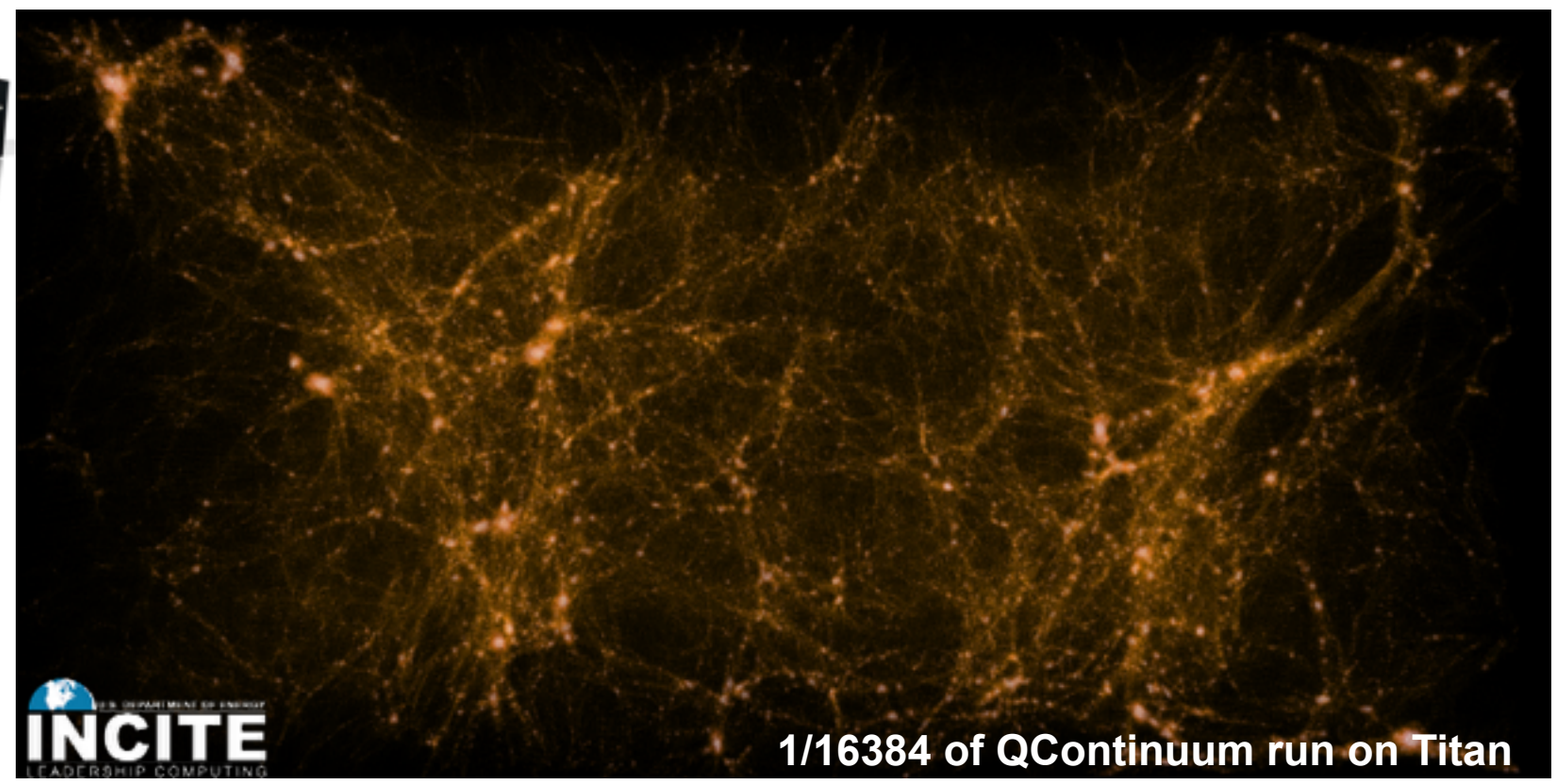
How to Stuff a Supercomputer inside a Laptop

Salman Habib
Argonne National Laboratory



1/16384 of QContinuum run on Titan

MIRA at Argonne

What are Supercomputers Good For?

- Quantitative predictions for complex systems
- Discovery of new physical mechanisms
- Ability to carry out 'virtual' experiments (system scale simulations)
- Solving large-scale inverse problems
- Ability to carry out very large individual runs as well as simulation campaigns
- New communities, new directions?
- Supercomputing + Large-Scale Data Analytics?

OLCF

OLCF Science

Q&A: Jack Wells, Director of Science, OLCF
Simulations Explore Next-Generation Fuels
Jaguar Accelerates Design of GE Turbomachinery
Ramgen Simulates Shock Waves, Makes Shock Waves across Energy Spectrum
Climate Scientists Compute in Concert
Supernovas Explode in 3D Detail at ORNL
Materials Modeling Shows Big Future for Boron Nitride Nanoribbons
Researchers Pinpoint How Copper Folds Protein into Precursors of Parkinson's Plaques

SCIENCE HIGHLIGHTS
BIOLOGICAL SCIENCES
Macromolecular Folding and Aggregation 30
Studies of Large Conformational Changes in Biomolecular Machines 32
CHEMISTRY
From LES to DNS of Explosions in Semi-Confined Domains 34
COMPUTER SCIENCE
Open Source Compiling for Supercomputers 36
Efficient Parallel Volume Rendering of Large-Scale Adaptive Mesh Refinement Data 38
ENERGY TECHNOLOGIES
Thermodynamics of Binding Biomass to Cellulases for Renewable Fuel 40
ENGINEERING
Direct Numerical Simulations of High Reynolds Number Turbulent Channel Flow 42
Enabling Green Energy and Propulsion Systems via Direct Noise Computation 44
MATERIALS SCIENCE
Dynamics of Conformational Transition in Thermo-Sensitive Polymers and Hydrogels 30
Materials Design and Discovery (Improving Cancer Drug Design) 31
Petascale Simulations of Stress Corrosion Cracking 32
PHYSICS
Ab Initio Quantum Liquid Water and Aqueous Ionic Solutions 34
Ab Initio Reaction Calculations for Carbon-12 35
Kinetic Simulations of Fusion Energy Dynamics at the Extreme Scale 36
Lattice QCD 37
MockBOSS: Calibrating BOSS Dark Energy Science with HACC 38

