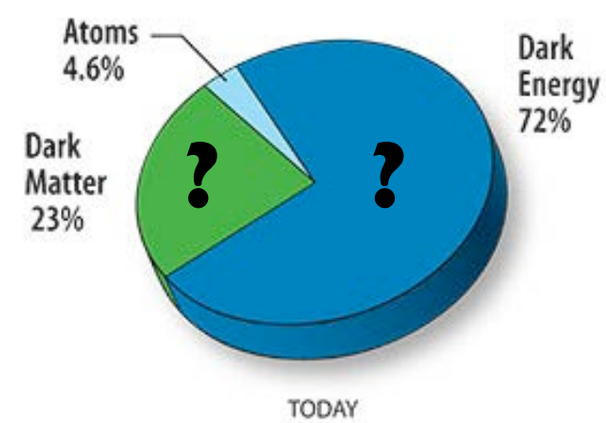
Computing at the Cosmic Frontier

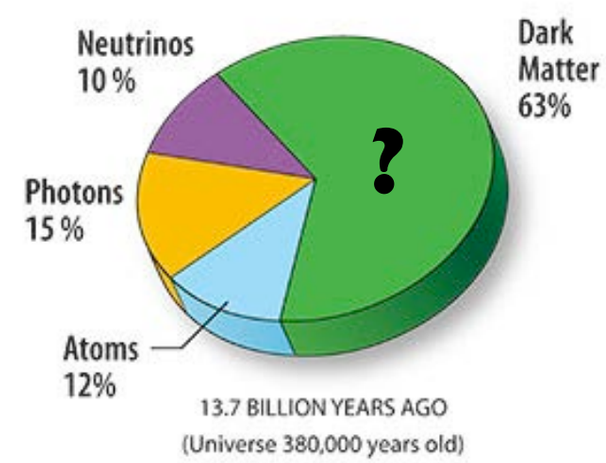
Dark Energy
Dark Matter
Neutrinos
Inflation

SciDAC-3 Project: Computation-Driven Discovery for the Dark Universe

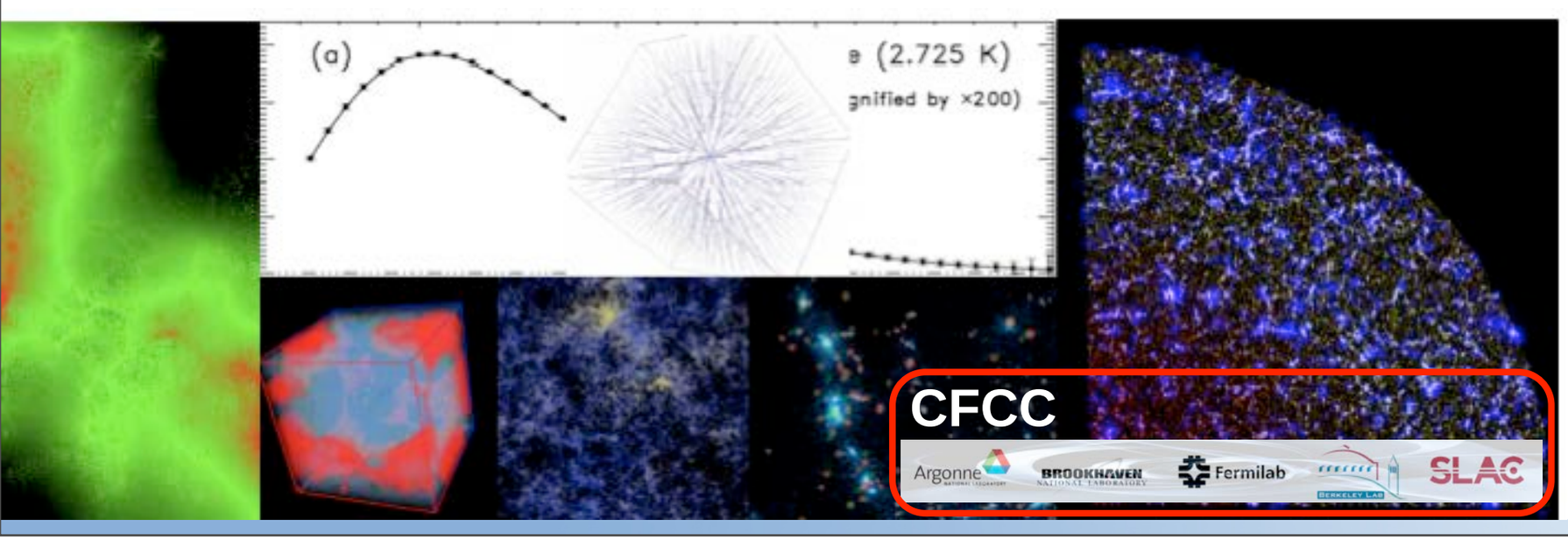
Salman Habib (Proj. Dir., ANL)
Jim Ahrens (PI, LANL)
Katrin Heitmann (PI, ANL)
Peter Nugent (PI, LBNL)
Anze Slosar (PI, BNL)
Risa Wechsler (PI, SLAC)

Institute Partnerships:
FASTMath (Ann Almgren, LBNL)
QUEST (Dave Higdon, LANL)
SDAV (Rob Ross, ANL)



