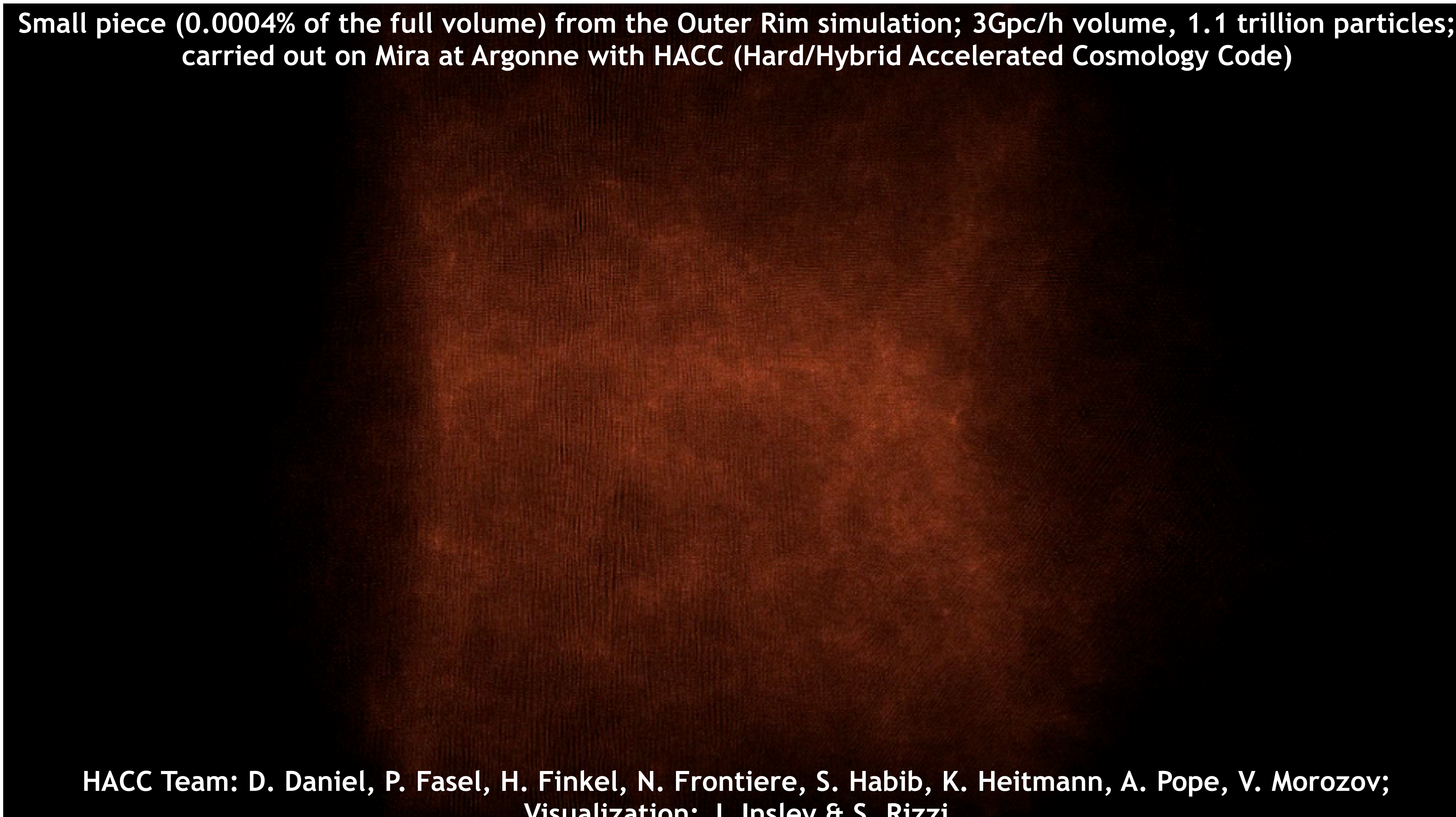
Cosmological Simulations: Under the Hood

Salman Habib
Argonne National Laboratory
Kavli Institute for Cosmological Physics



ool

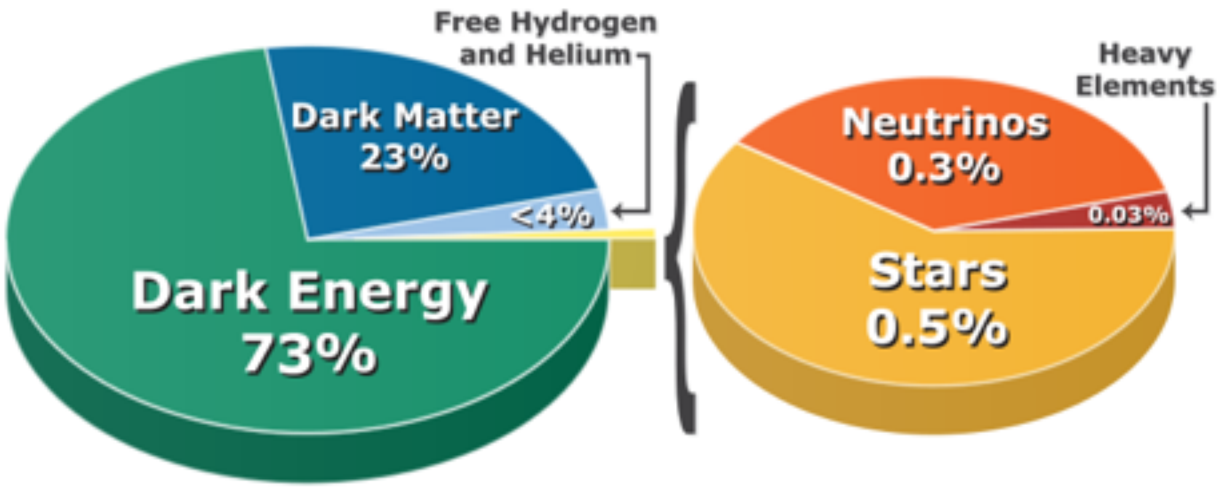
Cosmology: The Modern Context

- **Cosmology: The study of the Universe as a dynamical system**
  - Propose and test laws governing the dynamics of the large-scale Universe (matter and geometry)
  - Propose and test theories of initial conditions and primordial fluctuations
  - Sharpen cosmology as a tool for fundamental discoveries
  - Develop rigorous methodology for 'cosmological experiments'

Then --

Lev Landau

Now --



Cosmic Metrology

Cosmology is an Observational Science

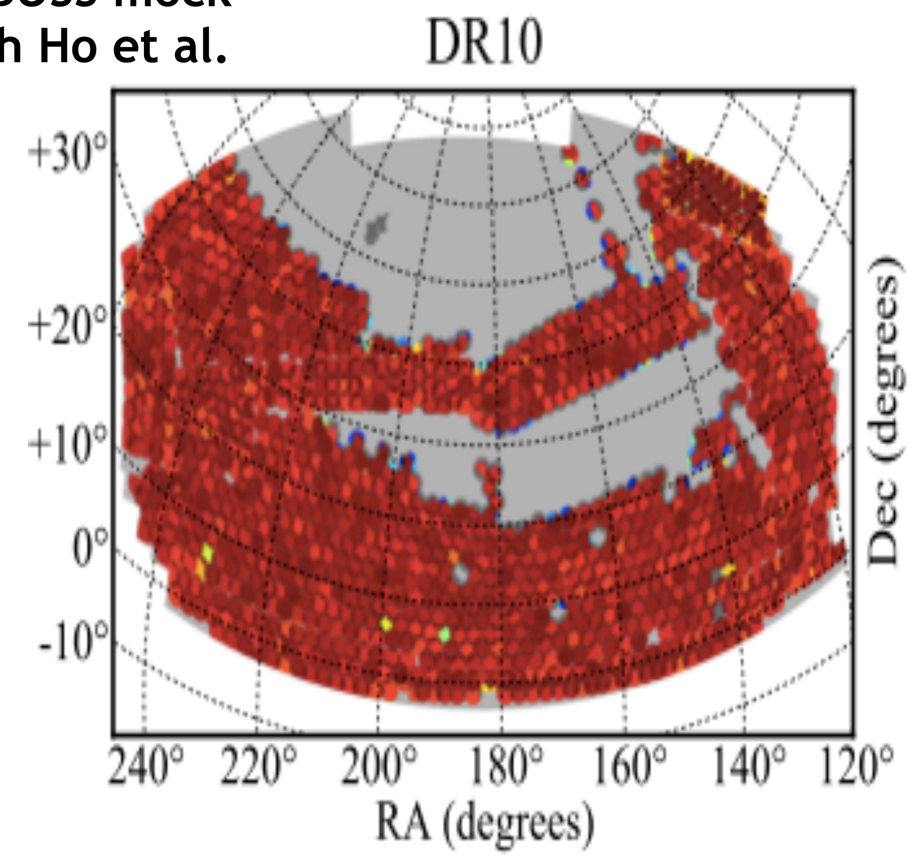
- **Not Particle Physics —**
  - Fundamental requirements of precision measurements — controllability, repeatability, and isolation — not available
  - Observations most often dictated by circumstances, not by optimal design for statistical inference
  - Cross-checks required to control a number of systematic errors
  - Theory and modeling limitations can be a serious source of bias in the inverse problem
  - Need to solve an initial value 'forward' problem (dynamics vs. scattering)
  - Aim is to systematically check and refine the cosmological 'Standard Model' even in the absence of fundamental guidance

"JUST CHECKING."

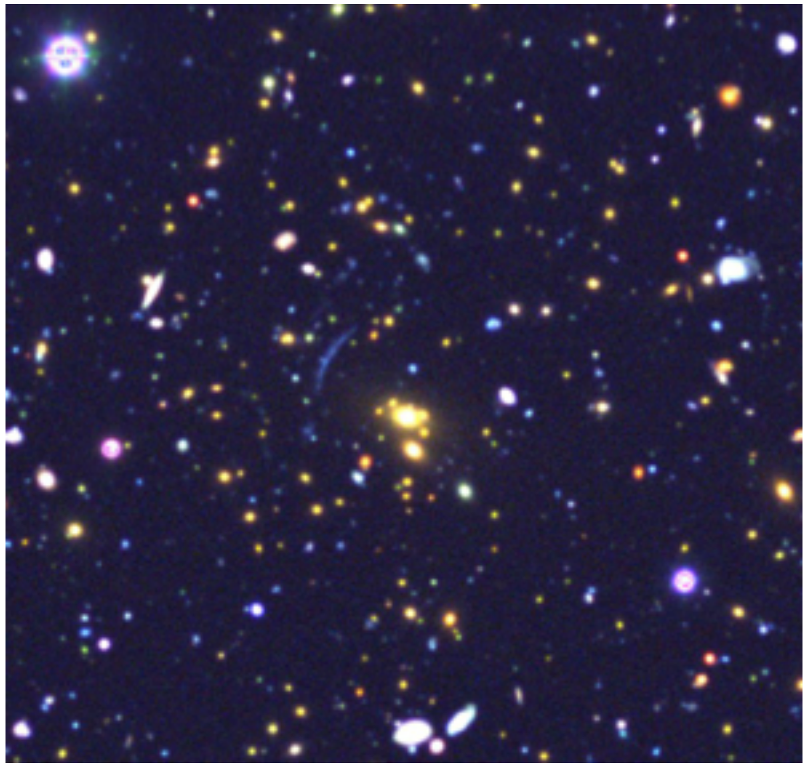
Context of Lectures

- **Cosmological Simulations**
  - These lectures: Large-scale simulations connected to cosmological observations (surveys/precision cosmology/cross-correlations) — will cover a blend of basic ideas, nuts and bolts issues, next-generation computing systems, and emulation methods
- **Connections with other lectures**
  - From simulations to galaxy catalogs (Peder Norberg)
  - Alternative models (Baojiu Li)
  - Smaller scales (Aldo Rodriguez-Puebla)
  - Significant background already covered in the pre-lectures (basic equations, methods, history)

