Simulating (and moving) the Dark Universe

Katrin Heitmann
High Energy Physics Division
Mathematics & Computer Science Division
Argonne National Laboratory

Collaborators:
S. Bhattacharya, D. Daniel, S. Deb, P. Fasel, H. Finkel,
N. Frontiere, S. Habib, D. Higdon, R. Hudson, K. Kumaran,
J. Kwan, E. Lawrence, Z. Lukic, V. Morozov, C. Nakhleh,
T. Peterka, A. Pope, A. Upadhye, C. Wagner, M. White,
B. Williams, and the ANL visualization team

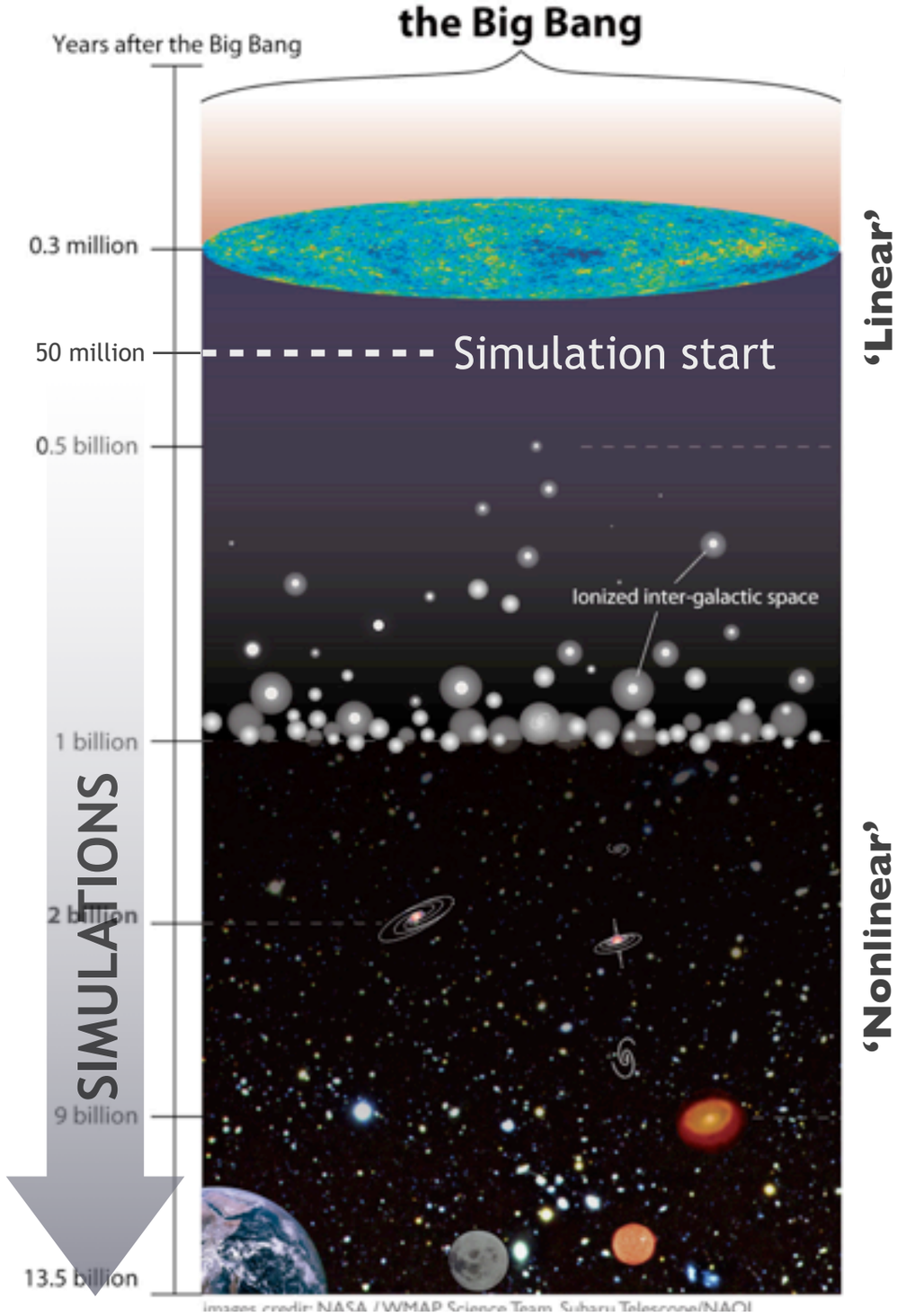
The Make-up of the Dark Universe

- **Dark Energy:** Multiple observations show that the expansion of the Universe is accelerating (first in 1998, Nobel prize 2011)
- Imagine you throw a ball in the air and instead of coming down it flies upwards faster and faster!
- Questions: What is it? Why is it important now? Being totally ignorant, currently our main task is to characterize it better and exclude some of the possible explanations
- **Dark Matter:** Observations show that ~27% of the matter in the Universe is "dark", i.e. does not emit or absorb light
- So far: indirect detection, aims: characterize nature of dark matter and detect the actual dark matter particle

