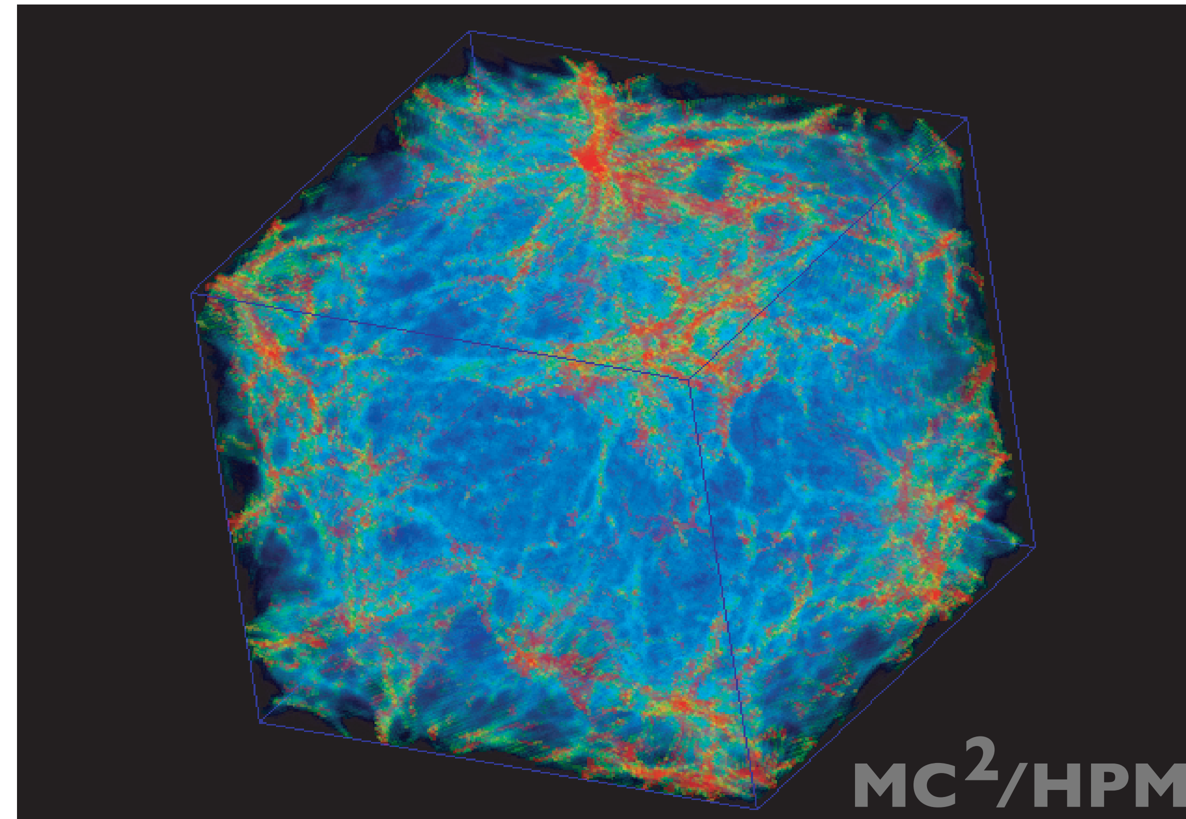
Computing the Universe

Adrian Pope
High Energy Physics Division
Argonne National Laboratory

Collaborators:
Jim Ahrens, Suman Bhattacharya, Jordan Carlson, David Daniel, Patricia Fasel, Nick Frontiere, Salman Habib, Katrin Heitmann, Zarija Lukic, Pat McCormick, Paul Sathre, Martin White, Jon Woodring

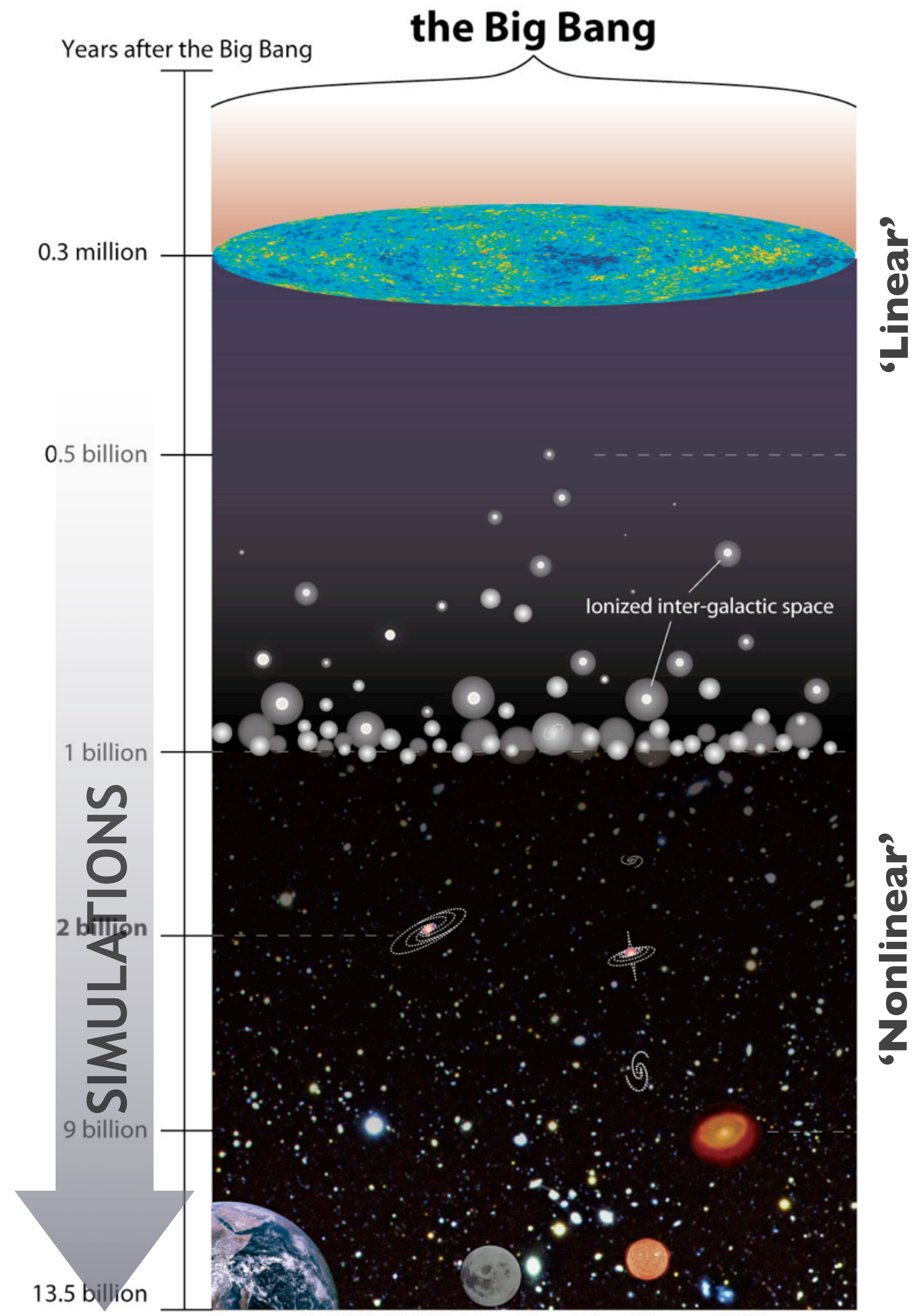
Computational Cosmology: A ‘Particle Physics’ Perspective