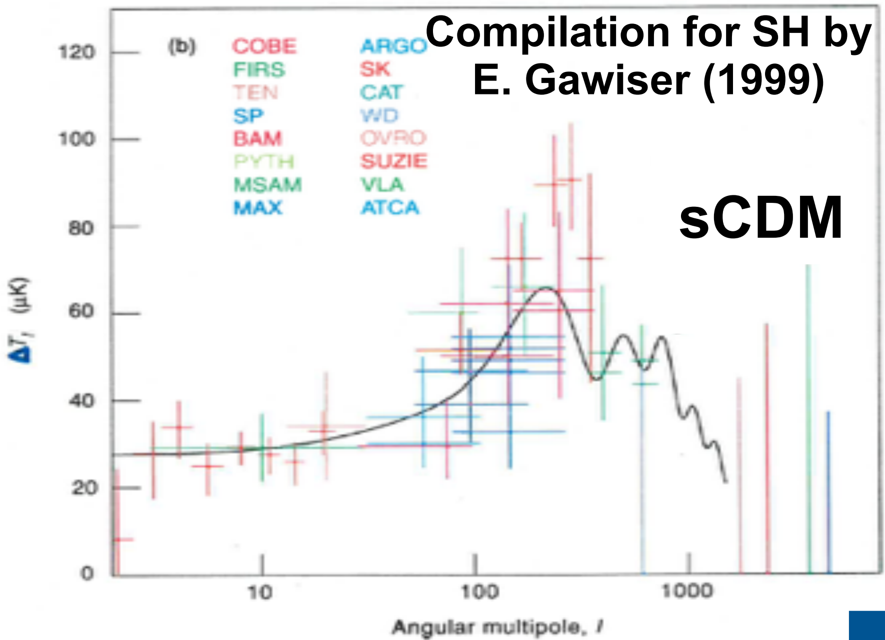
HOD based BOSS mock catalog, with Ho et al.

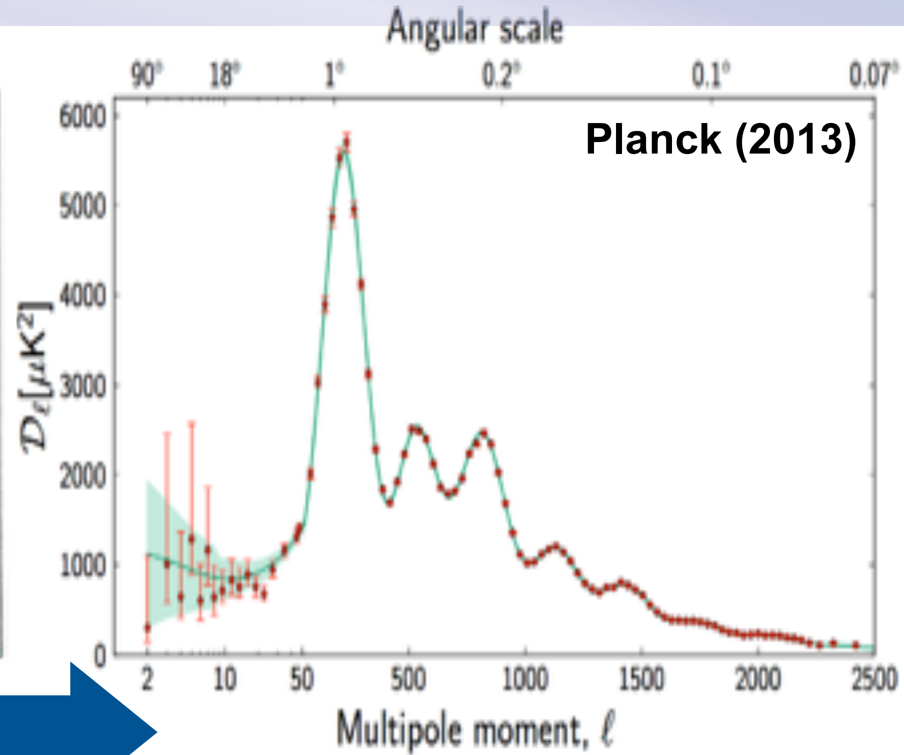


Strong lensing simulations with Li, Gladders et al.