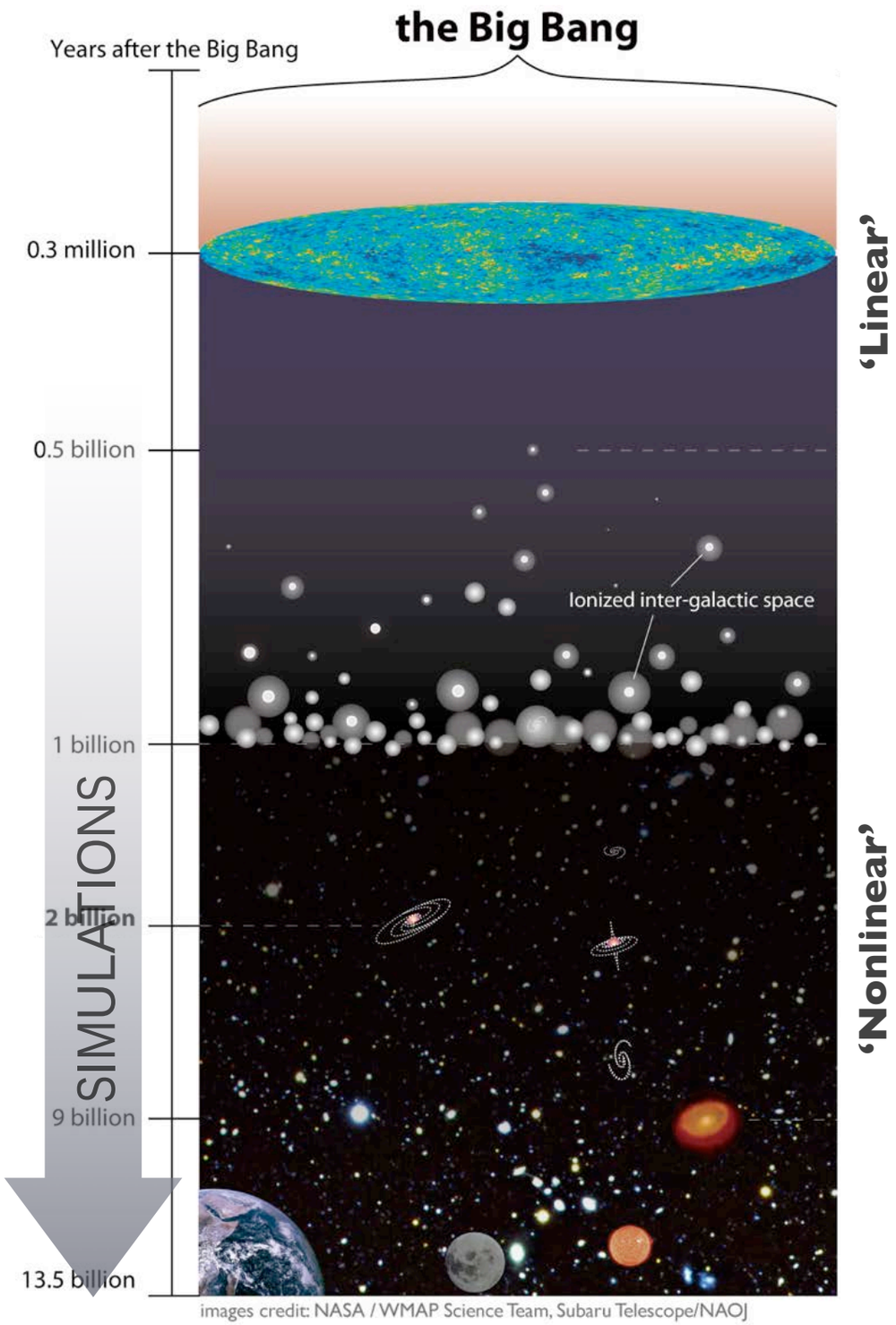


Cosmic content pie charts



Structure Formation in the Universe: The Basic Paradigm

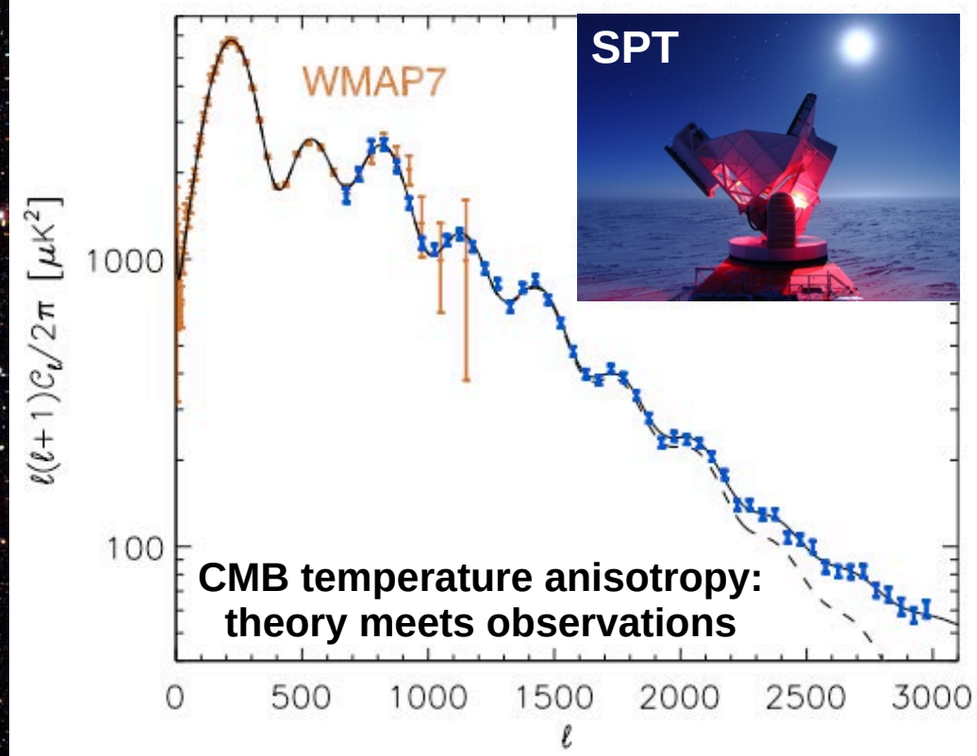
- **Solid understanding of structure formation underpins most cosmic discovery**
  - To high accuracy, initial conditions are given by a Gaussian random field
  - Initial perturbations amplified by gravitational instability in a dark matter-dominated Universe
  - Relevant theory is gravity and atomic physics (‘first principles’)
- **Early Universe: Linear perturbation theory very successful (Cosmic Microwave Background)**
- **Latter half of the history of the Universe: Nonlinear domain of structure formation, impossible to treat without large-scale computing**

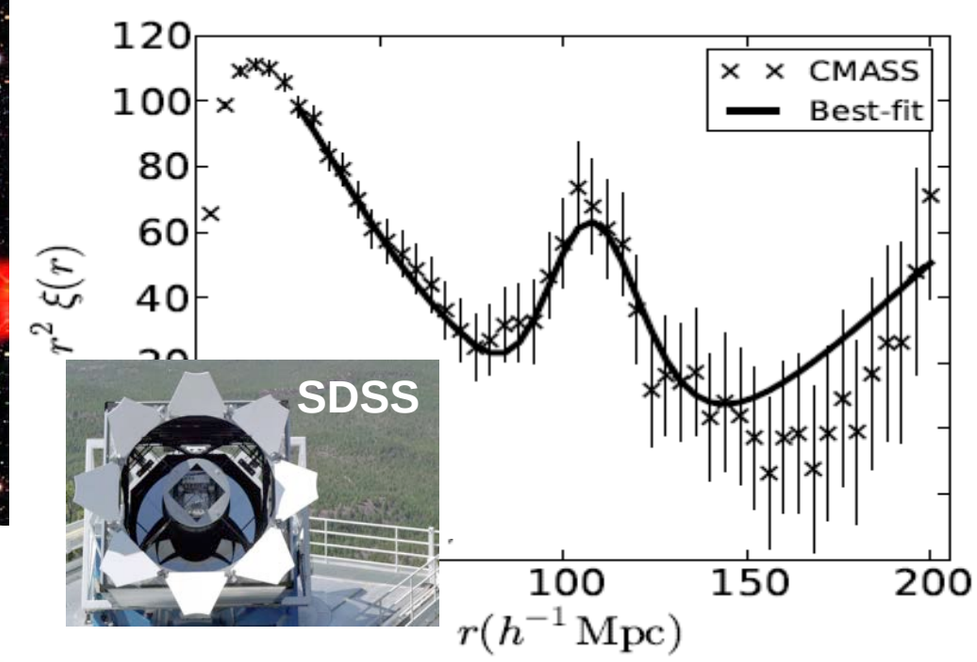


Observations of Cosmic Structure

- Cosmology=Physics+Statistics
  - Mapping the sky with large-area surveys across multiple wave-bands
  - Many different probes: abundances, clustering, weak lensing, redshift space distortions, cross-correlations --

Galaxies in a patch of sky with area roughly the size of the full moon as seen from the ground (Deep Lens Survey). LSST will cover an area 50,000 times this size (and go deeper)



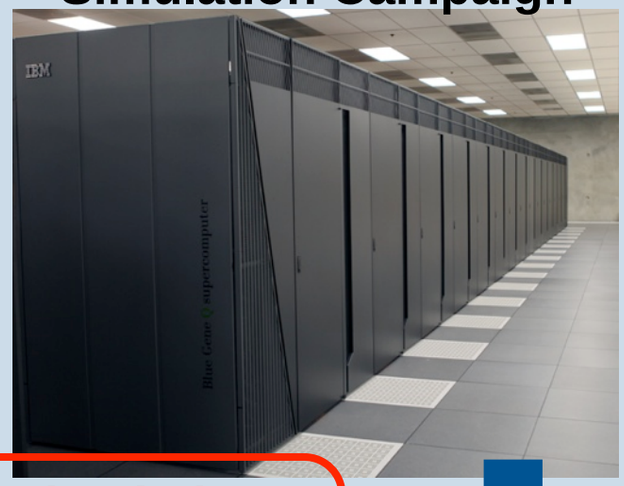


The same signal in the galaxy distibution

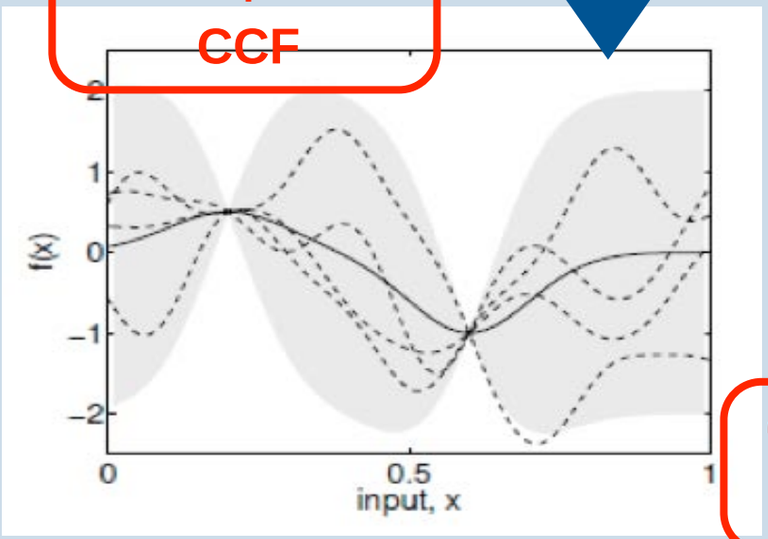
Setting the Stage: Precision Cosmology

SciDAC-3 Project

Supercomputer Simulation Campaign



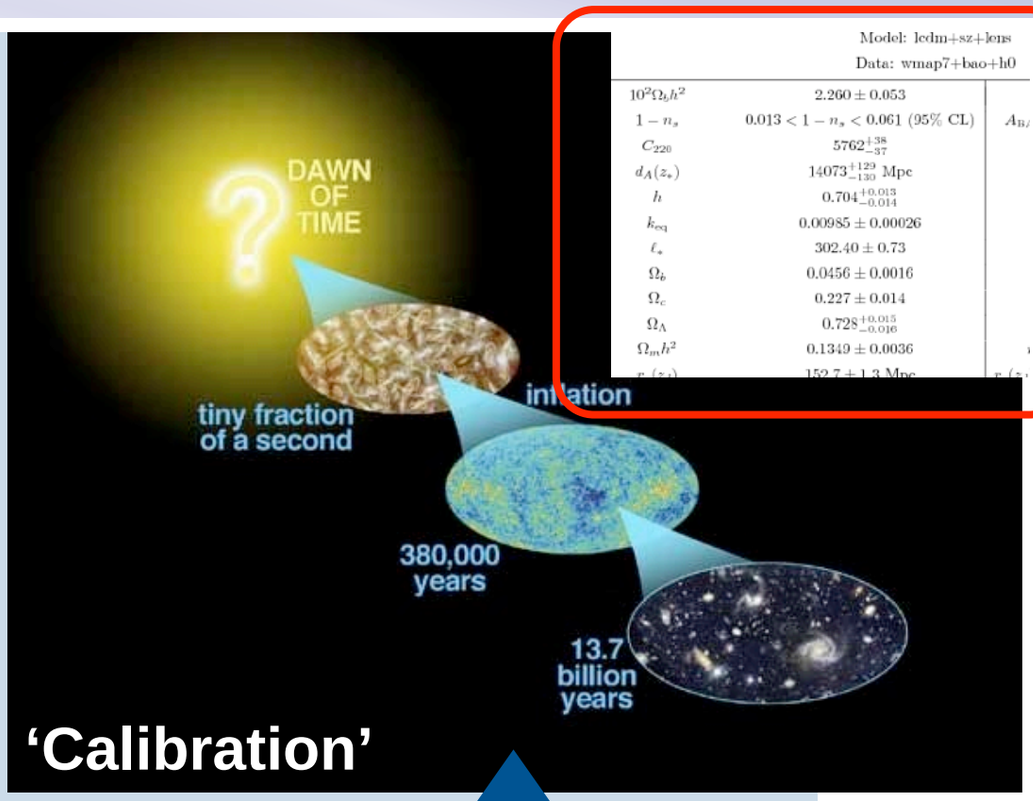
Simulations + CCF



Emulator based on Gaussian Process Interpolation in High-Dimensional Spaces

CCF= Cosmic Calibration Framework

'Precision Oracle'