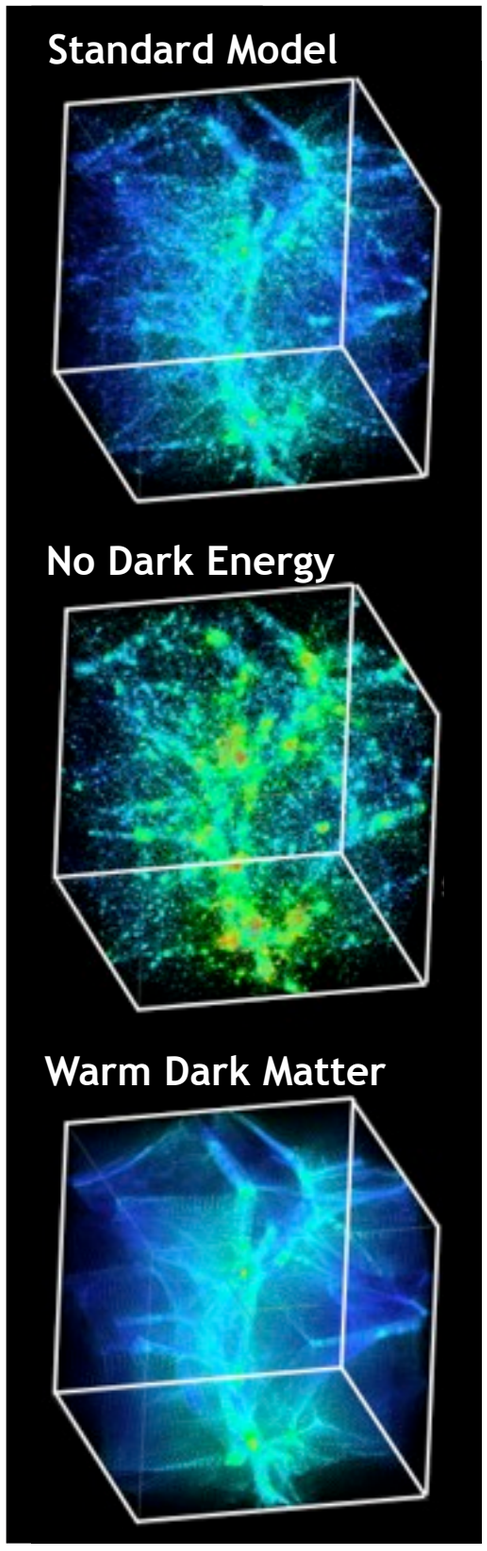
~95% of the Universe is "dark" -- we do not understand the nature and origin of dark energy and dark matter.

