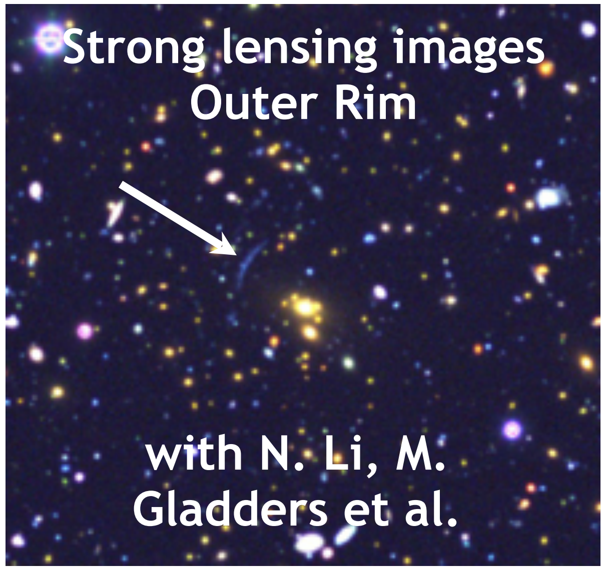












Mexican Numerical Simulations School
Lecture 3, October 5, 2016

More on Simulations —

- **Initial conditions**
  - Why can't we run the Universe backwards?
  - How do we know what the Universe is doing at early times?
  - How can we guard against initial condition artifacts?
- **Time-Stepping**
  - What is a symplectic integrator and why do we use it?
  - Hierarchical time-stepping
  - Time-stepping errors and shadowing
- **Considerations for running simulations**
  - Effect of the box size, initial redshift, force, and mass resolution, etc. etc.
- **Analysing simulations**
  - What are the kinds of analyses we need to carry out? (halos, subhalos, merger trees, density estimation, etc.)
  - Approaching the "virtual universe"

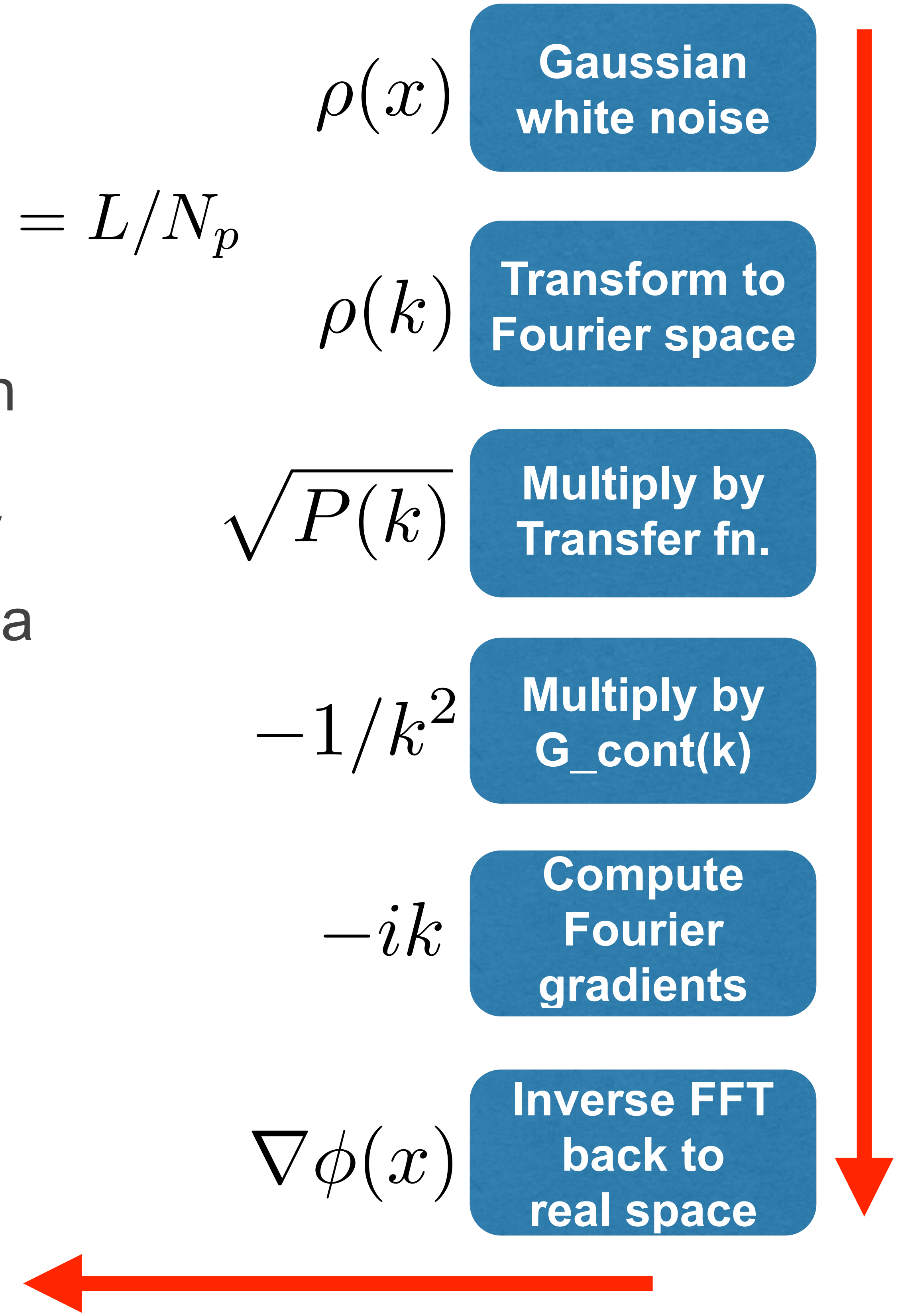
Basic Method for Initial Conditions

Linear regime of structure formation

- Given the box size and the number of particles, we have two scales,

$$k_{box} = 2\pi/L \text{ and } k_{Ny} = \pi/\Delta_p, \text{ where } \Delta_p = L/N_p$$

- Want the first to be large enough to avoid finite-size effects and z_in large enough such that the initial power spectrum is definitely linear up to the particle Nyquist wavenumber
- Can start particles on a grid (quiet start) or via a “glass” initial condition
- Generate a Gaussian random field as a realization of the desired linear power spectrum (can do this in real or k-space)
- Move particles using the Zel’dovich (or higher-order) approximation to obtain the IC at z_in (such that the linear P(k) is properly normalized at z=0)