ALCF

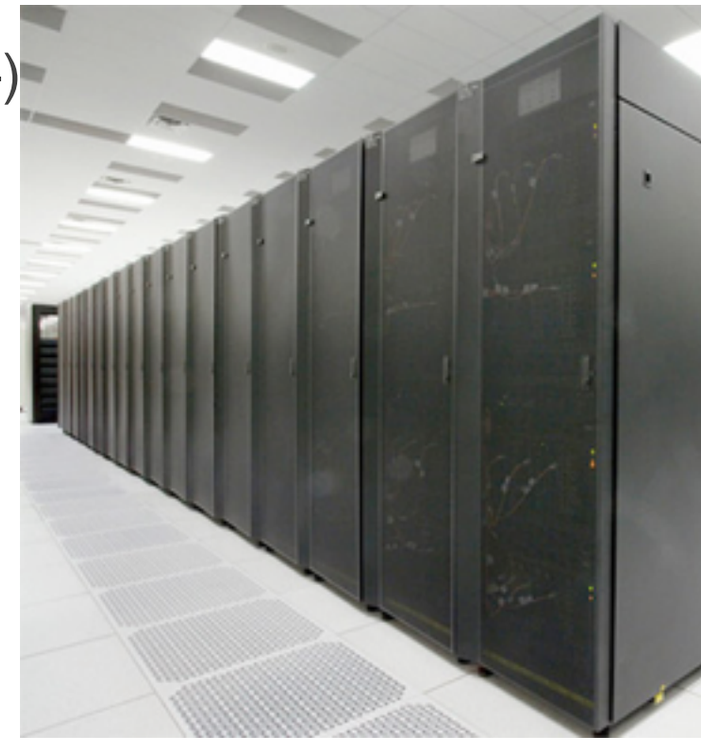
Table of Contents

4 The Year in Perspective
6 Research News
8 Edison Opens New Doors for Early Users
12 IceCube IDs 'Bert' and 'Ernie'
14 Spotting Hotspot Volcanoes
16 Unlocking the Secrets of Solar Wind Turbulence
18 With Nanoparticles, Slower May Be Better
20 Calculations Confirm 150 Years of Global Land Warming
22 Physicists ID New Molecules With Unique Features
24 Why Foams Foam and Bubbles Pop
26 Reading the Cosmic Writing on the Wall
28 The Life and Times of Massive Stars
30 Thinnest Solar Cells Ever?
32 Taming Plasma Fusion Snakes
34 Getting More from Natural Gas Reserves
36 Policing Air Quality in California
38 Turning Greenhouse Gases into Gold
40 Helping a Catalyst Reach its Full Potential
42 How Cells Interact With Their Surroundings
44 An Inside Look at an MOF in Action

NERSC

But —

- **Dealing with supercomputers is painful!**
  - HPC programming is tedious (MPI, OpenMP, CUDA, OpenCL, —)
  - Batch processing ruins interactivity
  - File systems corrupt/eat your data
  - Software suite for HPC work is very limited
  - Analyzing large datasets is frustrating
  - HPC experts are not user-friendly
  - Machine crashes are common
  - Ability to ‘roll your own’ is limited



Running Jobs | Queued Jobs | Reservations

Total Queued Jobs: 172

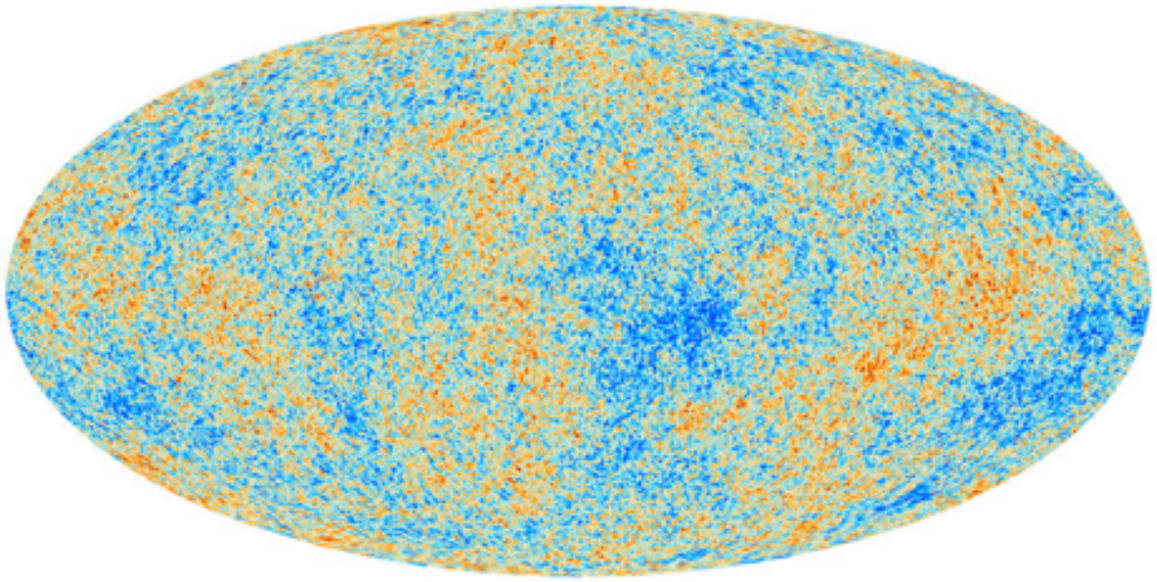
Job Id	Project	Score	Walltime	Queued Time	Queue	Nodes
307941	SkySurvey	8351.7	1d 00:00:00	5d 01:10:03	prod-capability	32768
307942	SkySurvey	8350.5	1d 00:00:00	5d 01:09:42	prod-capability	32768
309793	NucStructReact_2	7069.0	01:00:00	1d 19:13:34	prod-capability	32768
309794	NucStructReact_2	7065.1	01:00:00	1d 19:12:28	prod-capability	32768
309795	NucStructReact_2	7056.8	01:00:00	1d 19:10:04	prod-capability	32768
309271	LatticeQCD_2	6121.1	03:00:00	3d 03:40:34	prod-capability	12288
309314	LatticeQCD_2	5036.1	04:50:00	2d 22:51:59	prod-capability	12288
309315	LatticeQCD_2	5034.8	03:00:00	2d 22:51:38	prod-capability	12288
309316	LatticeQCD_2	5034.0	04:50:00	2d 22:51:24	prod-capability	12288
309317	LatticeQCD_2	5033.0	03:00:00	2d 22:51:08	prod-capability	12288
309318	LatticeQCD_2	5032.6	04:50:00	2d 22:51:01	prod-capability	12288