- **Primary Research Target:** Cosmological signatures of physics beyond the Standard Model
- **Structure Formation Probes:** Exploit nonlinear regime of structure formation
  - **Discovery Science:** Derive signatures of new physics, search for new cosmological probes
  - **Precision Predictions:** Aim to produce the best predictions and error estimates/distributions for structure formation probes
  - **Design and Analysis:** Advance ‘Science of Surveys’; contribute to major ‘Dark Universe’ missions: BOSS, DES, LSST, BigBOSS, DESpec --



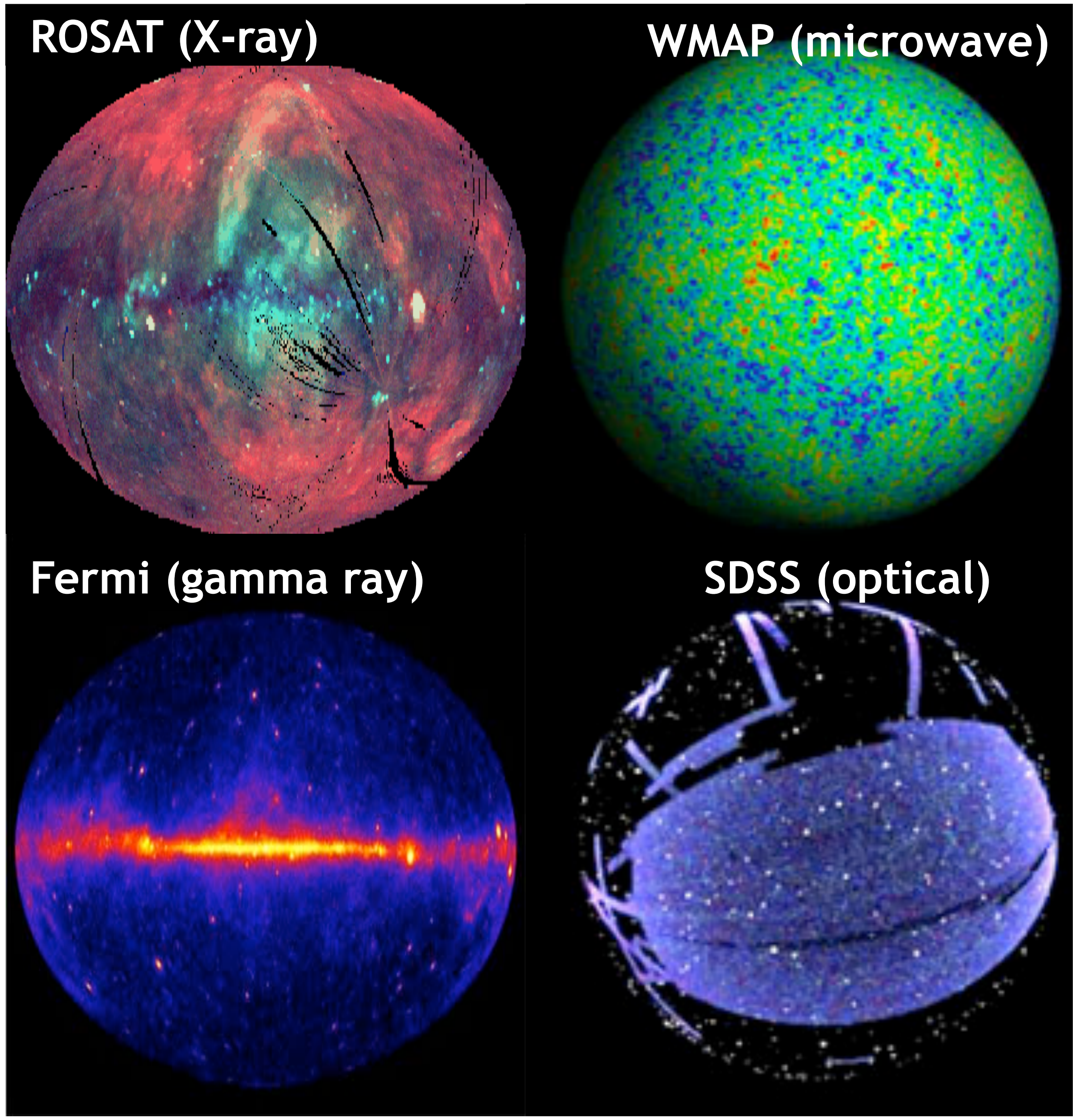
Structure Formation: The Basic Paradigm

- **Solid understanding of structure formation; success underpins most cosmic discovery**
  - Initial conditions laid down by inflation
  - Initial perturbations amplified by gravitational instability in a dark matter-dominated Universe
  - Relevant theory is gravity, field theory, and atomic physics ('first principles')
- **Early Universe:** Linear perturbation theory very successful (CMB)
- **Latter half of the history of the Universe:** Nonlinear domain of structure formation, impossible to treat without large-scale computing



Cosmological Probes of Physics Beyond the Standard Model

- **Dark Energy:** Properties of DE equation of state, modifications of GR, other models?
  Sky surveys, terrestrial experiments
- **Dark Matter:** Direct/Indirect searches, clustering properties, constraints on model parameters
  Sky surveys, targeted observations, terrestrial experiments
- **Inflation:** Probing primordial fluctuations, CMB polarization, non-Gaussianity
  Sky surveys
- **Neutrino Sector:** CMB, linear and nonlinear matter clustering
  Sky surveys, terrestrial experiments