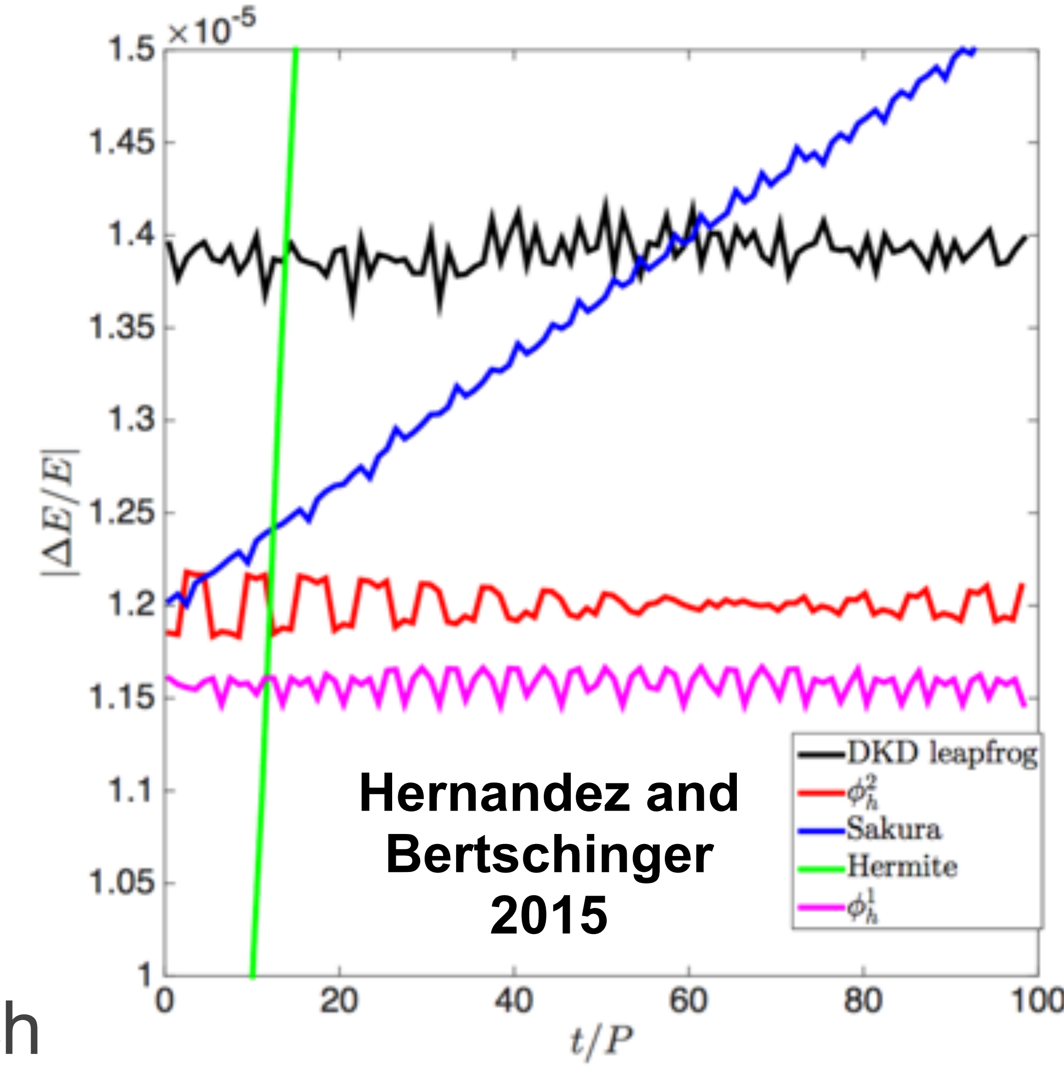
More on Initial Conditions

- **Technical issues**
  - Parallel random number generation — we use counter-based random number generators (CBRNGs) (Salmon, Moraes, Dror, and Shaw 2011) https://www.deshawresearch.com/resources_random123.html
  - Glass ICs have suppressed power at large k (‘frozen’ Coulomb gas instability under gravity)
  - Multiple-species ICs still not an unsolved problem
- **Dynamical issues**
  - The actual initial condition is wrong; does it hurt us? (The Universe does not look like a grid at z=z_in)
  - Known issues with high-k artifacts due to the grid (especially for cut-off initial power spectra; “string of pearls” effect)
  - Need time-varying smoothing of the density field in a way that does not compromise solution accuracy, still work in progress
  - For LCDM-like models most high-z IC problems do not appear to be a major issue at low-z (but one should still worry)

See Heitmann et al. 2010

Time-Stepping: Symplectic Integrators

- **Symplectic integration (1980's onwards)**
  - VPE is Hamiltonian; need to preserve key aspects of this essential structure
  - Naive time-stepping not compatible with Hamiltonian property of the VPE
  - History — plasma community tried 4th-order RK, failed to time-reverse Landau damping, tried 2nd-order leapfrog, worked (understanding came later)
  - Many different ways to think of symplectic integration, we use the split-operator approach (simple method with no storage overhead)
  - Instead of discretizing the equations of motion, we discretize formal solutions written in operator form (Lie-algebraic approach)
  - Long-time stability/Energy surface (Ge's theorem), respect global topology



Symplectic integrators evolve an exact Hamiltonian problem that is 'close' to the problem of interest, they have a number of attractive properties such as conservation of Poincare invariants, time-reversal symmetry, etc.

Basic Formalism I

- **Hamilton's equations**

$$\frac{dz}{dt} = -[H, z] \text{ where } z = (x, p)$$

- **Liouville's equation (also a Hamiltonian flow)**

$$\frac{df}{dt} = \frac{\partial f}{\partial t} - [H, f] = 0$$

- **Formal solution of Hamilton's equations in exponentiated form**

$$z(t) = \exp(-t : H :) z(0) \text{ where}$$

$$: H : g = [H, g] \text{ (Dragt notation)}$$

- **For a two-term Hamiltonian**

$$z(t) = \exp(-t(: H_1 : + : H_2 :)) z(0)$$

Basic Formalism II

- Note operators do not commute, so use Campbell-Baker-Hausdorff

$$\exp(A)\exp(B) = \exp\left(A + B + \frac{1}{2}\{A, B\} + \frac{1}{12}(\{A, \{A, B\} + \{\{B, A\}, B\}) + \cdots\right)$$

where $\{,\}$ denotes the commutator

- Applying CBH, we can show that

Composition of exactly known solutions if H_1 is the kinetic energy and H_2 is the potential energy

$$\exp(-t(:H_1: + :H_2:)) =$$

$$\exp\left(-\frac{1}{2}t:H_1:\right)\exp\left(-t:H_2:\right)\exp\left(-\frac{1}{2}t:H_1:\right) + O(t^3)$$

$$\boxed{\equiv \mathcal{M}_1(\frac{1}{2}t)\mathcal{M}_2(t)\mathcal{M}_1(\frac{1}{2}t) + O(t^3)}$$

By construction, the error term is also Hamiltonian

Basic Formalism III

- **H_1 and H_2**
  - If H_1 is the kinetic term of the Hamiltonian, then M_1 is the exactly known "drift" or "stream" map (D or S)
  - If H_2 is the potential piece, then M_2 is the exactly known "kick" map
  - Symplectic integrators can be either SKS or KSK (obvious symmetry)
  - Higher-order integrators can be easily found (but usually not needed in the curent cases of interest)
- **Hierarchical time-stepping**
  - Slow/fast split (Long-range force is "slow", short-range is "fast")

$$H = H_{sr} + H_{lr} \text{ where } H_{sr} = p^2/2m + \phi_{sr} \text{ and } H_{lr} = \phi_{lr}$$

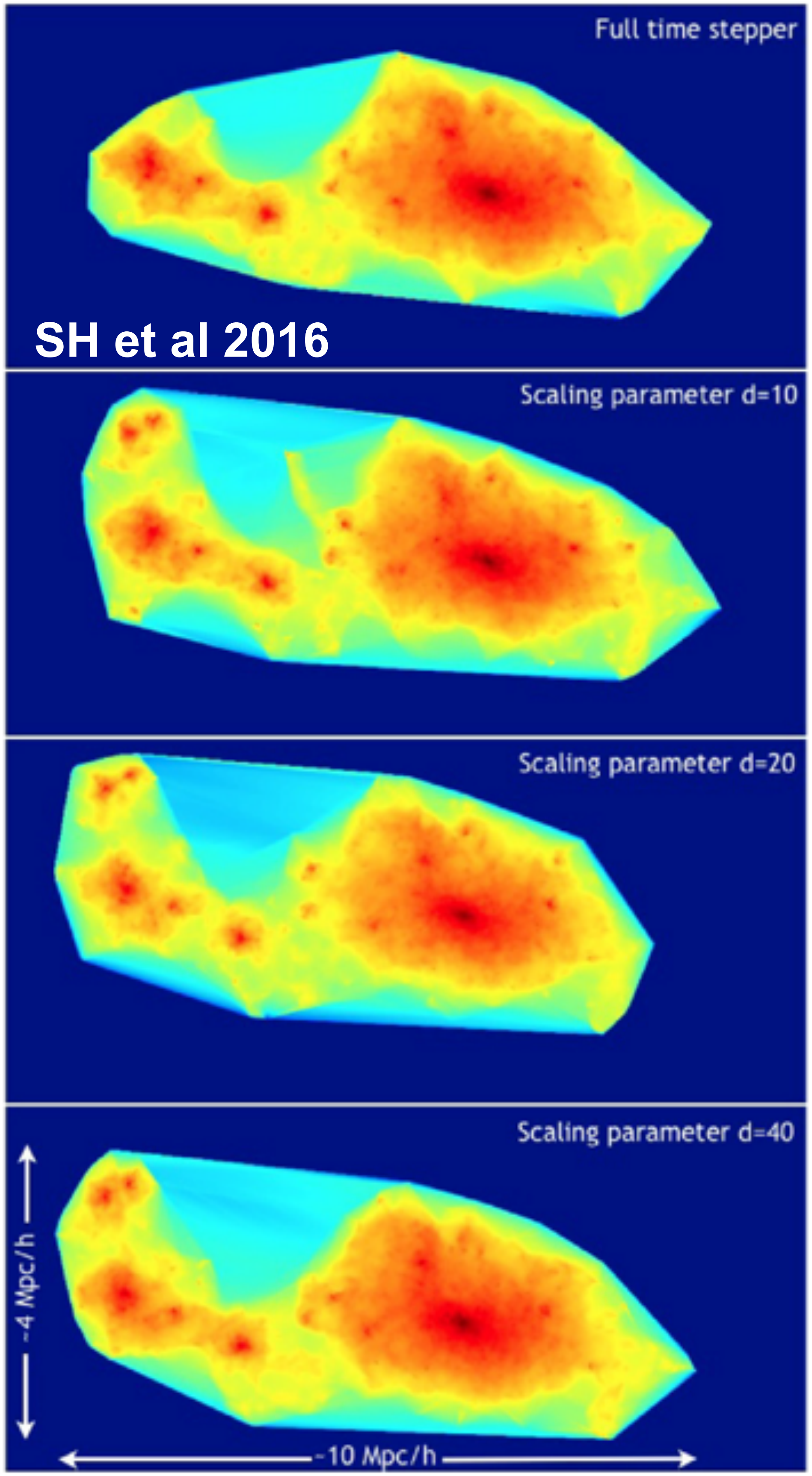
$\mathcal{M}_{sr}$ is the usual second-order SKS integrator

