Chicagoland Computational Cosmology Intitiative

Salman Habib (ANL)
HEP and MCS Divisions

Bob Rosner (UoC), Scott Dodelson (FNAL)

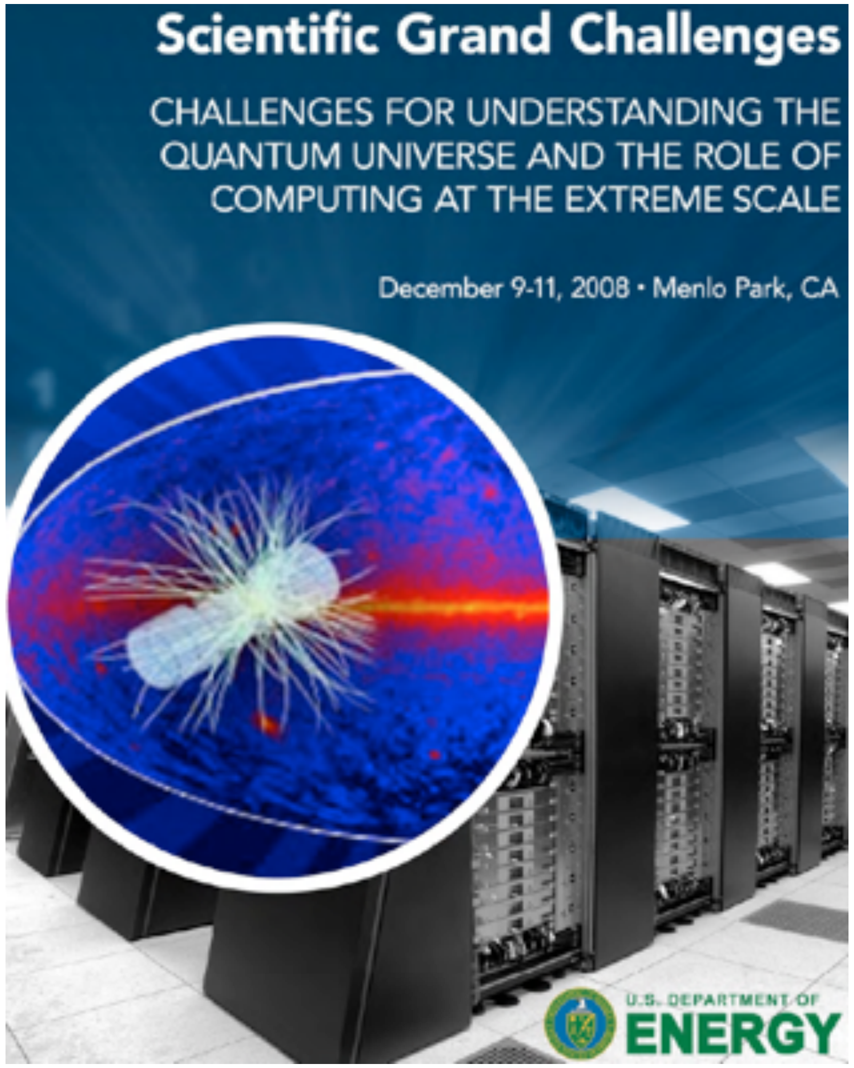
Updates:
1) Formation of new DOE HEP Labs collaboration
2) New Physics Frontier Center at KICP
3) New Research Computing Center at U Chicago
4) New opportunities (SciDAC, --)

Computational Cosmology and the Cosmic Frontier

- **Cosmic Frontier Science:**
  - **Dark Energy**
  - **Dark Matter**
  - **Inflation/Early Universe**
  - **Neutrinos**
- **Projects:** Use probes based on structure formation, the primary focus area of computational cosmology
  - **BOSS:** Baryon Oscillation Spectroscopic Survey
  - **DES:** Dark Energy Survey
  - **LSST:** Large Synoptic Survey Telescope