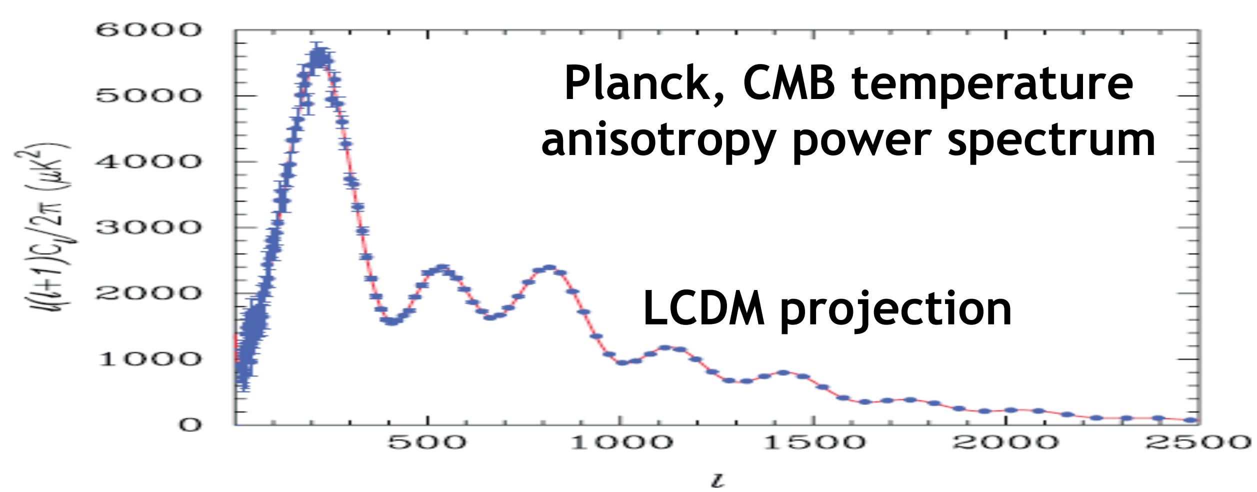


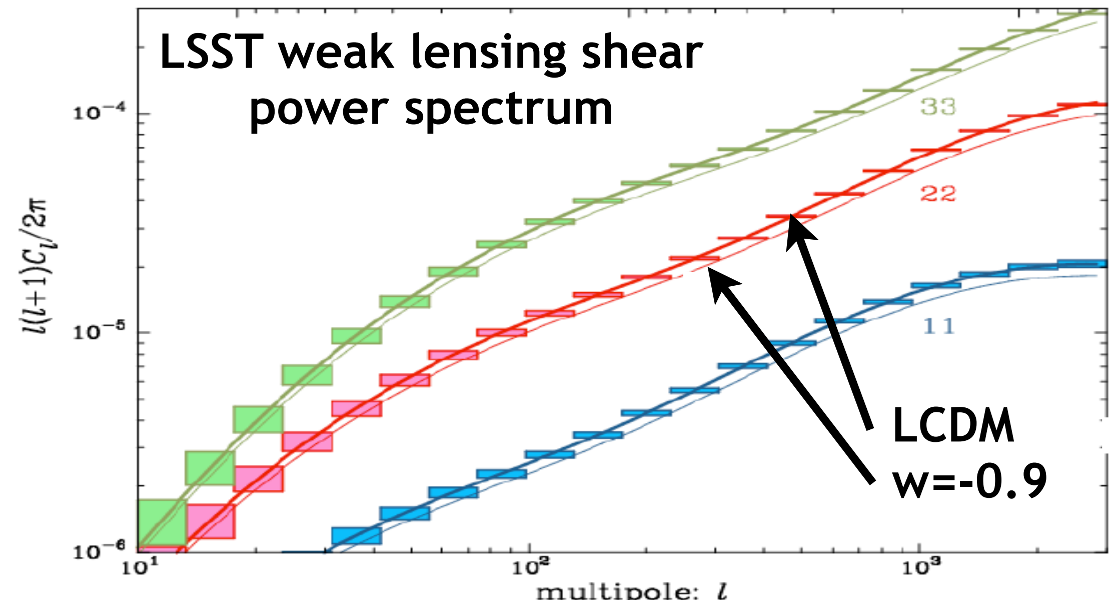
Explosion of information from sky maps: Precision Cosmology

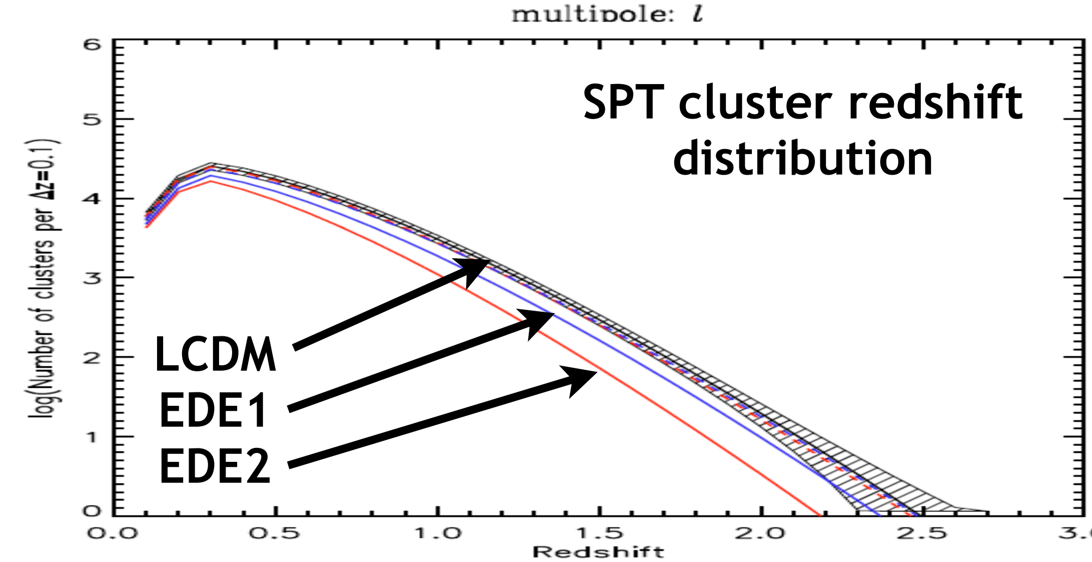
Precision Cosmology: "Inverting" the 3-D Sky

- **Cosmic Inverse Problem:** From sky maps to scientific inference
- **Cosmological Probes:** Measure geometry and presence/growth of structure (linear and nonlinear)
- **Examples:** Baryon acoustic oscillations (BAO), cluster counts, CMB, weak lensing, galaxy clustering, --
- **Cosmological Standard Model:** Verified at 5-10% with multiple observations

- **Future Targets:** Aim to control survey measurements to the ~1% level
- **The Challenge:** Theory and simulation must satisfy stringent criteria for inverse problems and precision cosmology not to be theory-limited!