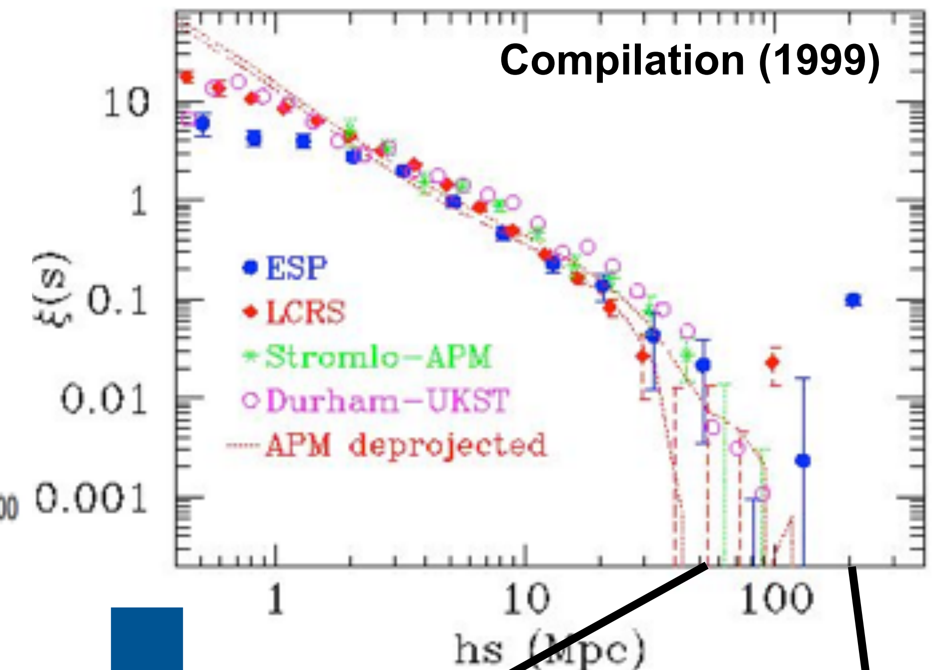


Precision Cosmology Progress

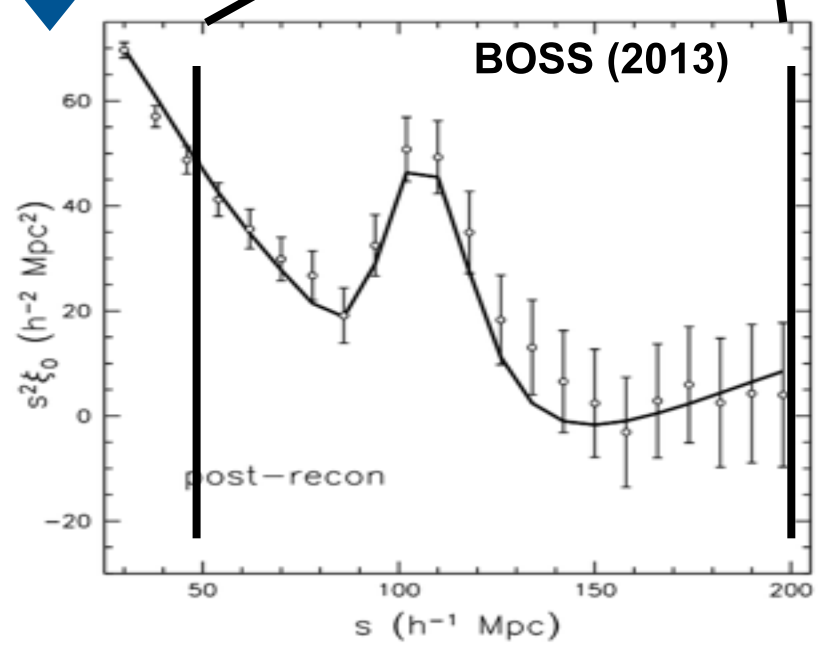


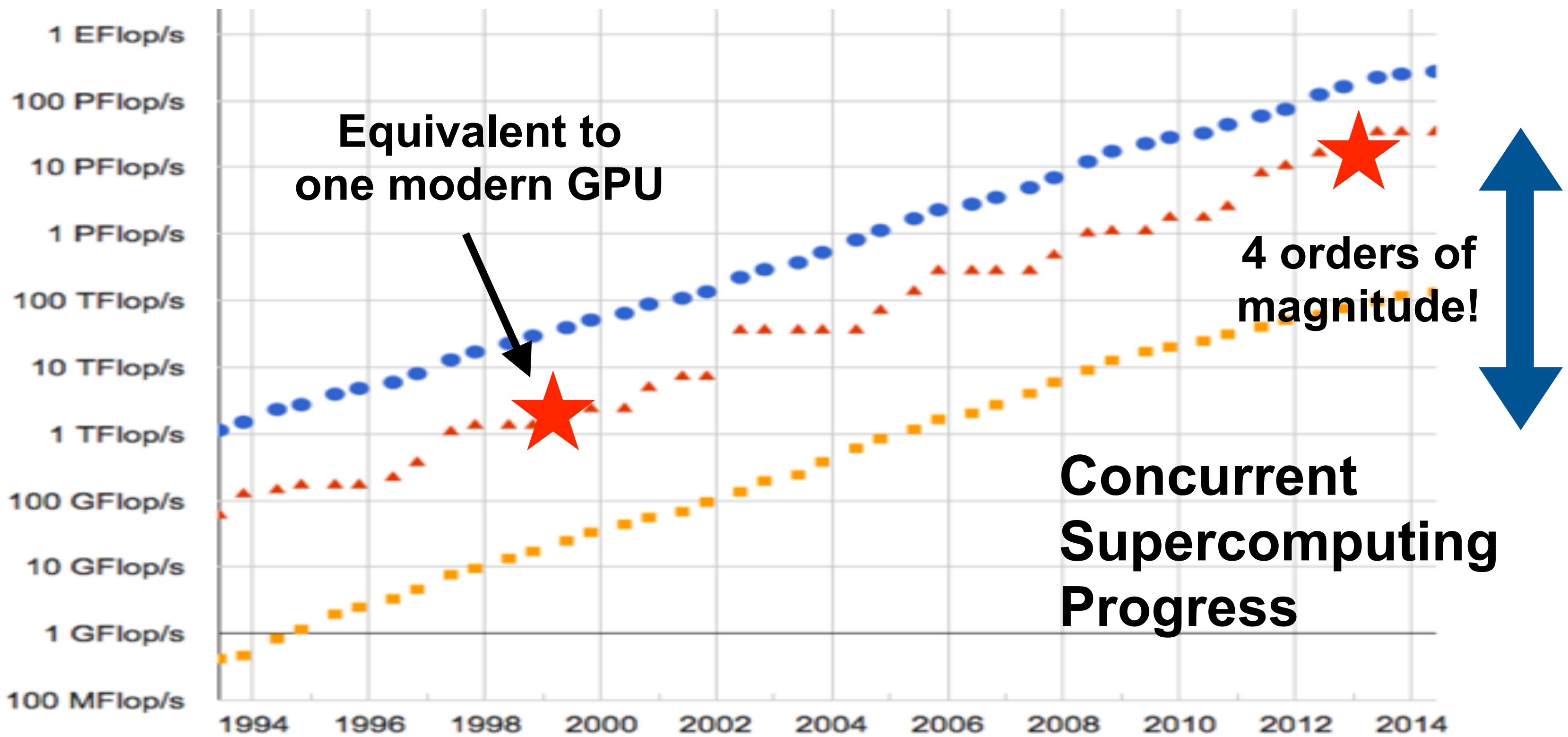


CMB



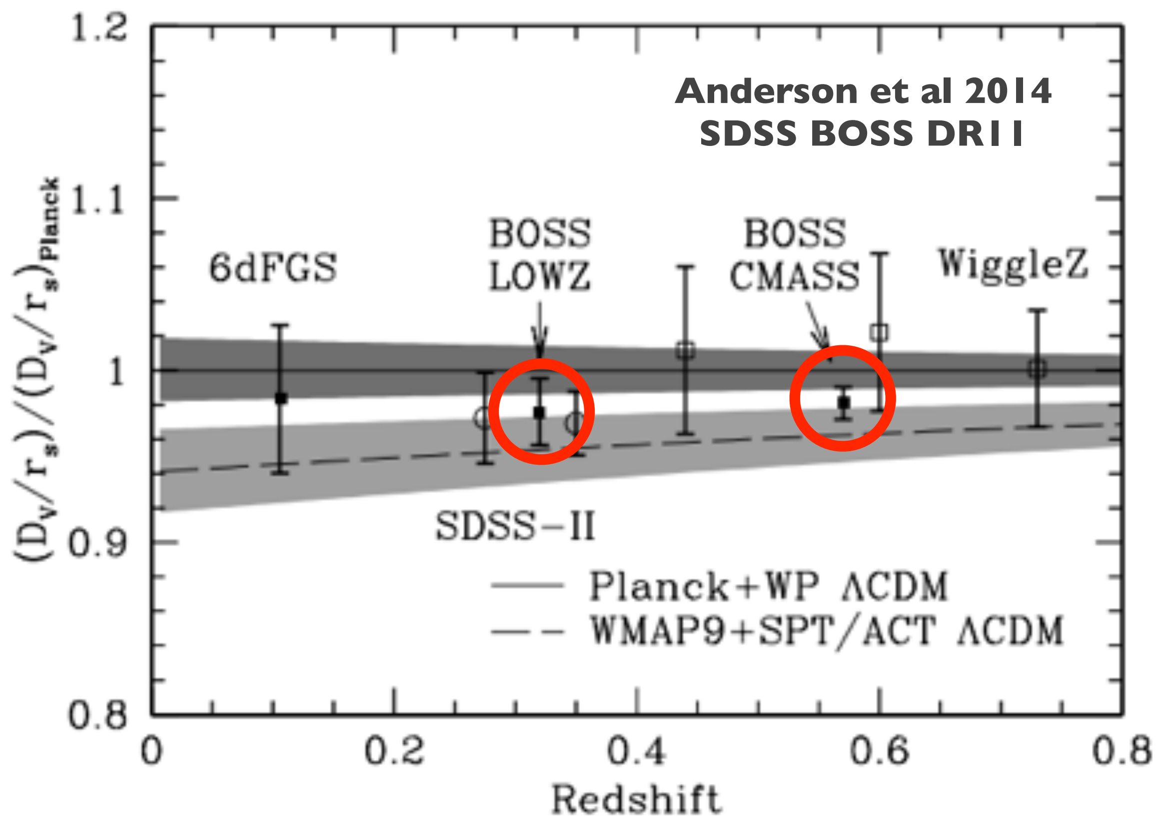
LSS

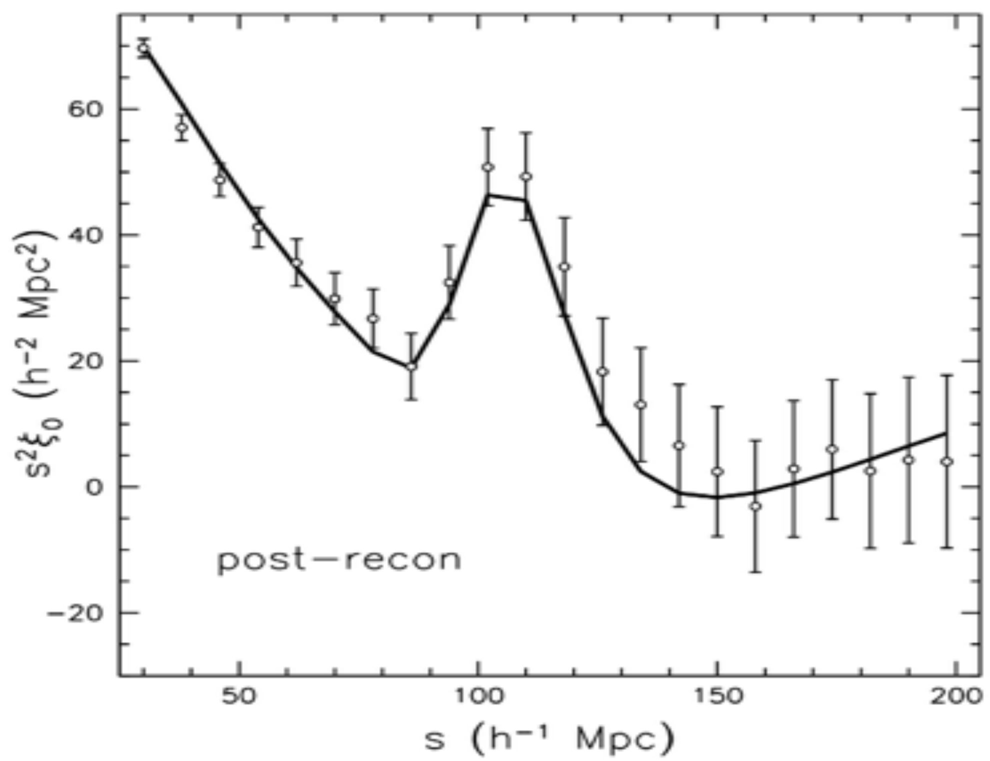




Cosmological Probes I: Galaxy Power Spectrum

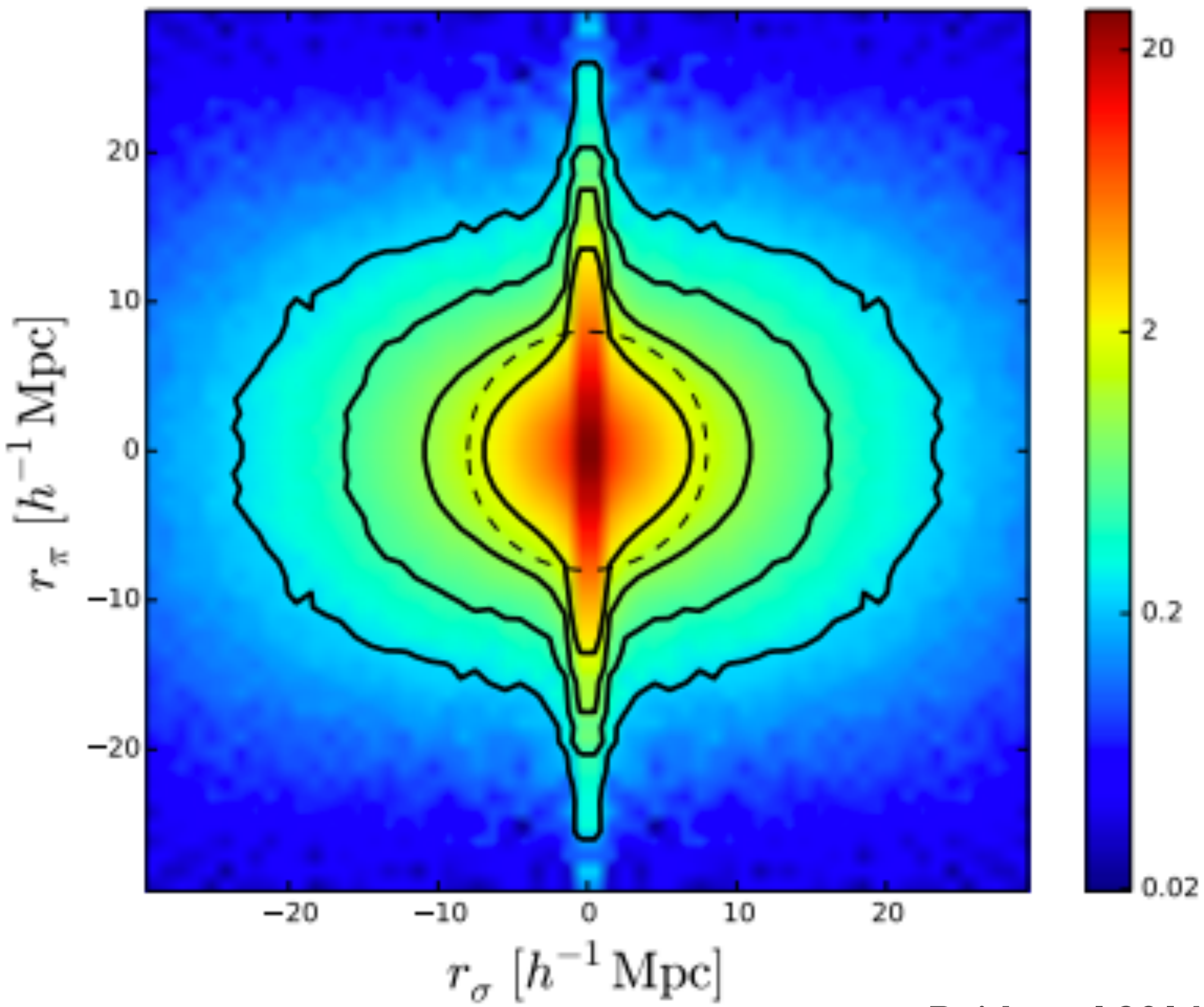
- **Baryon Acoustic Oscillations**
  - **Measurement:** Geometry at z~0.3-3
  - **P1:** Well-defined scale (peak) in the galaxy-galaxy correlation function or BAO 'wiggles' in P(k)
  - **P2:** Functions as a robust standard ruler to infer expansion history (sub-percent measurements, in principle)
  - **P3:** Large scale of ruler set by CMB acoustic scale (~150 Mpc) with width of ~15 Mpc (Silk scale)
  - **P4:** Very resistant to systematics effects because of large length scale, but requires very large survey volumes
  - **P5:** Broadband power constrains the spectral index, neutrino mass, dark matter properties (temperature)





II: Redshift Space Distortions

- **Redshift-space distortions**
  - **Measurement:** Growth of structure, tests of modified gravity
  - **P1:** Galaxies have local motions on top of the Hubble flow (they are not Hubble-fixed tracers)
  - **P2:** RSD arise from these motions, due to gravitational potentials, hence are sensitive to the matter distribution and modification of gravity
  - **P3:** Recent 2.5% growth rate measurement severely constrains possible modified gravity models