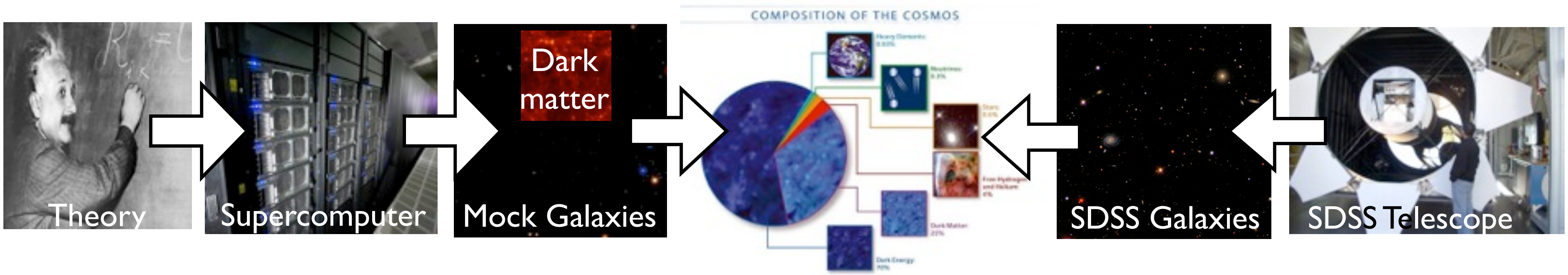
DARK ENERGY SURVEY

Large Synoptic Survey Telescope

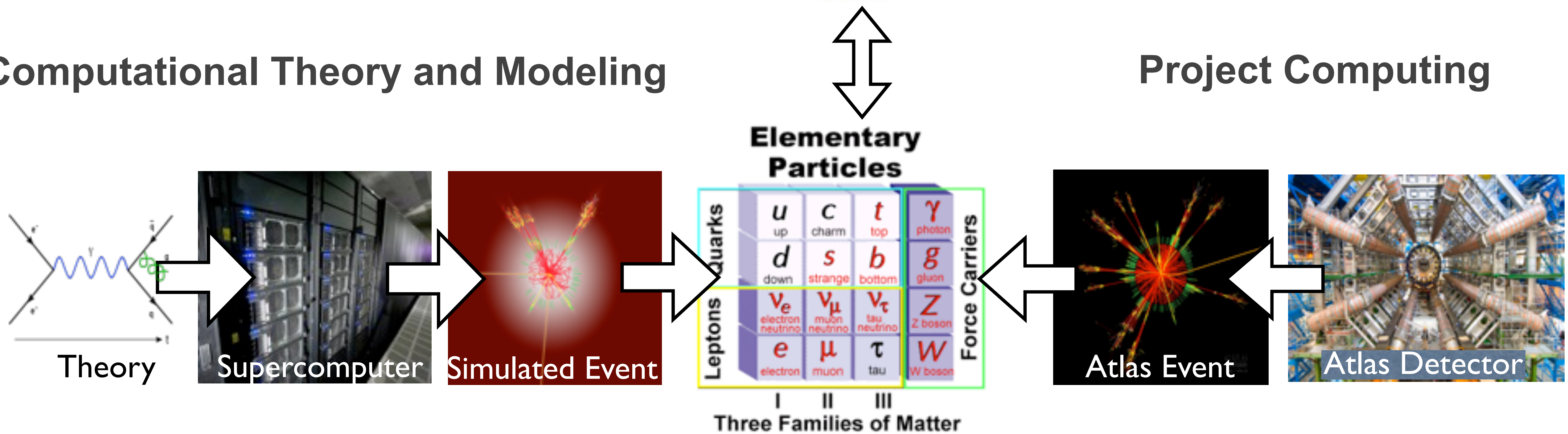


The Universe as a HEP experiment

Cosmic Frontier: **Uncontrolled experiments (Universe is the apparatus); role of computational theory and modeling is pervasive, complex, and crucial to the success of the entire enterprise**



