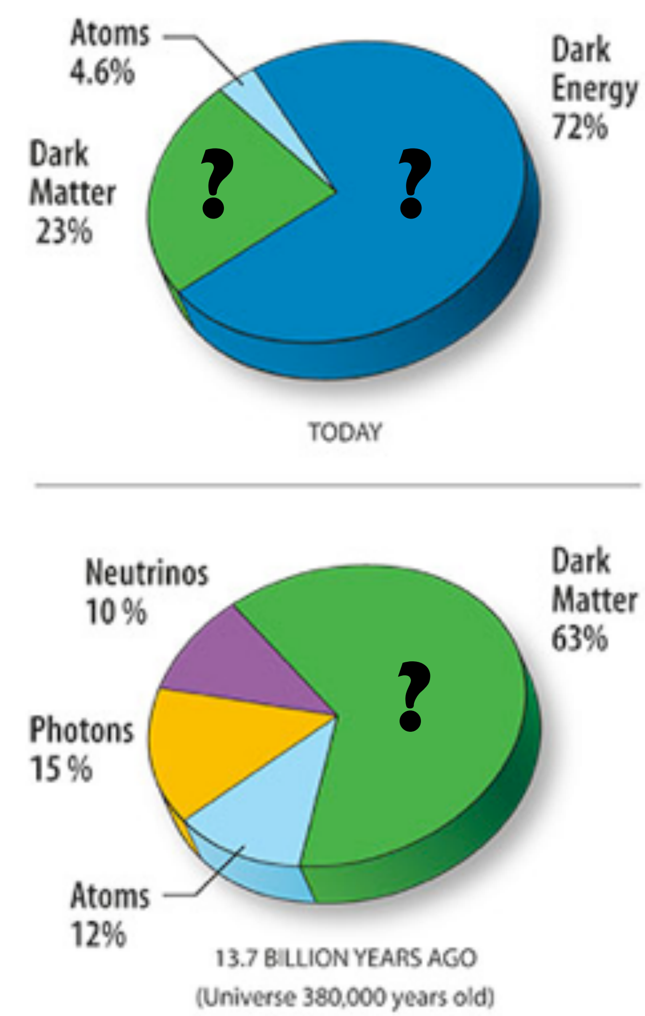
Computing at the Cosmic Frontier

Dark Energy
Dark Matter
Neutrinos
Inflation

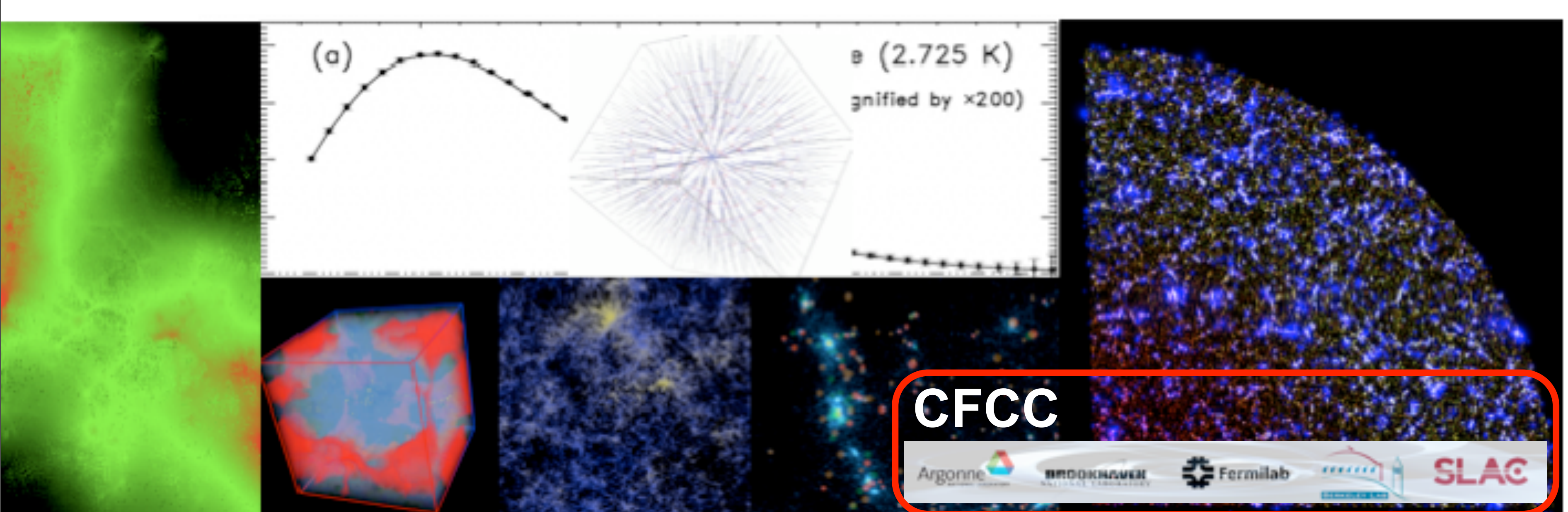
SciDAC-3 Project: Computation-Driven Discovery for the Dark Universe

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)