Supercomputers: Personal History

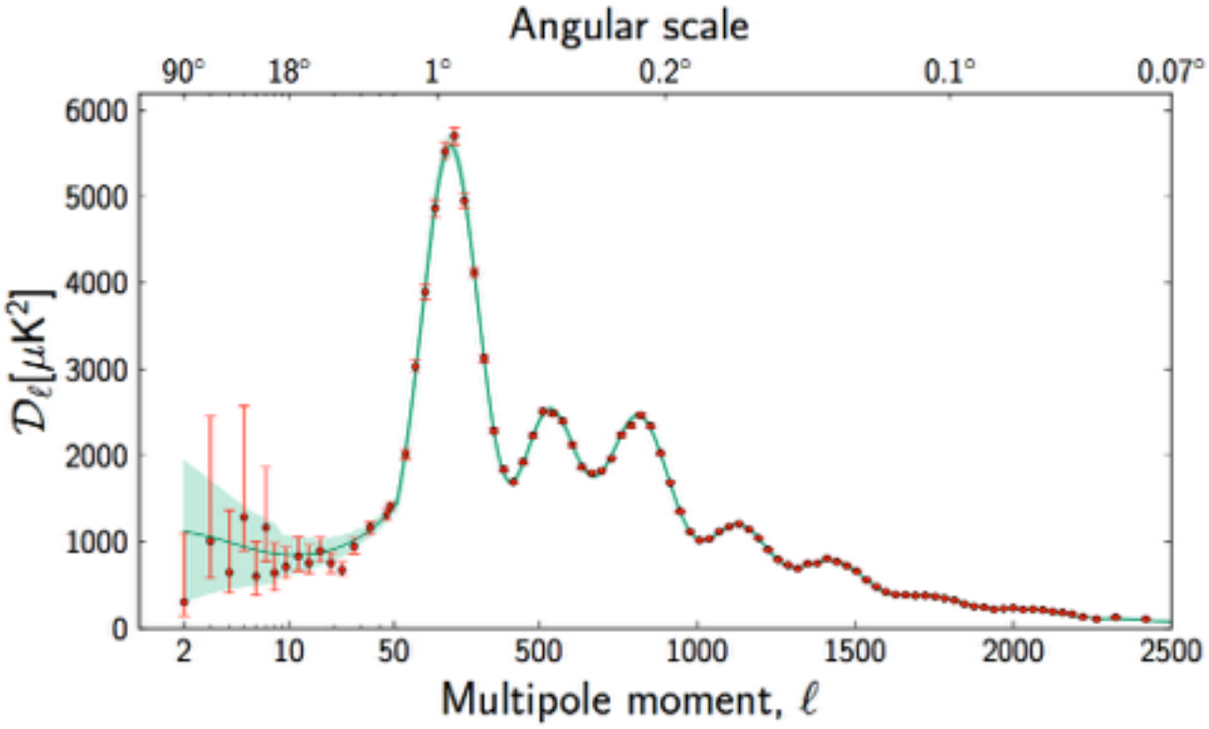
Supercomputing is not for everyone —

When are Supercomputers (almost) Replacable?

- Obviously if there's no hurry, problems are small enough, etc.
- More interestingly — what is the quantity of actual interest?
- Typically supercomputers 'create' information — how much of this is actually important?
- Key role of summary statistics/feature extraction — a form of data compression
- Emulation — The idea of using a finite ('small') number of supercomputer simulations to interpolate over parameter dependencies
- How smooth are the parametric dependencies?
- Can such a strategy work in a high-dimensional space?