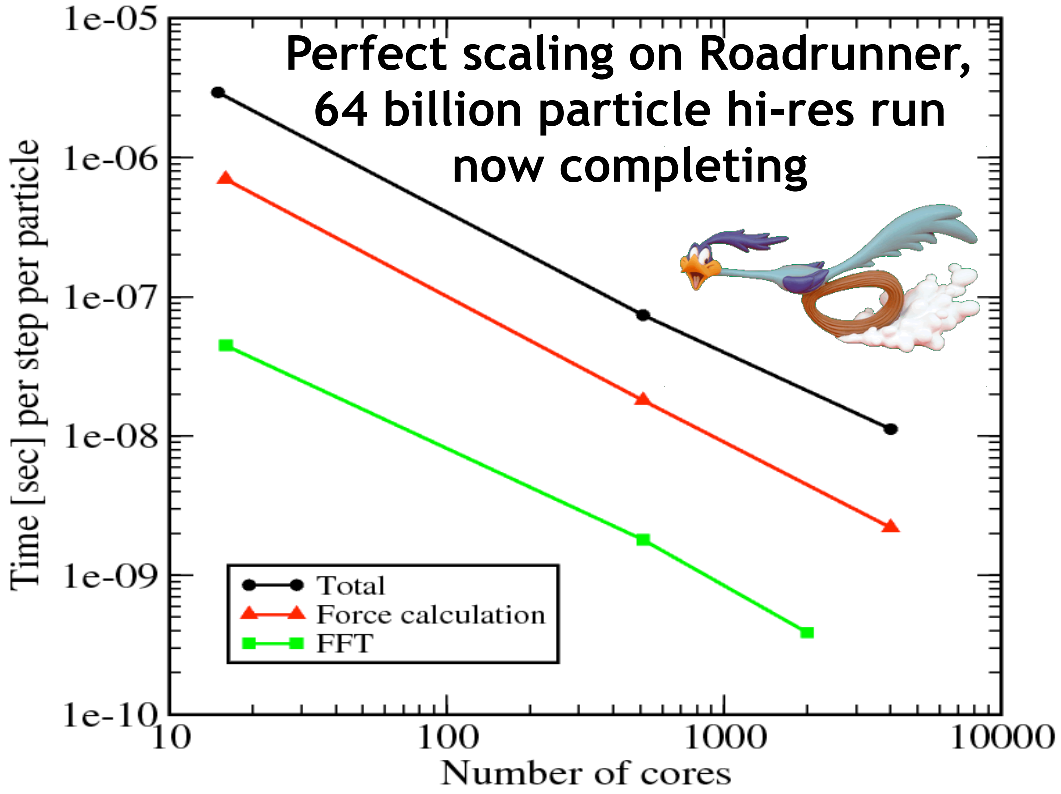


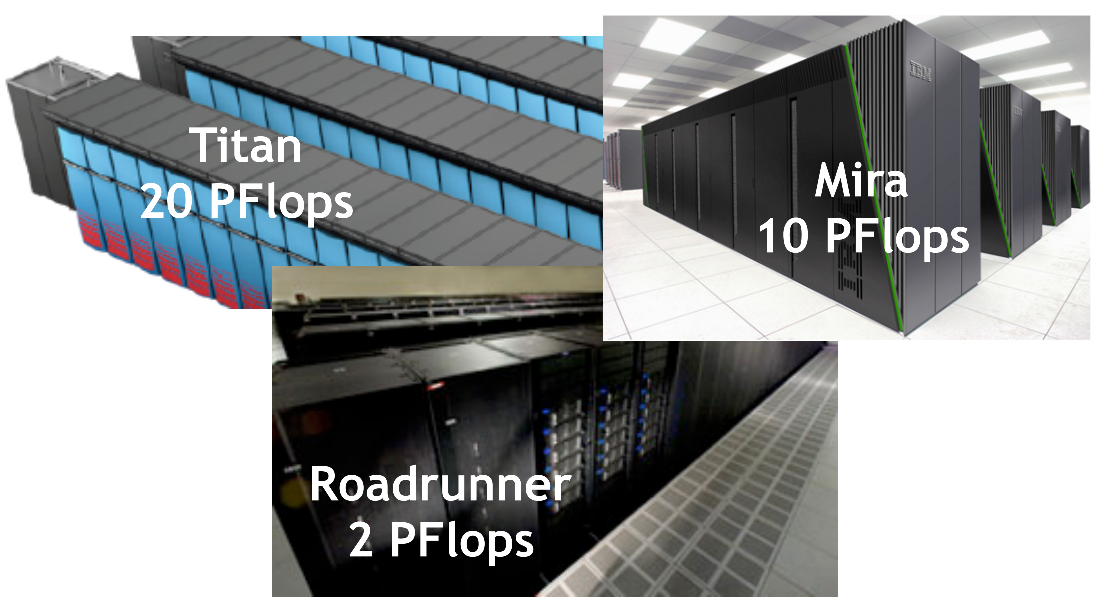
Computing the Universe: Simulating Surveys

- **Simulation Volume:** Large survey sizes impose simulation volumes ~ (3 Gpc)^3, with memory requirements ~**100** TB
- **Number of Particles:** Mass resolutions depend on ultimate object to be resolved, ~**10^8 --10^{10}** solar masses, N~**10^{11}--10^{12}**
- **Force Resolution:** ~kpc, yields a (global) spatial dynamic range of **10^6**
- **Hydrodynamics/Sub-Grid Models:** Phenomenological treatment of gas physics and feedback greatly adds to computational cost
- **Throughput:** Large numbers of simulations required (**100's --1000's**), development of analysis suites, and emulators; peta-to-exascale computing exploits
- **Data-Intensive-SuperComputing:** End-to-End simulations and observations must be brought together in a DISC environment (theory-observation feedback)

Hardware-Accelerated Cosmology Code (HACC) Framework

- **Architecture Challenge:** HPC is rapidly evolving (clusters/BG/CPU+GPU/MIC --)
- **Code for the Future:** Melds optimized performance, low memory footprint, embedded analysis, and cross-platform scalability
- **Implementation:** Long/short-range force matching with spectral force-shaping (long-range=PM, short-range=PP, tree)
- **Key Features:** Hybrid particle/grid design, particle overloading, high-order spectral operators, ~50% of peak Flops
- **Embedded Analysis:** High performance with low I/O and storage requirement
- **Early Science Project on Mira:** 150M CPU-hrs on ANL BG/Q (summer 2012)