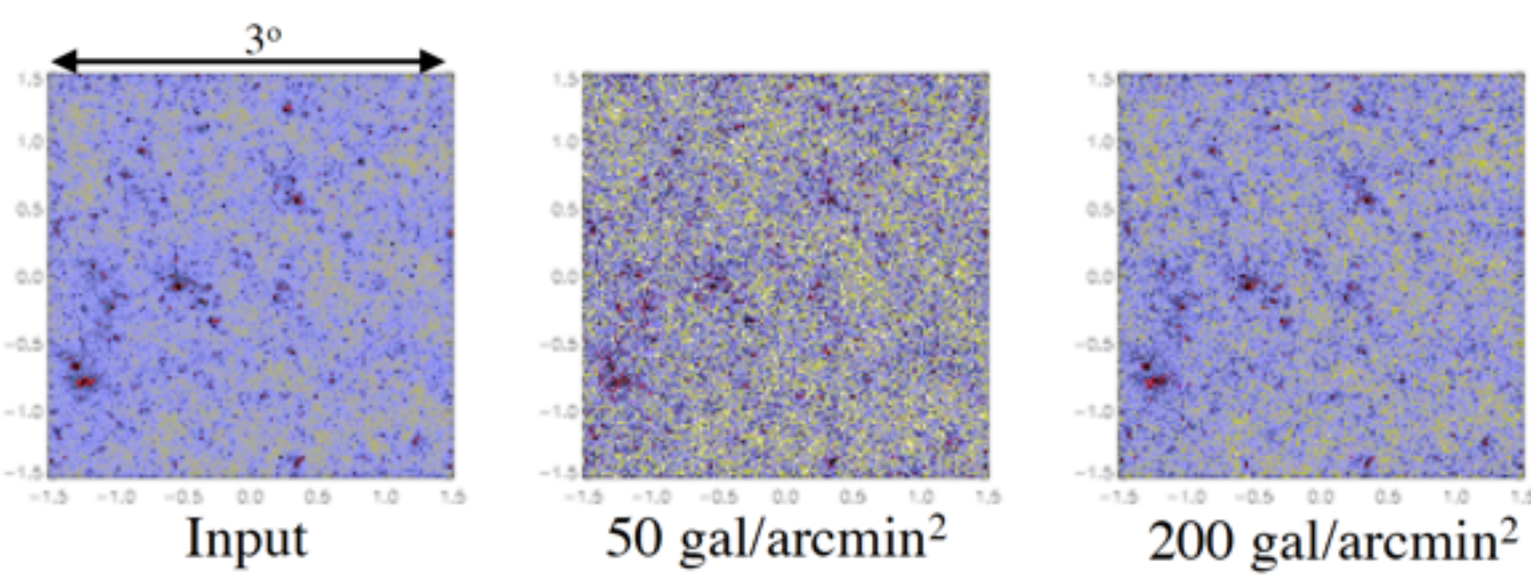


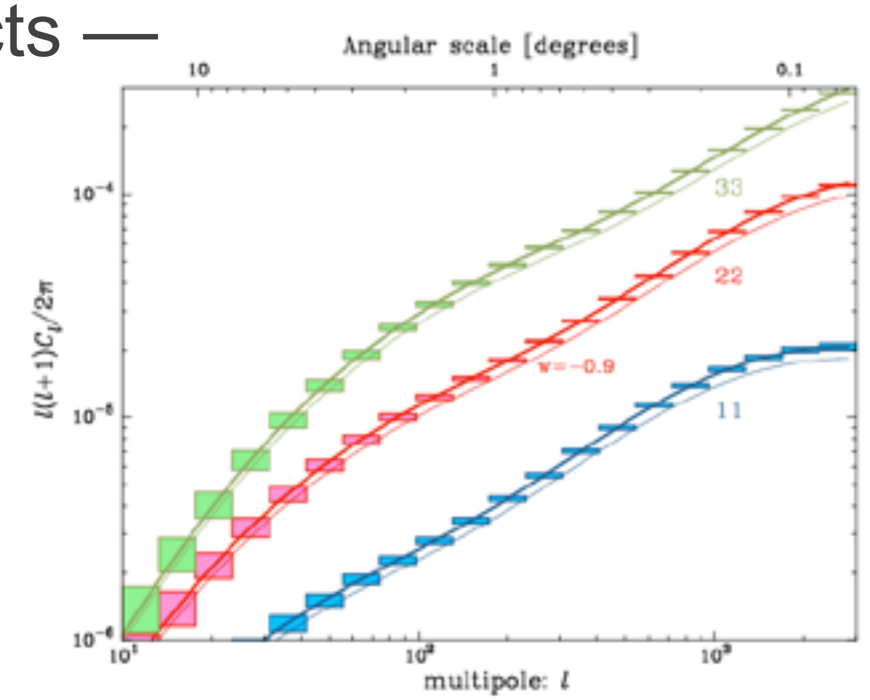
Reid et al 2014
SDSS-III CMASS

Two effects: 'Squashing' due to coherent infall and 'fingers of god' (FOG) due to random, virial motion at small scales; note isotropy broken by observer LOS

III: Cosmic Shear

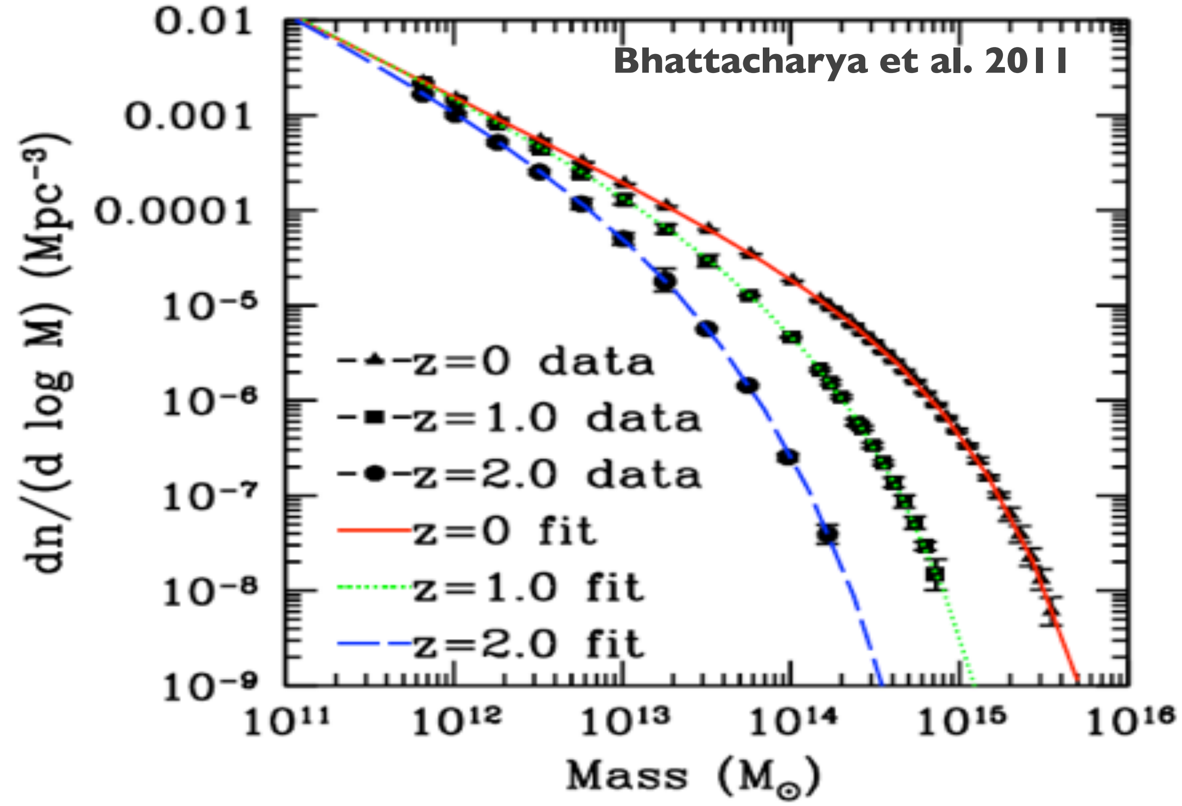
- **Weak lensing**
- **Measurement:** Multiple uses -- geometry, growth, mass
- **P1:** Distortion of shapes of background galaxies by foreground matter distribution
- **P2:** With sources at z~1 and lenses distributed between 0.1<z<1, the signal is weak, but statistically detectable
- **P3:** Theory is well understood
- **P4:** The shear signal is largely in the domain of nonlinear clustering
- **P5:** Systematics include shape measurement, PSF variations, intrinsic alignments, photo-z estimation, baryonic effects —

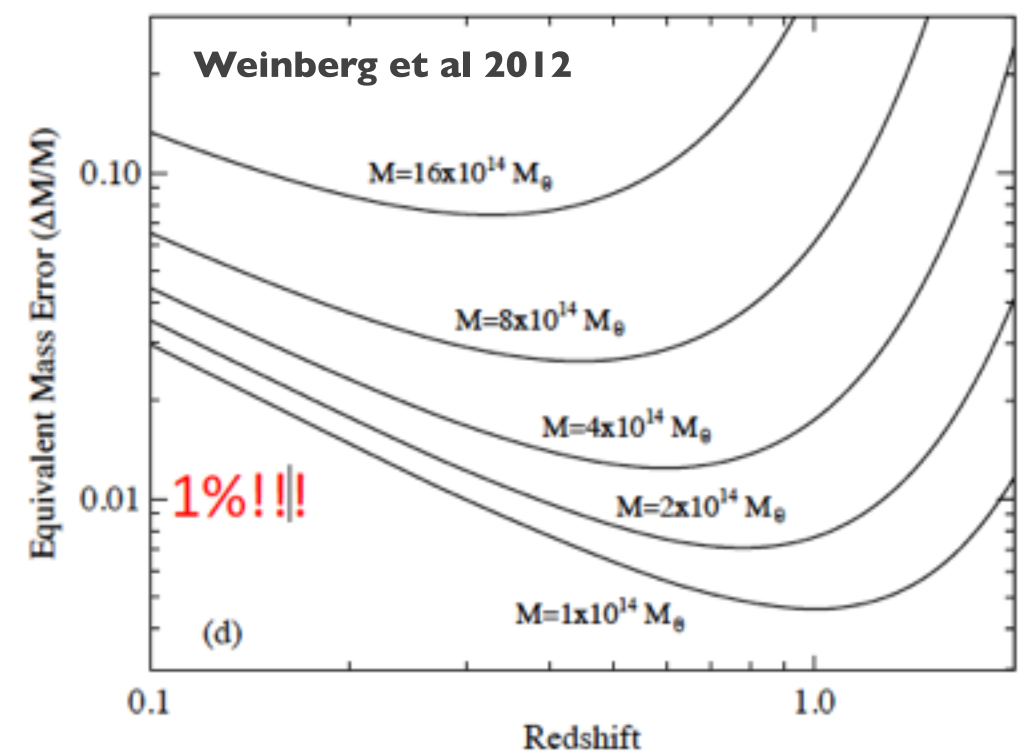




IV: Clusters

- **Cluster counts**
  - **Measurement:** Geometry and structure growth
  - **P1:** Clusters are the largest collapsed structures today (masses up to ~10^15 solar masses)
  - **P2:** Associated with the largest density concentrations in the initial density field
  - **P3:** Very sensitive probe for dark energy, potentially also for neutrino mass; can also probe dark matter EOS
  - **P4:** Systematic errors are a key concern (how to define and measure mass? Weak lensing?)