$\mathcal{M}_{lr}$ is a long-range "kick" map

$$\boxed{\mathcal{M}_{full}(t) = \mathcal{M}_{lr}(t/2)(\mathcal{M}_{sr}(t/n_c))^{n_c}\mathcal{M}_{lr}(t/2)}$$

Basic Formalism IV

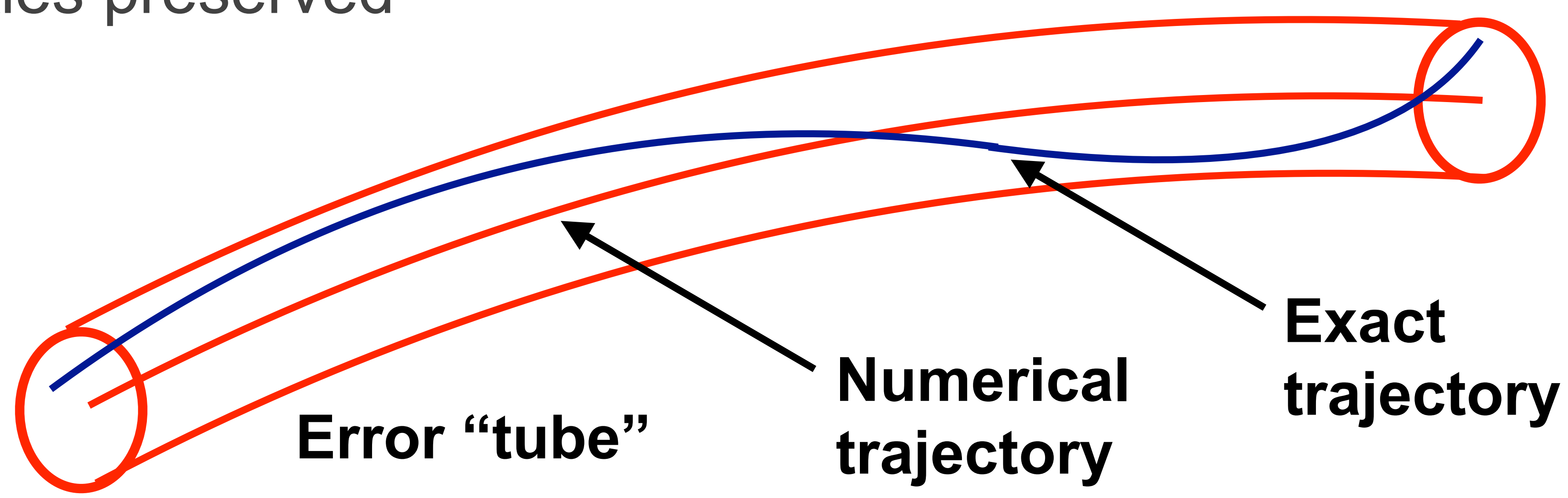
- **Individual Particle time-steps**
  - Depending on some error control criterion (local density, acceleration, —), each particle can be acted on by fewer or more short-range kick maps; to ensure global time synchronization all particles get the same number of stream steps
  - Even in non-extreme cases, this method can lead to time-savings of roughly a factor of two
  - Exercise: write a symplectic integrator for the gravitational 3-body problem



Coarse time-steps (individual particles)

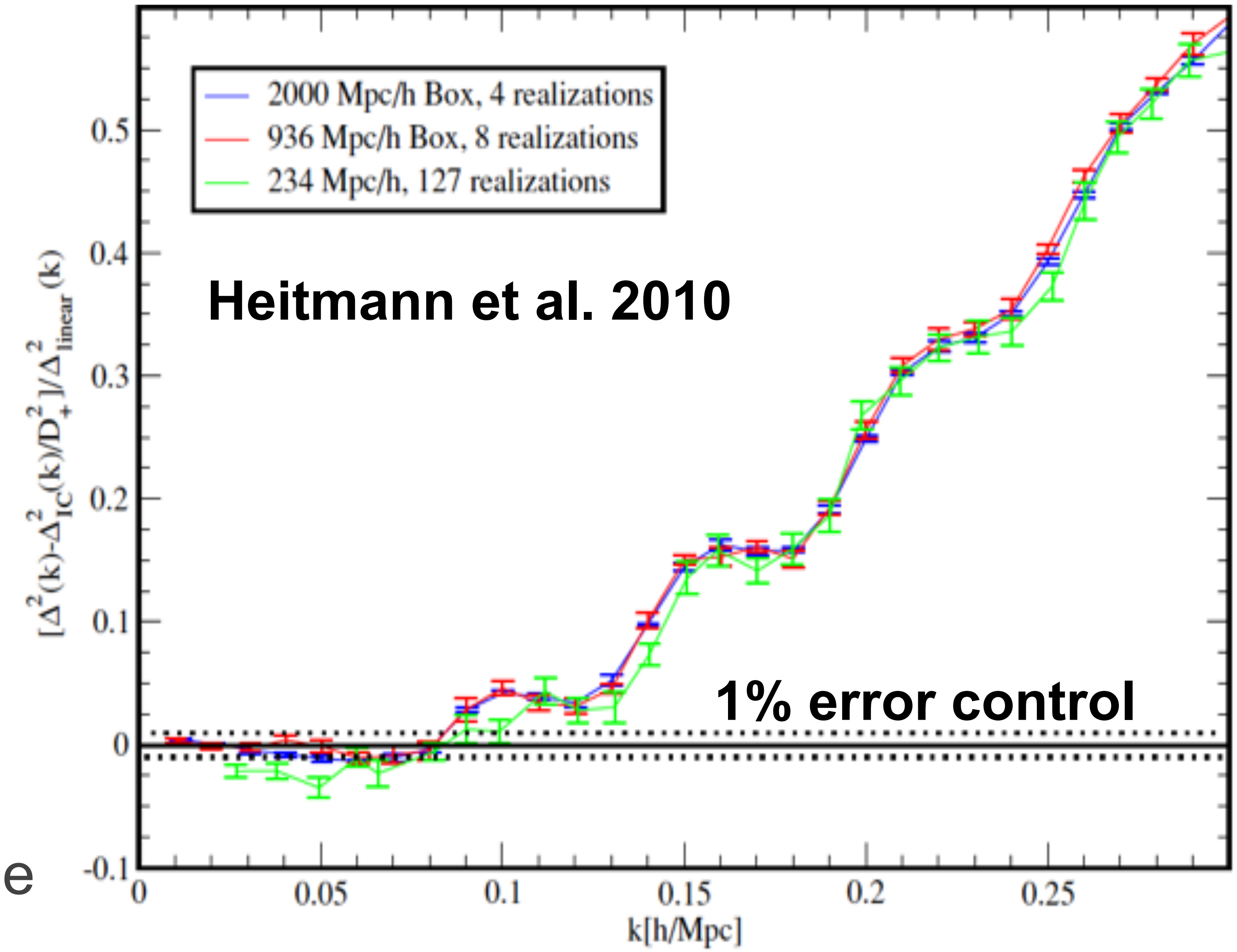
Open Issues

- **Symplectic integration is not perfect**
  - Energy errors vs. phase errors
  - Can change integrability properties of the original system
- **Shadowing**
  - What is the appropriate notion of error in our particle method approach?
  - Force errors don't have to be controlled to high accuracy (say 0.001, fractionally) because of the unavaoidable discreteness "noise"
  - It is not necessary to worry too much about the absolute time-stepping error (anyway, in high density regions, the trajectories are chaotic, so this would not make much sense); more important to have the global properties preserved