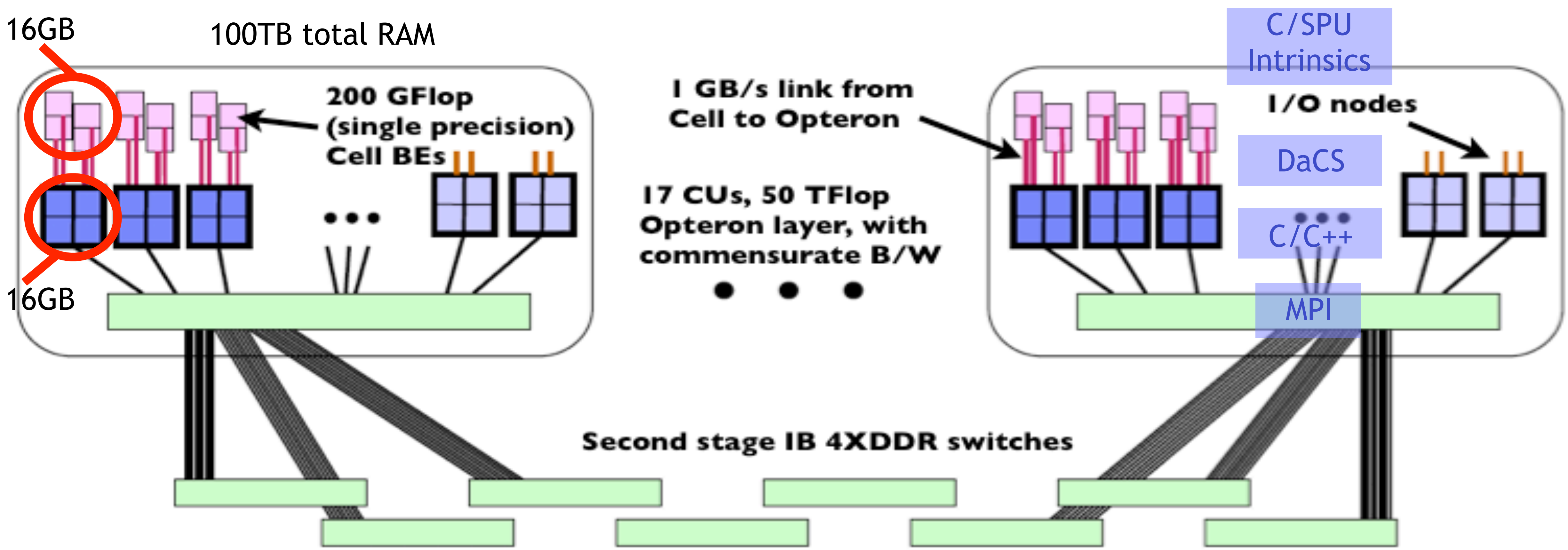




Habib et al. 2009, Pope et al. 2010

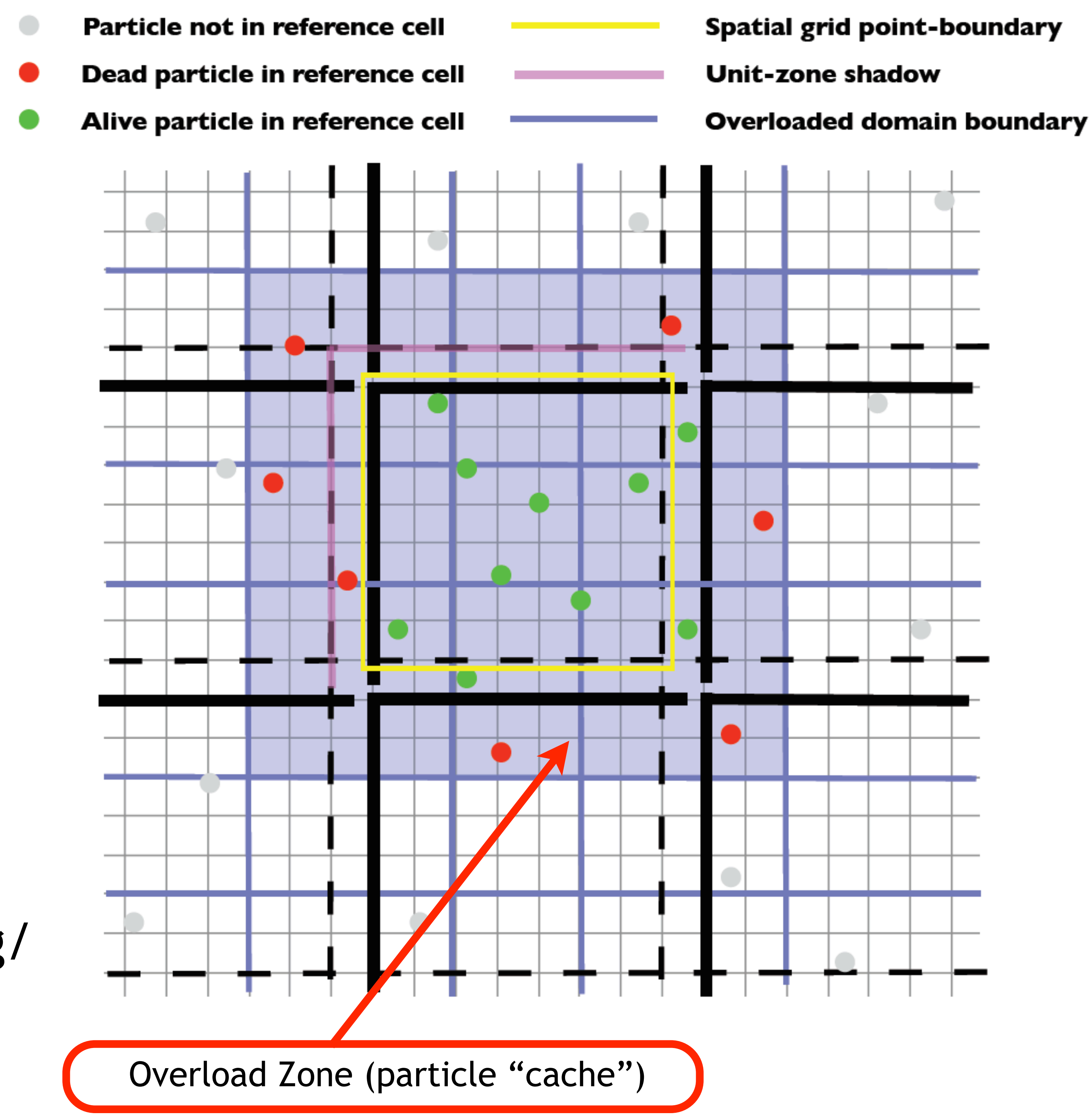
HACC Example I: Roadrunner (CPU+Cell)

- Hybrid machine architecture, out of balance communication (50-100) and performance (20)
- Multi-level programming paradigm
- 'On the fly' analysis to reduce I/O
- Prototype for exascale code design problems
- Scalable approach extensible to all next-generation architectures (BG/Q, CPU/GPU, --)