Structure Formation: The Basic Paradigm

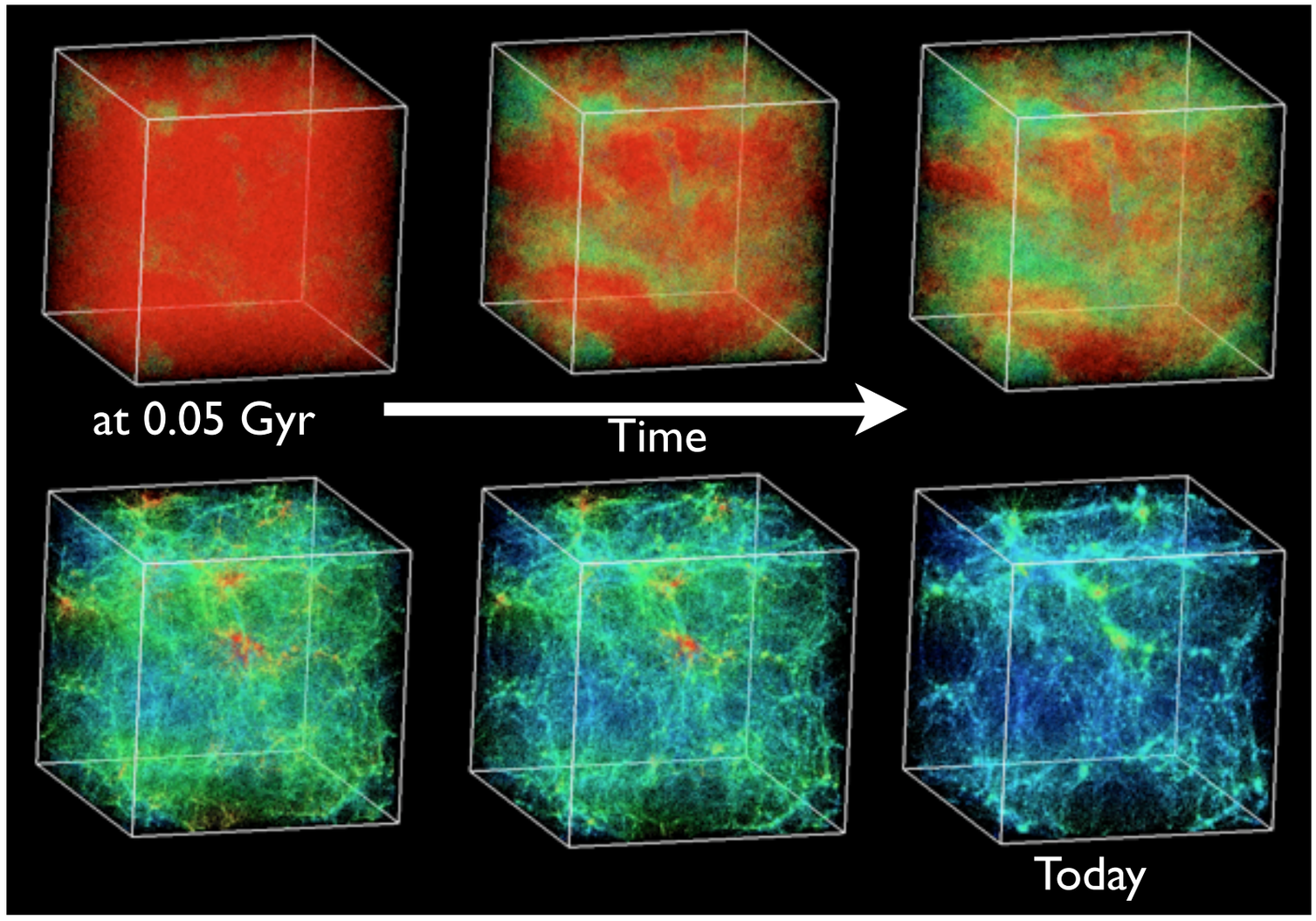
- **Solid understanding of structure formation; success underpins most cosmic discovery**
  - Initial conditions determined by primordial fluctuations
  - Initial perturbations amplified by gravitational instability in a dark matter-dominated Universe
  - Relevant theory is gravity, field theory, and 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