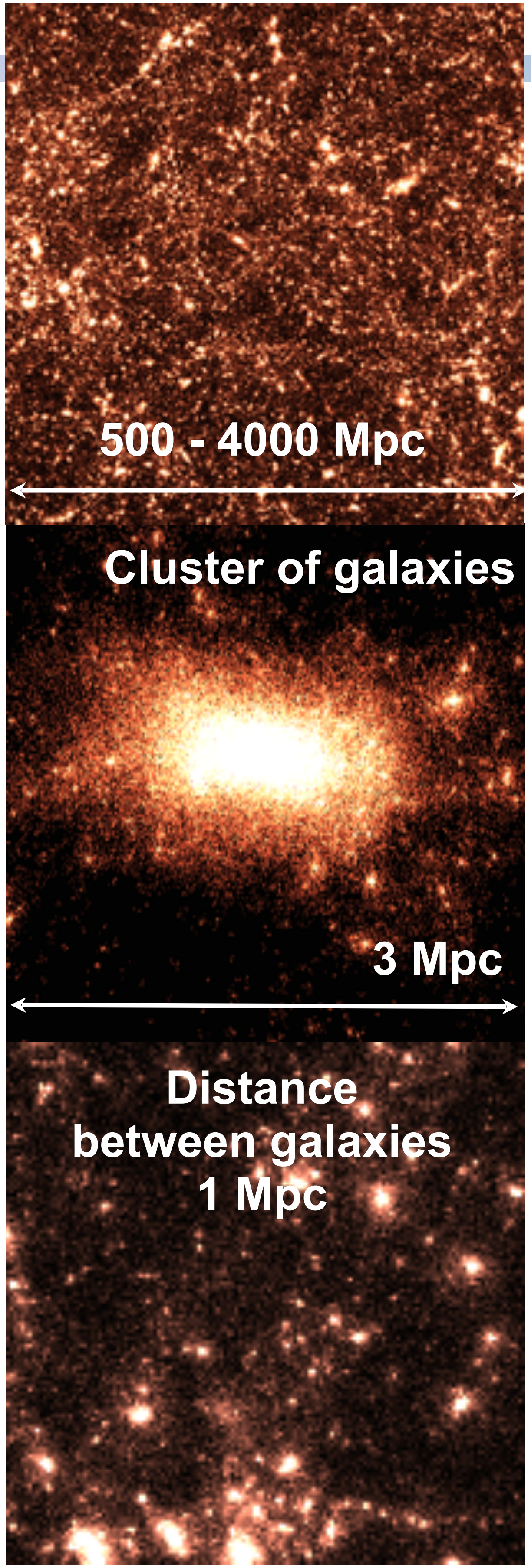




Error tolerance on mass measurement in order to remain statistics-limited

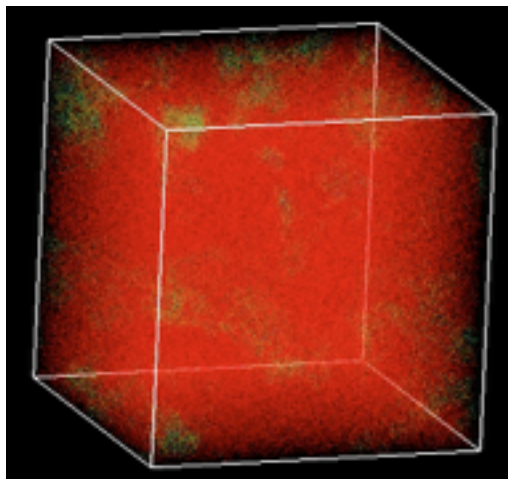
Simulating Surveys

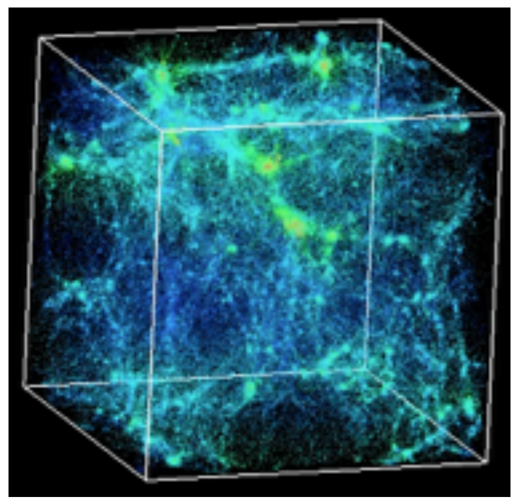
- **Simulation Volume:** Large survey sizes impose simulation volumes **~(4** Gpc)^3, memory required **~100**TB **- 1**PB
- **Number of Particles:** Mass resolution depends on ultimate object to be resolved, $\sim 10^8 M_\odot - 10^{10} M_\odot$ (subhalos to halos), N~$10^{11} - 10^{12}$
- **Easy to remember:** ~(1000 Mpc)^3 and ~(1000)^3 particles lead to a mass resolution of $\sim 10^{11} M_\odot$
- **Therefore:** ~(10,000)^3 particles in ~(1000 Mpc)^3 leads to $\sim 10^8 M_\odot$, (5Gpc)^3 at this mass resolution would require 125 trillion particles = 5PB per snapshot! (Mira has 0.75PB overall, currently largest US DOE open machine)
- **Force Resolution:** **~**kpc, yields a (global) spatial dynamic range of **~10**^{6}

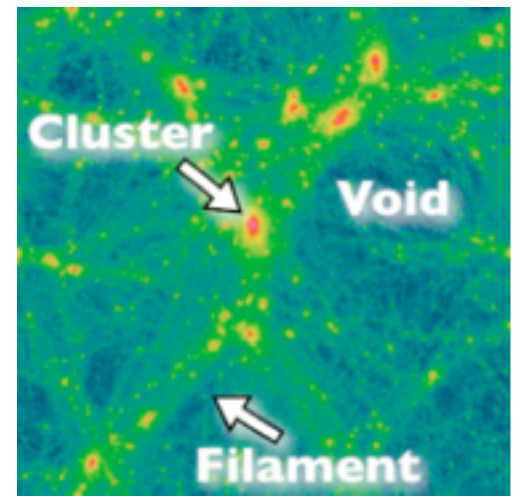


Analytics/Workflow Complexity

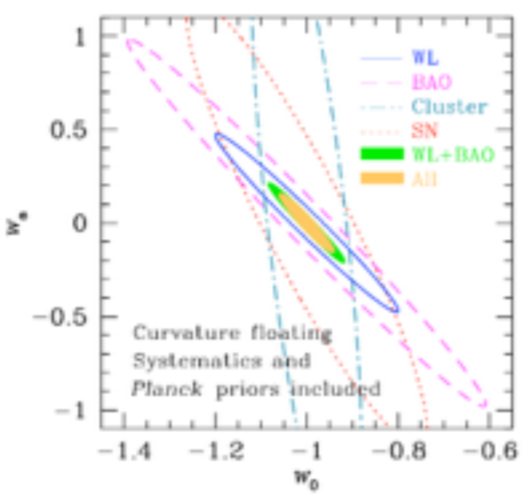
Gaussian Random Field Initial Conditions → High-Resolution N-Body Code (HACC) → Multiple Outputs Halo/Sub-Halo Identification → Halo Merger Trees







Halo Merger Trees → Semi-Analytic Modeling Code (Galacticus) → Galaxy Catalog → Realistic Image Catalog → Atmosphere and Instrument Modeling



- **Simulation Campaigns:** Statistics of virtual universes; construction of emulators
- **Modeling the Measurement:** End-to-End modeling necessary to understand crucial systematic errors
- **PDACS:** Custom workflow system under development
- **Data-Intensive Computing:** New architectures needed

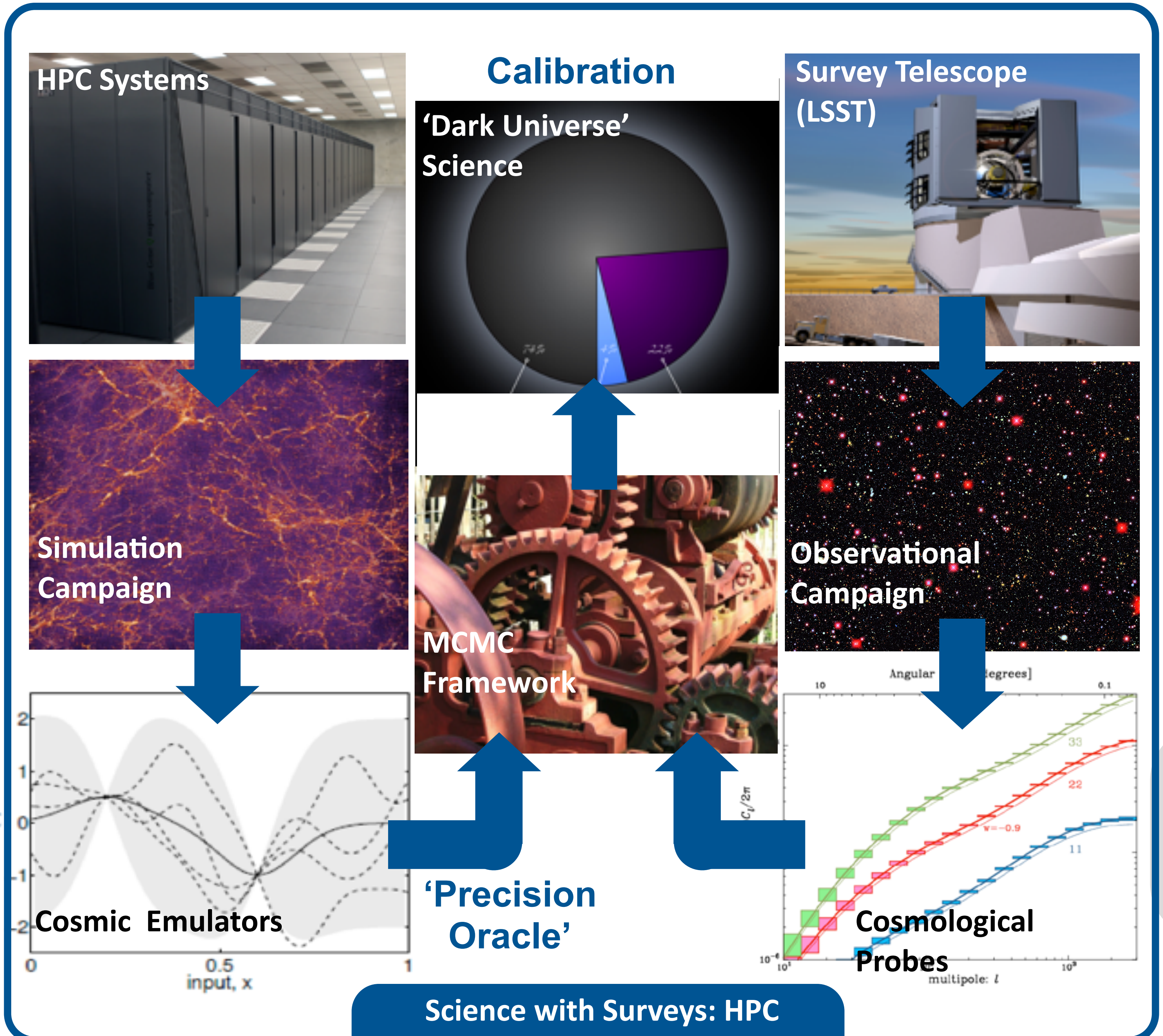
Atmosphere and Instrument Modeling → Data Management Pipeline → Data Analysis Pipeline → Scientific Inference Framework

Precision Cosmology: ‘Big Data’ Meets Supercomputing

Supercomputer Simulation Campaign

Simulations + CCF

Emulator based on Gaussian Process Interpolation in High-Dimensional Spaces



Mapping the Sky with Survey Instruments

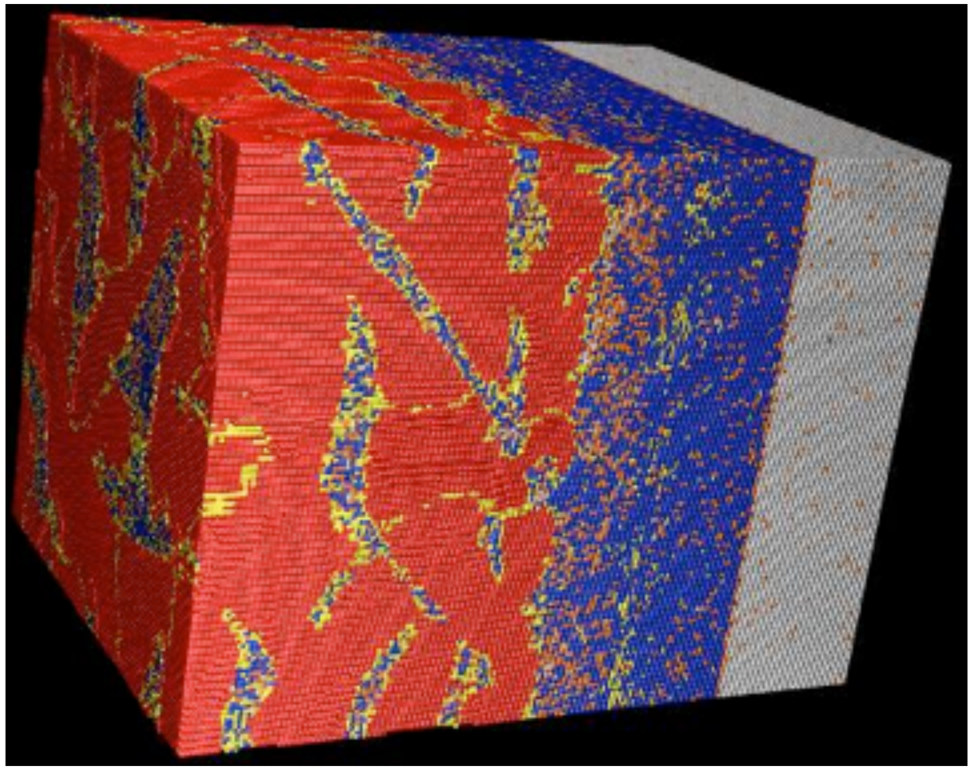
Observations: Statistical error bars will ‘disappear’ soon!