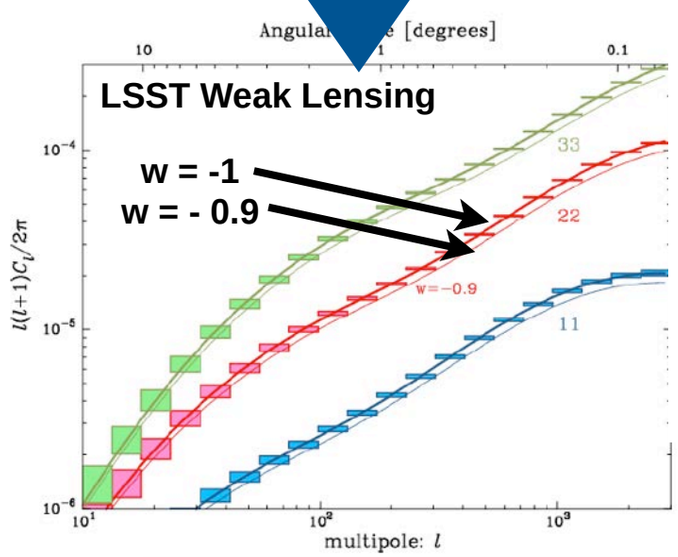


	Model: lcdm+sz+lens Data: wmap7+bao+h0
$10^2\Omega_b h^2$	2.260 ± 0.053
$1-n_s$	$0.013 < 1-n_s < 0.061$ (95% CL)
C_{220}	5762^{+38}_{-37}
$d_A(z_*)$	14073^{+129}_{-130} Mpc
h	$0.704^{+0.013}_{-0.014}$
k_{eq}	0.00985 ± 0.00026
ℓ_*	302.40 ± 0.73
Ω_b	0.0456 ± 0.0016
Ω_c	0.227 ± 0.014
Ω_Λ	$0.728^{+0.015}_{-0.016}$
$\Omega_m h^2$	0.1349 ± 0.0036

'Calibration'

Markov chain Monte Carlo

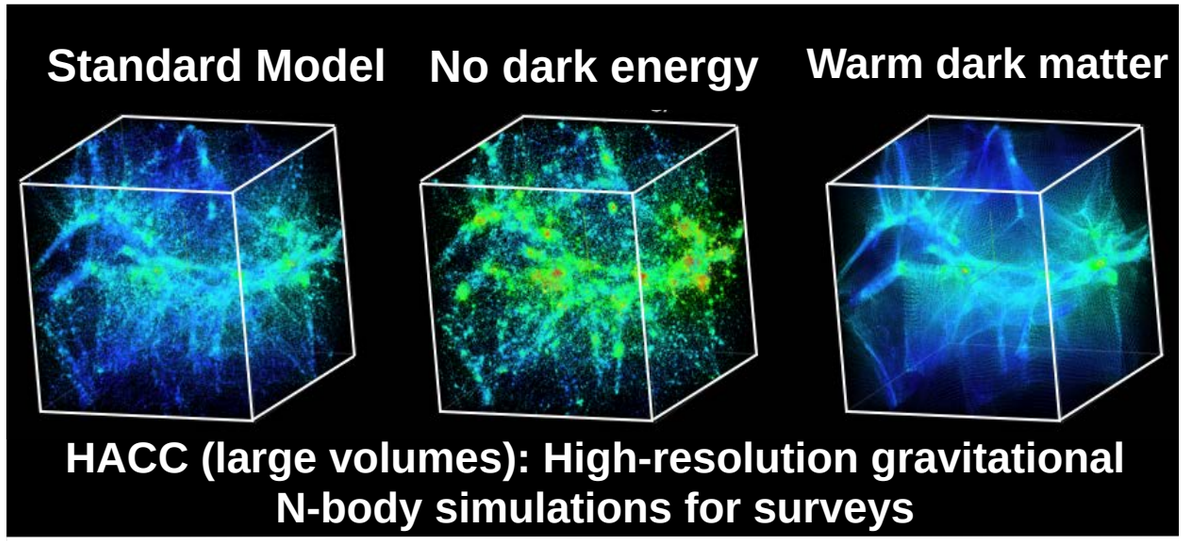
Mapping the Sky with Survey Instruments



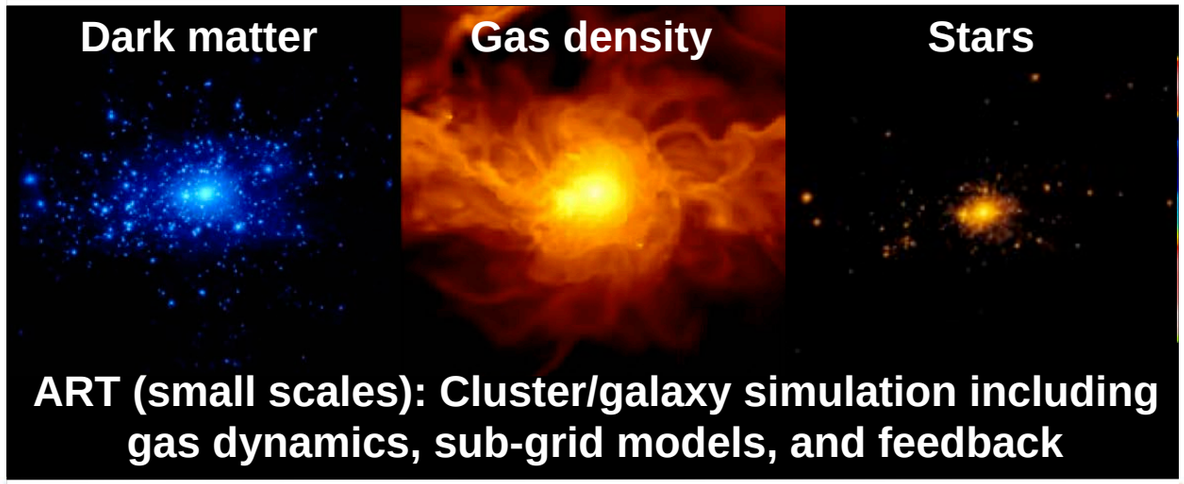
Observations: Statistical error bars will 'disappear' soon!

SciDAC Project: Overview

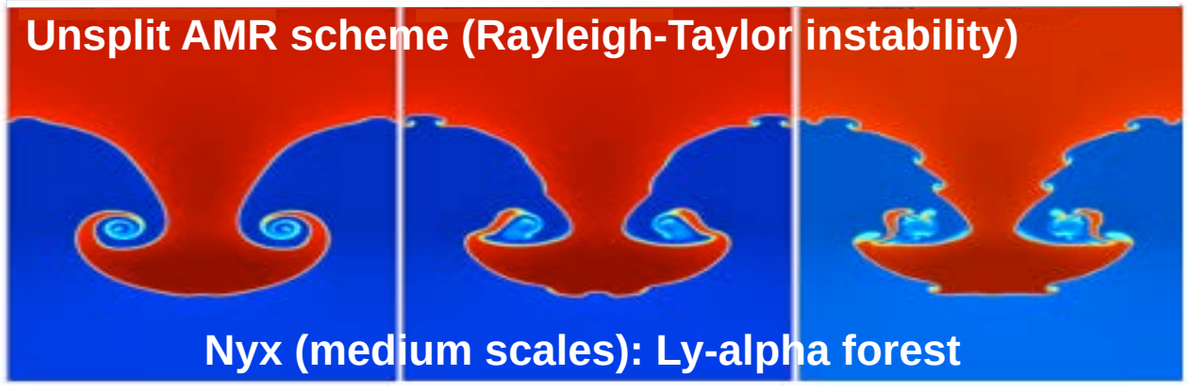
- **SciDAC-3 Project Aims**
  - Build next-generation computational cosmology prediction and analysis frameworks for current and future surveys
  - Explore the physics of dark energy, dark matter, neutrinos, and the early Universe via large-scale structure probes, maintain close relationship with observations
  - Further development of three large-scale high-performance cosmology simulation codes: HACC (@ANL), ART (@Fermilab/ UChicago), Nyx (@LBNL)
  - Make full use of DOE's Leadership class systems: Mira (ANL), Jaguar/Titan (ORNL), Hopper (NERSC)
  - In situ and post-processing analysis frameworks for direct comparison against observations



HACC (large volumes): High-resolution gravitational N-body simulations for surveys



ART (small scales): Cluster/galaxy simulation including gas dynamics, sub-grid models, and feedback