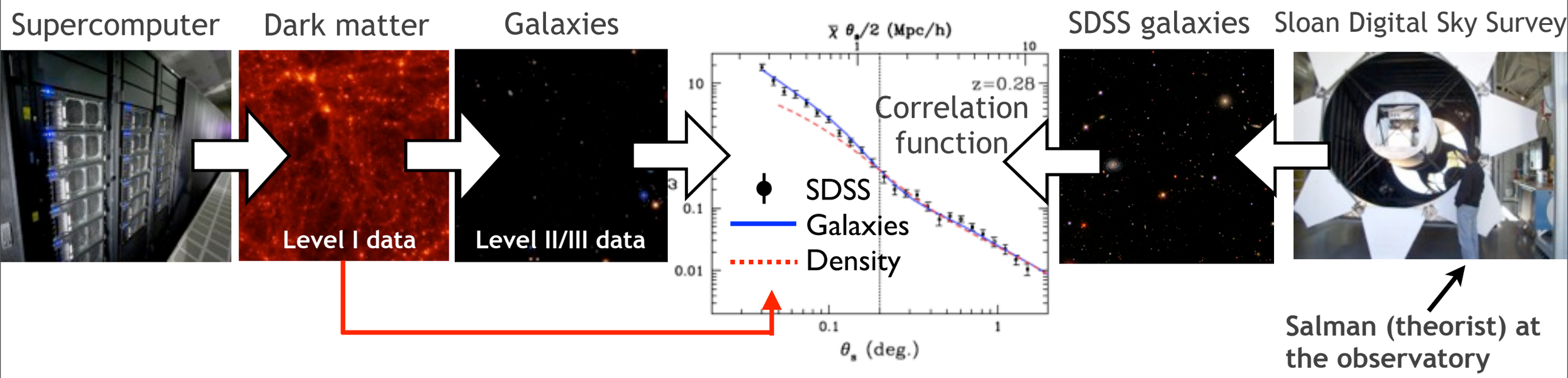
Computing the Universe



> **"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)*



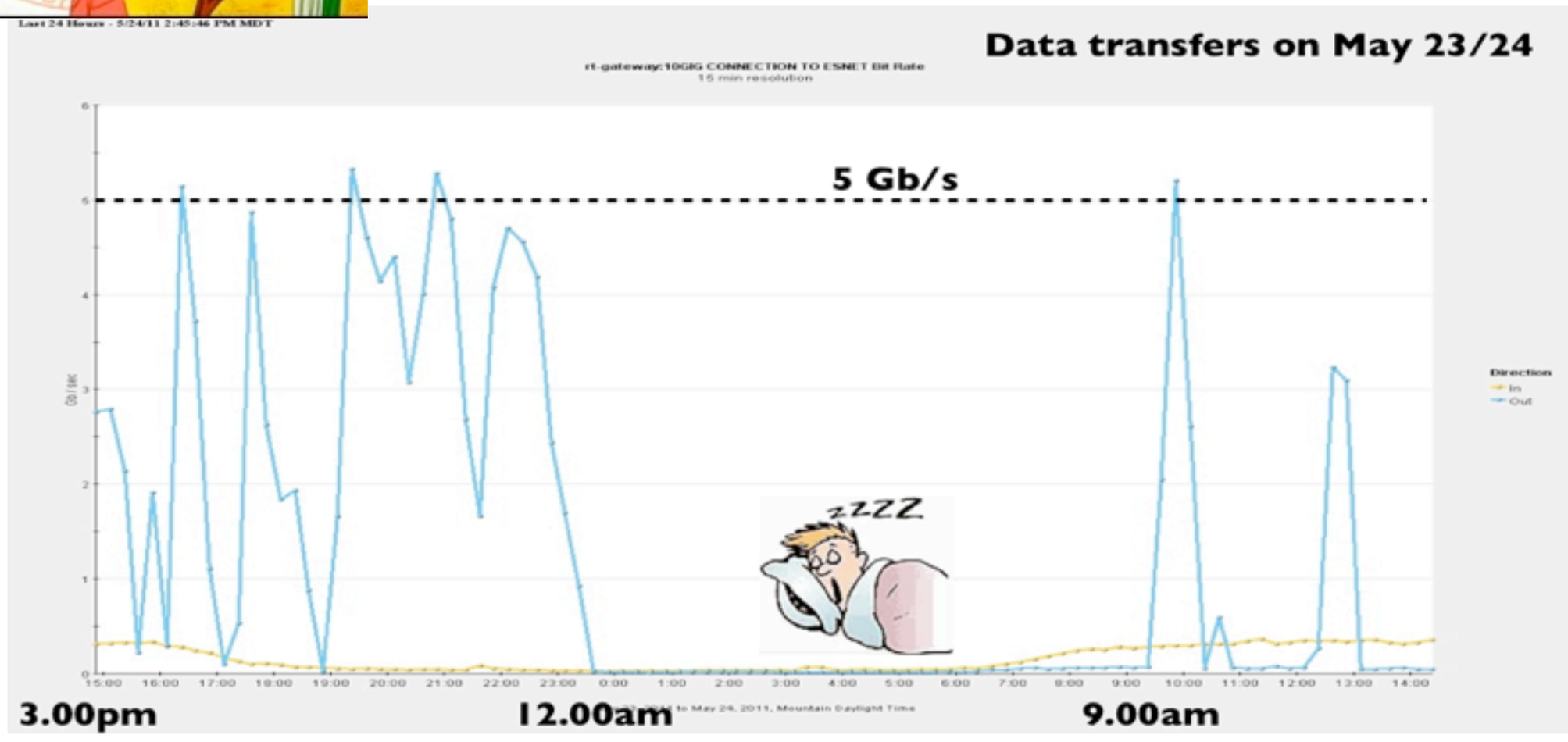
Connecting Theory and Observations



- Simulate the formation of the large scale structure of the Universe via dark matter tracer particles; Level I data, TB - PB data sets
- Take dark energy into account in the expansion history
- Measure the high-density peaks (dark matter halos) in the mass distribution; Level II data, factor 10-100 smaller than Level I
- “Light traces mass” to first approximation, therefore populate the halos with galaxies, number of galaxies depends on mass of halo, type of galaxies depend on formation history of dark matter halos; Level III data, again large reduction from Level I