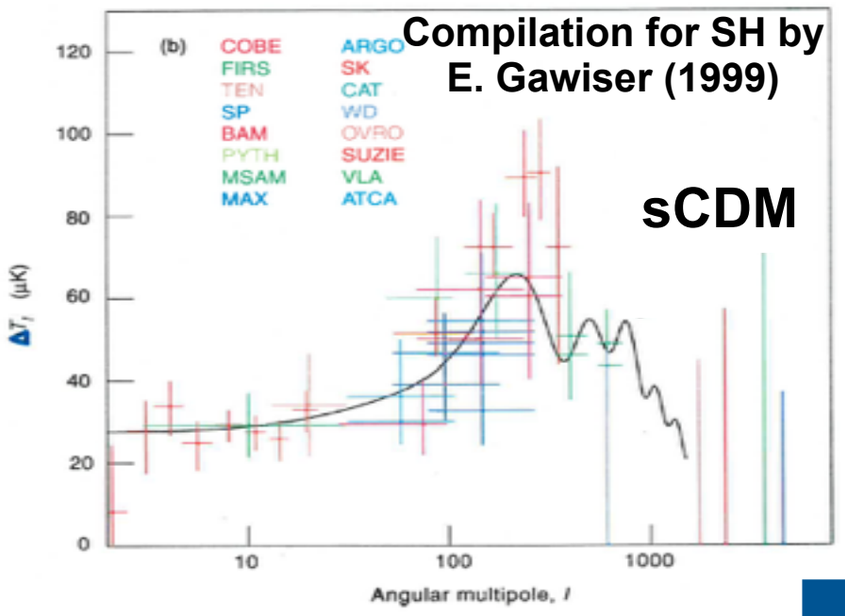


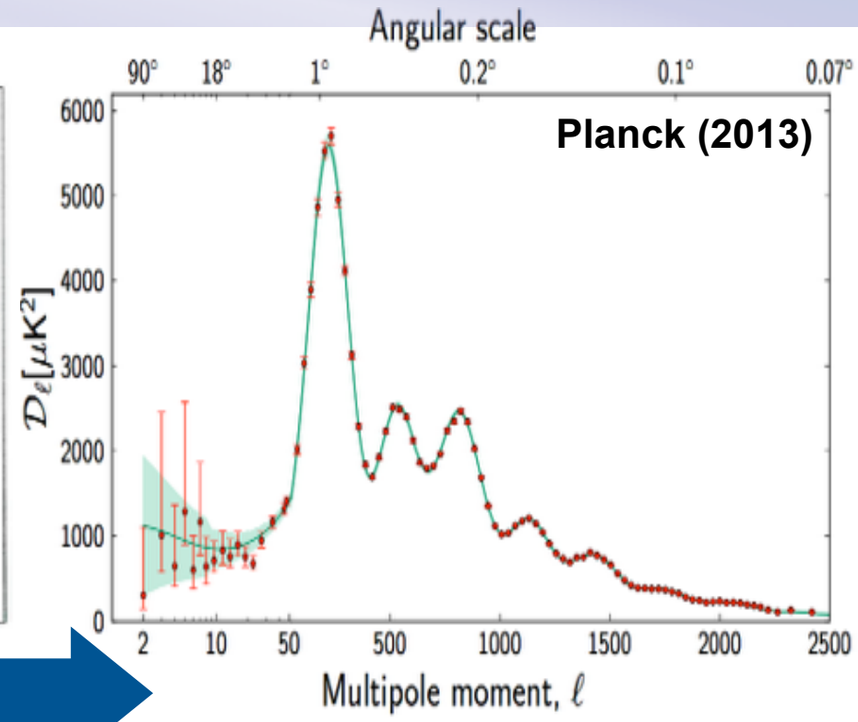
Temperature fluctuations in the CMB sky as measured by Planck, 2013

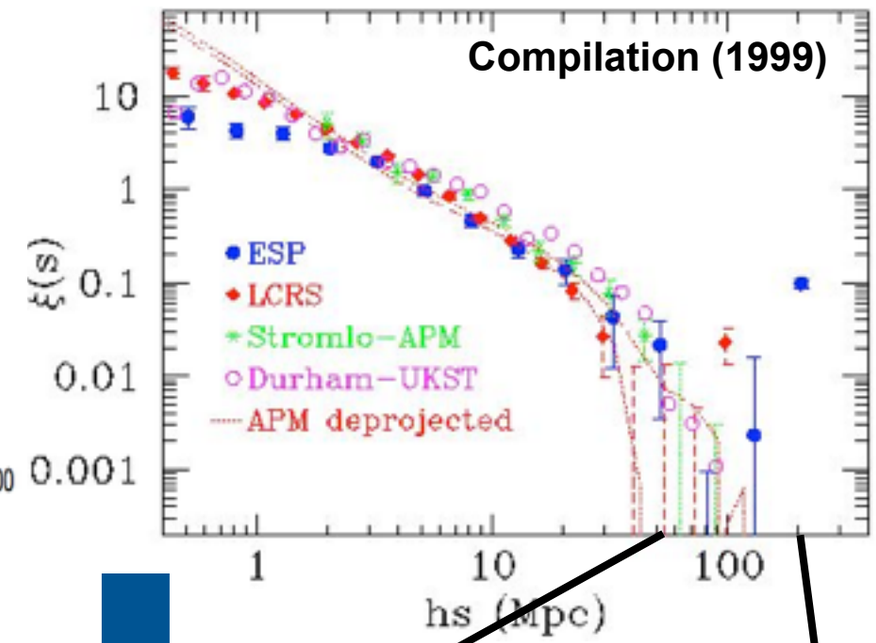


Summary Statistic: Temperature fluctuation angular power spectrum in multipole space