Science with Surveys: HPC meets Big Data

CCF= Cosmic Calibration Framework

Cosmological Simulations: Good News and Bad News

- **Cosmological simulations: Bad News**
  - Need large amounts of memory
  - Need large spatial dynamic range everywhere
  - Need large dynamic range in mass
- **Comparison with other fields**
  - Dynamic range in time is not extreme wrt to molecular dynamics or protein folding (picosec to sec; femtosec to millisec)
  - Computational work can be parallelized efficiently (closest to plasmas and unlike, say, dynamical systems)
  - Complexity is not extreme (unlike, say, climate simulations)
- **Good News:**
  - Next-generation/exascale systems well-suited for cosmological simulations



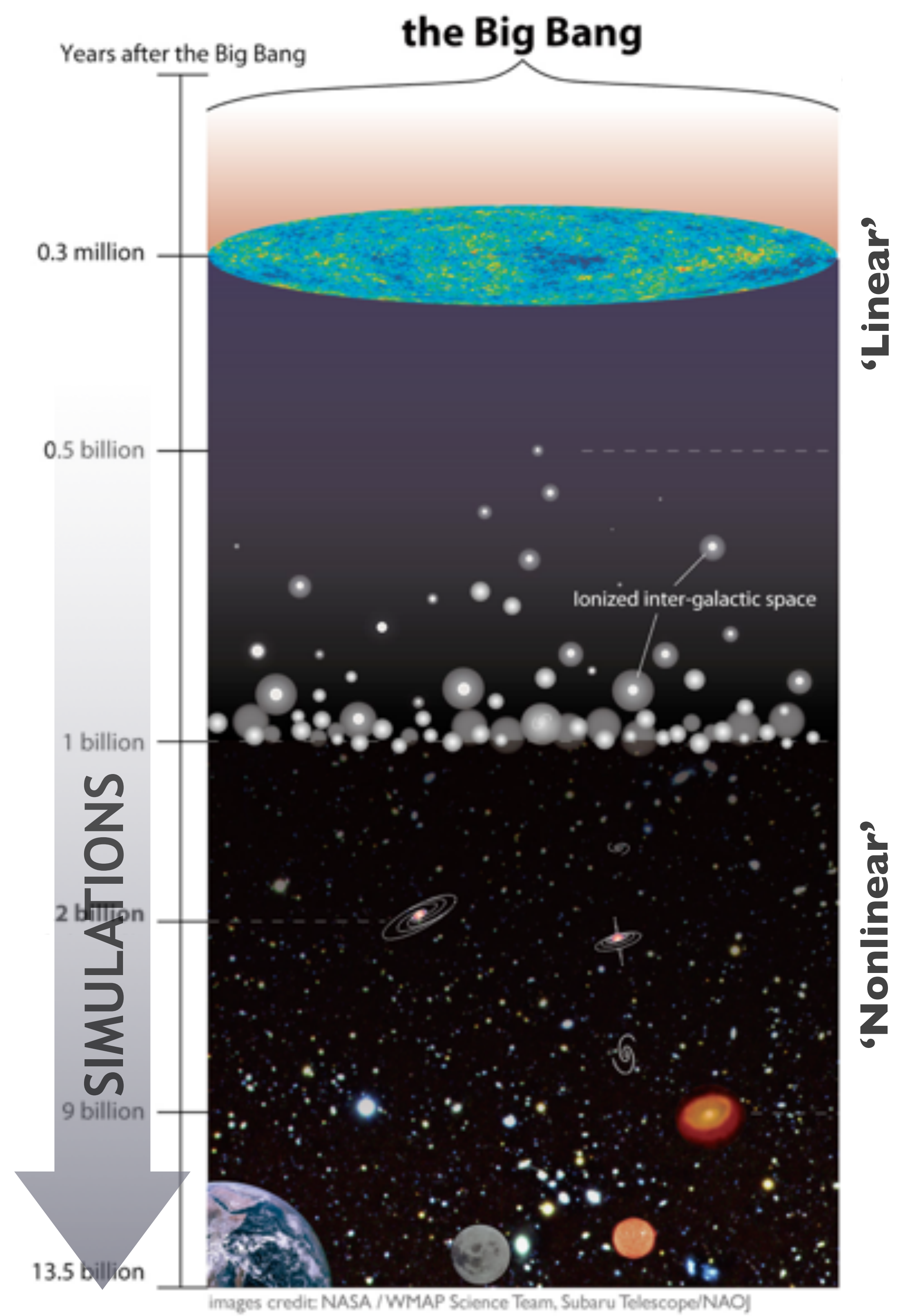
Molecular dynamics simulation of shock-induced material transformation (~9ps)

Next-Generation Systems in the US (DOE)

System attributes	NERSC Now	OLCF Now	ALCF Now	NERSC Upgrade	OLCF Upgrade	ALCF Upgrades	
Name Planned Installation	**Edison**	**TITAN**	**MIRA**	**Cori 2016**	**Summit 2017-2018**	**Theta 2016**	**Aurora 2018-2019**
System peak (PF)	2.6	27	10	> 30	150	>8.5	180
Peak Power (MW)	2	9	4.8	< 3.7	10	1.7	13
Total system memory	357 TB	710TB	768TB	~1 PB DDR4 + High Bandwidth Memory (HBM)+1.5PB persistent memory	> 1.74 PB DDR4 + HBM + 2.8 PB persistent memory	>480 TB DDR4 + High Bandwidth Memory (HBM)	> 7 PB High Bandwidth On-Package Memory Local Memory and Persistent Memory
Node performance (TF)	0.460	1.452	0.204	> 3	> 40	> 3	> 17 times Mira
Node processors	Intel Ivy Bridge	AMD Opteron Nvidia Kepler	64-bit PowerPC A2	Intel Knights Landing many core CPUs Intel Haswell CPU in data partition	Multiple IBM Power9 CPUs & multiple Nvidia Voltas GPUS	Intel Knights Landing Xeon Phi many core CPUs	Knights Hill Xeon Phi many core CPUs
System size (nodes)	5,600 nodes	18,688 nodes	49,152	9,300 nodes 1,900 nodes in data partition	~3,500 nodes	>2,500 nodes	>50,000 nodes
System Interconnect	Aries	Gemini	5D Torus	Aries	Dual Rail EDR-IB	Aries	2nd Generation Intel Omni-Path Architecture
File System	7.6 PB 168 GB/s, Lustre®	32 PB 1 TB/s, Lustre®	26 PB 300 GB/s GPFS™	28 PB 744 GB/s Lustre®	120 PB 1 TB/s GPFS™	10PB, 210 GB/s Lustre initial	150 PB 1 TB/s Lustre®

Evolution of Perturbations

- **Solid understanding of structure formation is necessary for precision cosmology probes**
  - Well-determined initial conditions (Gaussian random field)
  - Initial perturbations amplified by gravitational instability in an expanding Universe
  - Nonlinear evolution is challenging
- **Early Universe: Linear perturbation theory very successful (e.g., CMB)**
- **Latter half of the history of the Universe: Nonlinear domain of structure formation, impossible to treat accurately without computers (pre-lectures covered this)**



Basic Equations

- **Assumption: Single species**
  - Particles interact gravitationally
  - Neglect non-Newtonian GR corrections
  - Keep expansion of the universe
- **What is the correct level of description?**
  - Exact Klimontovich and Liouville (ensemble picture) equations not useful
  - Argue that two-body collisional effects are negligible (some subtleties here)
  - Obtain Vlasov-Poisson Eqn.

$$\frac{\partial f_i}{\partial t} + \dot{\mathbf{x}}\frac{\partial f_i}{\partial \mathbf{x}} - \nabla\phi\frac{\partial f_i}{\partial \mathbf{p}} = 0, \qquad \mathbf{p} = a^2\dot{\mathbf{x}},$$

$$\nabla^2\phi = 4\pi G a^2(\rho(\mathbf{x},t) - \langle\rho_{\rm dm}(t)\rangle) = 4\pi G a^2\Omega_{\rm dm}\delta_{\rm dm}\rho_{\rm cr},$$

$$\delta_{\rm dm}(\mathbf{x},t) = (\rho_{\rm dm} - \langle\rho_{\rm dm}\rangle)/\langle\rho_{\rm dm}\rangle),$$

$$\rho_{\rm dm}(\mathbf{x},t) = a^{-3}\sum_i m_i\int d^3\mathbf{p}\, f_i(\mathbf{x},\dot{\mathbf{x}},t).$$

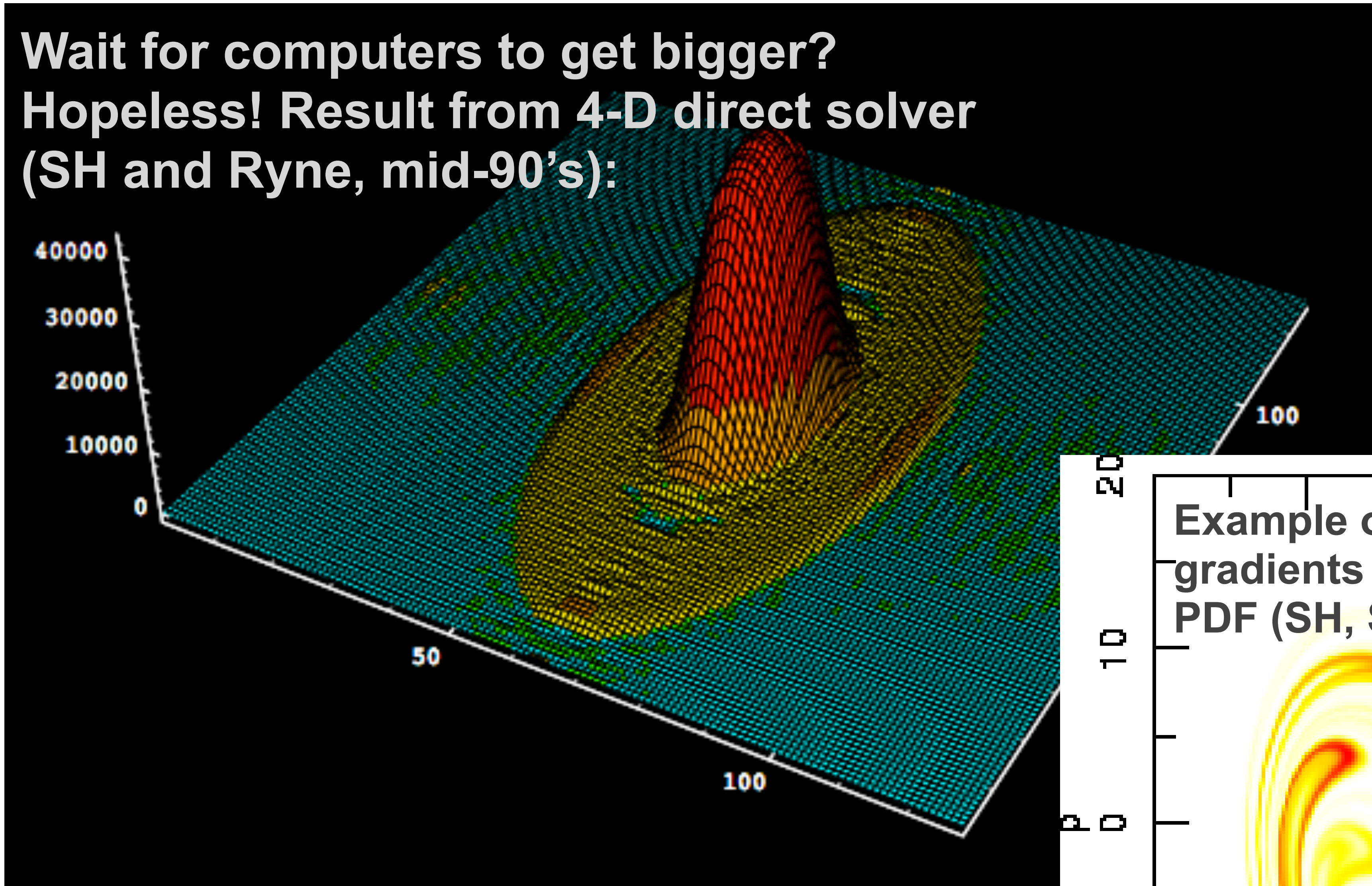
- **Features of the VPE:**
  - f(x,v,t) is a smooth phase space distbn. function
  - The potential is the mean field potential
  - All higher-point correlations follow from the properties of f(x,v,t)
  - 6-D nonlinear PDE (memory nightmare)
  - Very complex solutions (resolution disaster)
  - Evolution is Hamiltonian, symplectic wrt the Morrison bracket