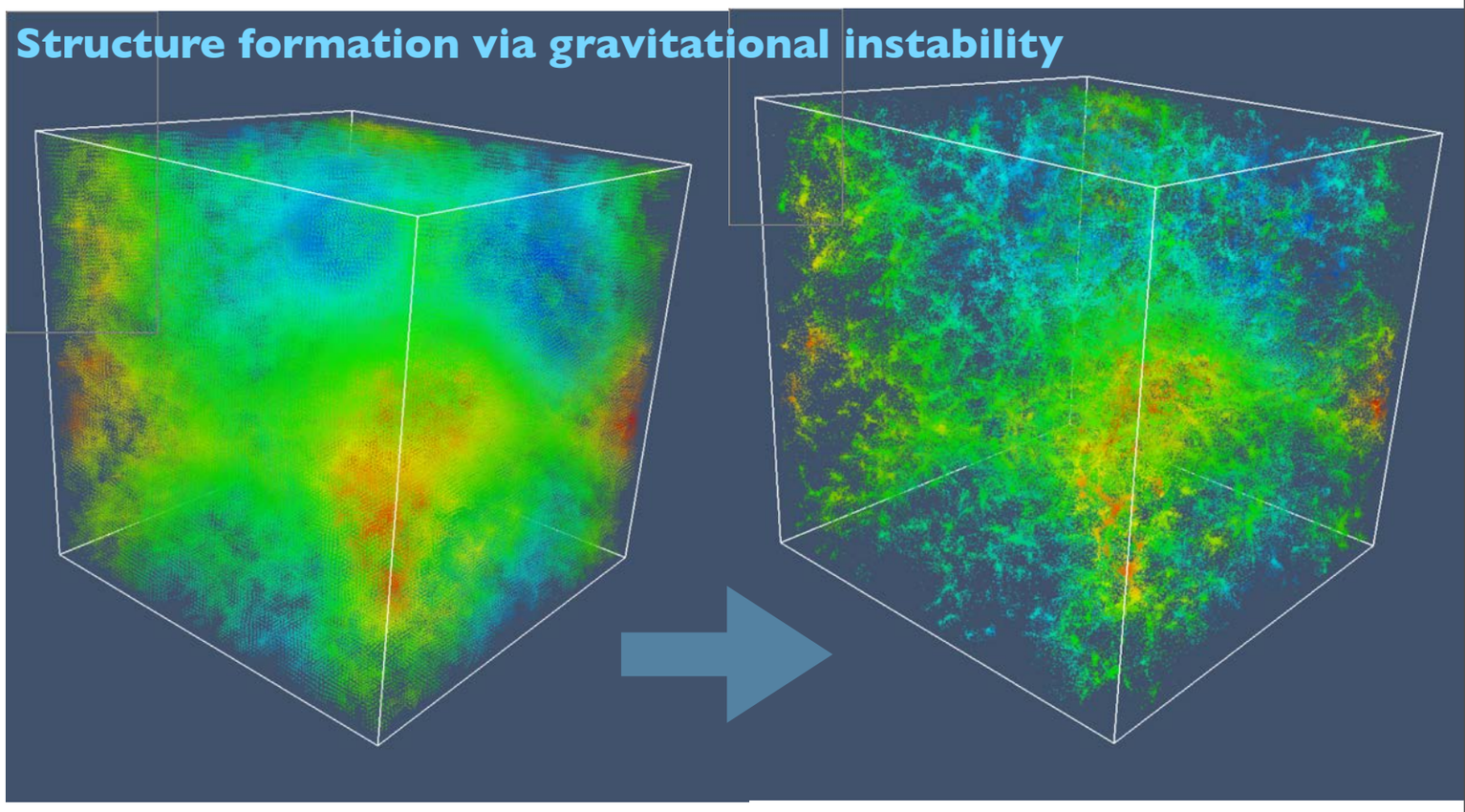
Nyx (medium scales): Ly-alpha forest

Forward Problem: Simulating the Universe

- **Gravity dominates at large scales:** Solve the Vlasov-Poisson equation (VPE)
- **Nonlinearity:** 6-D VPE cannot be solved as a PDE, must use N-body methods
- **Errors:** How well can they be controlled? **'Math'/UQ**
- **Astrophysics:** At small scales need to add gas physics, feedback, 'sub-grid' physics (AMR/SPH gasdynamics) **'Math'**
- **Phenomenology:** Calibrate theory and simulations against observations ('self-calibration') **Stats/UQ**

$$
\begin{aligned}
\frac{\partial f_i}{\partial t} + \dot{\mathbf{x}}\frac{\partial f_i}{\partial \mathbf{x}} - \nabla\phi\frac{\partial f_i}{\partial \mathbf{p}} &= 0, \qquad \mathbf{p} = a^2\dot{\mathbf{x}}, \\
\nabla^2\phi &= 4\pi G a^2(\rho(\mathbf{x},t) - \langle\rho_{\rm dm}(t)\rangle) = 4\pi G a^2 \Omega_{\rm dm}\delta_{\rm dm}\rho_{\rm cr}, \\
\delta_{\rm dm}(\mathbf{x},t) &= (\rho_{\rm dm} - \langle\rho_{\rm dm}\rangle)/\langle\rho_{\rm dm}\rangle), \\
\rho_{\rm dm}(\mathbf{x},t) &= a^{-3}\sum_i m_i \int d^3\mathbf{p} f_i(\mathbf{x},\dot{\mathbf{x}},t).
\end{aligned}
$$

Gasdynamics/feedback equations do not fit on a slide!



Machine learning

Prehistory: An Early Simulation

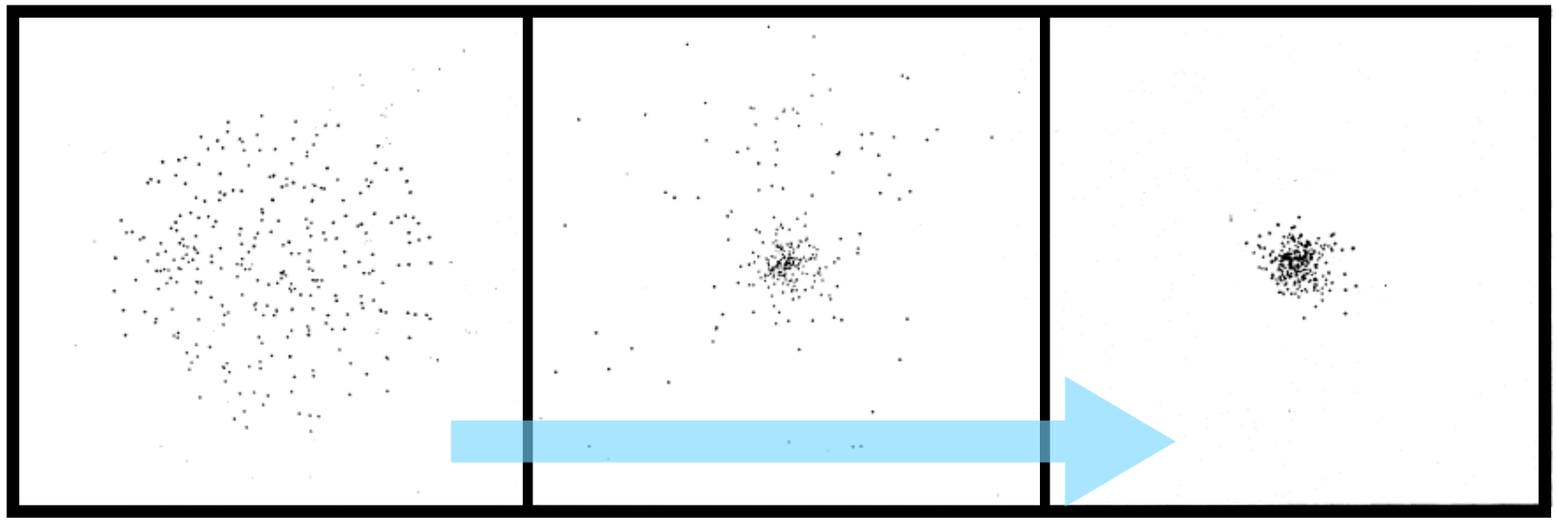
- Suite of 300 (and less) particle simulations
- Run on a CDC 3600, ~1Mflops, 32KB+ at LANL
- Is nine-ten orders of magnitude improvement in both performance and memory good enough for precision cosmology?

THE ASTRONOMICAL JOURNAL VOLUME 75, NUMBER 1 FEBRUARY

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.