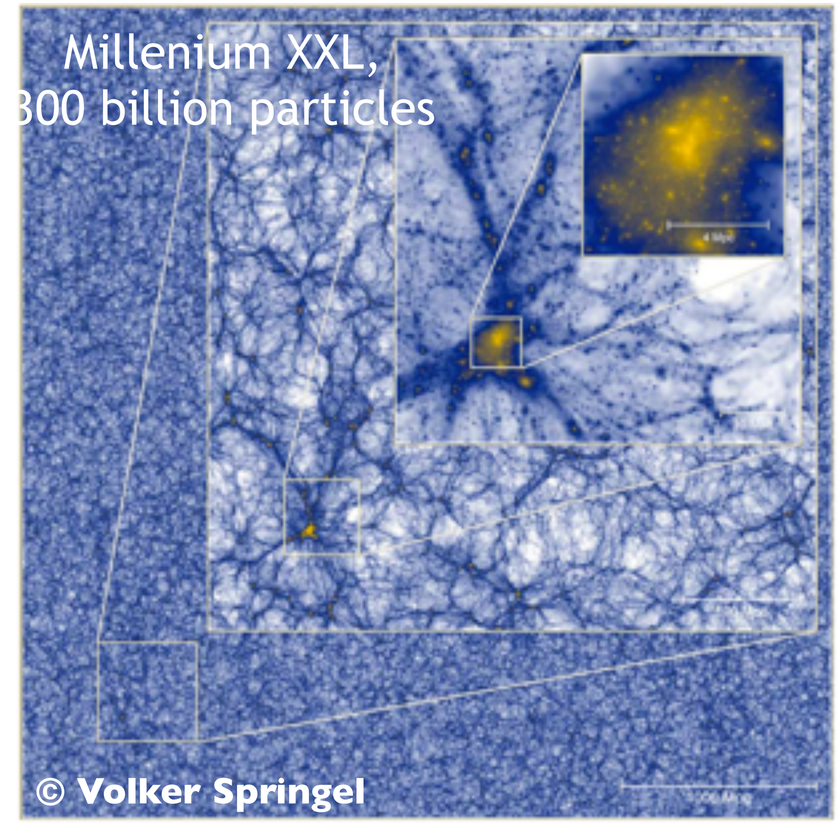
Computational Theory and Modeling | **Project Computing**



Energy Frontier: **Controlled experiments, role of theory can be targeted and focused on specific critical calculations**

Computational Paradigm I

- **N-body, gravity only:** At large scales (>~1 Mpc), gravity is the dominant force and remains important down to the smallest scales
  - **Methods:** Modern codes are parallel solvers for the Vlasov-Poisson equation, employing particle and grid techniques, often together
- **'Hydrodynamics':** Addition of baryonic physics
  - **Methods:** Eulerian (typically AMR) or Lagrangian (typically SPH); feedback processes and sub-grid models (chemistry, star formation), other physics (e.g., MHD) introduced as appropriate
- **Scales:** Code runs can cover scales as small as resolving individual galaxies to box sizes at the horizon scale (several Gpc to a side)
- **Analysis:** Many compute-intensive techniques (real-time/post-processing): ray-tracing, N-point statistics, halo/sub-halo finders, merger-trees, --



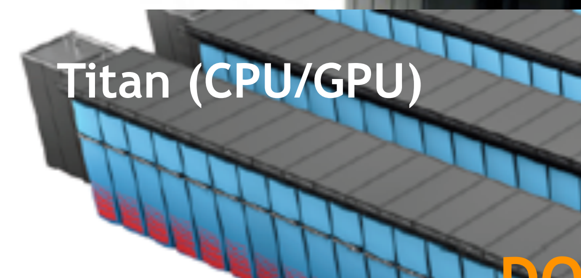
Simplified DES catalog flow chart

- simulation
- halo finding
- merger trees
- assign galaxies
- compute shear
- lens galaxies
- photometric errors
- photo-z
- sky images