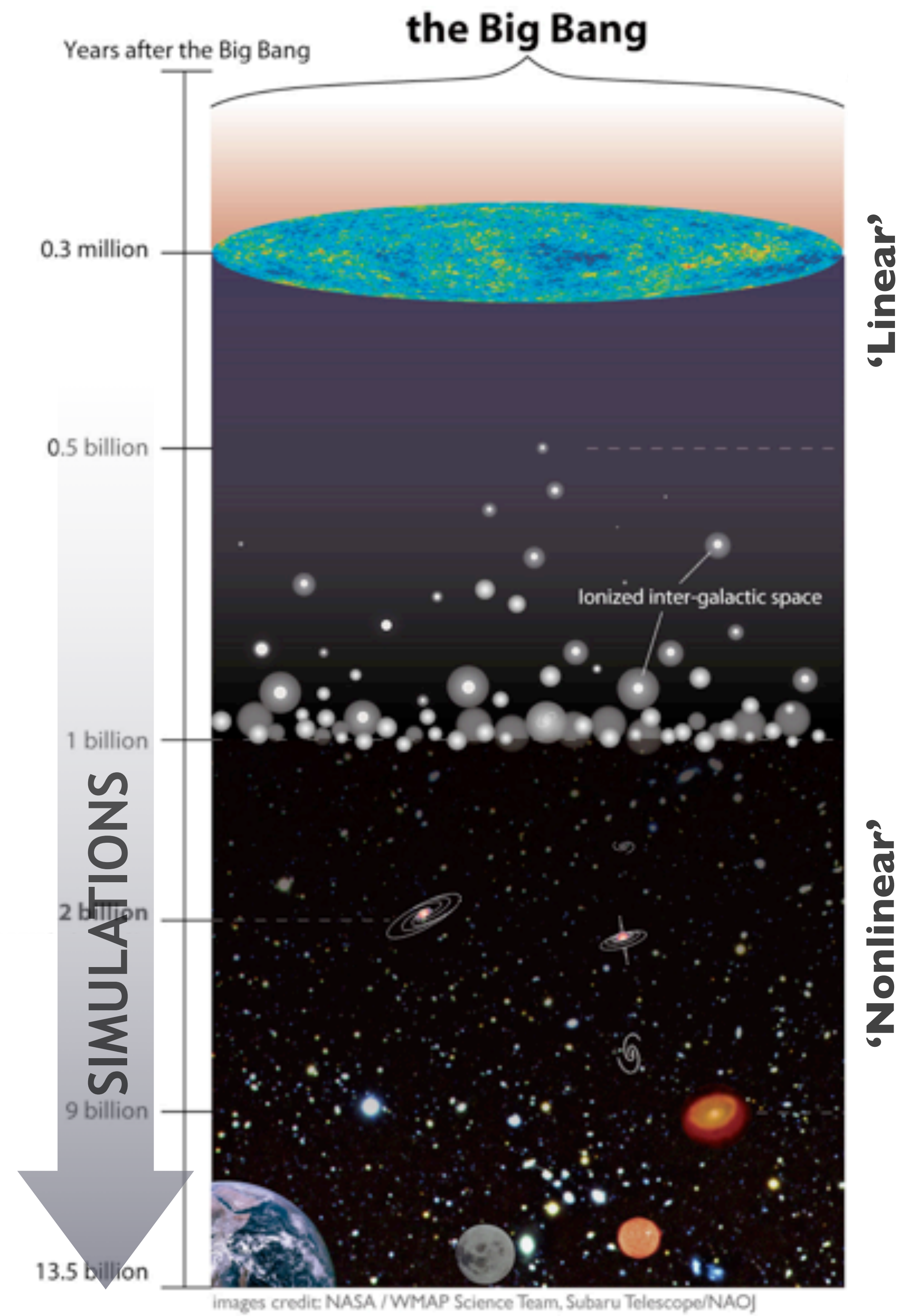


Cosmic content pie charts



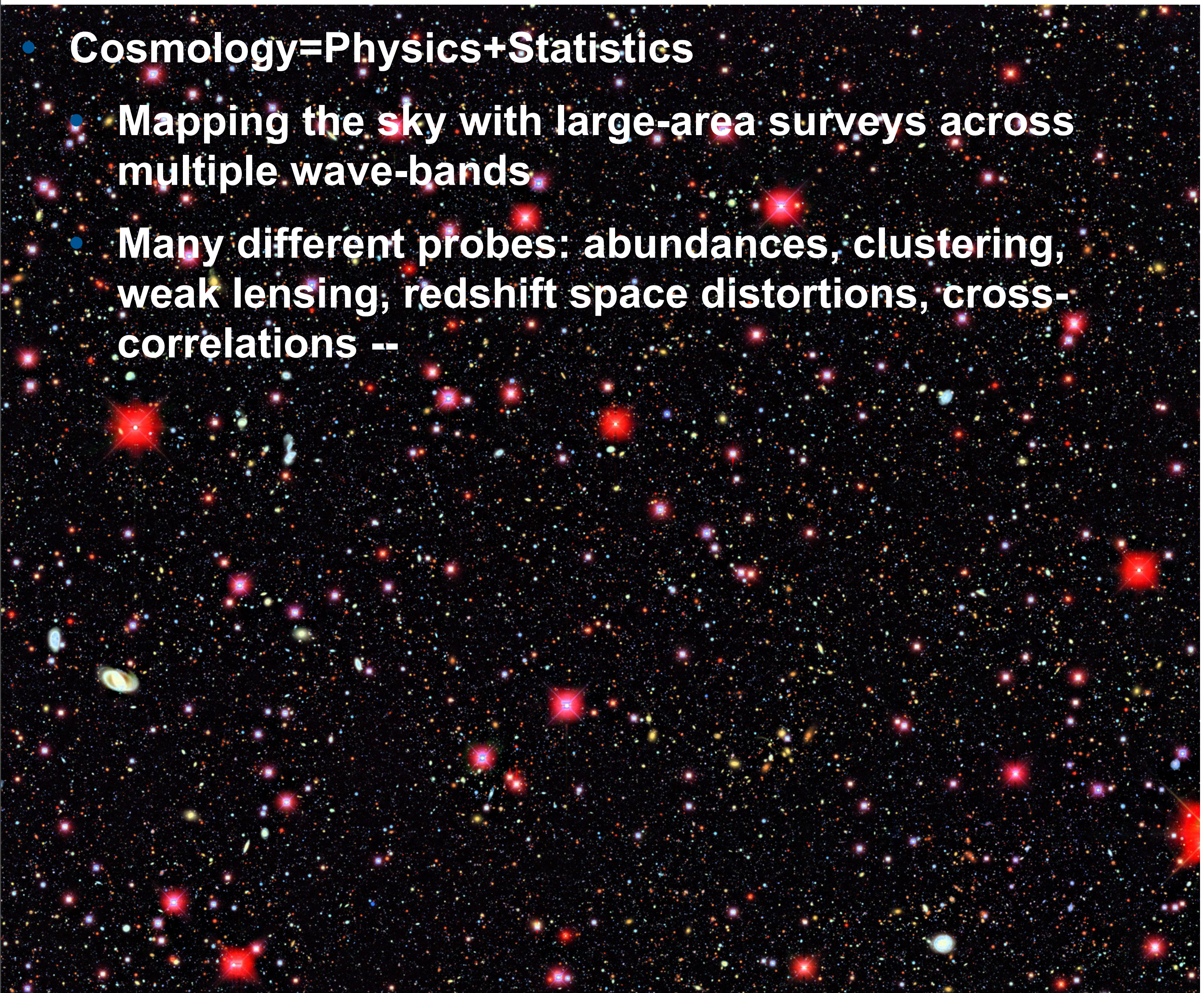
Structure Formation in the Universe: The Basic Paradigm