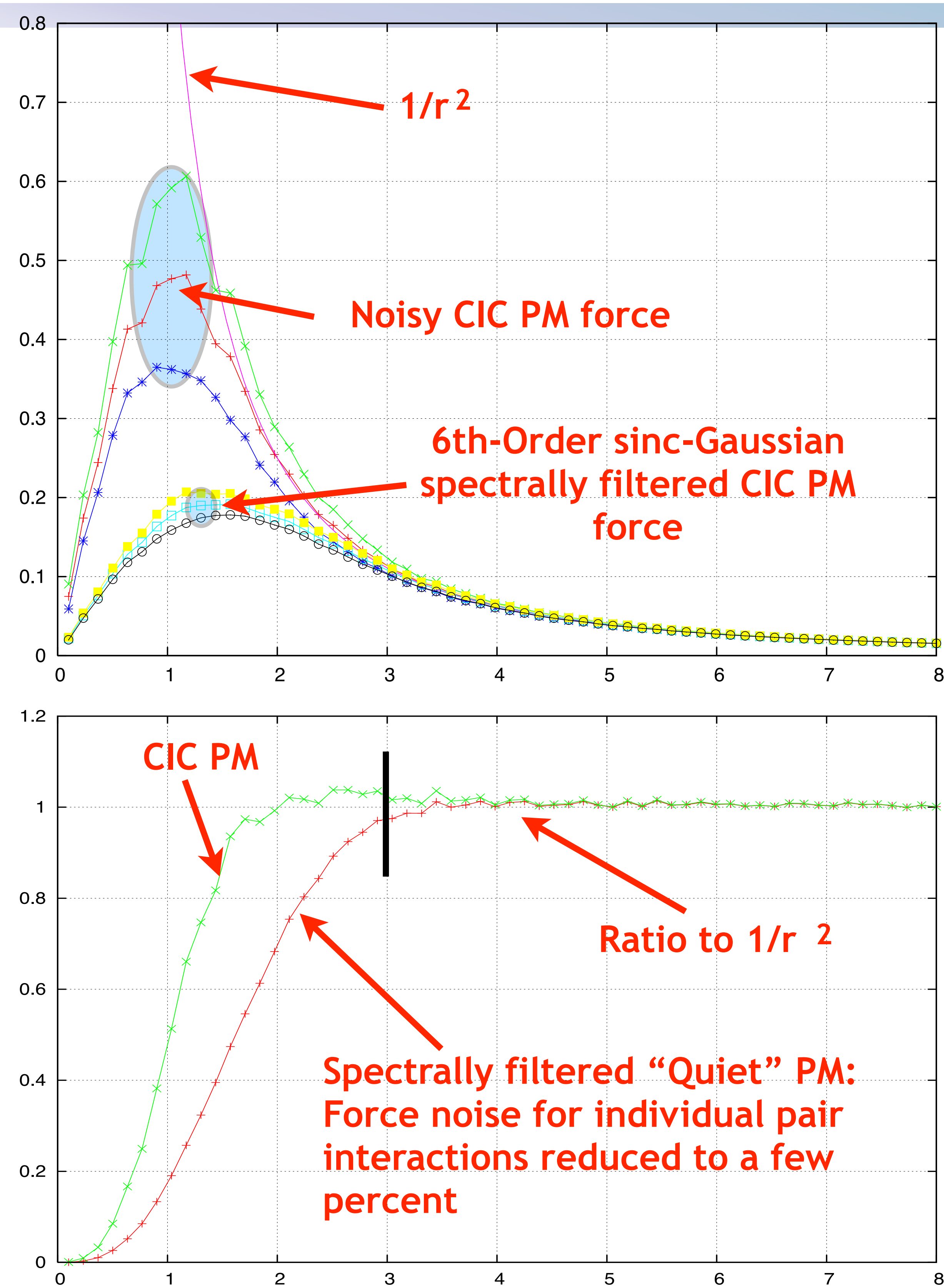
HACC Algorithmic Details 1

- Solve compute imbalance: Split problem into long-range and short-range force updates
- Long-range handled by a grid-based Poisson solver
- Direct particle-particle short-range interactions
- Simplify and speed-up Cell computational tasks
- Reduce CPU/Cell traffic to avoid PCIE bottleneck: use simple CIC to couple particles to the grid, followed by spectral filtering on the grid
- Reduce inter-node particle communication: particle caching/ replication (ghost zone analog)
- ‘On the fly’ analysis and visualization to reduce I/O



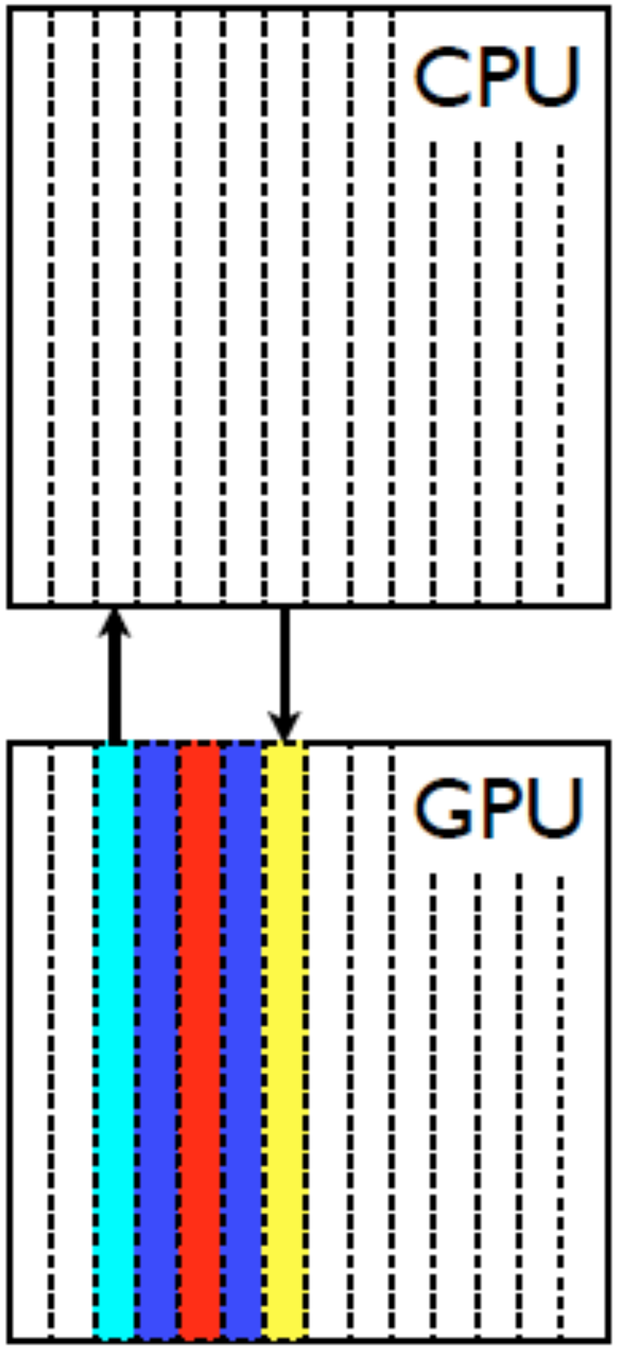
HACC Algorithmic Details 2

- Spectral smoothing of the CIC density field allows 6-th order Green function and 4th order super-Lanczos gradients for high-accuracy Poisson-solves
- Short-range force is fit to the numerical difference between Newtonian and long-range force (not conventional P^3M)
- Short-range force time-steps are sub-cycled within long-range force kicks via symplectic algorithm
- Short-range computations isolated as essentially 'on-node', replace or re-design for different architectures (e.g., BG/Q or GPU)