The Coyote and Roadrunner Universe

- Coyote Universe: 38 different cosmological models (1 billion particles each), 11 time snapshots per simulation
- Roadrunner Universe: Two 8-billion particle simulations
- With great help from Globus Online team and Los Alamos network folks, moved ~20TB of data from LANL to ANL, Level I data
- Extraction of more science and extension with additional simulations carried out at ANL on Intrepid, NERSC on Hopper, LANL on Mapache



HACC Port to the BG/Q Systems

- H(ardware) A(ccelerated) C(osmology) C(ode) port is making ***very*** good progress
- Very good performance on VEAS (Very Early Access System), 1 rack, 1024 nodes
- Test runs up to 16 racks at IBM:

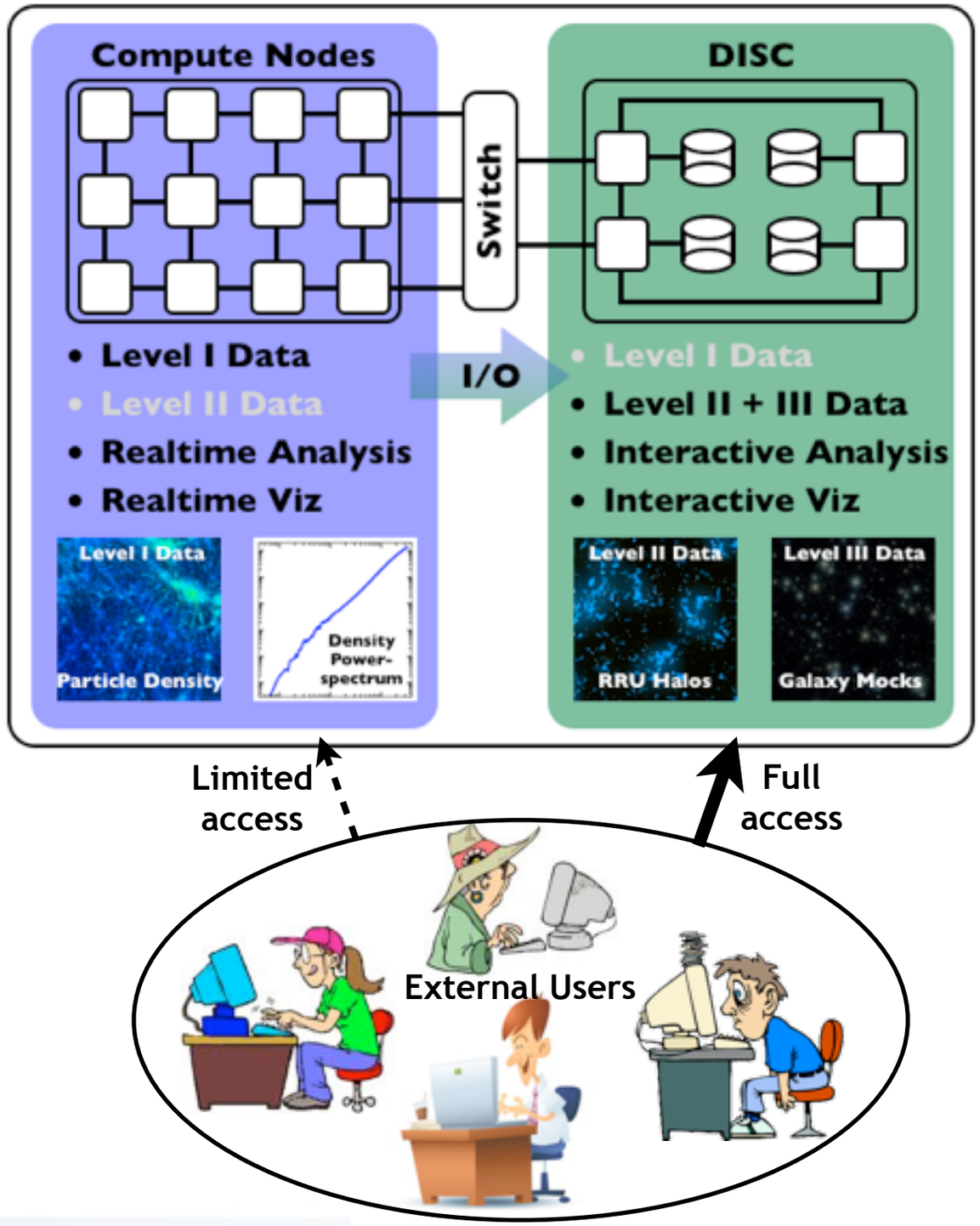
```
Creating rho(x,y,z)  3        39.77        17.78
Poisson solve        3        51.94        23.23
Particle move        3         2.04       0.9123
Output               3         0.53        0.237
================================================

PXCH buffer = 2*1943590 = 3.707104 MB
InitialExchange TotalAliveParticles 1073741824000
```