- **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**

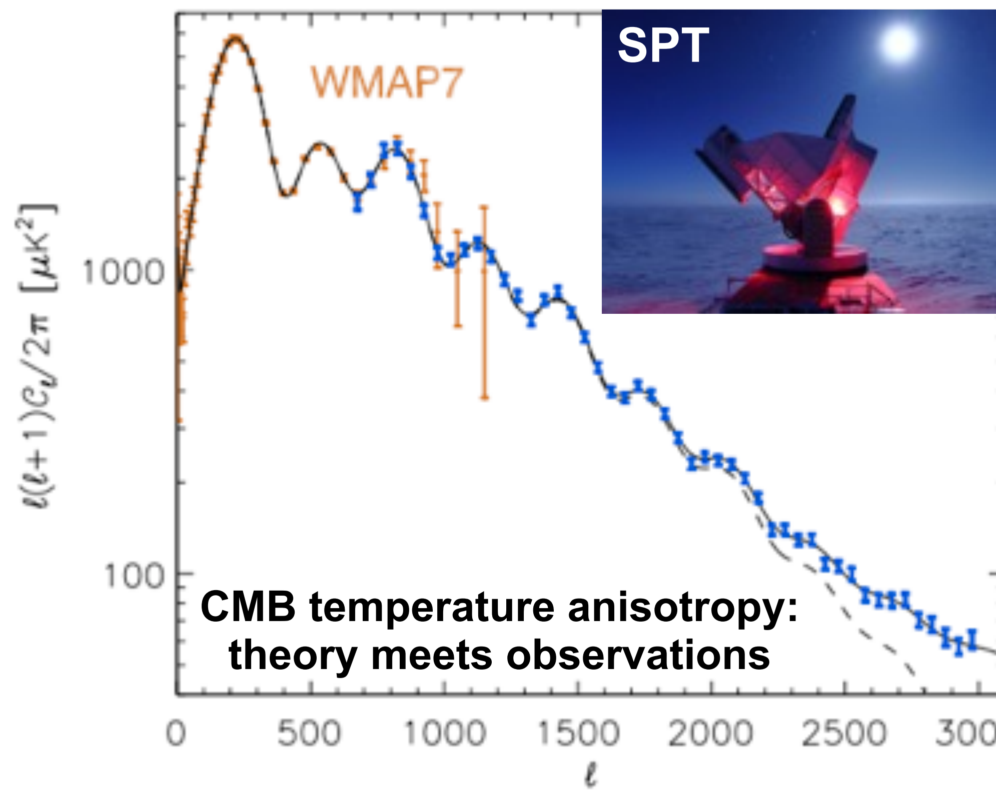


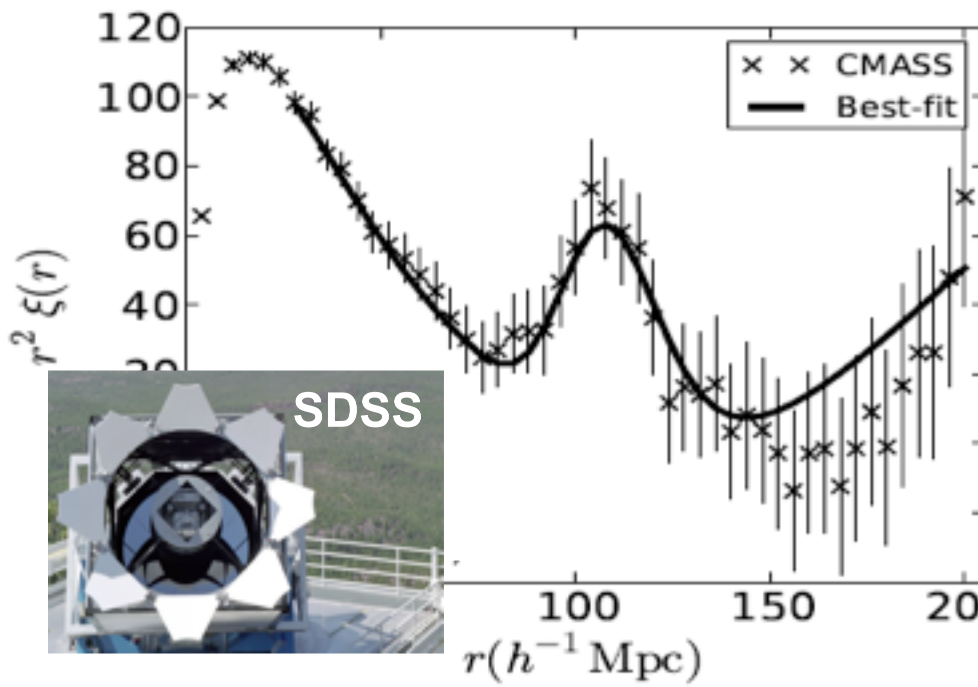
Observations of Cosmic Structure

- 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)



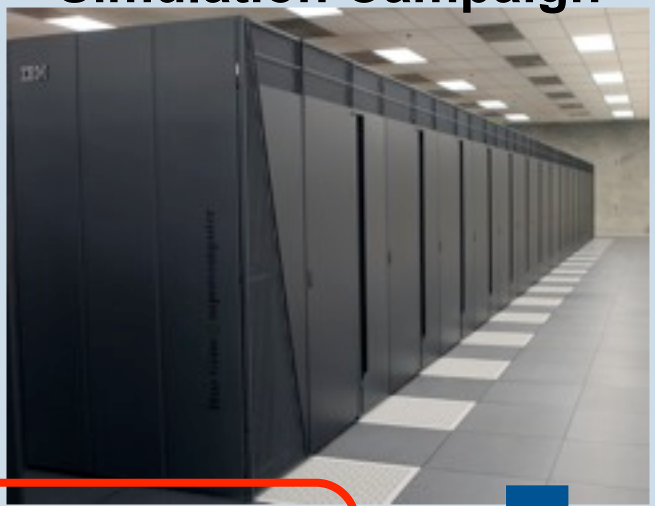


The same signal in the galaxy distibution

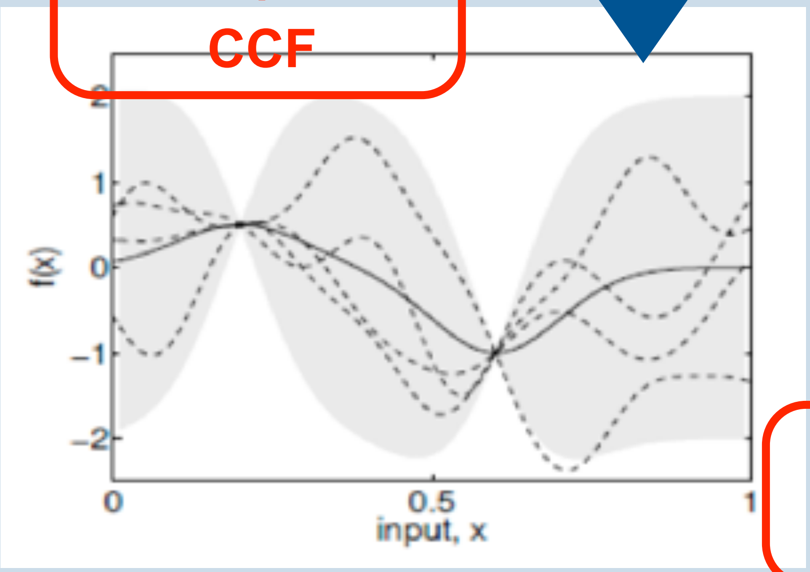
Setting the Stage: Precision Cosmology

SciDAC-3 Project

Supercomputer Simulation Campaign



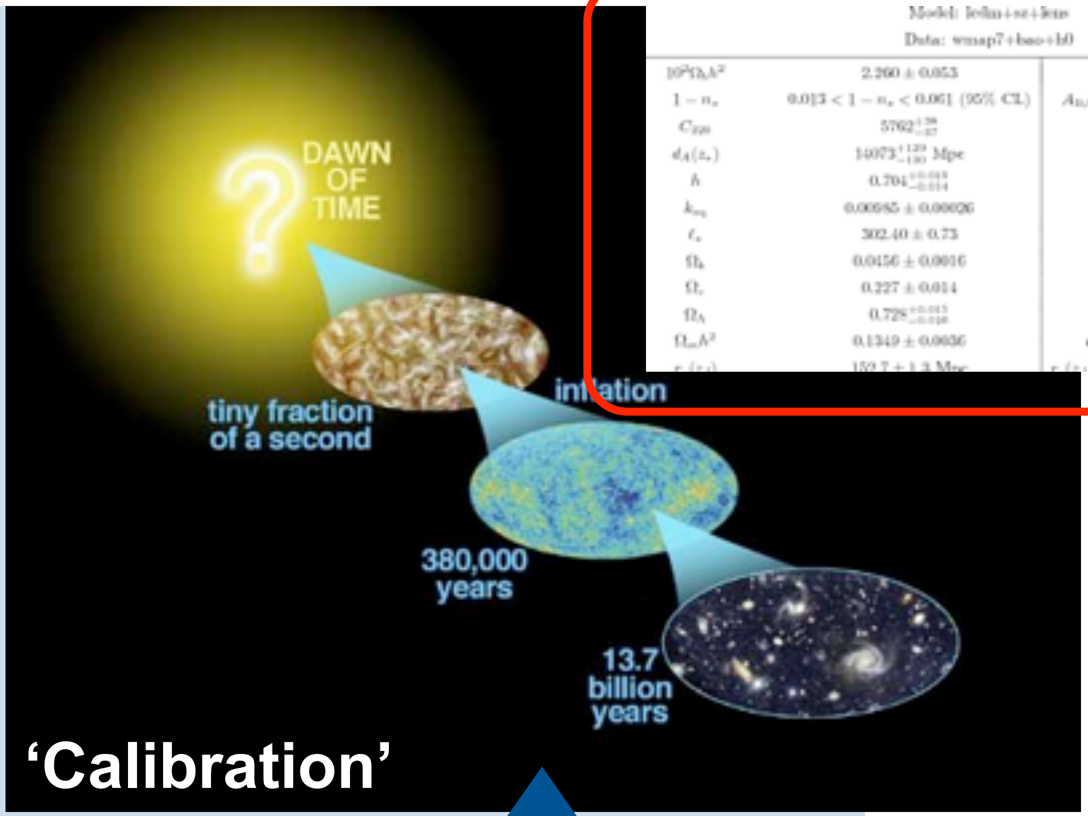
Simulations + CCF



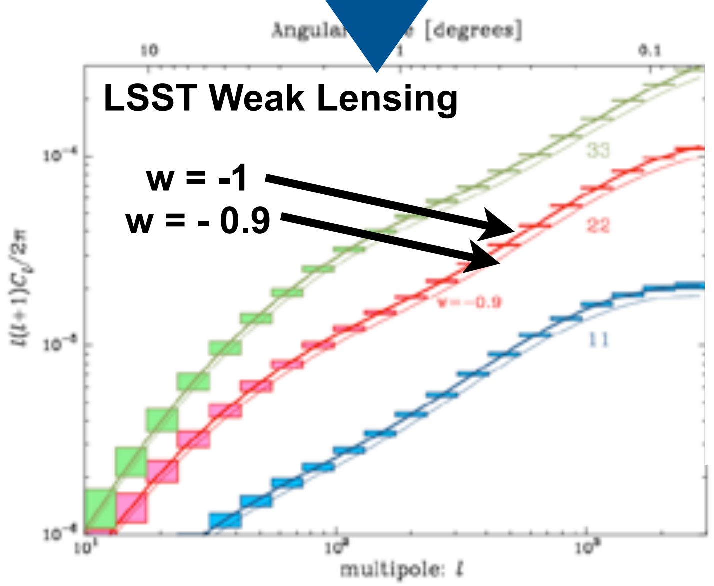
Emulator based on Gaussian Process Interpolation in High-Dimensional Spaces