Running Simulations: Practicalities

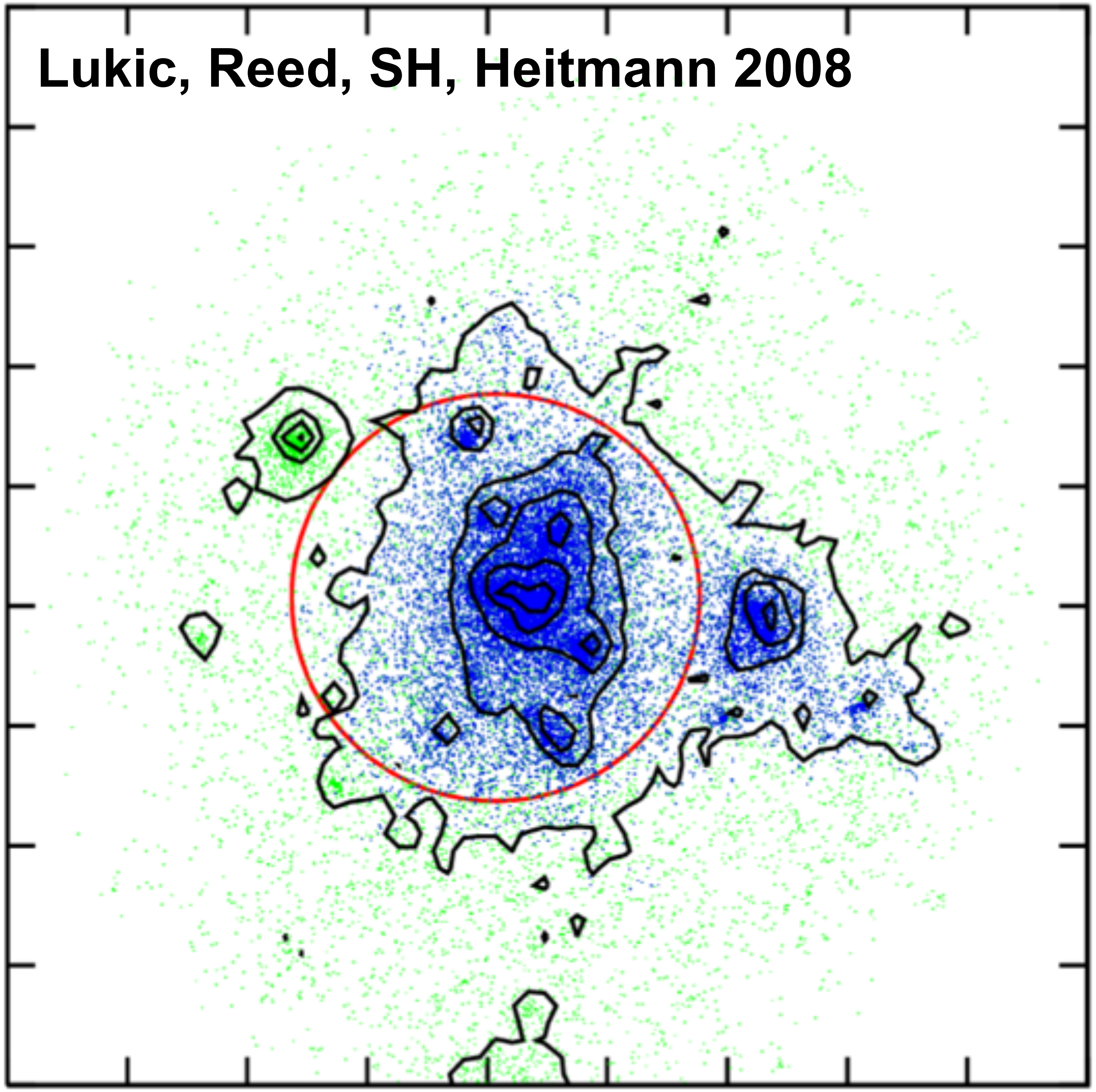
- **Things to keep in mind**
  - If you want to get accurate answers in the quasi-linear regime (percent level) box size is important (~1Gpc)
  - Mass resolution is very important for accurate P(k) determination (and obviously for tracking halos down to lower masses)
  - Initial redshifts should be large enough
  - Time-stepping should show convergence on k-scales of interest



High-statistics error tests for box-size limits, note ~200Mpc/h is too small ; the two larger boxes show good convergence

Analysis of Simulations: Quantities/Objects of Interest

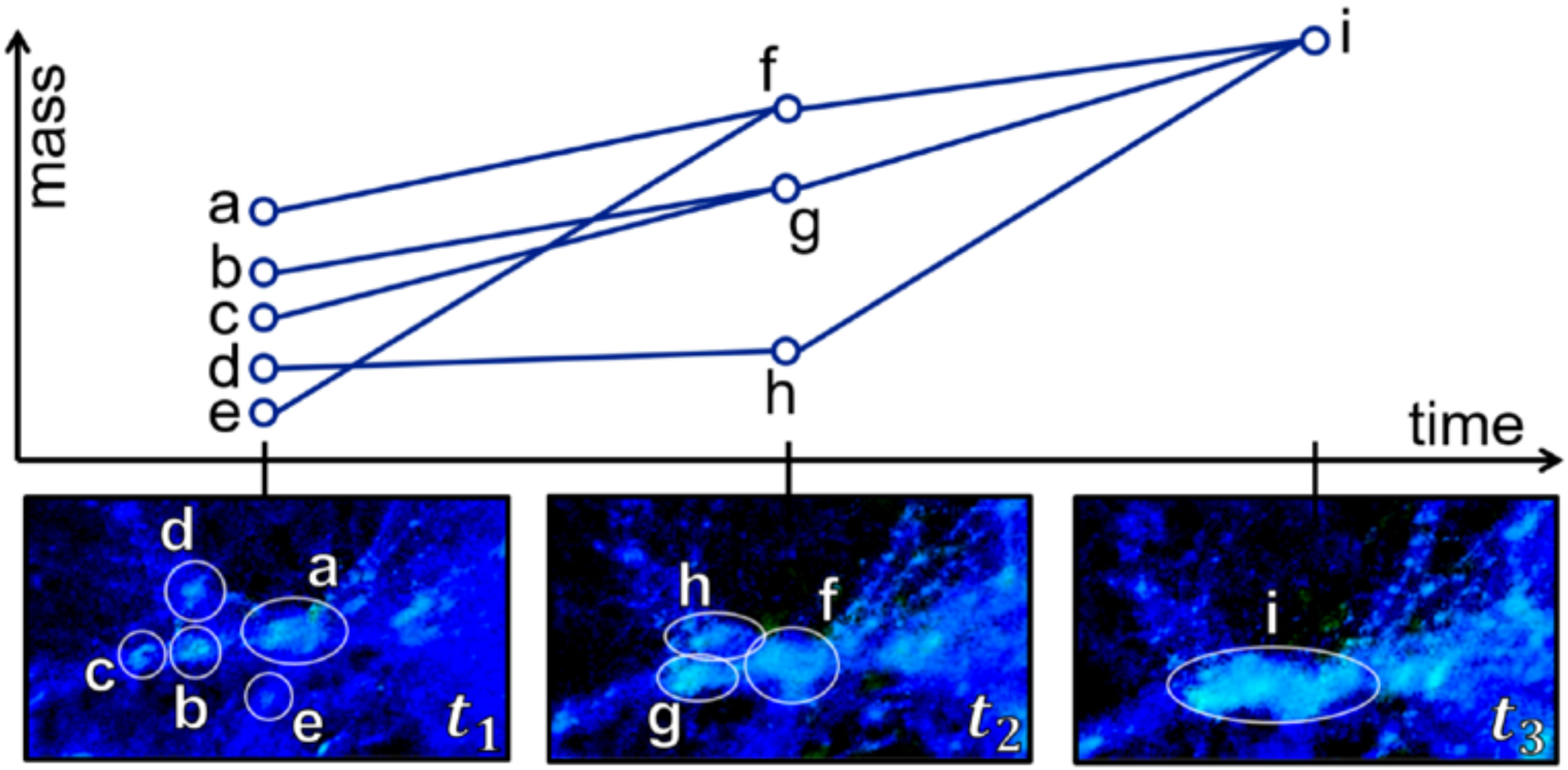
- **Many things to compute**
  - Light-cones (Peder)
  - Halos and merger trees
  - Halo properties
  - Power spectra, correlation functions
  - Mass functions
  - Synthetic catalogs (Peder, Aldo)
  - Weak lensing shear
  - Galaxy-galaxy lensing
  - Cluster strong lensing
  - CMB tSZ and kSZ maps
  - Etc.