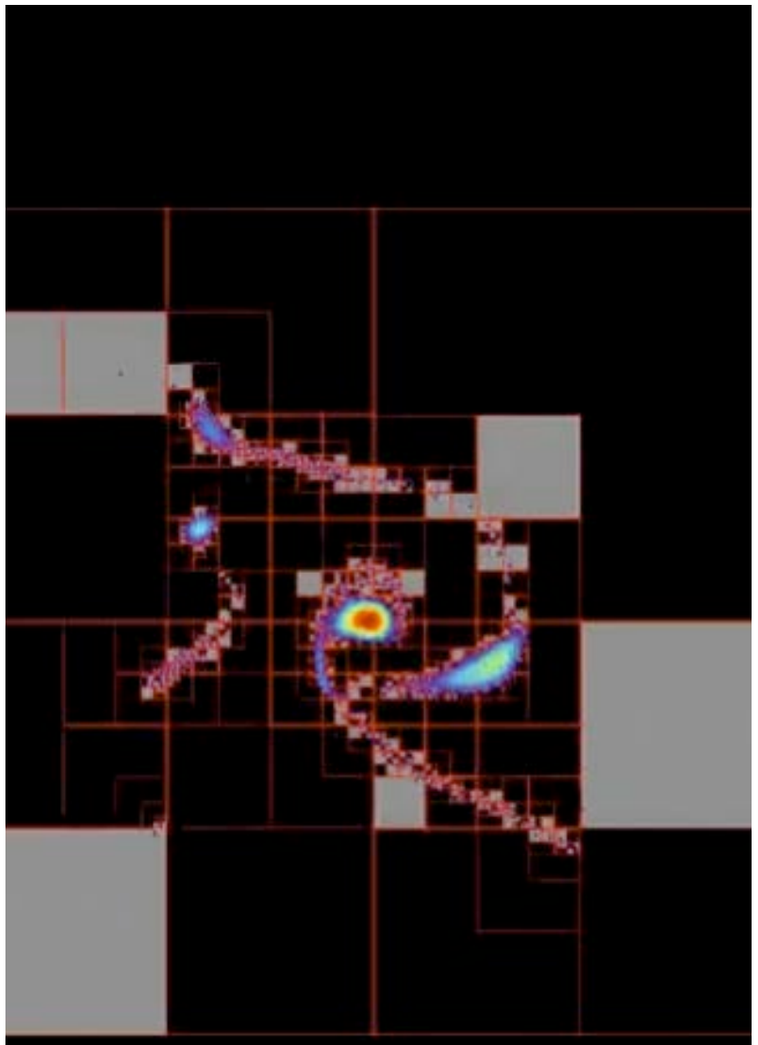


"The Universe is far too complicated a structure to be studied deductively, starting from initial conditions and solving the equations of motion."

Robert Dicke (Jayne Lectures, 1969)

Project Context: Computational Cosmology and HPC

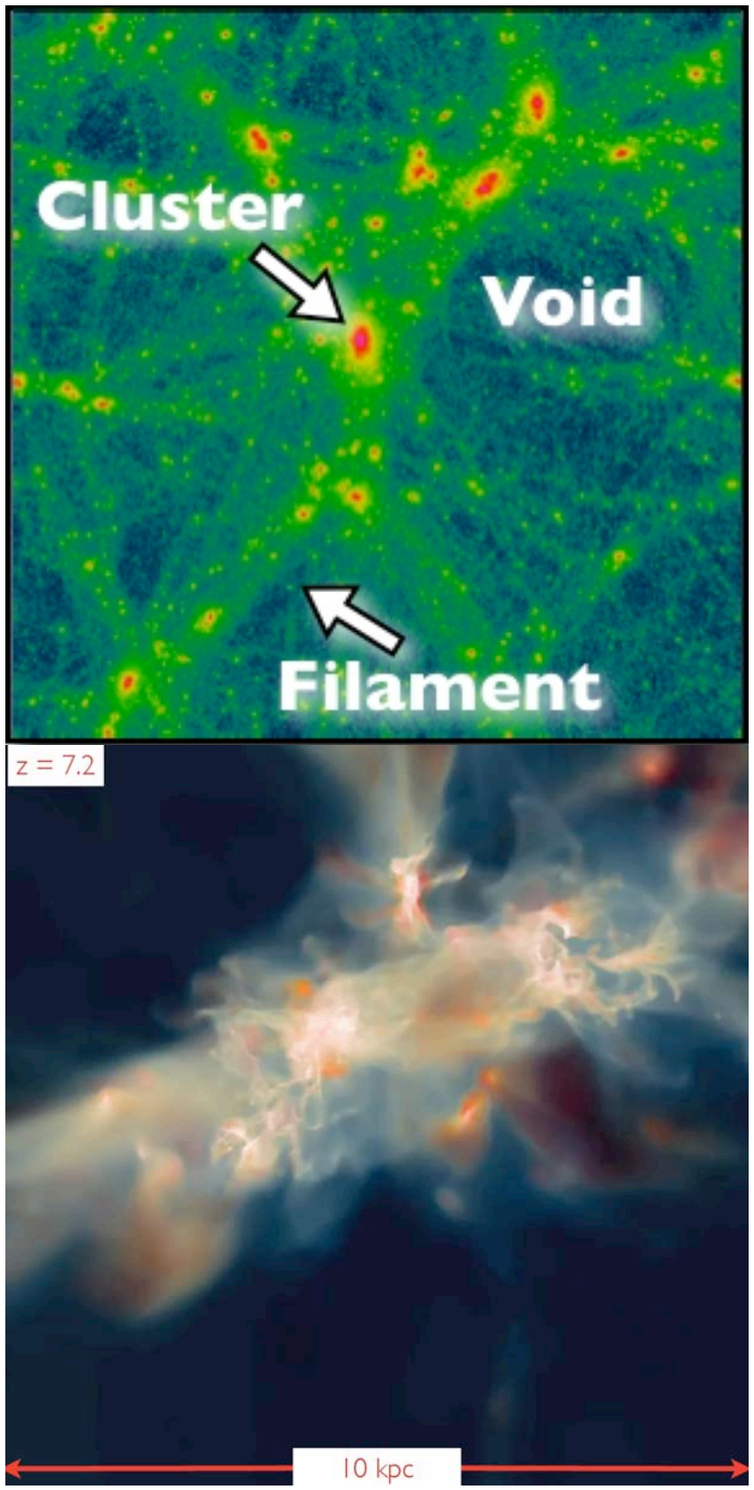
- **Early Era**
  - Early use of vector systems -- PM, adaptive P3M, AMR -- plus tree/FMM algorithms; ‘power user’ history continues to this day
- **Parallel Simulations**
  - Early parallelization of all codes, efficient utilization of supercomputing resources (Gordon Bell awards, including custom hardware)
- **More Physics: Multi-Physics/Multi-Resolution**
  - Some cosmology codes have complexity similar to combustion codes, others optimize for ‘at-scale’ performance and memory (e.g., survey scale volume simulations)
- **Next-Generation Architectures**
  - Increased concurrency with communication bottlenecks and costs -- codes must evolve to deal with these issues; power-awareness next



Barnes-Hut oct-tree algorithm on YouTube

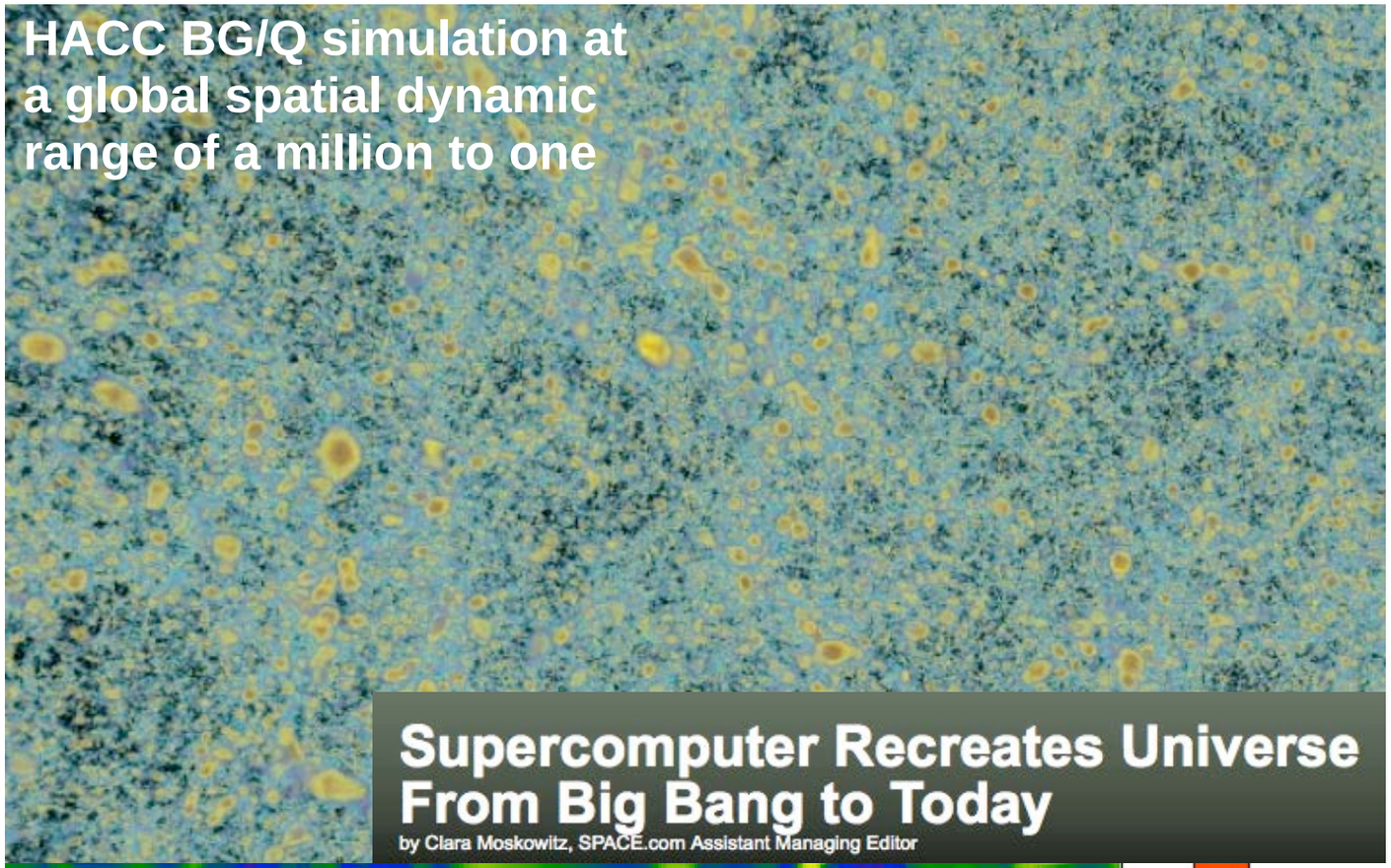
Cosmological Simulations: Science Achievements