CCF= Cosmic Calibration Framework

'Precision Oracle'



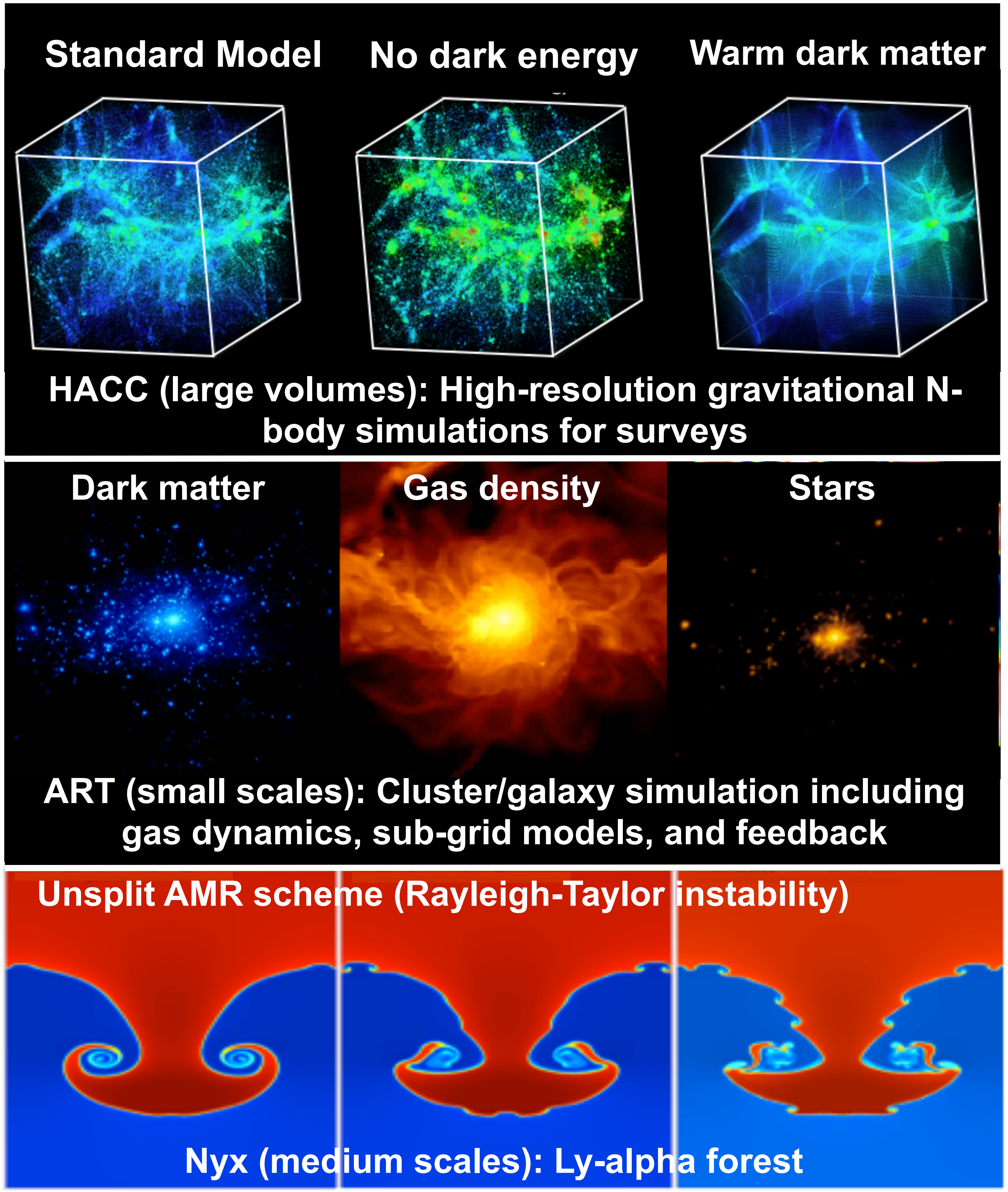
Mapping the Sky with Survey Instruments



Observations: Statistical error bars will 'disappear' soon!