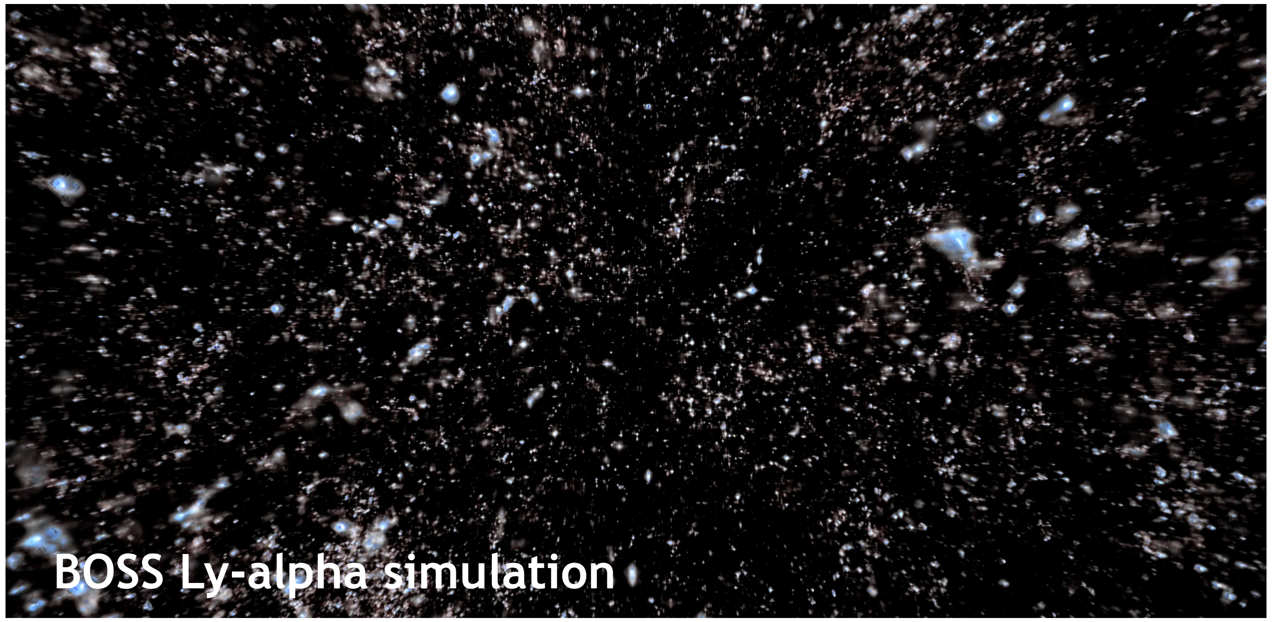
HACC Example 2: CPU+GPU

- CPU/GPU performance and communication out of balance, unbalanced memory (CPU/main memory dominates)
- Multi-level programming (mitigate with OpenCL)
- Particles in CPU main memory, CPU does low flop/byte operations
- Stream slabs through GPU memory (pre-fetches, asynchronous result updates)
- Data-parallel kernel execution
- Many independent work units per slab -- many threads, efficient scheduling, good performance achieved (improves on Cell)
- Scalability of HACC is the same across all 'nodal' variants

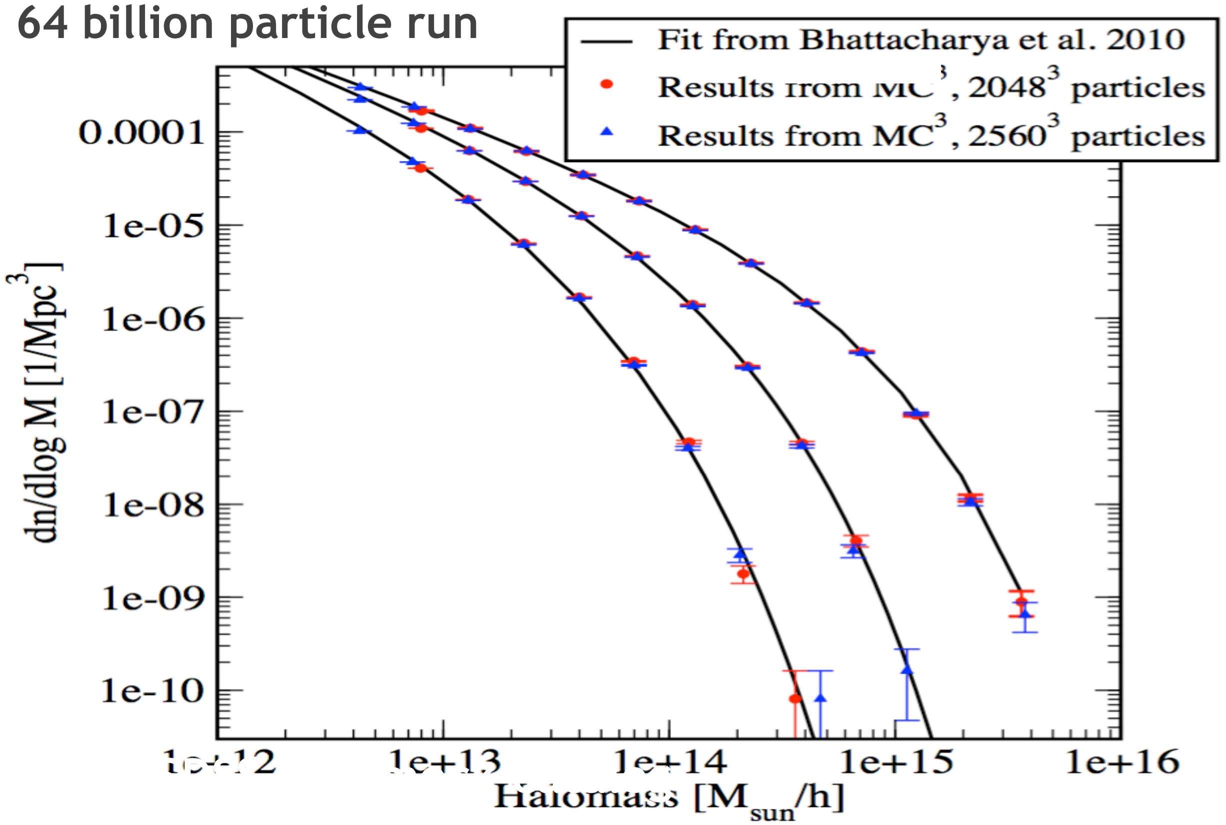


HACC Science Results, Present and Future

- Part of 'Open Science' program on LANL's Roadrunner
- Nine medium-resolution 64 billion particle simulations over one weekend
- Theoretical predictions for BOSS (Baryon Oscillations Spectroscopic Survey) Ly-alpha analysis
- Large volume simulations for halo statistics
- Currently porting to ANL's BG/Q 'Mira' for 'Dark Universe' project encompassing several cosmological probes (lensing, galaxy clustering, clusters, --)