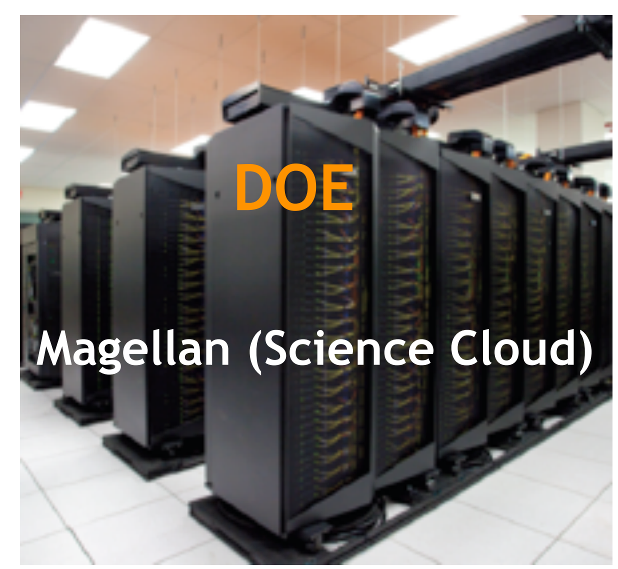
© Risa Wechsler

Computational Paradigm II

- **Algorithms:** Improvements are focus of major effort by the computational community (not just cosmologists, e.g., beam physics, CFD, plasma physics --)
- **Hardware evolution:** Radical architecture changes, new programming models and imperatives in the near future
  - **Code future:** How will codes evolve, community response needed
  - **Ecology:** Current code mix may shrink as development goes over to the team model
  - **Job mix:** Leadership-class vs. large numbers of large/medium job ensemble runs
- **Data-intensive applications:** Data-specific platforms already available ('cloud')
  - **Optimal mix:** Very important for future surveys, especially LSST (HEP/ASCR report)



Divergence of supercomputer architectures



Data-intensive systems

Computational Cosmology and ASCR

- **DOE ASCR:** Focus on scientific discovery through computing, cutting edge computer science and applied math as relevant to HPC
  - **'Big Iron':** Traditional HPC, INCITE/SciDAC
  - **Co-design:** Exascale pathway
  - **Algorithms:** Linear solvers, PDE solvers (CFD), uncertainty quantification, multiphysics problems, --
  - **Computer Science:** Programming models, frameworks, compilers, file systems, networks, visualization, ACTS collection, --
- **Computational cosmology:** Due to broad nature of the field, there are several strong connection possibilities, primarily HPC/algorithms/tools; data issues not yet a high priority within ASCR (but starting --)
- **Collaboration benefits:** Access to top-level CS expertise, programming models work, early access to next-generation architectures, --

ASCR projects

Computational Cosmology and Large Data

- **Observational datasets:** Large, high-dimensional datasets; ~10 TB in the SDSS-era, ~1 PB for DES, ~100 PB for LSST, serious computational requirements for data-intensive computing
- **Simulation datasets:** Currently simulation data generation is constrained only by storage and I/O bandwidth, ~PB datasets will be available in the near future
  - **In situ analysis:** Large-scale analysis tasks on the compute platform; data compression
  - **Post-processing:** Post-run analyses on host system or associated ‘active storage’
- **Joint strategy:** Analysis and workflow needs for computational cosmology and survey data will have substantial overlap; joint program should be developed (HEP/ASCR report)

5th Extremely Large Databases Conference

October 18-19, 2011
Panofsky Auditorium
SLAC National Accelerator Laboratory
Menlo Park, California

Summer 2011 Week 4: Large Datasets in Astrophysics and Cosmology

August 6, 2011 to August 13, 2011
The Canyons Resort, 4000 Canyons Resort Drive, Park City, Utah