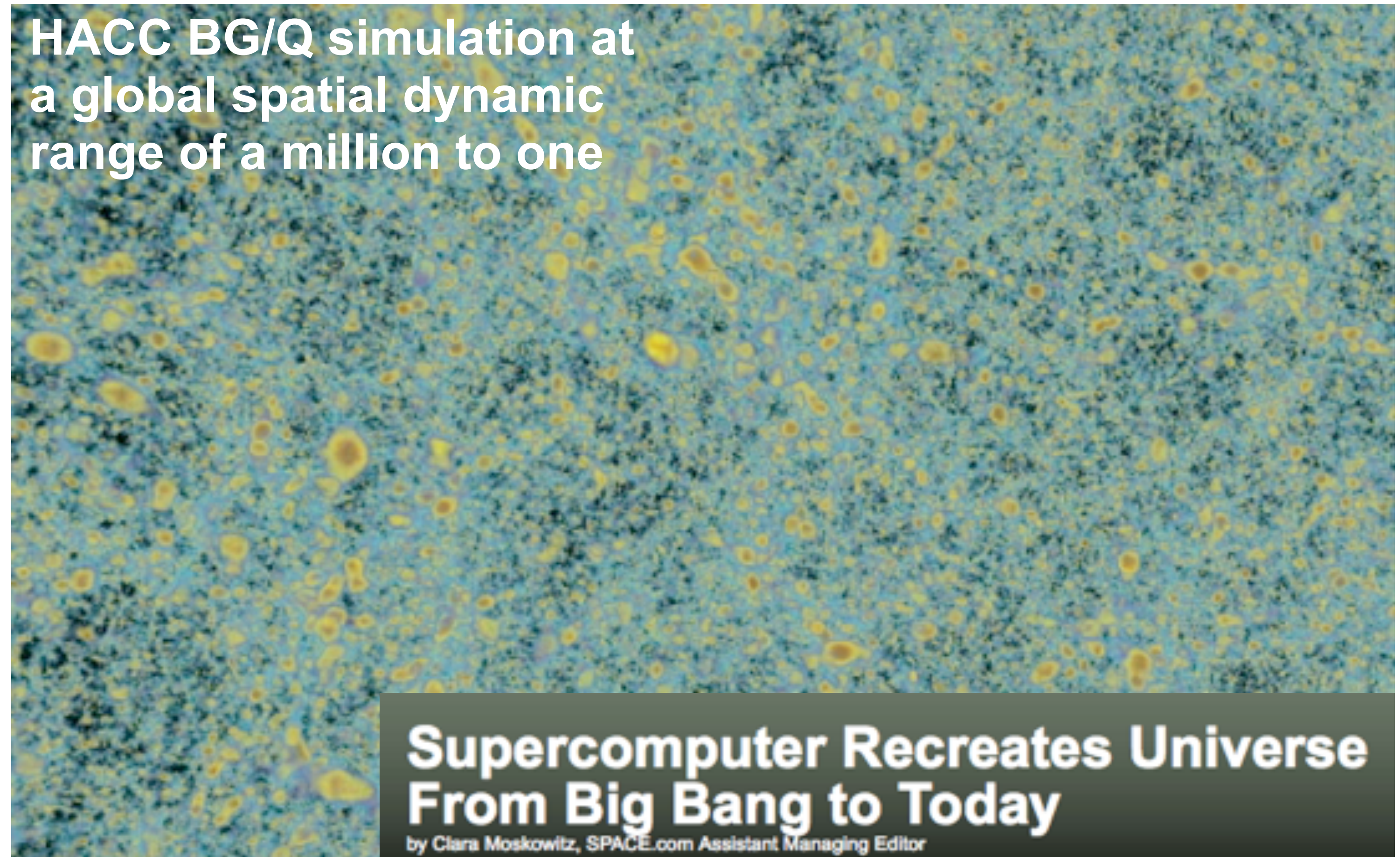
SciDAC-3 Project: Overview

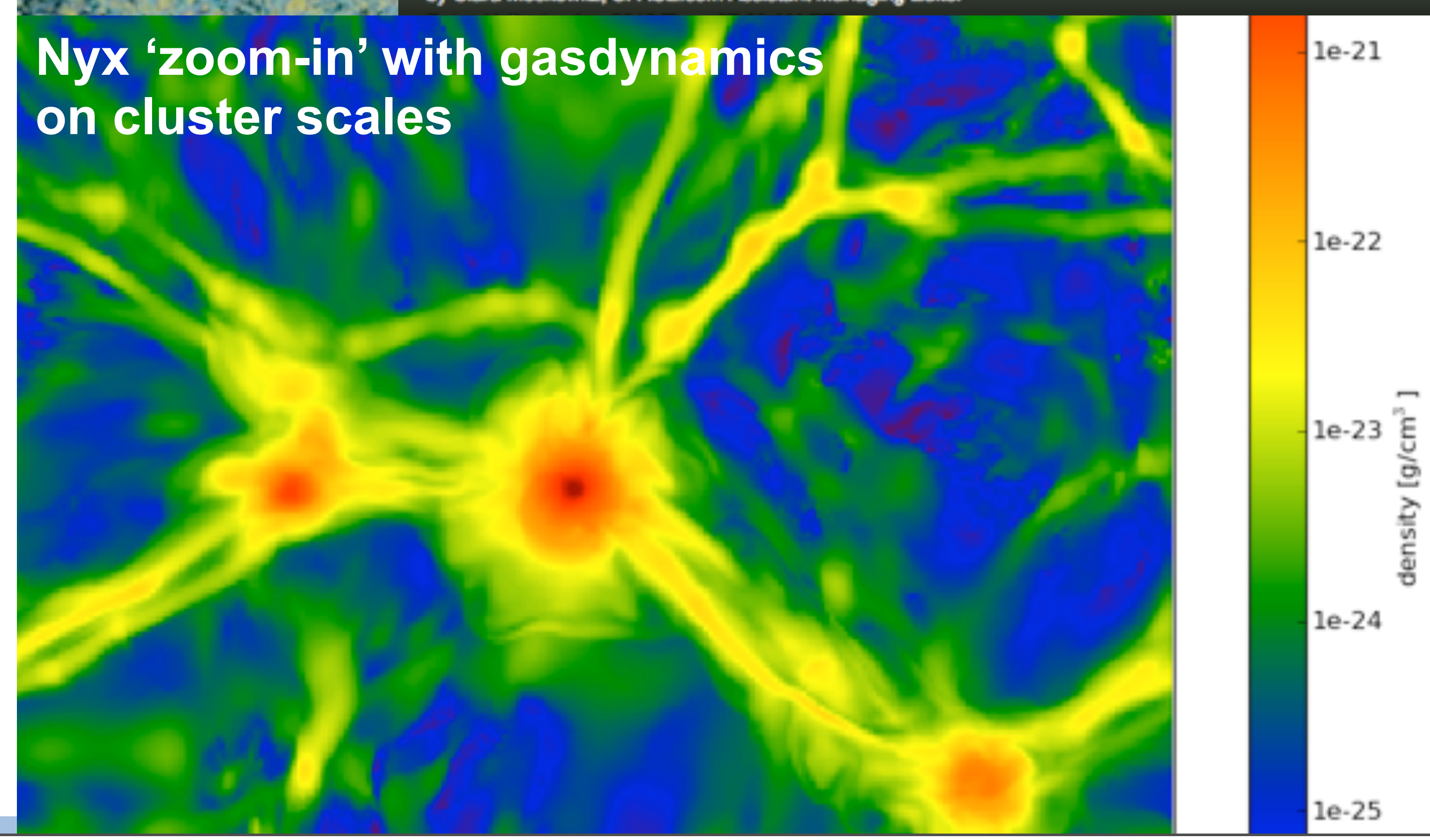
- **SciDAC-3 Project Aims**
  - Build next-generation computational cosmology prediction and analysis frameworks for current and future surveys
  - Explore the physics of dark energy, dark matter, neutrinos, and the early Universe via large-scale structure probes, maintain close relationship with observations
  - 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



Simulation Requirements

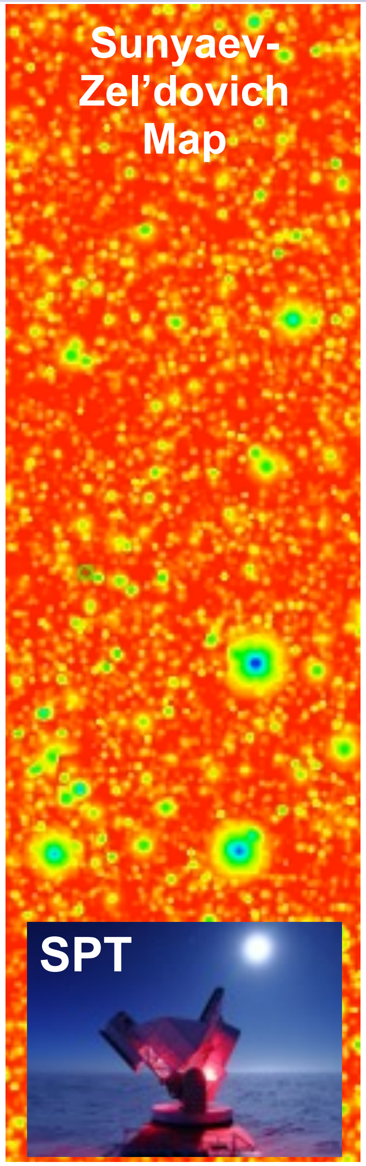
- **Size:**
  - Need to cover volumes in the 100's of Gpc cubed (1 pc=3.26 light-years)
  - To capture individual galaxy mass concentrations over this volume, need trillions of particles
- **Resolution:**
  - Force resolution has to be ~kpc, a global dynamic range of a million to one, also controls time-stepping
  - 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