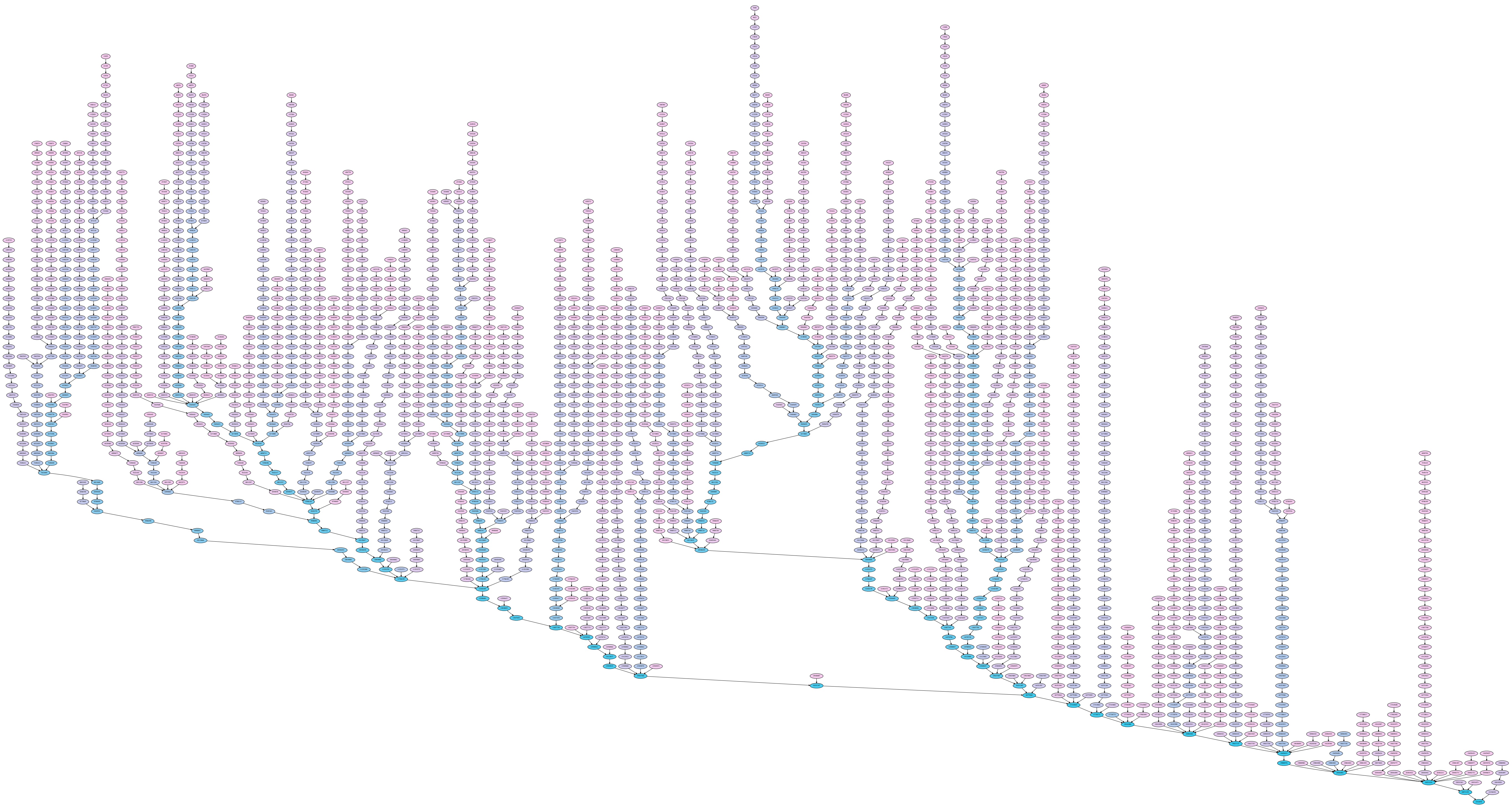
FOF vs. SO halos

Merger Trees

- **Importance/Issues**
  - Very important for a number of reasons (in our case, galaxy modeling)
  - Many tricky technical issues
  - In large simulations, can be exceedingly complex
  - How to disentangle complex halo “graphs”
  - Subhalos vs. “cores” (where do galaxies live)
  - And so on —