Summary Statistics: Precision Cosmology Example

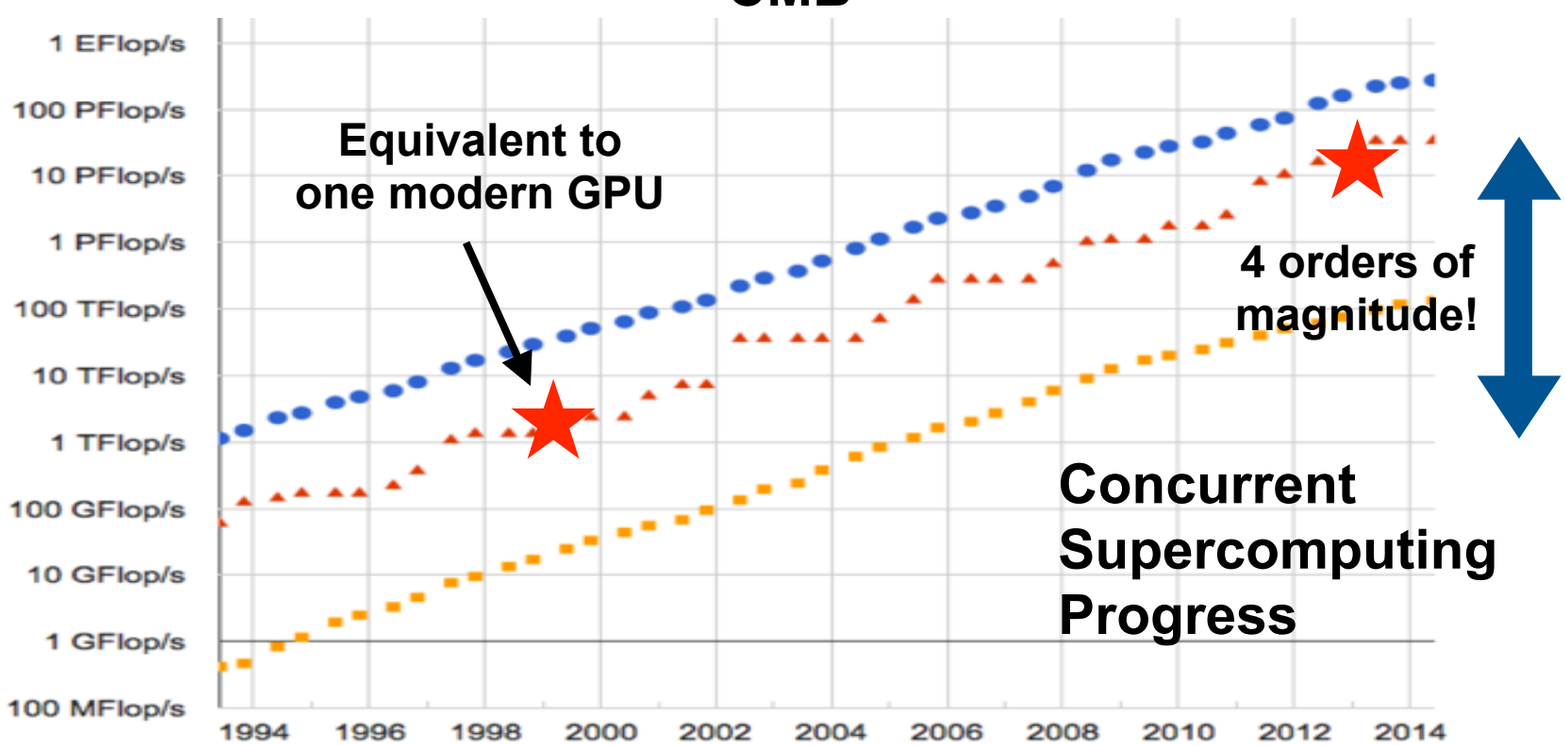


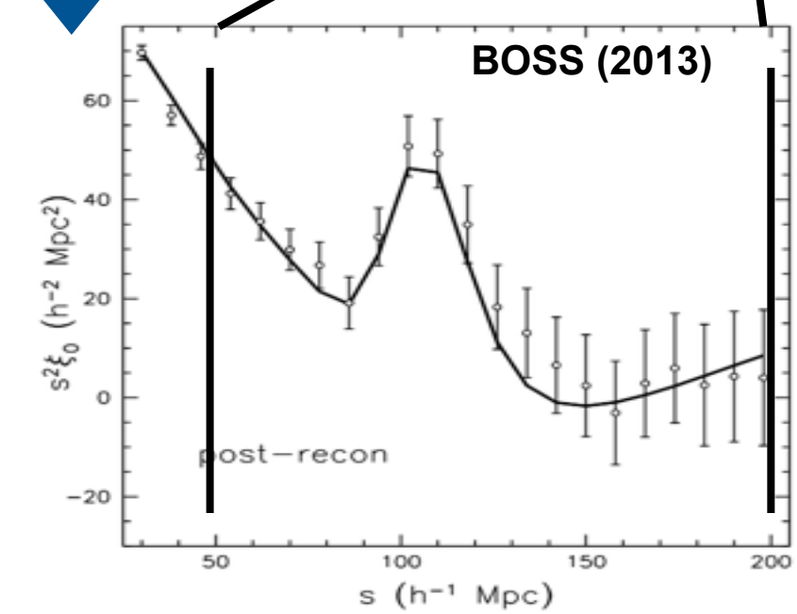




CMB

LSS

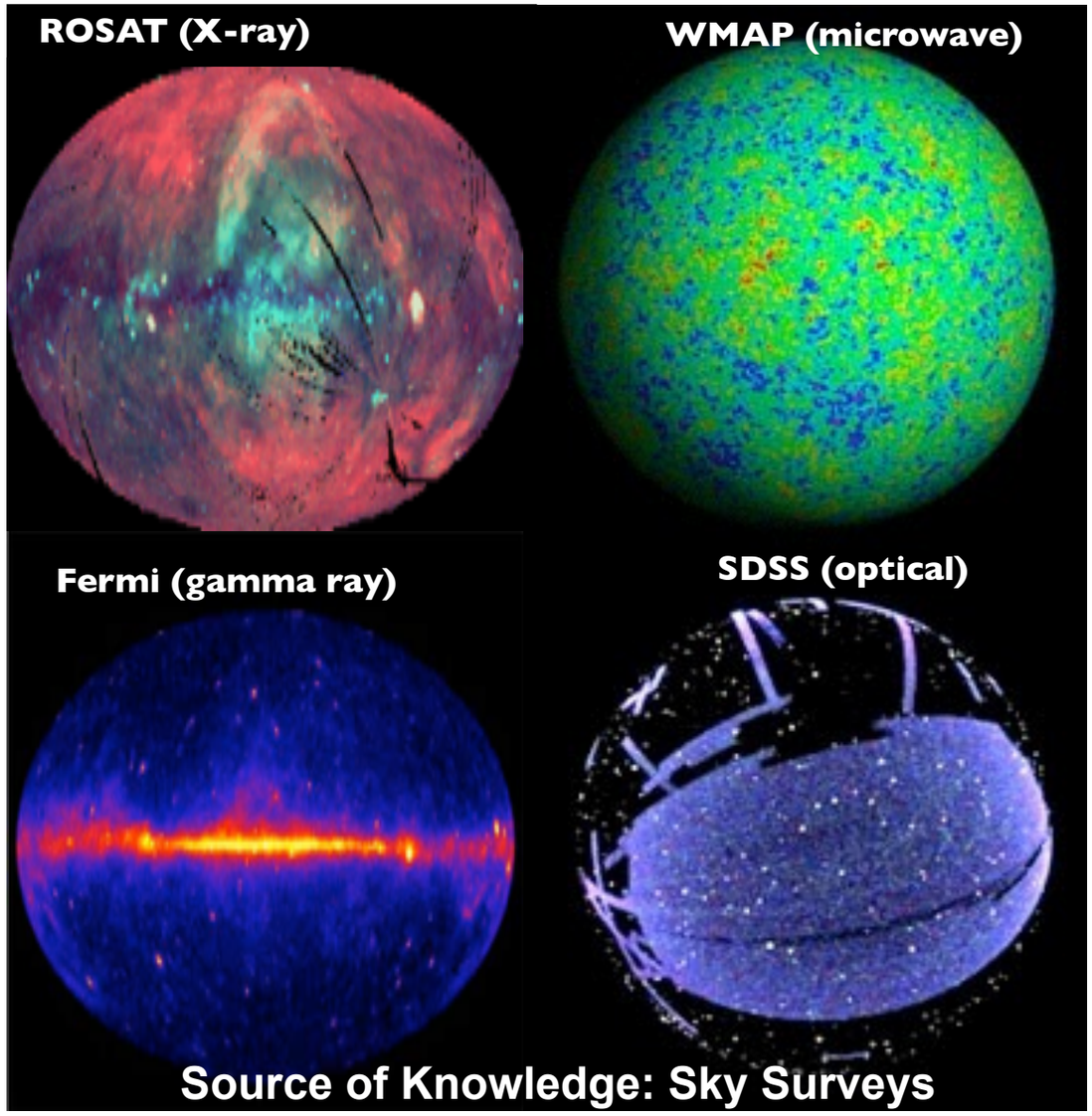




Emulation and Cosmic Calibration: People

Science at the Cosmic Frontier: 'Precision' Cosmology

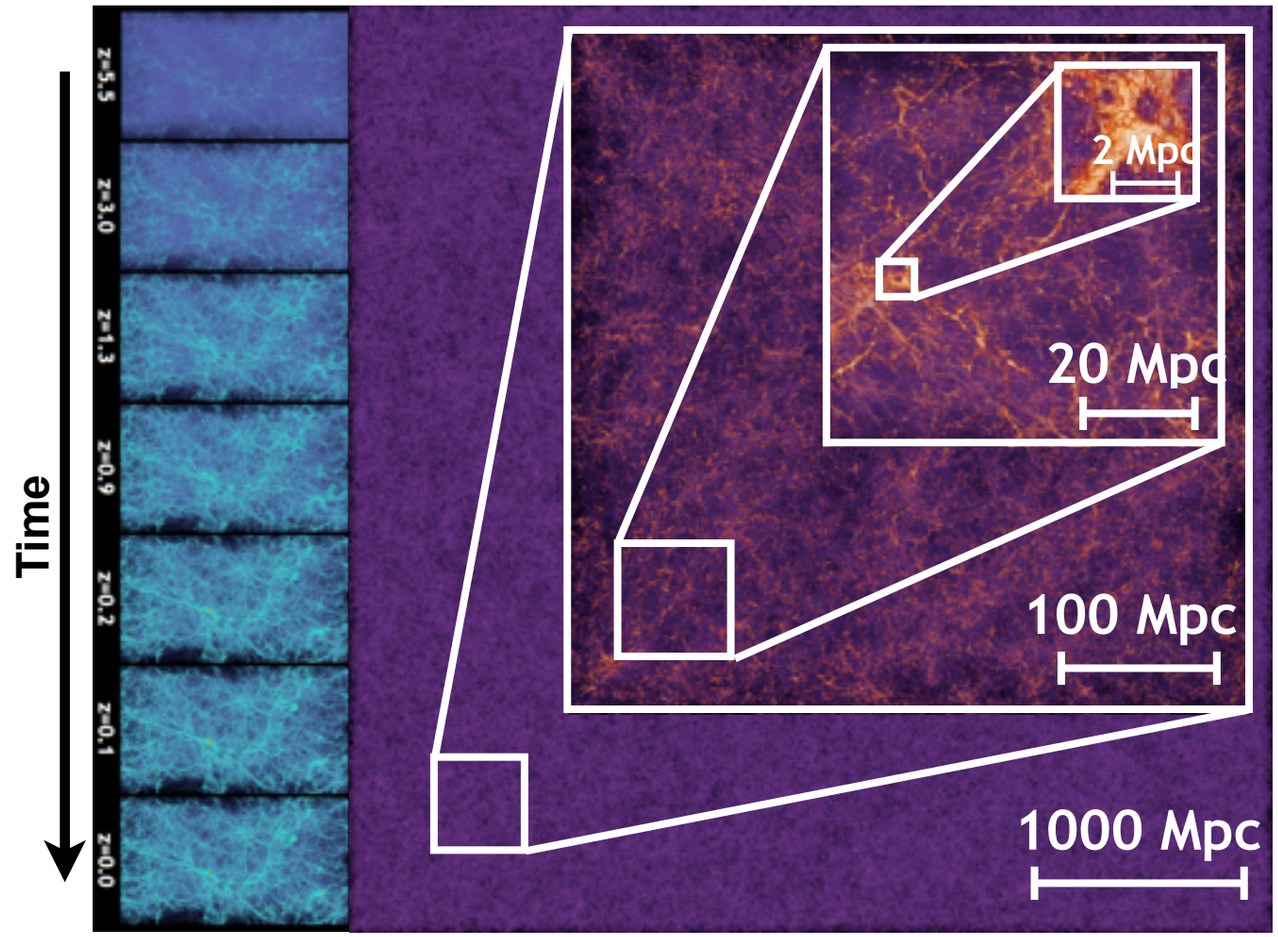
- **Instrumentation Advances:** Wide/deep sky coverage, multi-wavelength observations no longer statistics-limited (near-term target, ~1% errors on cosmological parameters or better)
- **Precision Cosmology Science:**
  - Nature of cosmic acceleration
  - Nature and interactions of dark matter
  - Primordial fluctuations and tests of inflation
  - Probes of fundamental physics (neutrino sector)
  - Physics of structure formation
- **Theory Advances:** Predictive theory and modeling becoming crucially important **('theory gap')**
- **Computational Cosmology:**
  - Basic predictions ('forward modeling')
  - Cosmological inference ('inverse problems')
  - End-to-end modeling (system-scale control of systematics)
  - Large-scale computing requirements (as for lattice QCD)