Beyond VPE: The Landau Equation

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \frac{\mathbf{F}}{m} \cdot \frac{\partial f}{\partial \mathbf{v}} = -\frac{\partial}{\partial \mathbf{v}} \cdot \underbrace{\mathbf{F}_{\mathbf{d}} f}_{\text{Dynamical Friction}} + \frac{1}{2} \overbrace{\frac{\partial^2}{\partial \mathbf{v} \partial \mathbf{v}} : \mathbf{D} f}^{\text{Diffusion}}$$

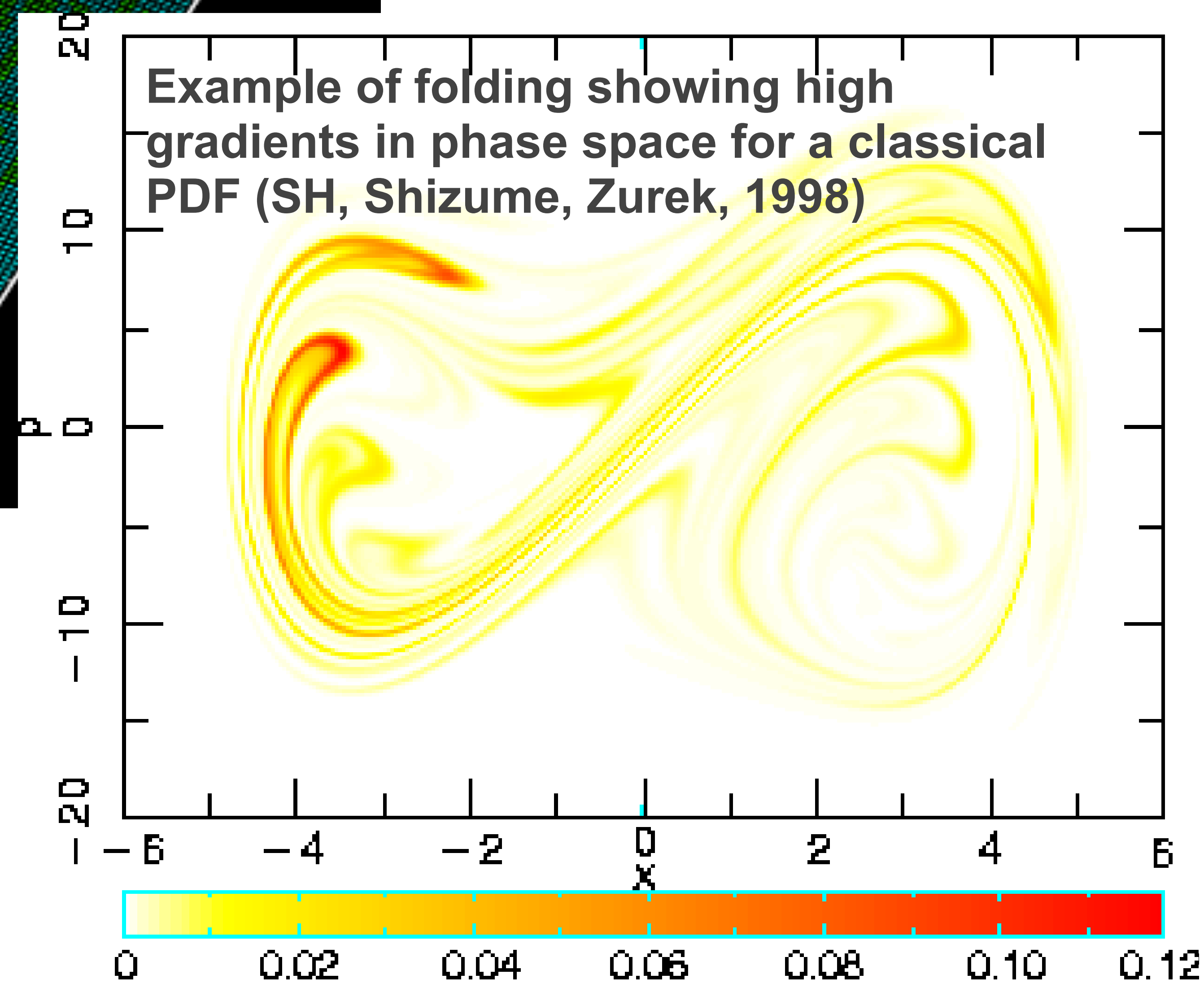
- **Features of the Landau equation:**
  - The long-range nature of the Coulomb force means that the first fluctuation correction to the mean field description arises from the collective contribution of a large number of small-angle collisions
  - Derivation follows BBGKY or more heuristic approaches
  - The LE is a Fokker-Planck equation of a special type — it conserves energy (because of the self-consistent nature of the derivation) but not entropy, unlike the VPE, which conserves both
  - Irreversibility — Unlike the VPE, the LE is irreversible -- it satisfies an H-theorem and describes an approach to thermal equilibrium: entropy monotonically increases.
  - Very little is known about the mathematical properties of this equation, however, like the VPE, it too can be solved by particle methods (Qiang, Ryne, SH, 2000)

How to Solve the Vlasov-Poisson Equation?



Wait for computers to get bigger? Hopeless! Result from 4-D direct solver (SH and Ryne, mid-90's):

- **Requirement:**
  - Need a scheme that fails "gracefully"
  - Small-scale errors should not feed back and crash the code



Example of folding showing high gradients in phase space for a classical PDF (SH, Shizume, Zurek, 1998)

- **PDE approaches:**
  - Low-resolution because of high dimensionality
  - Catastrophic errors due to resolution loss (aliasing, etc.)

Particle Methods

- **Particle-based approach:**
  - Sample phase space with particles (earlier techniques used a mixed grid and particle based method, but this ran into trouble)
  - Particles are not physical but are viewed as tracers of a collisionless Liouville flow
  - Derive evolution rules for the particles that are consistent with the fact that they are tracers for a smooth phase space distribution
  - Particle methods for solving the VPE should not be confused with "N-body" solvers, despite formal similarities, they live in a different physical space (in the same sense that SPH is not an "N-body" method)
  - Particle methods were first developed in a fluids context at Los Alamos in the 1950's, were then applied in plasma physics in the 1960's, and later adapted by cosmologists
  - A number of tricky issues arise in particle-based methods, such as force softening, force resolution, discreteness noise, time evolution systematics and errors, lack of error control theory, lack of (sufficiently) nonlinear exact solutions as benchmarks, etc.

Particle Methods: Some Pros and Cons

- **Particle-based approach:**
  - The fundamental advantage is that codes based on particle methods will not crash under the same circumstances as a grid-based solver, but the cost is some form of error — controlling this error is key
  - Because the fundamental representation is discrete, particle methods suffer from shot noise, particularly troubling if accuracy is required in low-density regions
  - Particle methods, as employed in cosmology, are naturally Lagrangian — particles go where they are needed — so no extra adaptive mesh scaffolding is needed as in grid-based fluid solvers
  - Because particle methods are a form of dynamical Monte Carlo algorithm, standard applied mathematics methods for error and convergence don't directly apply to them, a lot of current knowledge is based on a posteriori experience ("if we do this, bad things happen")
  - Particle methods have much simpler data structures than hierarchical grid methods, which makes them very suitable for modern high-perfromance computers

The Pioneers: CFD with "Particles"

Frank Harlow

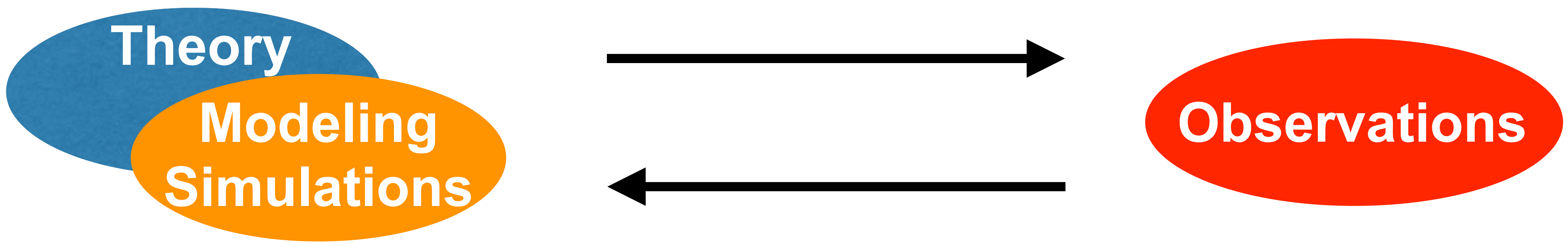
Oscar Buneman

Roger Hockney

Applied to Frank, unusual is an understatement. He is singular. I have known him for almost 40 of the 50 years he has been a staff member at Los Alamos, so mine is not an entirely objective point of view. However, even a completely detached observer would be struck by the breadth of his interests. His paintings line the walls of the Laboratory, his books on Southwest Indian pottery fill a shelf, and his monograph on brachiopods, in which he identifies 24 new species and 3 new genera, is still the definitive reference on the lowest parts of the Pennsylvanian period. He has grown orchids, learned silversmithing, and stopped riding a Harley because it was not fast enough. He knows an incredible number of people in all walks of life by their first name, yet can be met in person only in Los Alamos because he refuses to travel. It all seems to work for him, because he brings to his work an undiminished energy and focus.

J. Brackbill on Frank Harlow

Where do Numerical Simulations Fit?



Many of these results were published in reports and journal articles (see Bibliography), but not without some opposition. There was, for example, a conversation that was reported to me, which took place around 1955 between two of the upper-level managers of the Theoretical Division, one of whom said, "Look, what is this stuff? This PIC method that they're trying to publish a paper on will never work! There will never be computers big enough and fast enough to solve problems with all of the particles that you need in order to resolve what's going on. I don't think we should publish this." And the other person said, "Hey look, maybe it's not useful or won't work all that well. But on the other hand, it's something that was tried, something that people will learn from. We probably should mark it down and it'll go into the annals of things that were tried and didn't work very well but at least it will have been recorded." I'm happy that the second point of view prevailed and that the ideas found their way into print.

Another type of opposition occurred in our interactions with editors of professional journals, and with scientists and engineers at various universities and industrial laboratories. One of the things we discovered in the 1950s and early 1960s was that there was a lot of suspicion about numerical techniques. Computers and the solutions you could calculate were said to be the playthings of rich laboratories. You couldn't learn very much unless you did studies analytically. The truth, of course, has turned out to lie in the complementary interaction between calculations and analysis, and in the validation of both through comparisons with experiments. I think some of these early suspicions were partially justified, although we found as time went by that many universities, industries, and governmental agencies began to solve problems of interest using numerical techniques.