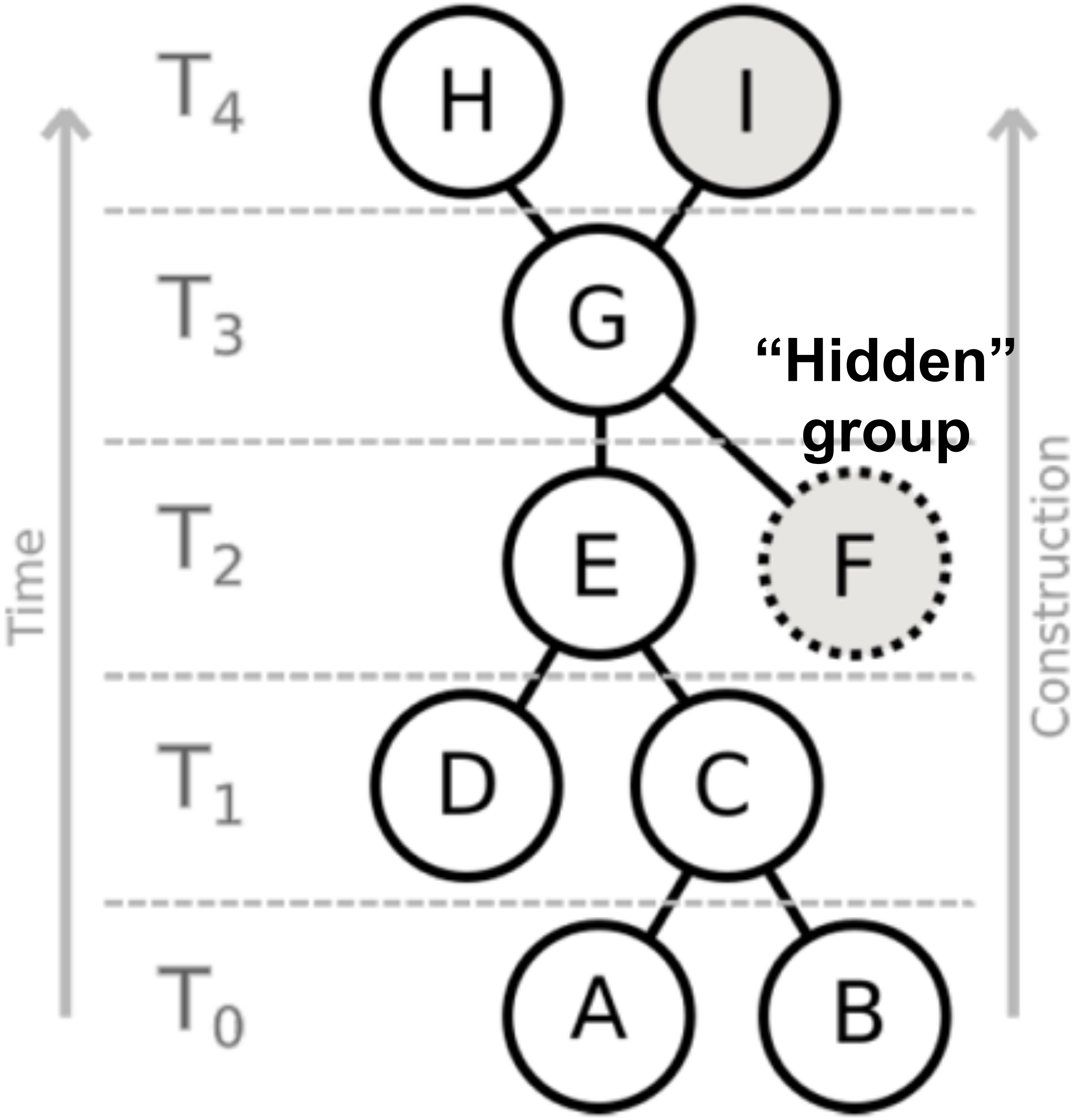


Potential Complexity —

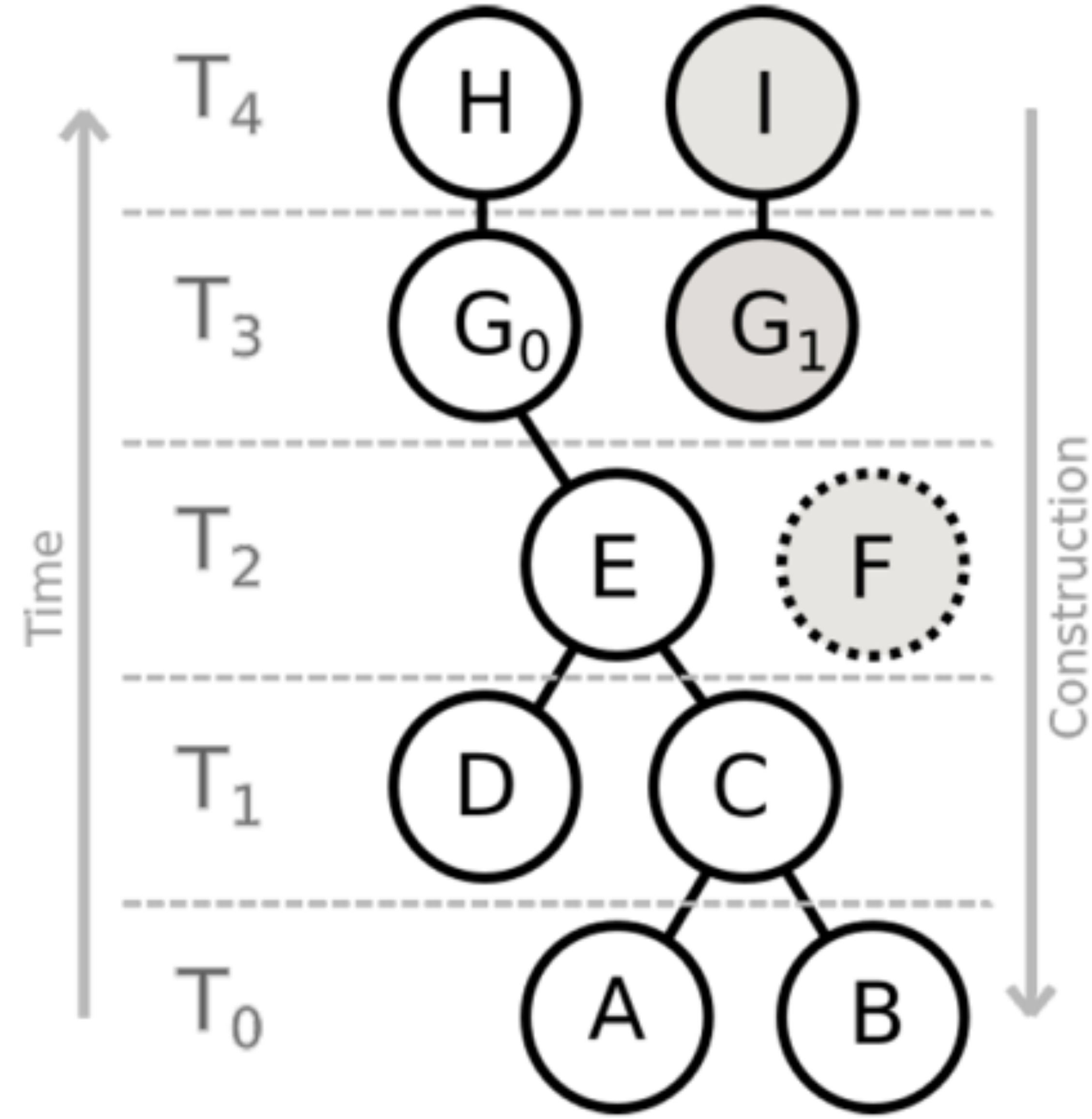


Typical merger tree in a high resolution simulation, statistics of merger trees deserve a real theory — don't have one yet!

Disentangling “Braids”