Emulation strategies work here

Large Scale Structure Simulation Requirements

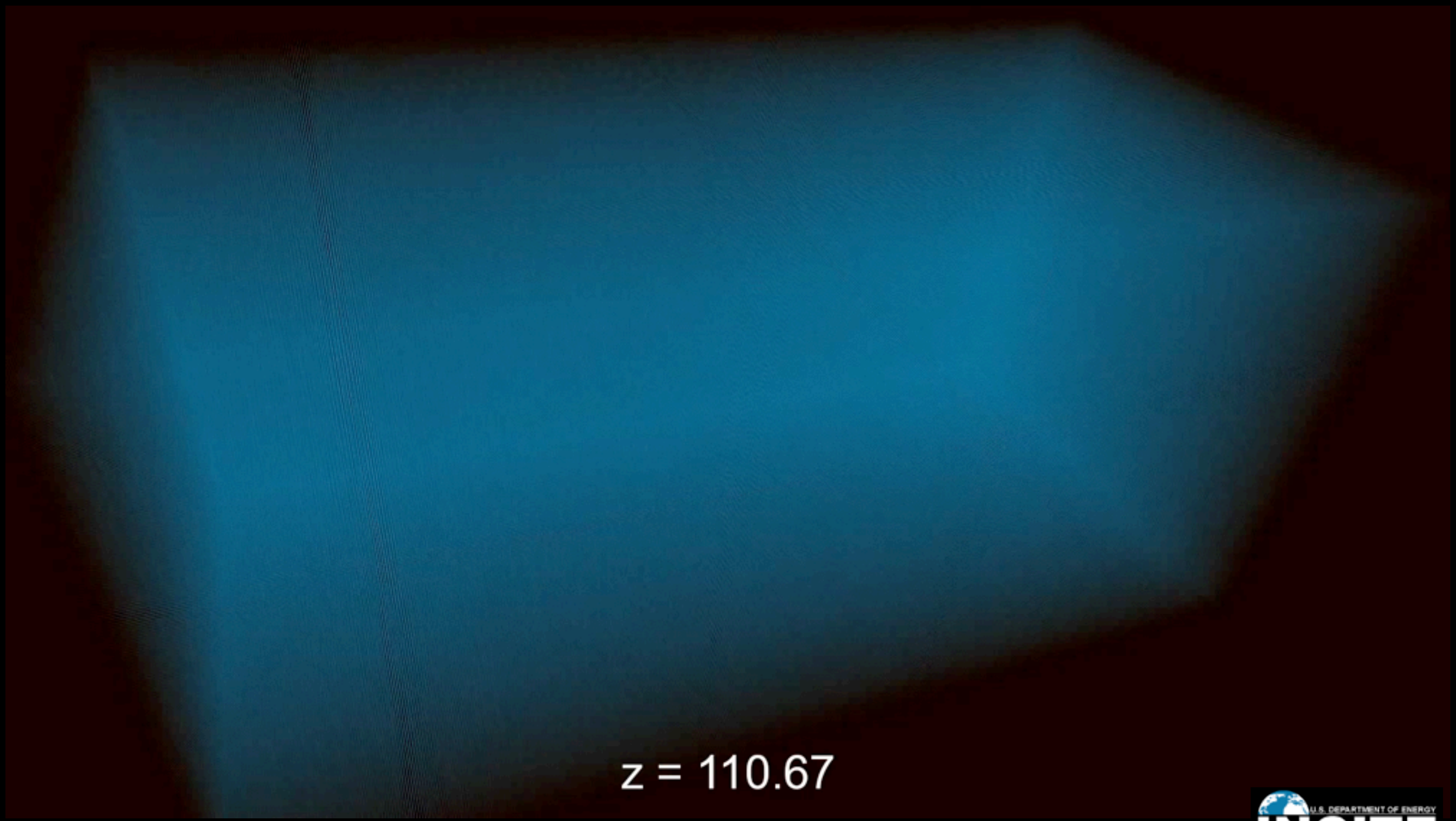
- **Force and Mass Resolution:**
  - Galaxy halos ~100kpc, hence force resolution has to be ~kpc; with Gpc box-sizes, a dynamic range of a million to one
  - Ratio of largest object mass to lightest is ~10000:1
- **Physics:**
  - Gravity dominates at scales greater than ~0.1 Mpc
  - Small scales: galaxy modeling, semi-analytic methods to incorporate gas physics/feedback/star formation
- **Computing 'Boundary Conditions':**
  - Total memory in the PB+ class
  - Performance in the 10 PFlops+ class
  - Wall-clock of ~days/week, in situ analysis



Gravitational Jeans Instability: 'Outer Rim' run with 1.1 trillion particles

Key motivation for HACC (Hardware/ Hybrid Accelerated Cosmology Code): Can the Universe be run as a short computational 'experiment'?