Chicagoland Opportunities

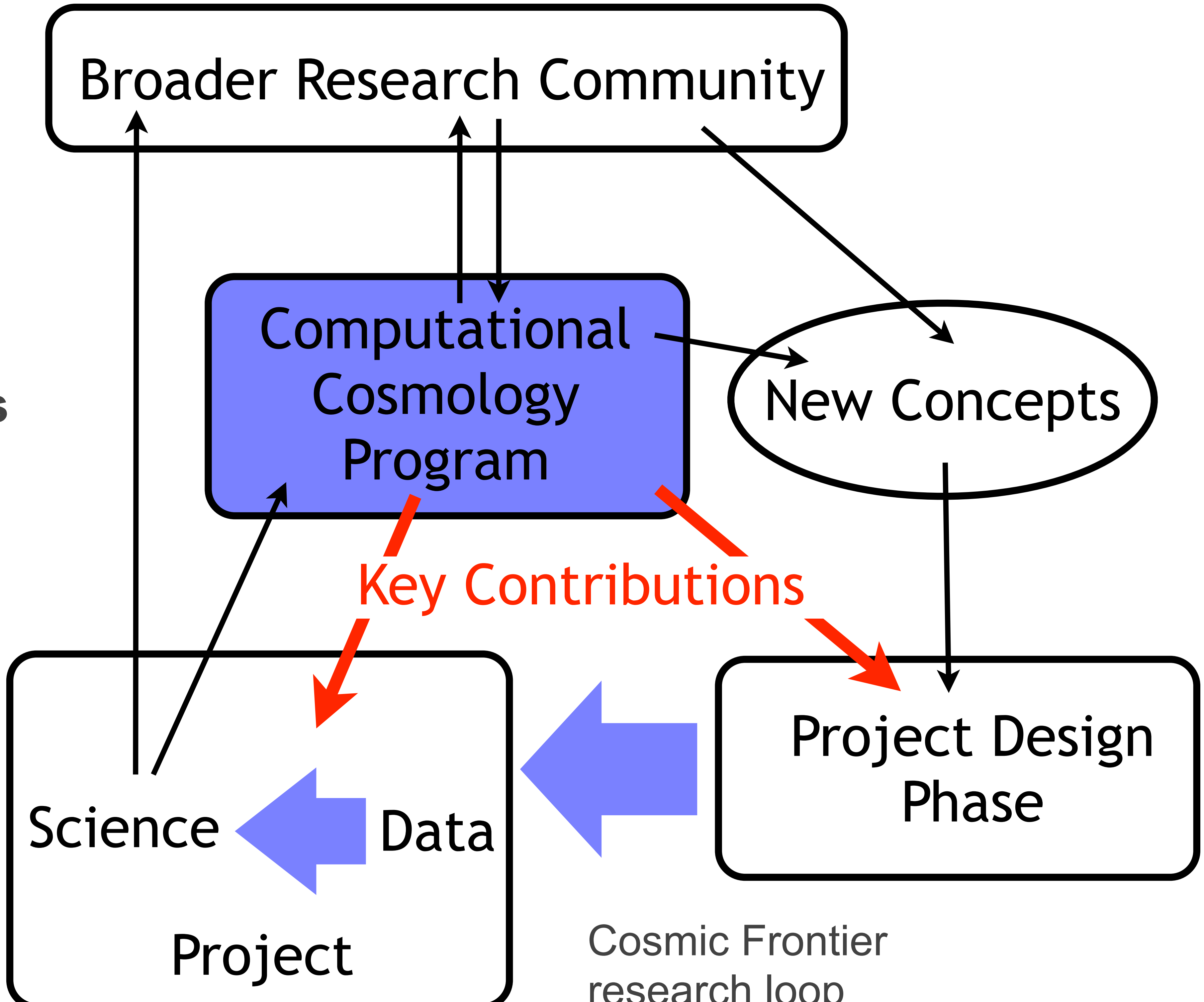
- **Fundamental Science**
  - Beyond the Cosmological Standard Model
  - Precision Predictions for Cosmological Probes
- **Survey Science**
  - Interpretation of data from surveys (DES, SPT, --)
  - Future surveys (DESpec, LSST, --)
  - Simulation database accessible by local and external community
- **High Performance Computing and Data-Intensive Supercomputing (DISC)**
  - HPC: Petascale to Exascale at Argonne
  - High-throughput computing at Fermilab
  - Data-oriented science at Computation Institute
  - New Research Computing Center at UoC
- **Education and Outreach**
  - Student and post-doc involvement
  - Opportunities for cross-disciplinary collaborations

Chicagoland: Computational Capabilities

- **Complementary Simulation Codes**
  - Coverage of gravity and hydrodynamics and 'sub-grid' physics
  - Adaptive Refinement Tree (ART) at UofC/Fermilab
  - Hardware Accelerated Cosmology Codes (HACC) framework at Argonne
  - FLASH at UoC in collaboration with Argonne
- **Data-Intensive Computing**
  - Grid computing at all three institutions
  - Petascale Active Data Store (PADS) at CI
  - High-speed data transfers available locally
- **Next-Generation Computational Platforms and Associated Computer Science**
  - IBM BG/Q Mira (10 PF, 750 TB) coming to Argonne in 2012
  - New medium-scale cluster at UoC; Research Computing data initiative
  - Remote and in situ visualization/analysis (CI/ANL)