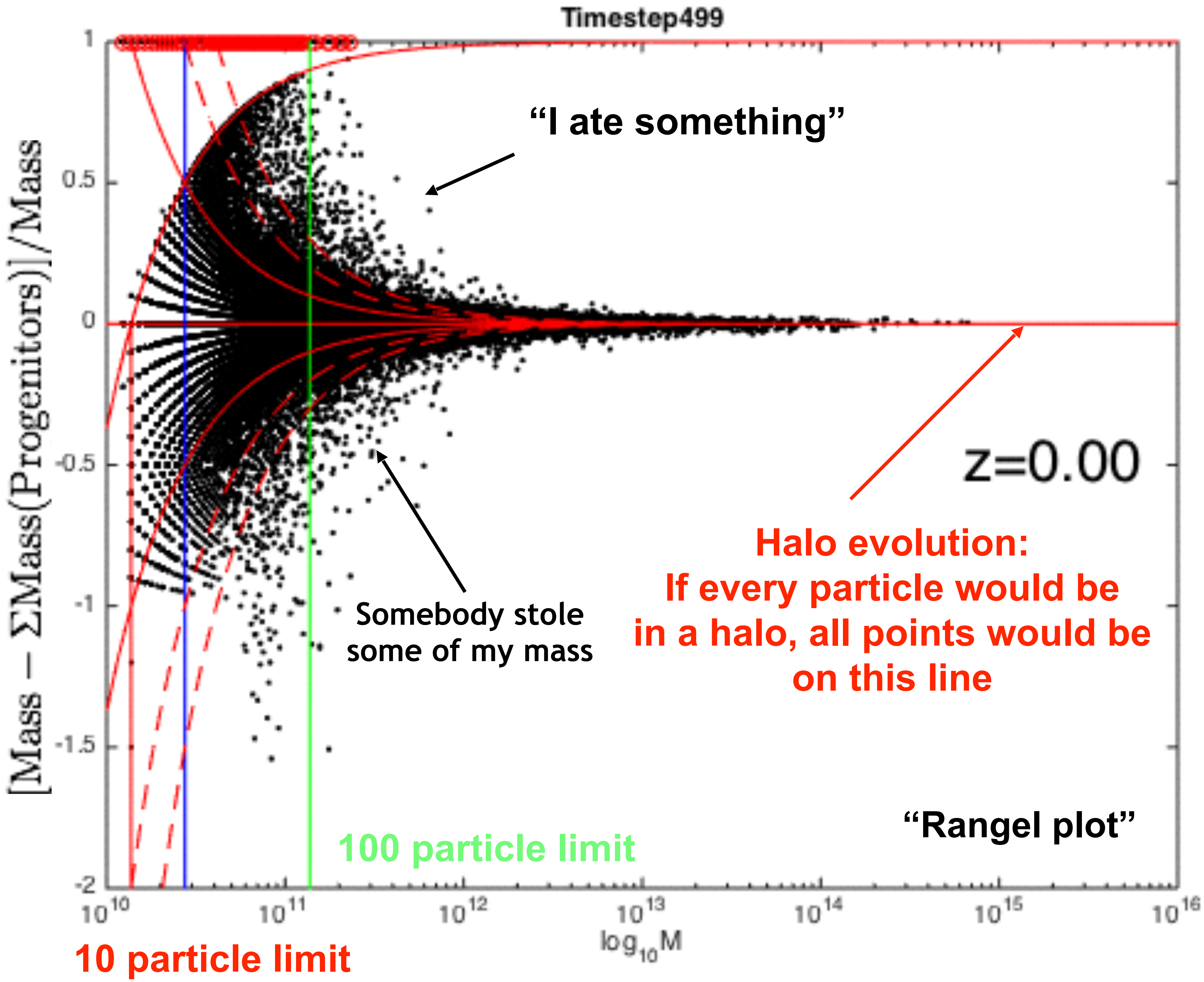


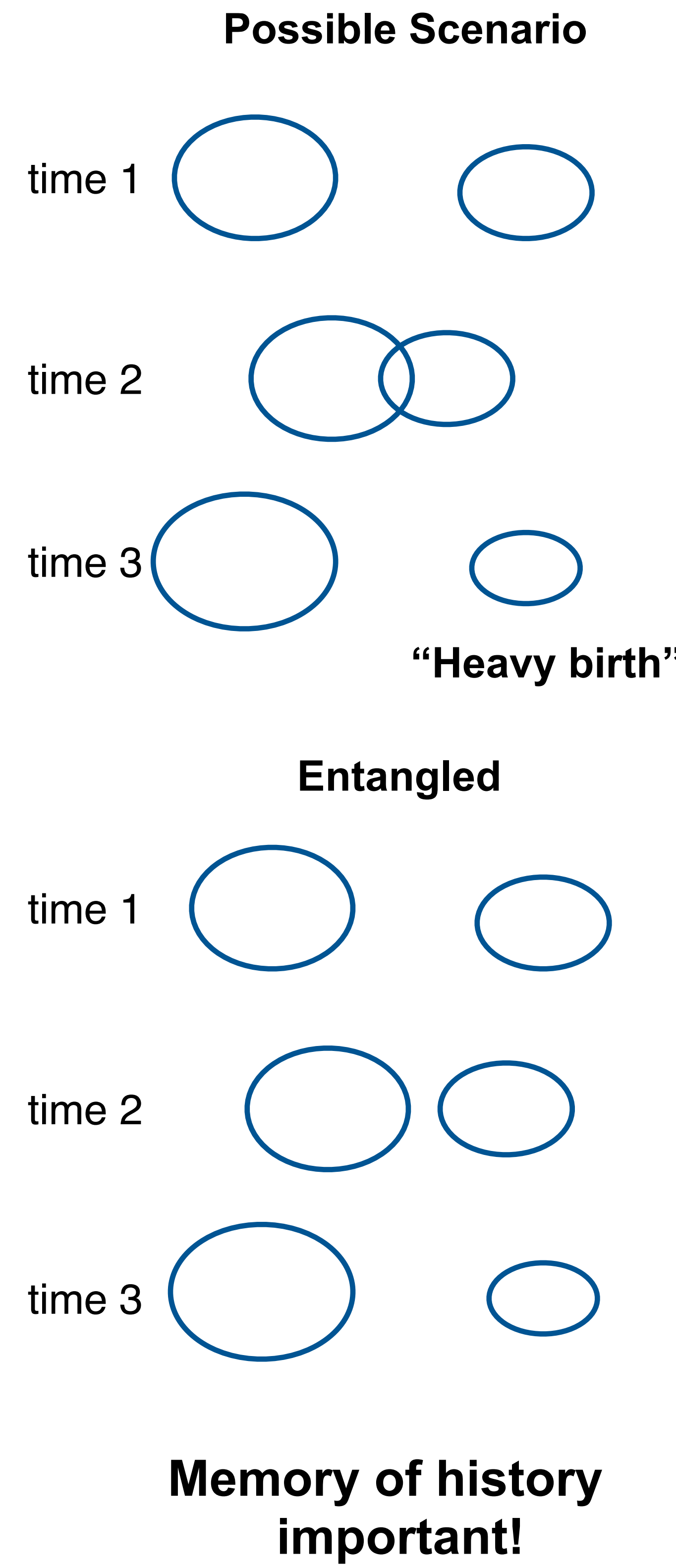
Forward in time construction requires storage of past time-steps