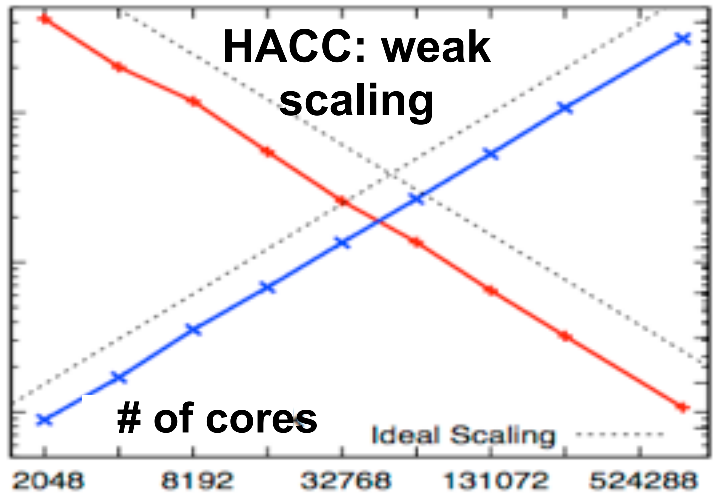


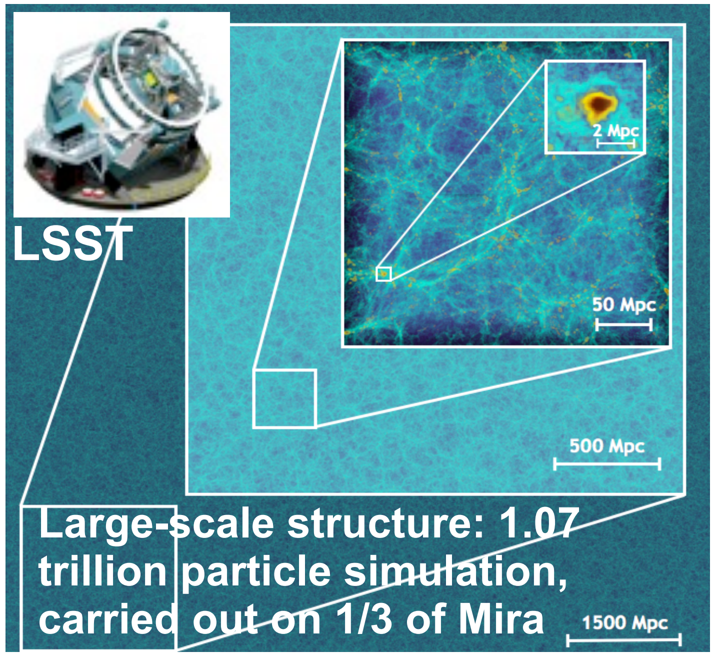


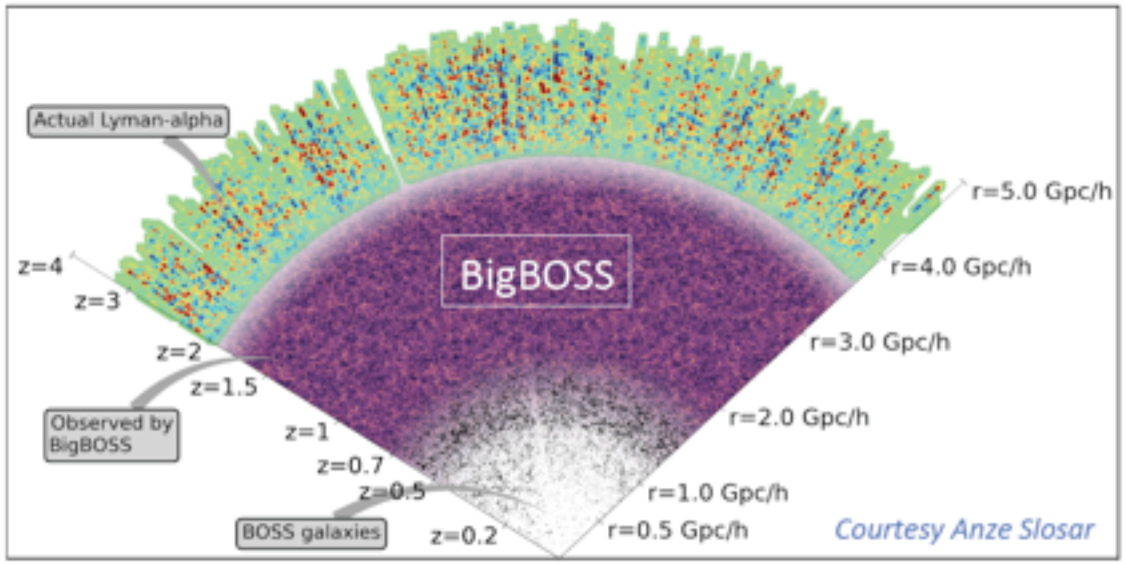
Simulation Status

- **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 (PPTreePM) code framework; runs everywhere: MPPs, Cell/ GPU-accelerated, Blue Gene systems, Intel MIC, --
  - 2012 Gordon Bell finalist, > 50% of peak on Mira/ Sequoia at > 90% parallel efficiency (>1 million cores), ~3.2 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