One trillion+ particles!
(remember, Coyote runs were one billion, largest cosmology run in the past: ~300 billion particles)

- One snapshot: ~36TB, want to keep 10-50, 0.5-2PB per run, Mira=48 racks...
- Impossible to store the raw data

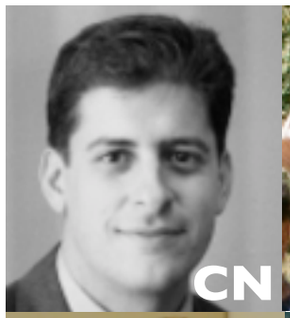
Serving the Mira Universe

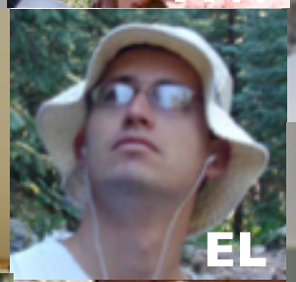
- **Simulation datasets: Currently simulation data generation is constrained only by storage and I/O bandwidth, ~PB datasets will be available in the near future**
  - Level I Data: Raw particle data, generated on main supercomputer (e.g Mira)
  - In-situ analysis leads to Level II Data: Halo catalogs, density fields (2-D), generated on supercomputer but live mainly on DISC system
  - Level III Data (derived from Level II): Simulated galaxy catalogs, X-ray “observations” etc, live on DISC system
- **How can we efficiently share data?**
  - Simulation campaigns are carried out at very few places (supercomputer centers)
  - Outputs are very science rich, many people can contribute to the analysis
  - Moving raw data is impractical (at some point impossible), analysis often takes a lot of computing power
  - Need for making data *and* analysis opportunity available to the community

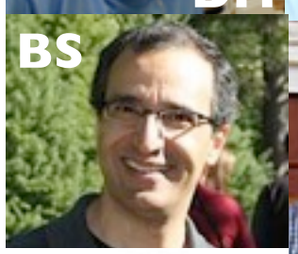


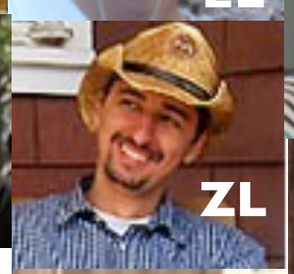
Thanks to all collaborators:

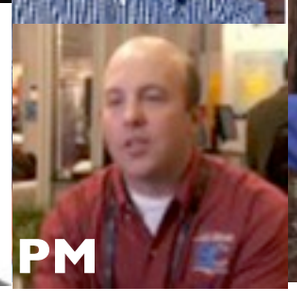
Our outstanding external collaborators (some pictures missing)

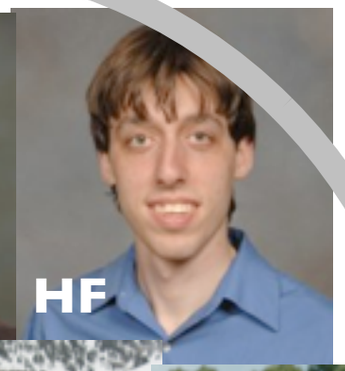


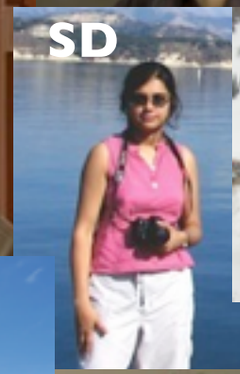


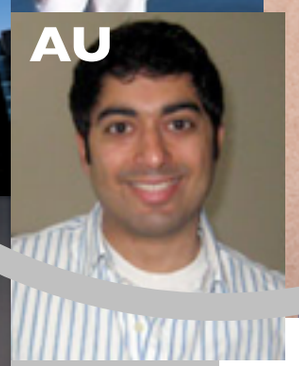












Our new group at Argonne National Laboratory