F. Harlow, J. Comp. Phys. 195, 414 (2004)

An Early Simulation

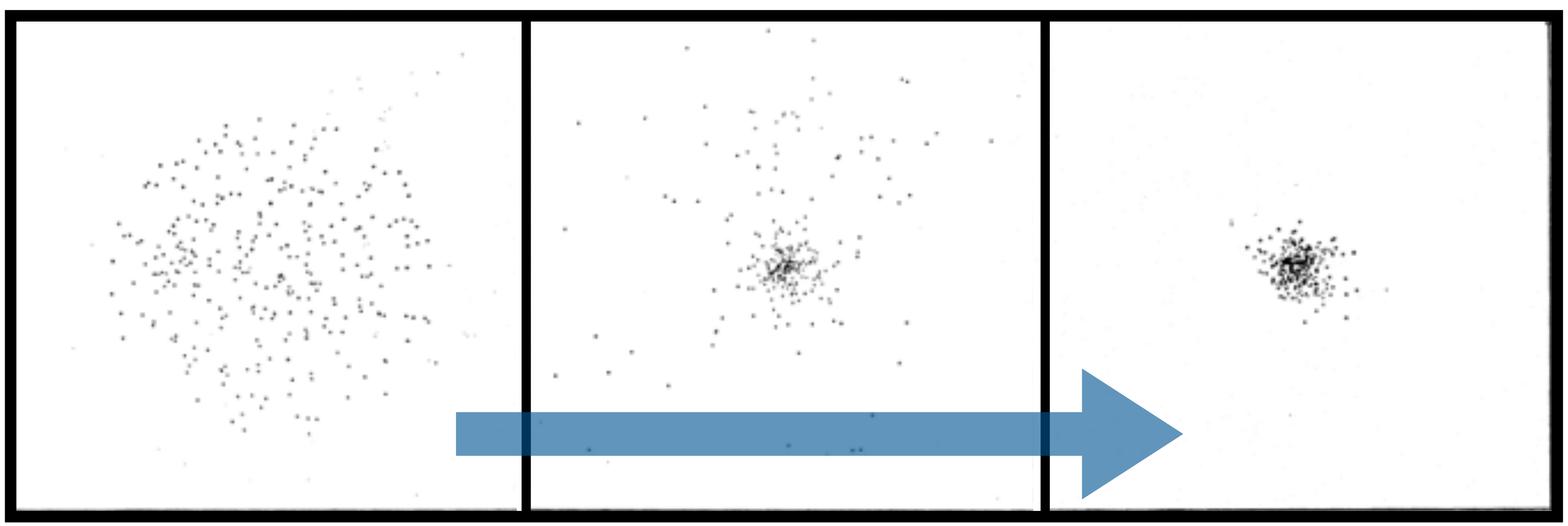
ASTRONOMICAL JOURNAL VOLUME 75, NUMBER 1 FEBRUARY 1970

Structure of the Coma Cluster of Galaxies*

P. J. E. Peebles†
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

(Received 7 October 1969)

In some cosmologies, a cluster of galaxies is imagined to be a gravitationally bound system which, in analogy with the formation of the Galaxy, originated as a collapsing protocluster. It is shown that a numerical model based on this picture is consistent with the observed features of the Coma Cluster of galaxies. The cluster mass derived from this model agrees with previous values; however, an analysis of the observational uncertainty within the framework of the model shows that the derived mass could be consistent with the estimated total mass provided by the galaxies in the cluster.



“The Universe is far too complicated a structure to be studied deductively, starting from initial conditions and solving the equations of motion.”
Robert Dicke (Jayne Lectures, 1969)

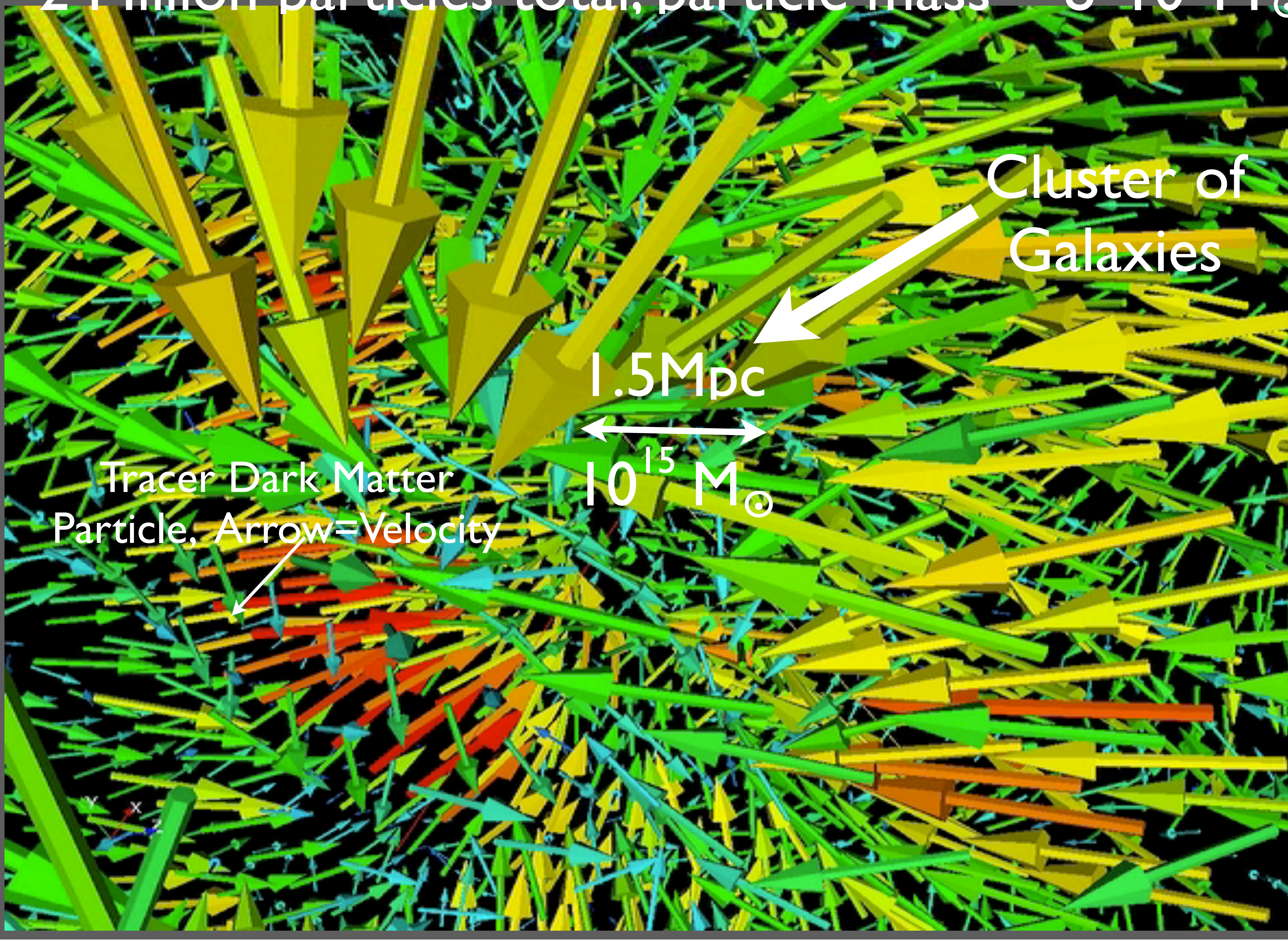
- Suite of 300 particle simulations
- Run on a CDC 3600, ~1Mflops*, 32KB+ memory at Los Alamos
- Is 9 orders of magnitude in improvement in both performance and memory enough for precision cosmology?

* Today’s laptop is ~1000 X faster

Dynamics of Structure Formation

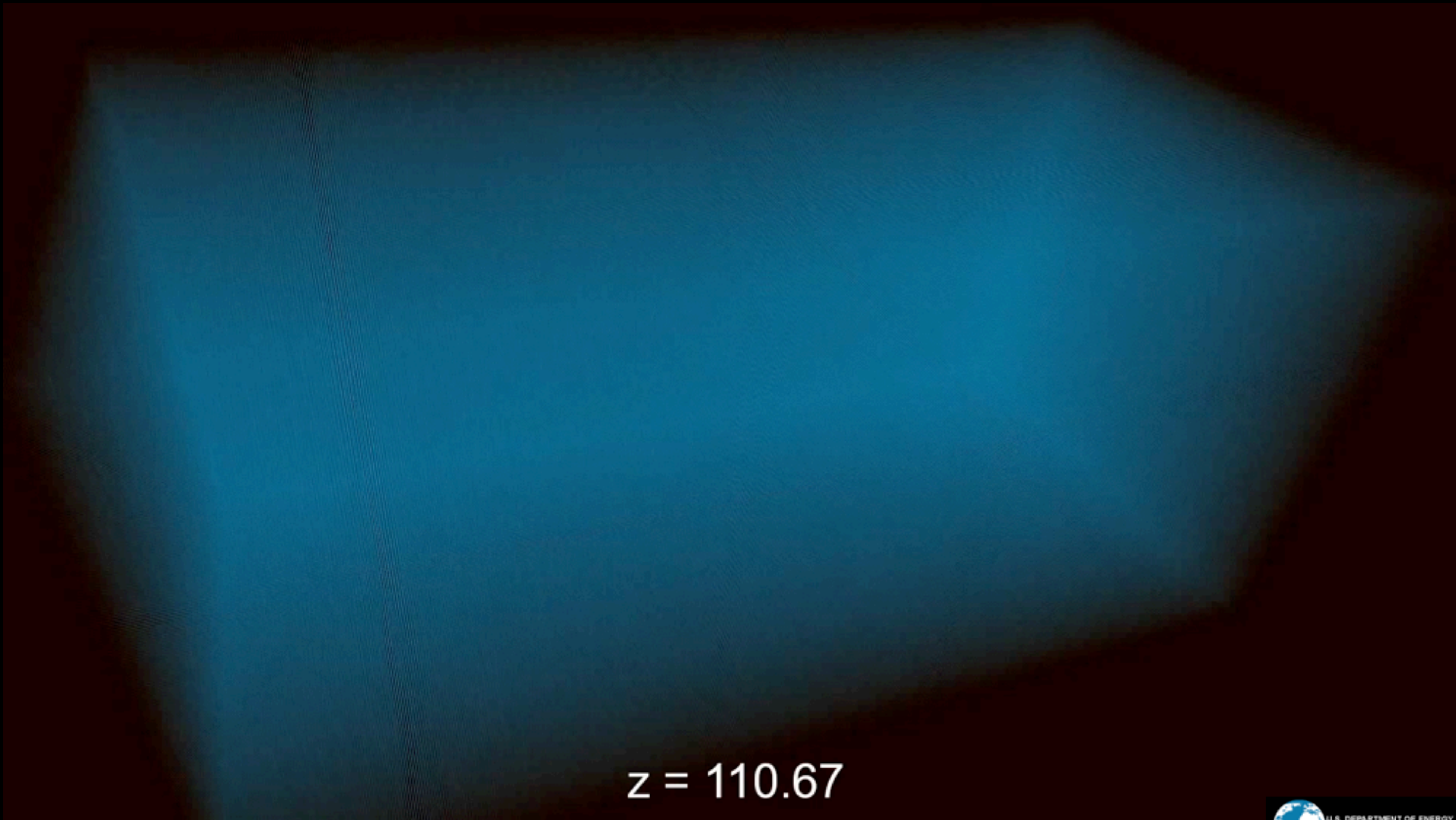
2 Million particles total, particle mass = $8*10^9 M_\odot$

Dark Matter Particles in a Cosmological Simulation



64 Mpc

The high resolution Q Continuum Simulation, finished July 13, 2014 on ~90% of Titan under INCITE, evolving more than half a trillion particles. Shown is the output from one node (~33 million particles), 1/16384 of the full simulation



U.S. DEPARTMENT OF ENERGY INCITE LEADERSHIP COMPUTING

Q Continuum: Extradimensional plane of existence

Visualization: Silvio Rizzi, Joe Insley et. al., Argonne

Computational Context: Important Before we Proceed

- **The computational world is changing**
  - Parallelism at all levels
  - Big disconnect between "python-world" and HPC-world
  - Future is murky — multiple architectures and programming models
  - Large-scale simulation projects are getting very hard to manage; what's the best way to collaborate?
  - Is the small code team on the way out?
  - How to make advanced computation an easy to use tool for theorists?
  - Will the Cloud solve all our problems?

- **What should you do?**
  - Good question!
  - Hopefully these lectures will help —