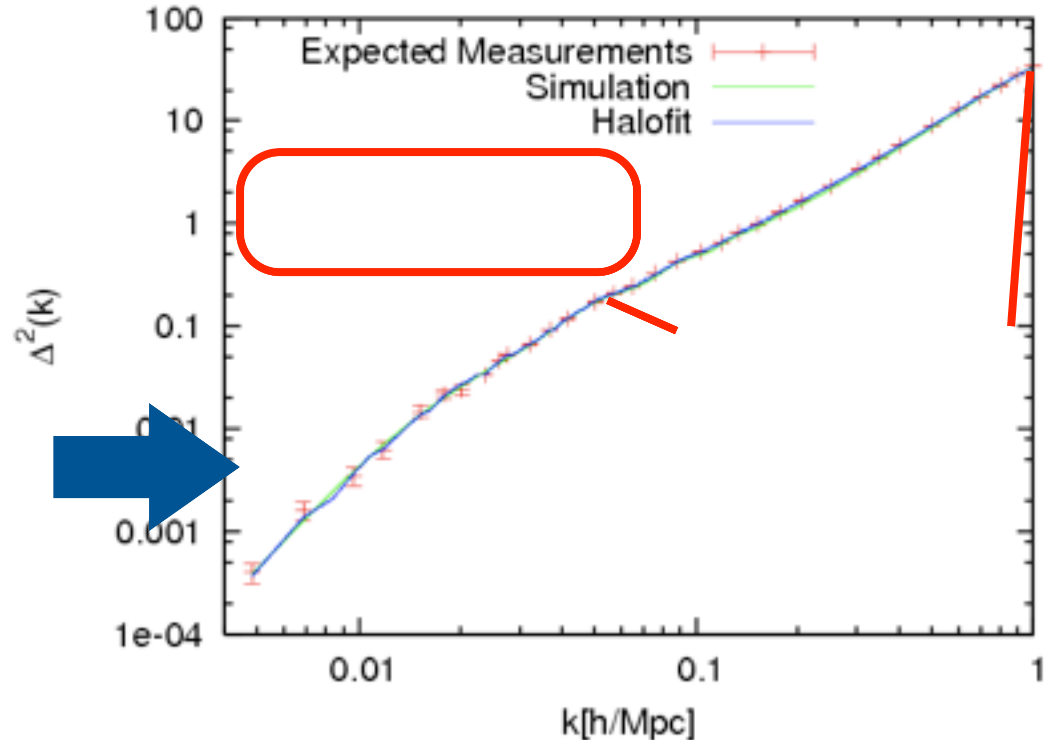


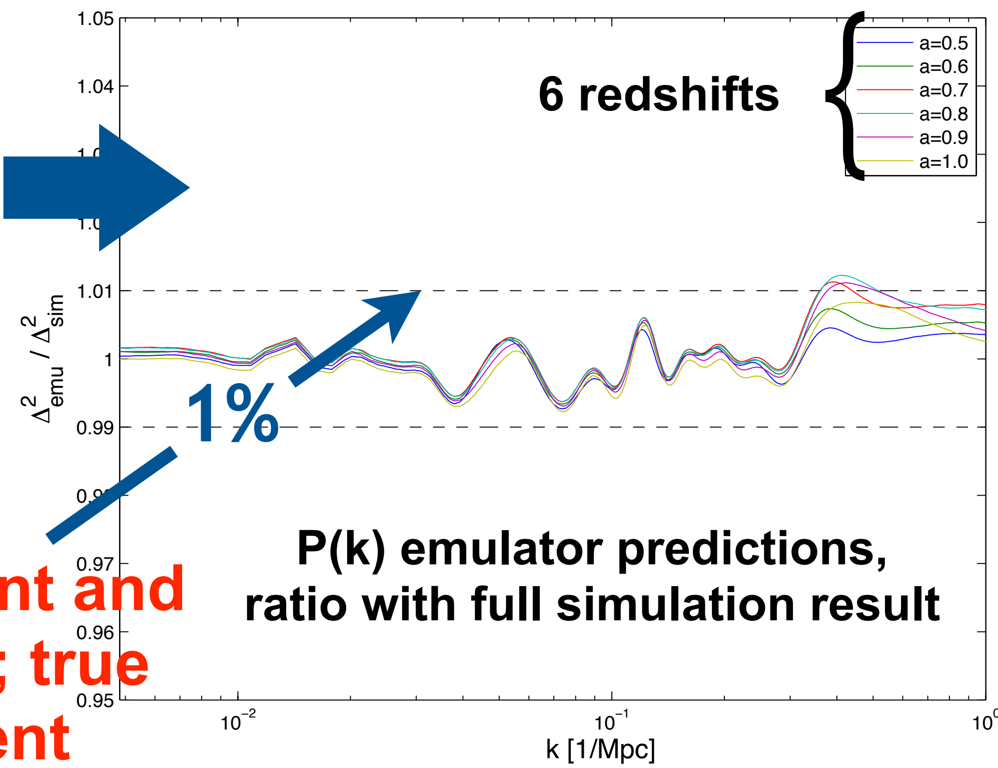




Precision Cosmology: UQ Example

- **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

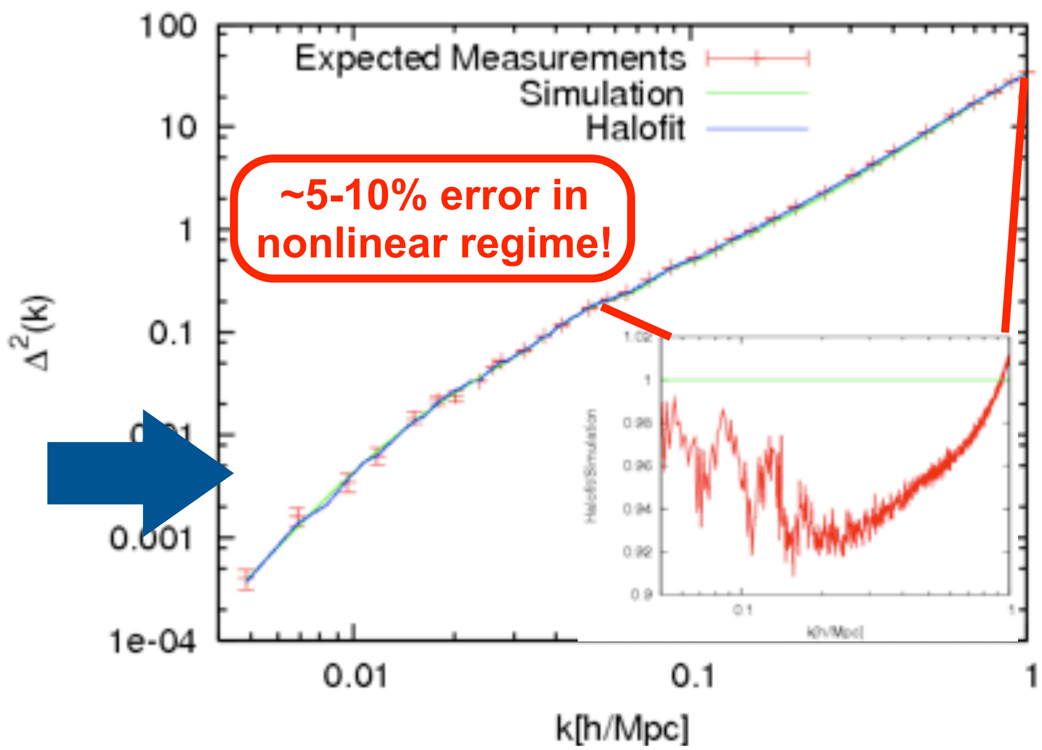


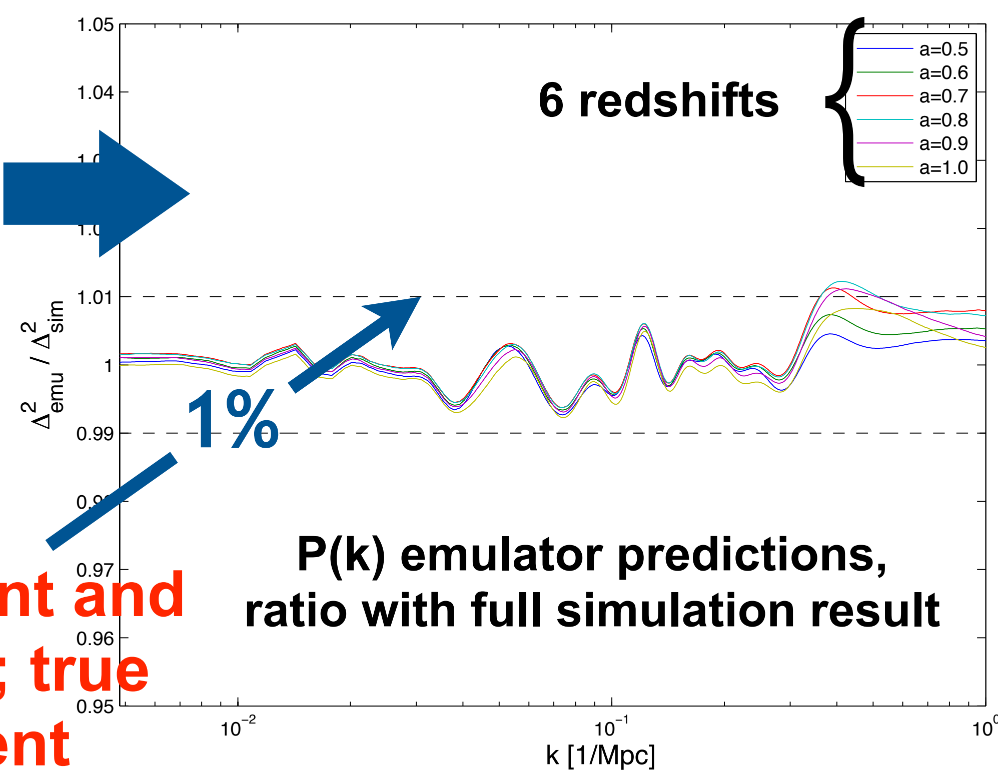


Example of successful development and application of UQ methodologies; true order of magnitude improvement