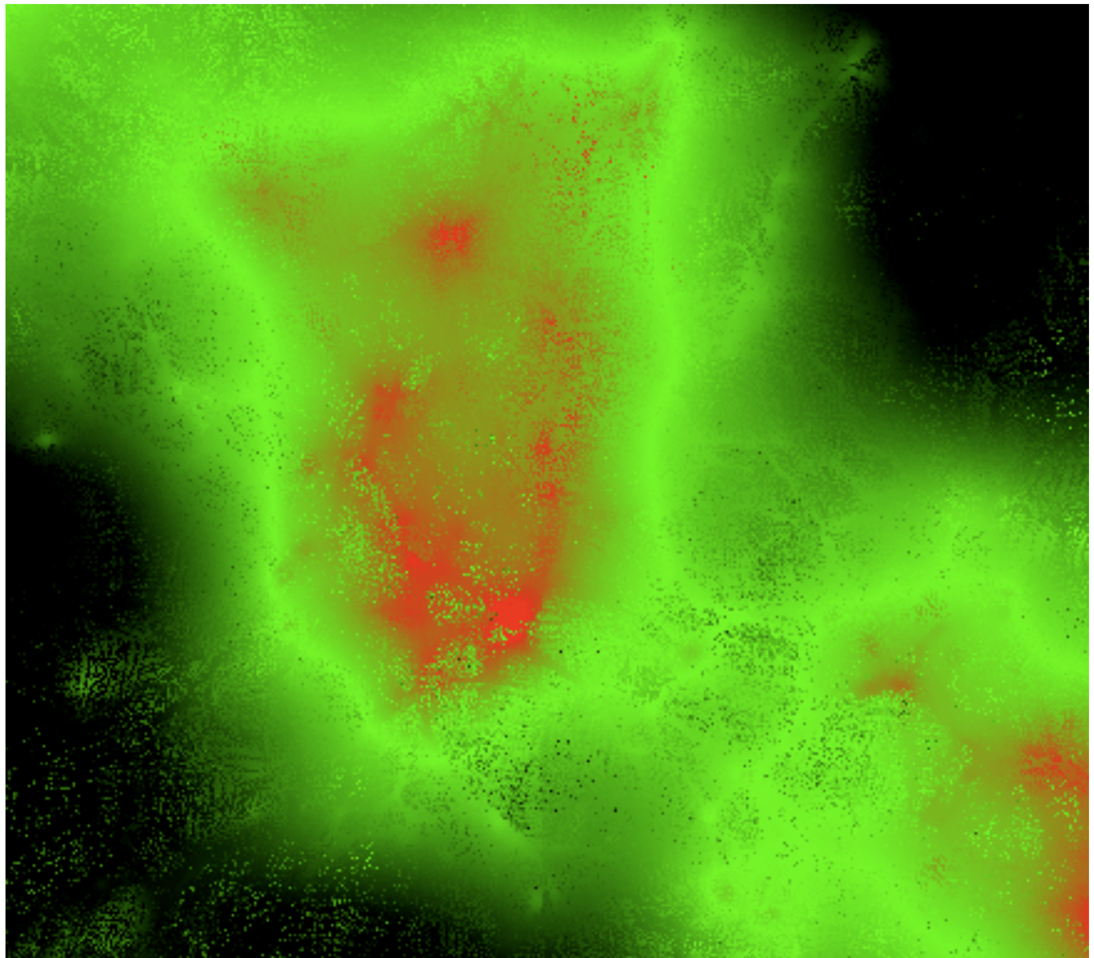
Roadrunner view (halos) of the Universe at z=2 from a 64 billion particle run



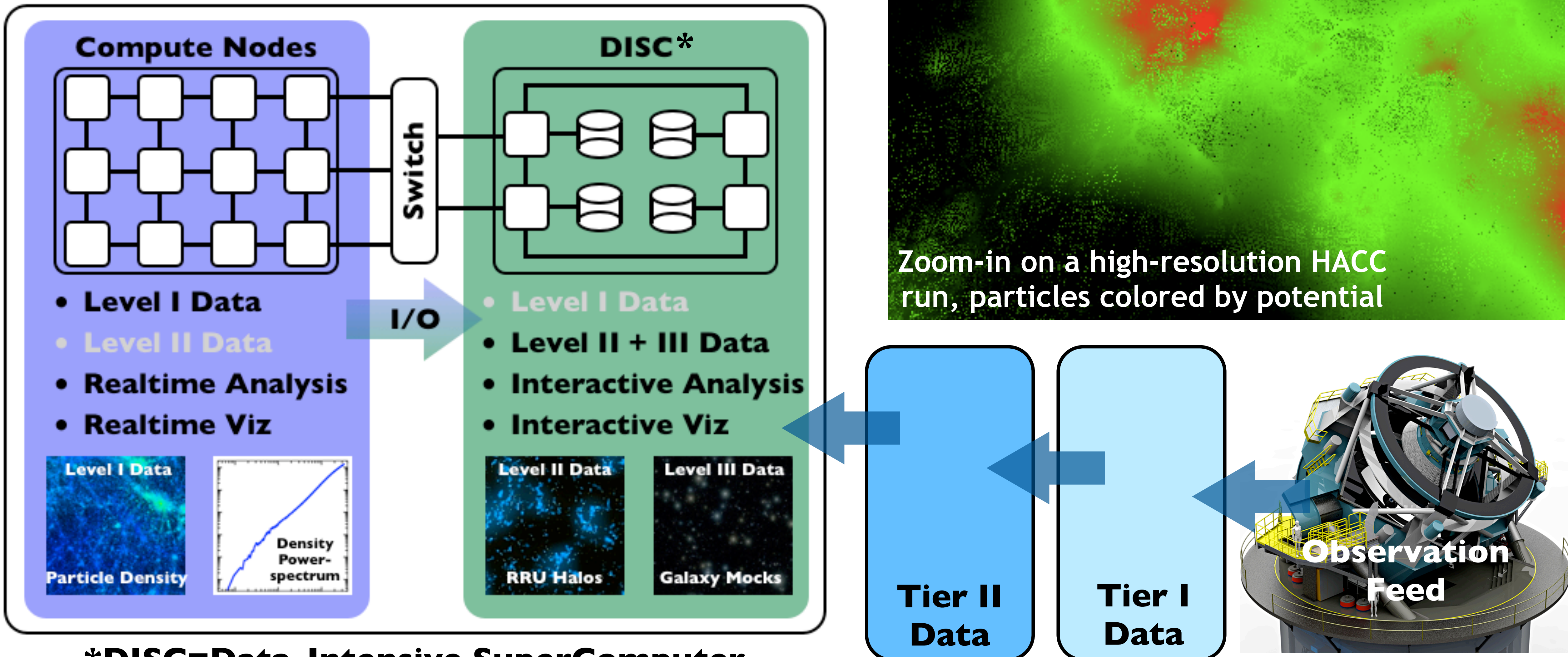
Halo mass function at z=0,1,2 and the corresponding phenomenological fit

HACC in the HPC/DISC Future

- **HACC as Exascale Co-Design Driver:** Most codes cannot meet future science requirements and HPC constraints, HACC capabilities already demonstrated on Cell and GPU-accelerated systems



Zoom-in on a high-resolution HACC run, particles colored by potential



*DISC=Data-Intensive SuperComputer