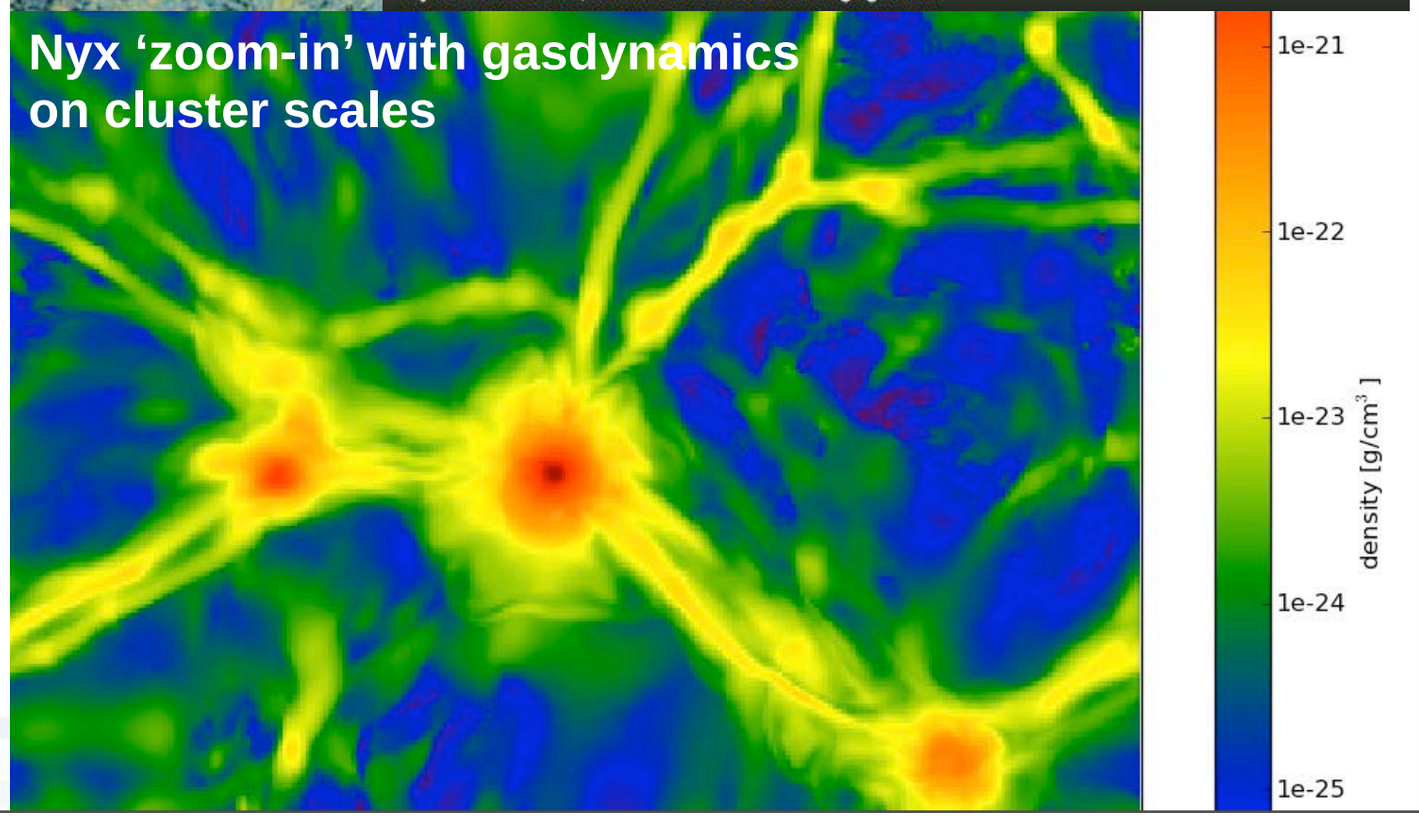
- **Paradigm of Structure Formation**
  - "Cosmic Web" unifies a wide array of observations, from galaxy surveys to the Ly-alpha forest
  - Predictions for halo shapes, substructure, spatial statistics, halo bias (clustering relative to underlying density field), and their evolution
- **Cosmic Probes**
  - Halo abundance and mass distribution
  - Nonlinear domains of clustering probes
- **Physics of the IGM, galaxies, clusters, reionization --**
  - Multi-resolution, multi-physics simulations essential tools for building more detailed picture, understanding of empirical relations and their scatter



Simulation Requirements

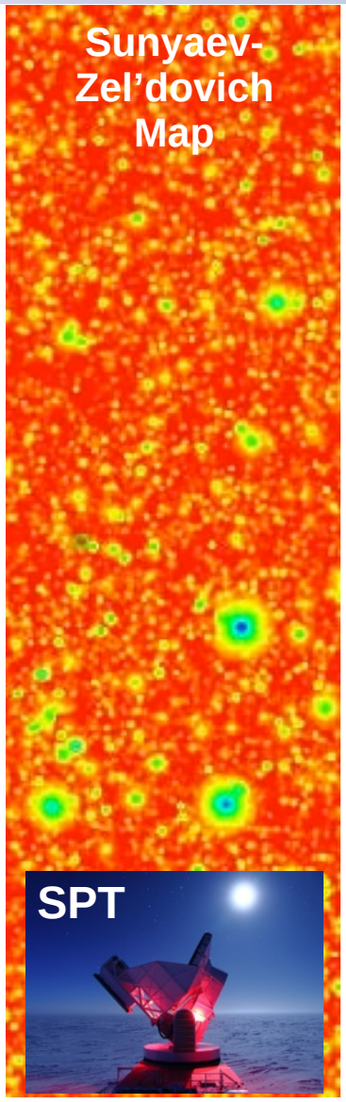
- **Size:**
  - Need to cover volumes in the 100's of Gpc cubed (1 pc=3.26 light-years)
  - To capture individual galaxy mass concentrations over this volume, need trillions of particles
- **Resolution:**
  - Force resolution has to be ~kpc, a global dynamic range of a million to one, also controls time-stepping
  - Local overdensity variation is a million to one
- **Physics (computationally intensive):**
  - Gravity dominates at scales greater than several Mpc
  - Below this scale, pressure forces become important; gasdynamics, feedback processes, radiative processes, star formation, etc. all play important roles