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

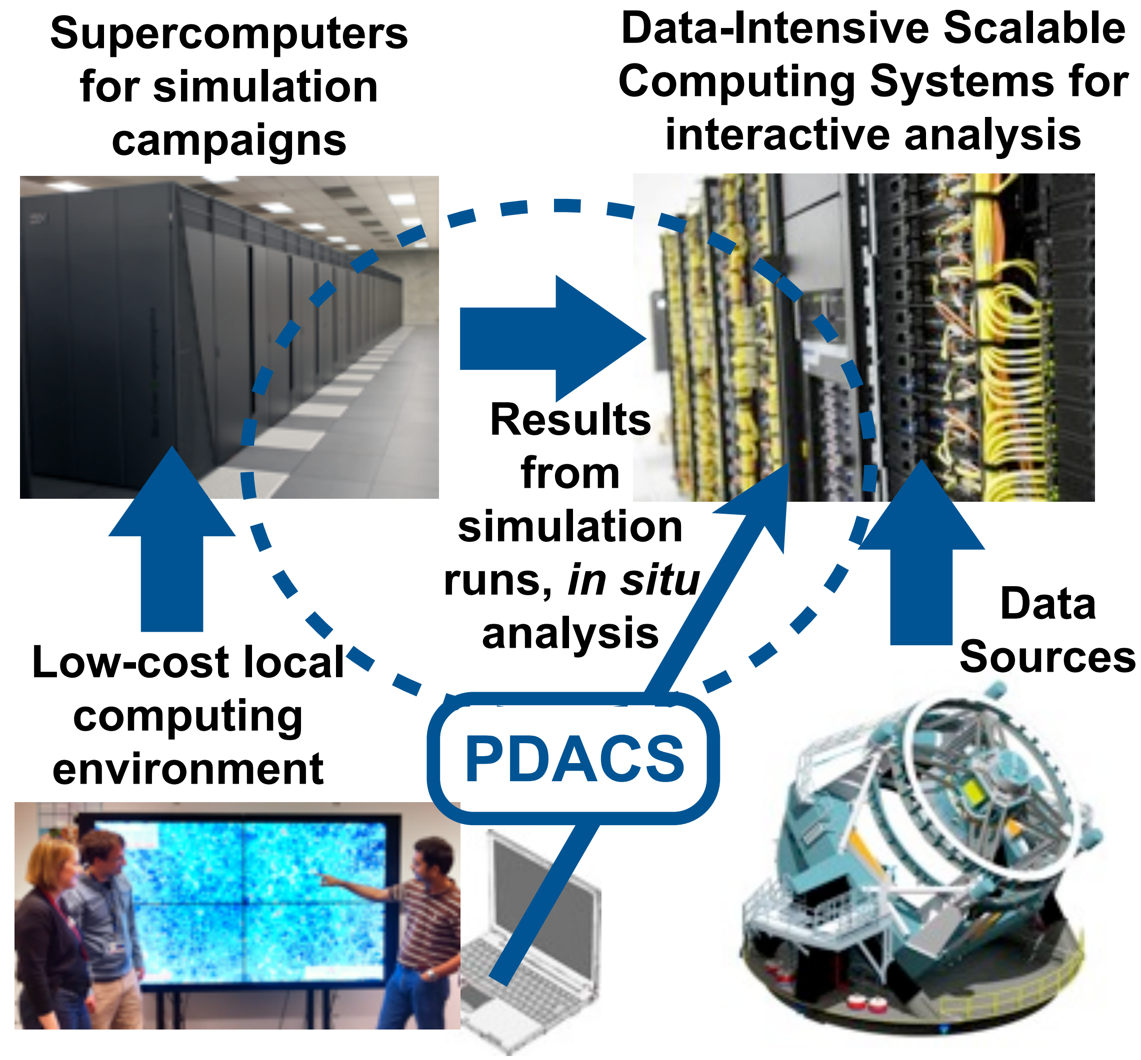


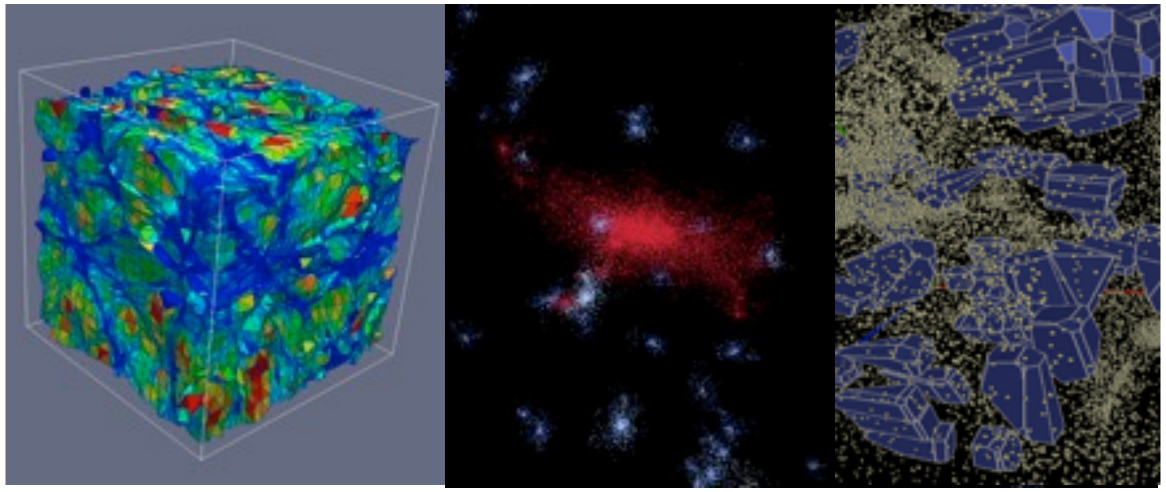


Example of successful development and application of UQ methodologies; true order of magnitude improvement

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)





Institutional Milestones

- **ANL Year 1 (ASCR):**
  - Complete integration of in situ analysis toolkit within HACC framework; establish formats and standards to include multiple tools within an HPC analysis system (w/ all Labs)
  - High-performance HPC I/O tools; profiling, data staging, and performance improvement
  - Begin development of extreme (multi-) resolution visualization framework
- **ANL Year 1 (HEP):**
  - Port HACC framework to next-generation hardware (Intel MIC, next-gen GPUs, ARM?)
  - Emulators for weak lensing, Ly-alpha, parameterized discrepancy functions (w/ BNL, FNAL, LANL, LBNL)
  - Development of error control and UQ for emulation methodology (w/ LANL)
  - Large-volume HACC campaign, integration of Galacticus within HACC analysis system (w/ LBNL)
- **BNL Year 1 (HEP):**
  - Code comparison (Nyx vs. Gadget) and establish standard data structures (w/ LBNL)
  - Baseline grid of Ly-alpha simulations - parameter space and sensitivity for emulator (w/ ANL, LBNL)
- **FNAL Year 1 (HEP):**
  - Investigate viability of ART port to BG/Q, carry out test program (w/ ANL)
  - Development of 4-D domain decomposition algorithms in ART; complete 50% of the final software implementation
  - Halo model framework for estimating baryonic effects; connection to N-body simulations

Institutional Milestones

- **LANL Year 1 (ASCR):**
  - Optimal sampling methods for large dataset analytics and viz/remote viz (w/ ANL)
  - Cosmic emulator framework development
  - Error control and UQ for emulation methodology using experiment design, data compression, and Gaussian processes (w/ ANL)
- **LBNL Year 1 (ASCR):**
  - Investigate grid and particle resolution requirements and validate Nyx for Ly-alpha simulations
  - Code comparison of Nyx and Gadget; assess systematics of grid vs. SPH codes (w/ BNL)
  - Adapt diagnostics for Ly-alpha simulations to Nyx data structures
  - Implement new unsplit PPM w/ reference states to make hydro algorithm more robust for Ly-alpha simulations
- **LBNL Year 1 (HEP):**
  - Ly-alpha systematics study - effects of large-scale power, initial condition prescriptions, on resulting Ly-alpha spectra
  - Compare Ly-alpha diagnostic statistics for Nyx vs. Gadget runs, using matched assumptions (w/ BNL)
  - Run grid of approximate simulations to define the parameter space and parameter sensitivity for simulation grid for Ly-alpha emulator (w/ ANL, BNL)
  - Use a suite of N-body simulations as a skeleton for observational mock catalogs (w/ ANL)

Institutional Milestones

- **SLAC Year 1 (HEP):**
  - Develop public version of multi-scale initial conditions generator (MUSIC)
  - Integrate halo finder/merger tree code into analysis framework
  - Studies for producing simulation-based model galaxy populations using different methods

	N-body					Hydrodynamics				Data/Vis/I/O				UQ				HPC	
	BAO	RSD	WL	CL	FA	BAO	P(k)	WL	AMR	DIY	Vis	I/O	Data	Emu	Error	Discr	Calib	MPI	Performance
T. Abel						●	●	●	●										
J. Ahrens											●		●						
A. Almgren									●										●
J. Cohn				●															
S. Dodelson			●					●											
H. Finkel		●			●													●	●
N. Gnedin								●	●										●
S. Habib	●		●	●	●									●	●	●			●
K. Heitmann	●	●	●								●			●	●				
M. Hereld											●	●	●						
D. Higdon			●					●						●	●	●	●		
E. Kovacs										●			●						
~~H. Krishnan~~											●								
J. Kwan		●																	
R. Latham												●							
E. Lawrence			●											●	●	●	●		
Z. Lukic						●	●		●										
~~B. Norris~~																			
P. Nugent									●		●		●						
T. Peterka										●		●							●
A. Pope	●	●			●							●							
A. Slozar						●	●												●
~~R. Thakur~~																		●	
R. Wechsler				●						●			●						
M. White	●	●				●	●												
J. Woodring											●		●						

Figure 14: *Responsibility Assignment Matrix.*

Charge for this Meeting

- **SciDAC Project Organization Logistics:**
  - Telecons, quarterly reports to Lali Chatterjee and Randall Laviolette (responsible DOE HEP and ASCR program managers), face to face meetings, 6-month report to Oversight Committee (Roger Blandford, David Brown, Bob Cahn, Craig Hogan, Morgan May, Marc Snir, Harry Weerts, Andy White)
  - Website, wiki, repos, etc.
  - Computing allocations, data storage and transfer, common software
  - Collaboration issues, mechanisms for connecting to SciDAC Institutes
- **Milestones:**
  - Convert milestones into a reasonable annual activity plan to establish whether progress is being made or not; guard against project creep and change of direction
- **Starting Issues:**
  - Manpower holes, anything holding back local efforts
- **Wider Issues:**
  - Connection to HEP future planning (Snowmass process, projects, --)
  - CFCC connection to projects (traction w/ BOSS, LSST, LSST-DESC)