DOE HEP Labs Computational Cosmology Program

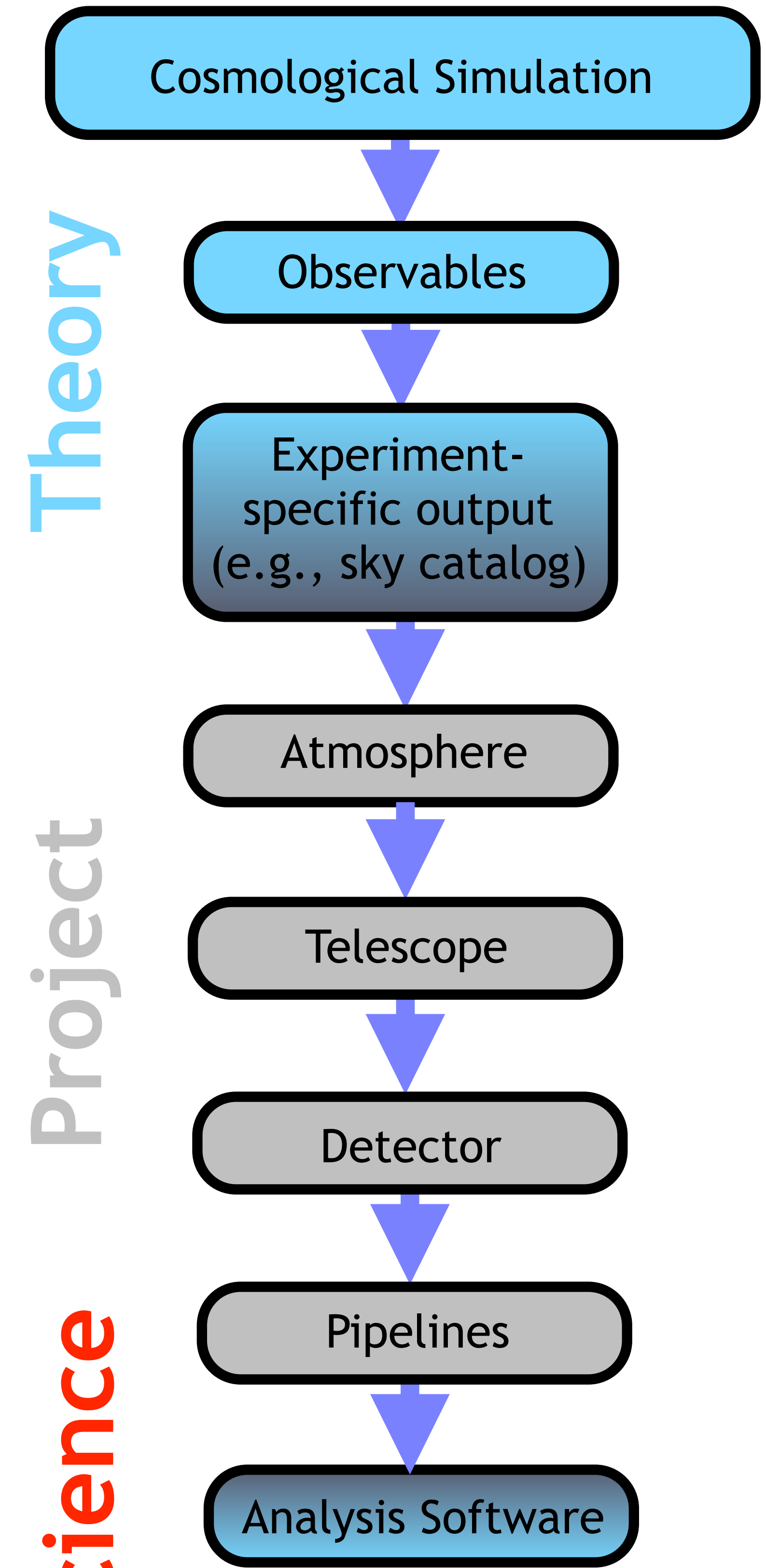
- **Resides as a core capability program within DOE HEP**
- **Contributes to 'discovery space'**
- **Catalyzes development of concepts into projects**
- **Plays a key role in project optimization**
- **Is an essential component of the 'Data to Science' step for projects**
- **Functions as a major community resource**