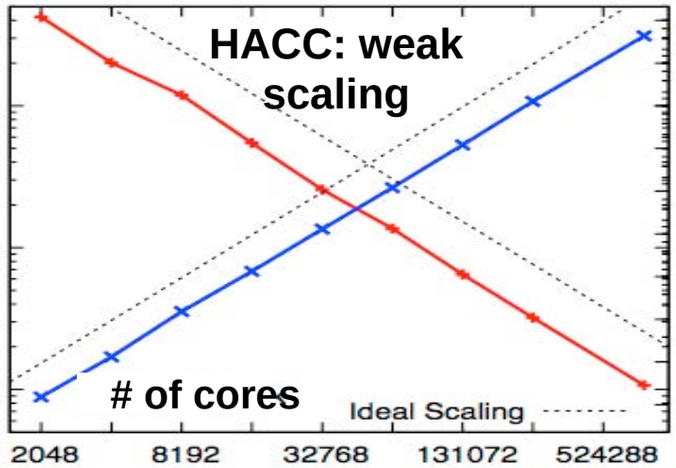


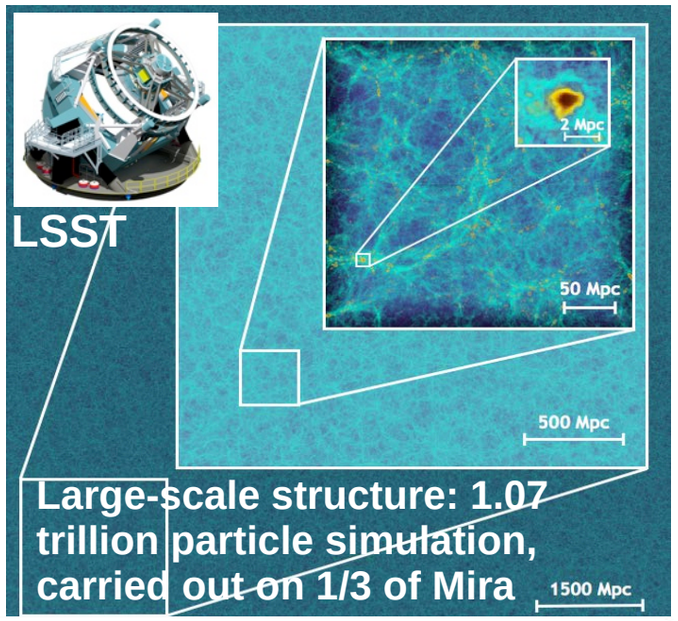


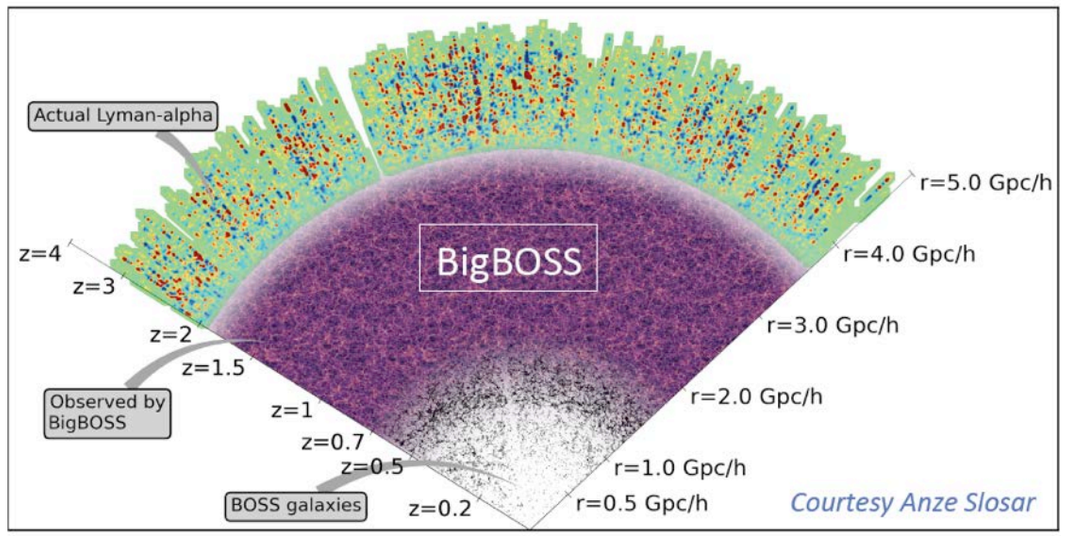
Simulation Status

- **ART:**
  - State of the art physics (cell-structured) gravity/ gasdynamics AMR code, work on improving scalability to beyond ~10K cores
  - Baryonic effects on weak lensing probes, sub-grid models for incorporation in N-body codes
- **HACC:**
  - High-resolution cosmological N-body (PPTPM) code framework; runs on every architecture: MPPs, Cell/ GPU-accelerated, Blue Gene systems, Intel MIC, --
  - 2012 Gordon Bell finalist, > 50% of peak on Mira/ Sequoia at > 90% parallel efficiency (million way), ~2 trillion particles
- **Nyx:**
  - Next-generation (block-structured, BoxLib-based) AMR code undergoing tests, first code paper submitted, physics modules being added
  - Large-scale runs for analysis of BOSS Ly-alpha forest observations



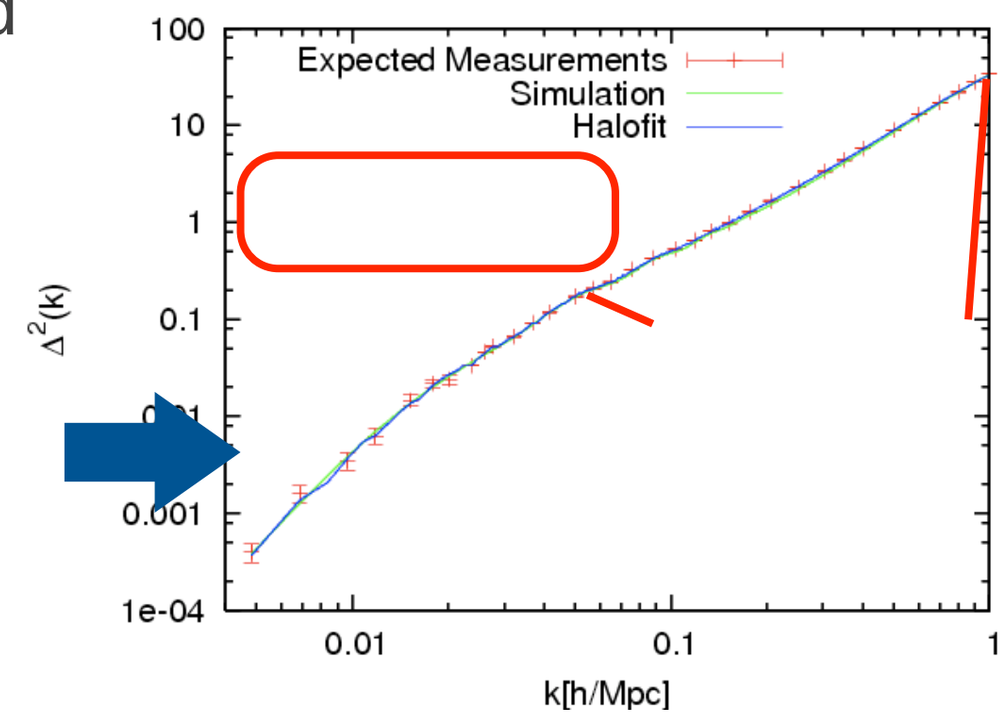


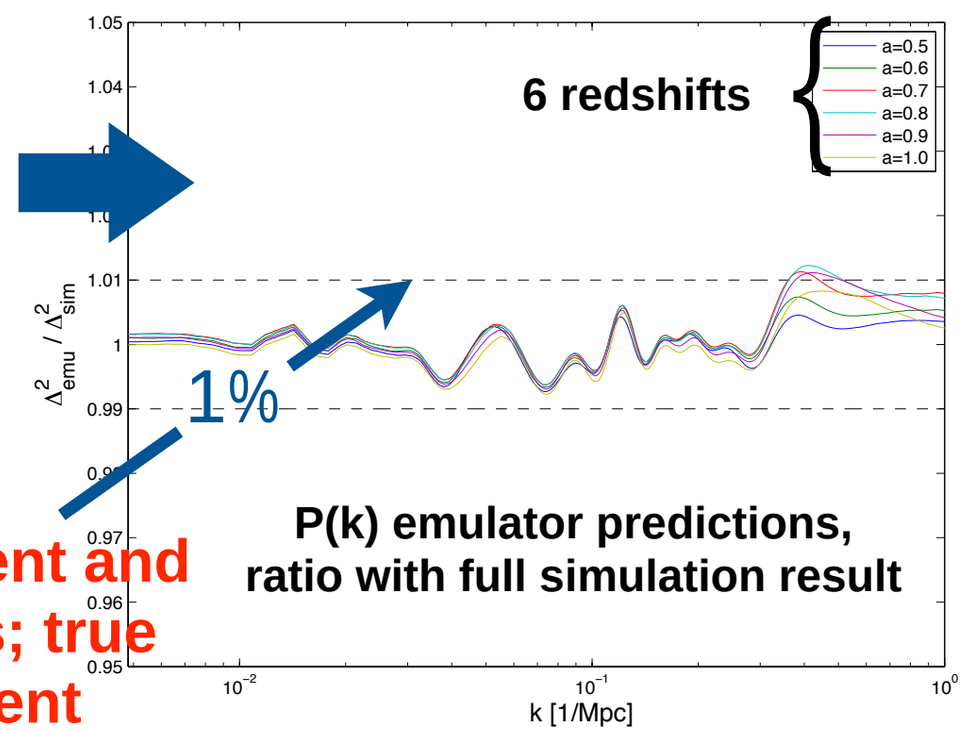




Precision Cosmology: UQ Example

- **Early Period**
  - Access to large data choked (technology did not exist), insignificant computing
  - Characterized by small datasets, ‘eyeball’ comparisons, simple statistics
- **Intermediate Phenomenology (~10%)**
  - Simulation-based intermediate, simplified theoretical model; use this to interact with observations (HOD models, HaloFit, scaling relations from simulations, --)
- **‘Direct’ Numerical Phenomenology (~1%)**
  - ‘Theory’ = interact with observations via sophisticated simulations (or via emulators); understand systematic errors (missing/wrong physics), bias