Reverse chronological order construction identifies G as a split node

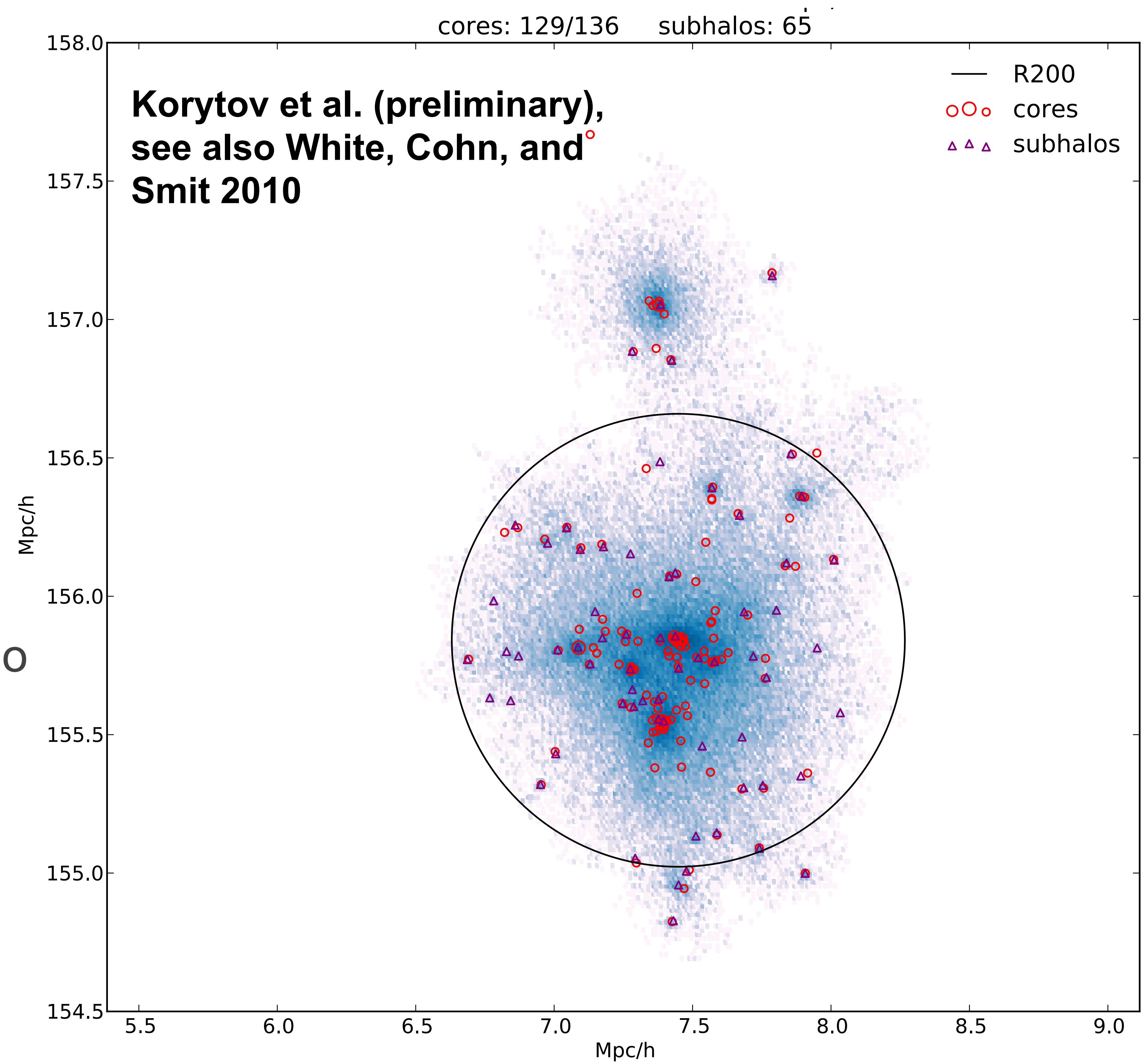
Merger Trees: Checking for Consistency





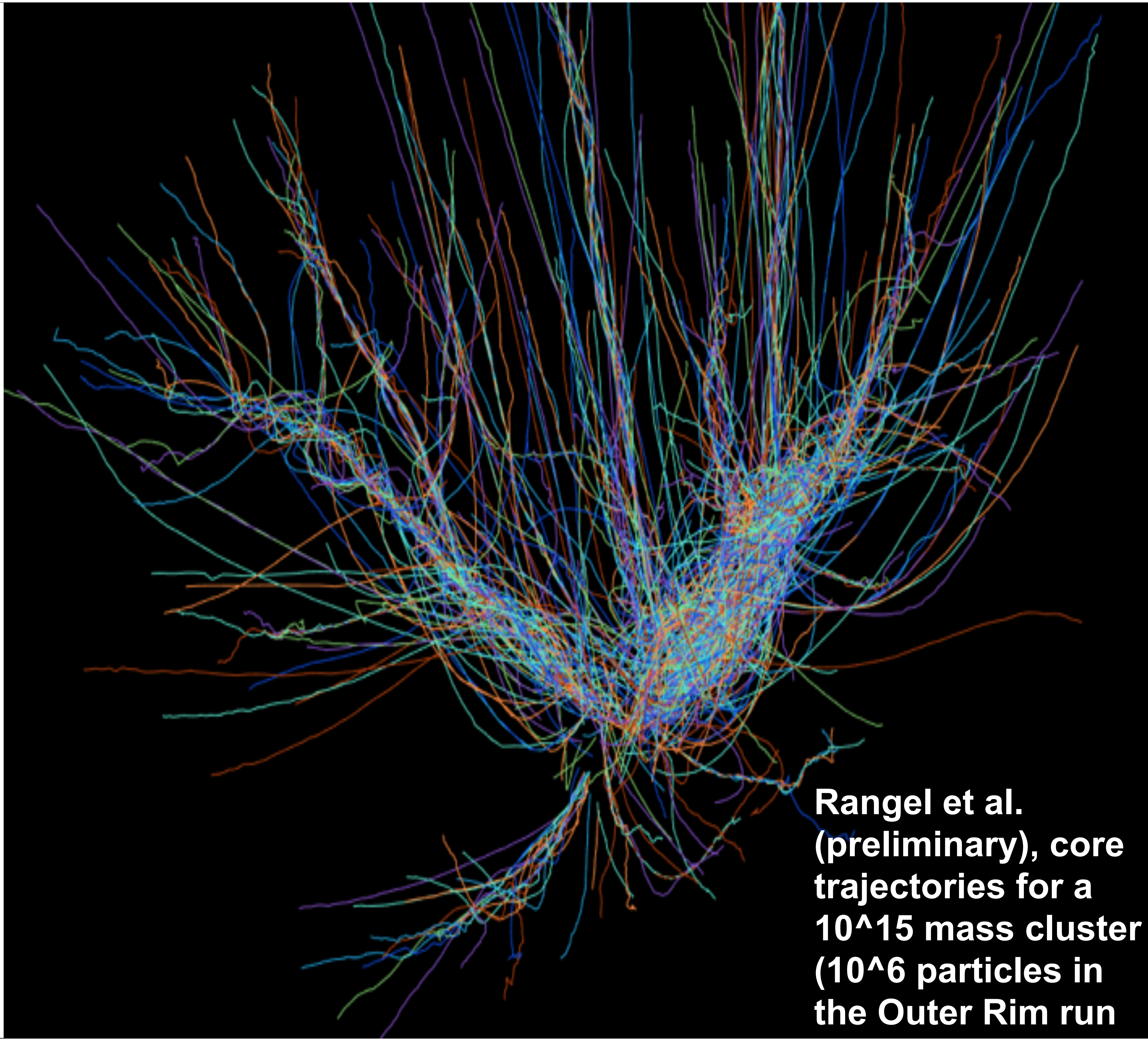
Core-Tracking I

- **Core-tracking**
  - Subhalos can lose most of their mass after infall
  - But galaxies are still there —
  - Need to track halo "cores" to see where galaxies go
  - Note subhalos typically found on halo outskirts
  - Cores have a more NFW-like distribution
  - Core distribution agrees with SDSS clusters (in prep)



Core-Tracking II

- **Status**
  - Completed building of merger trees and core-tracking for Outer Rim run (250 X Millennium)
  - Enormous data set now being analyzed
  - Lots of galaxy science projects possible —

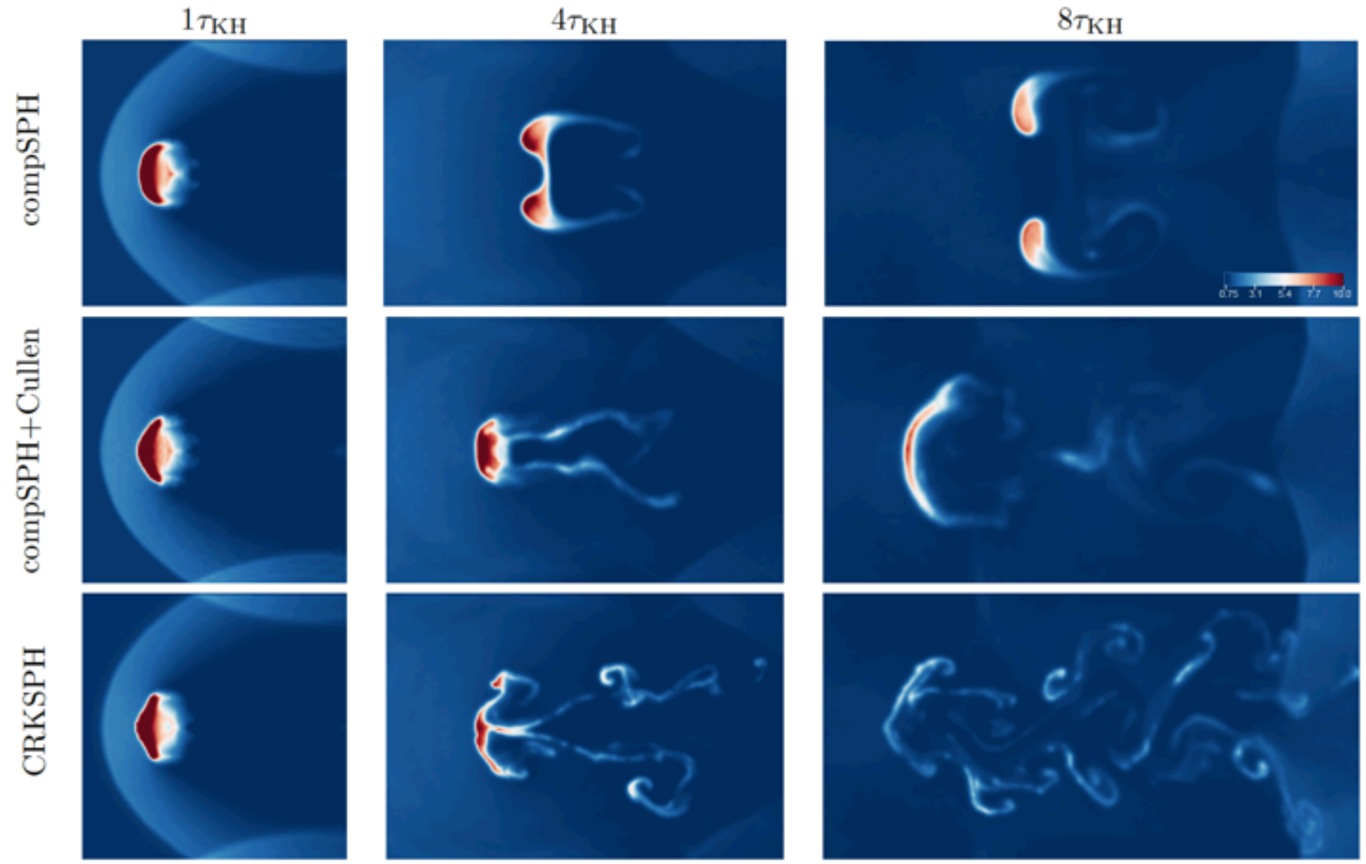


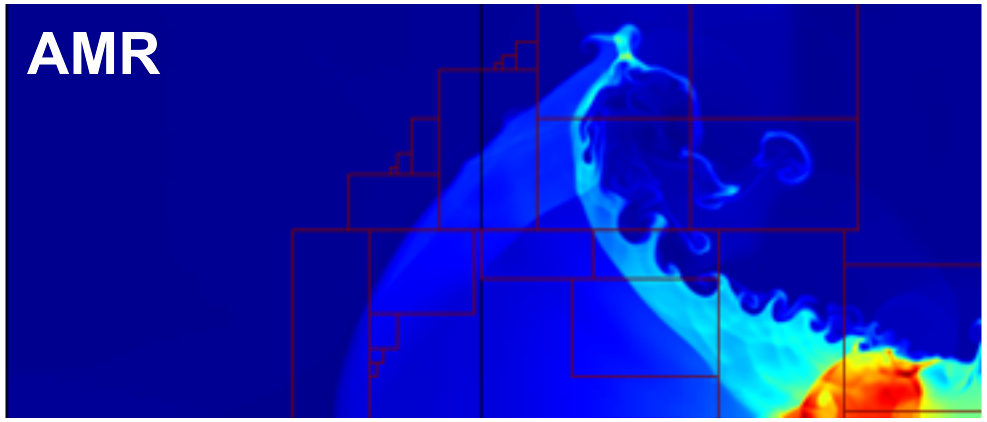
Rangel et al. (preliminary), core trajectories for a 10^15 mass cluster (10^6 particles in the Outer Rim run

Towards a “Virtual” Universe

- **Aim of our cosmological simulation effort**
  - Better understanding of physics in the simulations
  - Go from simulations directly to observations
  - Requires major effort on next-generation systems
  - Adding hydro — CRKSPH (Conservative Reproducing Kernel SPH)
  - Now supported by DOE ECP

ECP
EXASCALE COMPUTING PROJECT





PICS Example: Generating Strong Lensing “Observations”

PICS: Pipeline for Images of Cosmological Strong lensing

Simulated strong lens image (left) to match SPT cluster observations taken with the MegaCAM camera on Magellan (right), work by L. Bleem, M. Florian, S. Habib, K. Heitmann, M. Gladders, N. Li, S. Rangel

N. Li et al., arXiv:1511.03673, ApJ (in press)