Cosmic Frontier research loop

September Planning Meeting: Overall Concept

- **Day I: September 13**
  - Essential role of computing in DOE HEP's cosmology program
  - Computing enters at three levels (i) underlying theory, (ii) computing within experiments (design/execution), and (iii) analysis of data
  - Wide range of computing is needed, from the most significant HPC resources to data-intensive supercomputing
  - Day 1 outlines the case in a broad sense
- **Day 2: September 14**
  - Focused presentations on DOE HEP Lab-based computational cosmology collaboration, discuss path forward, including connections within and outside DOE
  - Discuss points to address and other requirements for meeting report



U.S. DEPARTMENT OF ENERGY | Office of Science

DOE HEP Lab Collaboration Established

- **Collaboration**
  - Cosmic Frontier Computing Collaboration (CFCC)
  - Member Institutions: ANL, BNL, FNAL, LBNL, SLAC
  - Institutional Representatives: S. Habib (spokesperson), H. Weerts (ANL), A. Slozar, M. May (BNL), S. Dodelson, C. Hogan (FNAL), P. Nugent, B. Cahn (LBNL), R. Wechsler, R. Blandford (SLAC)
- **Timeline**
  - Presentation to DOE HEP management (Washington DC, June 2011)
  - Institutional Planning Meeting (LBNL, August 2011)
  - DOE HEP-sponsored Planning Meeting (Washington DC, September 2011)
- **Opportunities**
  - SciDAC proposal call (HEP/ASCR partnership), due in January 2012
  - ALCC/INCITE proposals due in first half of 2012
  - Computation time awarded at NERSC

SciDAC
Scientific Discovery through Advanced Computing

Chicagoland Planning

- **Progress to Date**
  - Broad science case made and targets identified
  - Scale of computations understood (previous studies and white papers --)
  - Software and algorithmic development requirements identified (driven by next-generation hardware mix and expected 'data deluge')
- **Boundary Conditions**
  - DOE/NNSA exascale co-design initiative (unclear)
  - DOE HEP has initiated the CFCC
  - Need to get inputs from surveys (DES, LSST, --), increase interaction level
  - Possible cross-agency collaboration (DOE, NSF, --), more interesting now
- **Suggested Actions**
  - Need to initiate/continue and investigate local science collaborations and interactions in detail
  - Suggestion for a computational cosmology data center

Chicagoland Computational Cosmology Workshop

- **One-Day Meeting held at ANL**
  - Sponsored by ANL and KICP
  - Bring together astrophysicists, computer scientists, cosmologists, mathematicians, physicists, --
  - 40+ participants, 4 sessions:
    1) Computational Cosmology
    2) Astro/Cosmology Interface
    3) High Performance Computing
    4) Big Data
  - Tremendous intellectual resources available (potentially), state of the art hardware resources as well as unique ‘software as service’ expertise
  - Key Question: How to put this together?

Argonne NATIONAL LABORATORY

KICP

Chicagoland Simulation Data Center

Contact: K. Heitmann (ANL, heitmann@anl.gov)

- **Motivation**
  - Simulation data is large and complex, generation requires significant expertise
  - Data and simulation analysis should be in the same location
  - Connection to sky surveys (DES, SPT, LSST, others, --)
  - Large potential user base, can be expanded to a national facility
- **Possible Configuration(s)**
  - Analysis cluster and separate storage
  - Computing/storage combined in ‘active storage’ or data-intensive computer
  - Cloud-based systems allow for user-defined analysis
- **Staging**
  - Precomputed/analyzed data only
  - Ability to run canned queries and analysis routines (real-time/batch)
  - Ability to run specialized queries and analyses

What's Next?

- **Coordination**
  - Different institutions have their own priorities (projects, local needs, --)
  - Can use projects (DES, SPT, LSST) as drivers, other motivations are important as well (uniform needs for workflows, HPC, viz/analysis, etc.)
- **Resources**
  - Limited kick-start opportunities, how to overcome this?
  - Need to leverage current possibilities (SciDAC, INCITE, --)