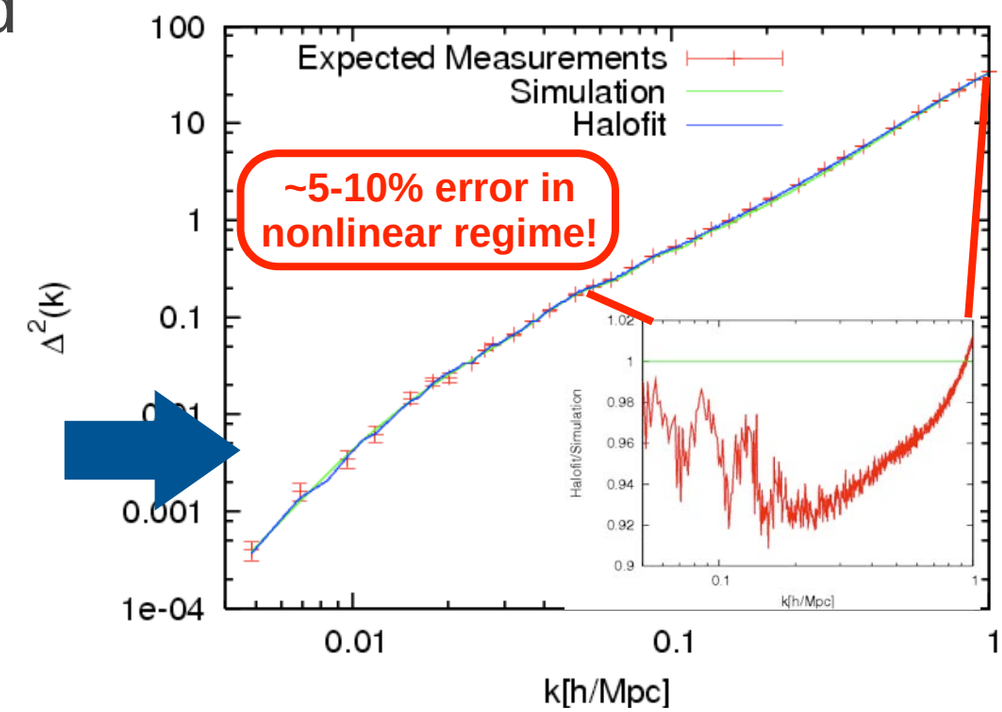


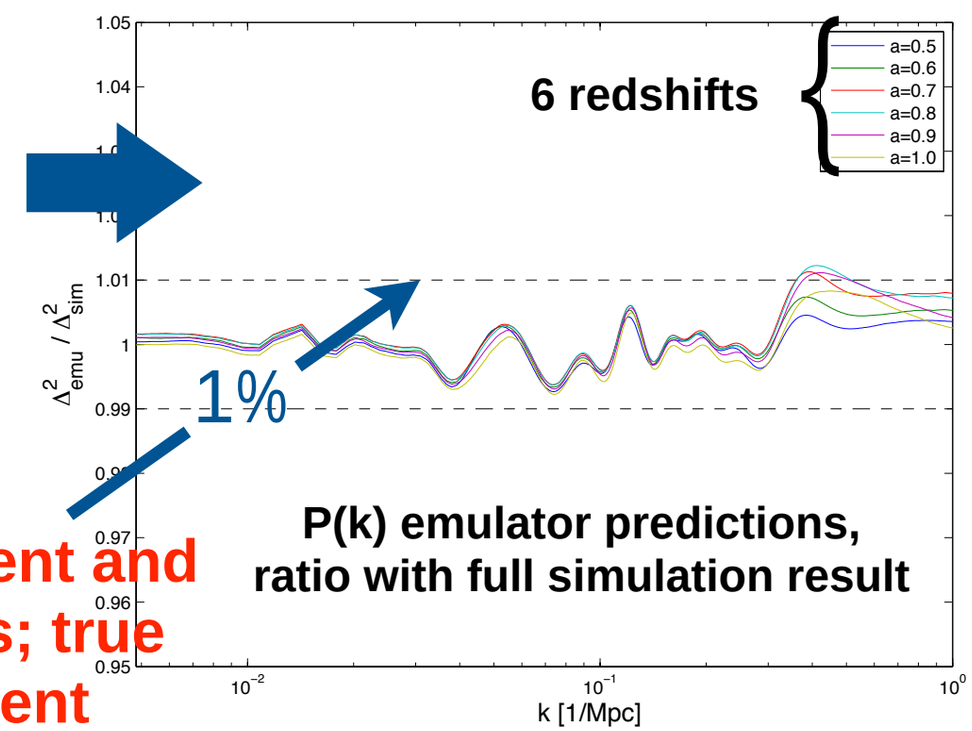


Example of successful development and application of UQ methodologies; true order of magnitude improvement

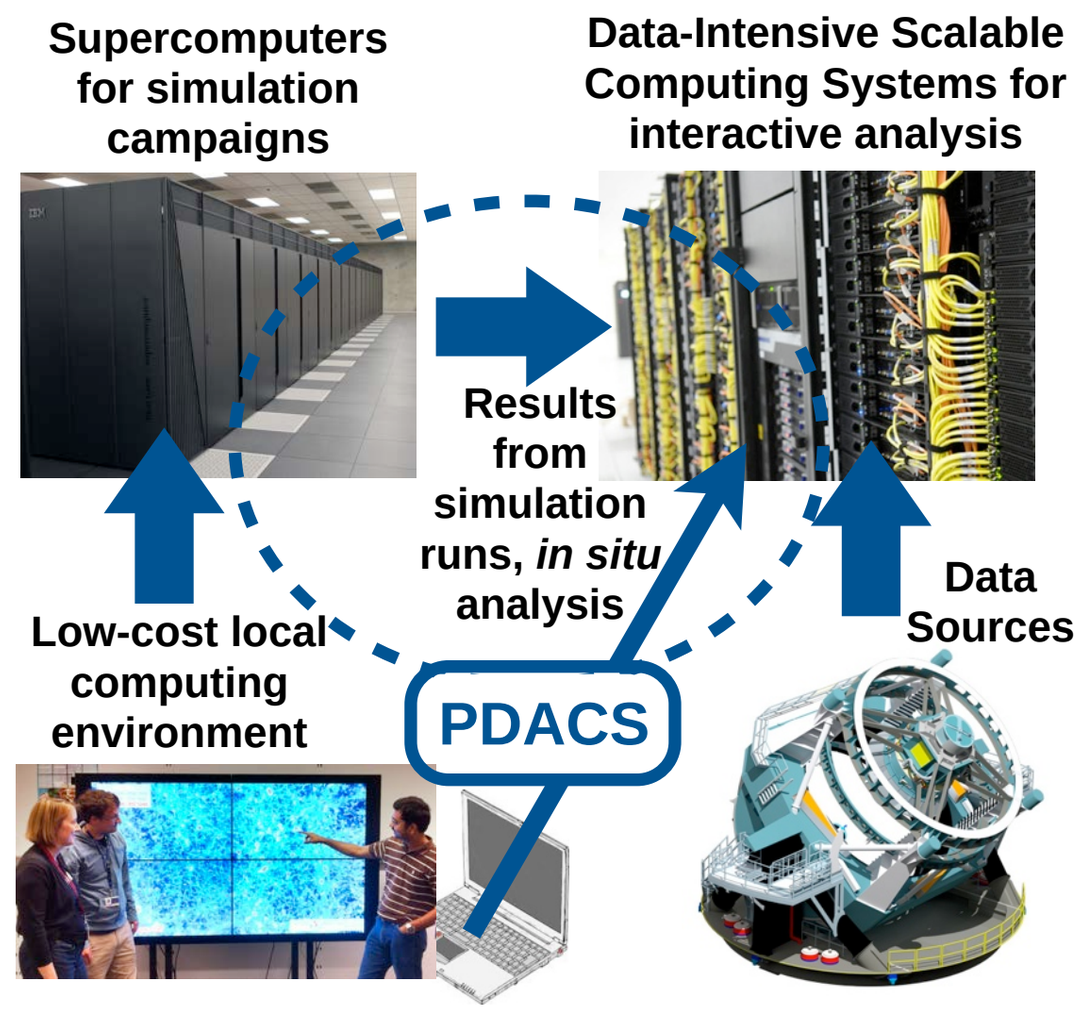
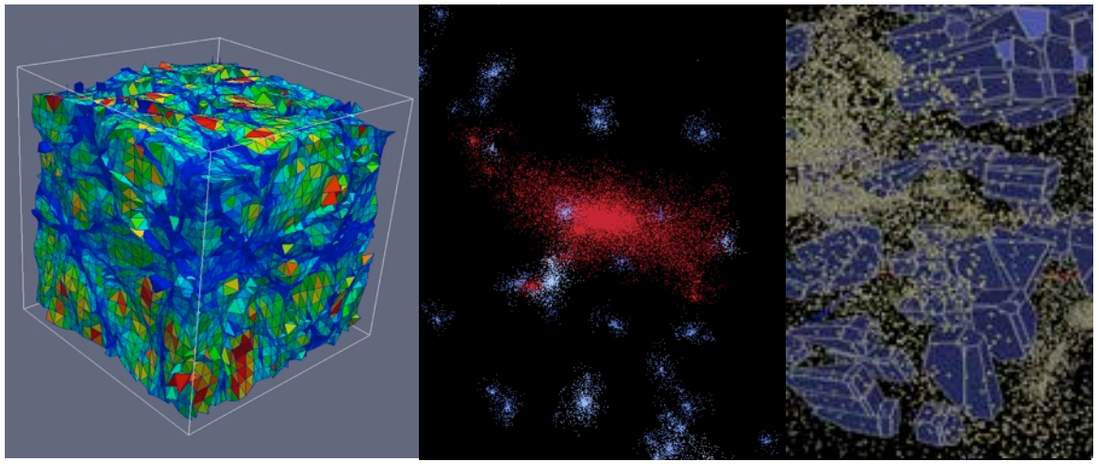
Precision Cosmology: UQ Example

- **Early Period**
  - Access to large data choked (technology did not exist), insignificant computing
  - Characterized by small datasets, ‘eyeball’ comparisons, simple statistics
- **Intermediate Phenomenology (~10%)**
  - Simulation-based intermediate, simplified theoretical model; use this to interact with observations (HOD models, HaloFit, scaling relations from simulations, --)
- **‘Direct’ Numerical Phenomenology (~1%)**
  - ‘Theory’ = interact with observations via sophisticated simulations (or via emulators); understand systematic errors (missing/wrong physics), bias





Example of successful development and application of UQ methodologies; true order of magnitude improvement

Interactions with Institutes

- **Simulation Methods (w/ FASTMath):**
  - AMR/particle method(s) development (dynamic range), multi-physics modeling
  - Design simulation methods for next-to-next-generation platforms
- **Data Issues (w/ SDAV):**
  - High-performance in situ and post-processing analysis framework
  - Parallel I/O, data containers, databases
  - Visualization-aided analysis, including remote/real-time visualization
- **UQ (w/ QUEST):**
  - Next-gen. emulator development; error covariances
  - Discrepancy functions and self-calibration
  - MCMC alternatives and error studies
- **Other (w/ SUPER):**
  - Performance/power prediction
  - Resilience (local compute and network)