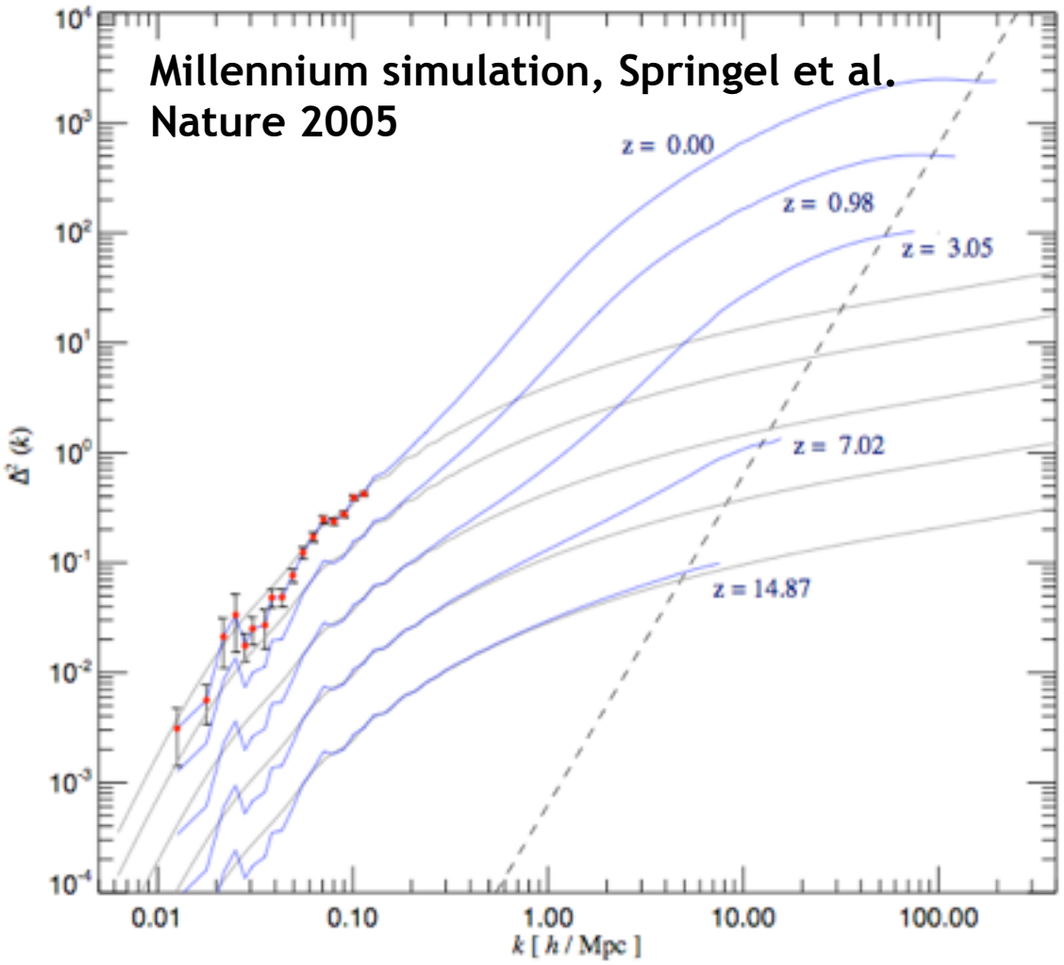
The high resolution Q Continuum Simulation, finished July 13 on ~90% of Titan under INCITE, evolving more than half a trillion particles. Shown is the output from one node (~33 million particles), 1/16384 of the full simulation

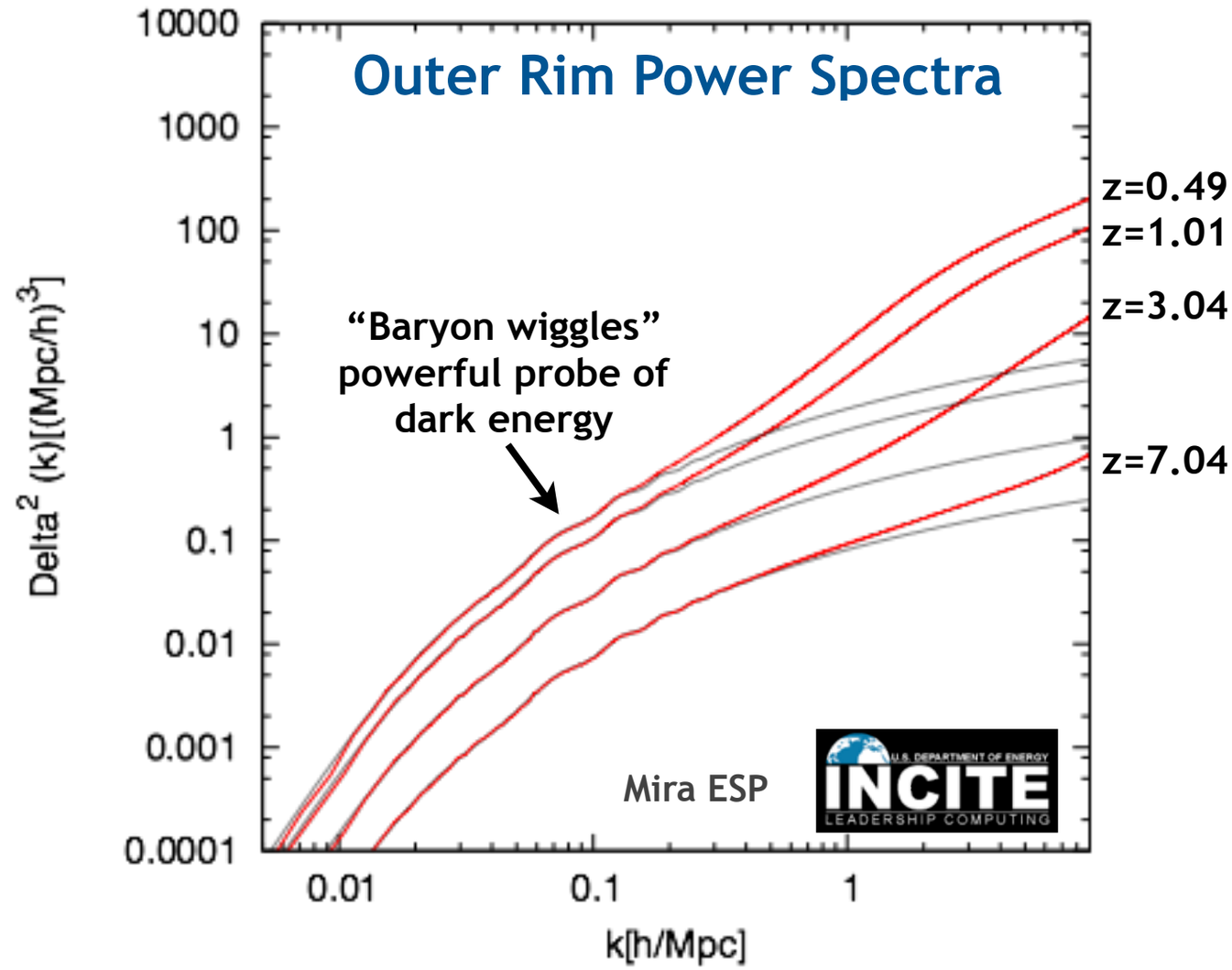


Q Continuum: Extradimensional plane of existence

Visualization: Silvio Rizzi, Joe Insley et. al., Argonne

Cosmology with HACC: Exquisite Statistics





- Millennium simulation and Outer Rim run are very similar in force and mass resolution, but volume larger by a factor of 216
- Exceptional statistics at high resolution enable many science projects

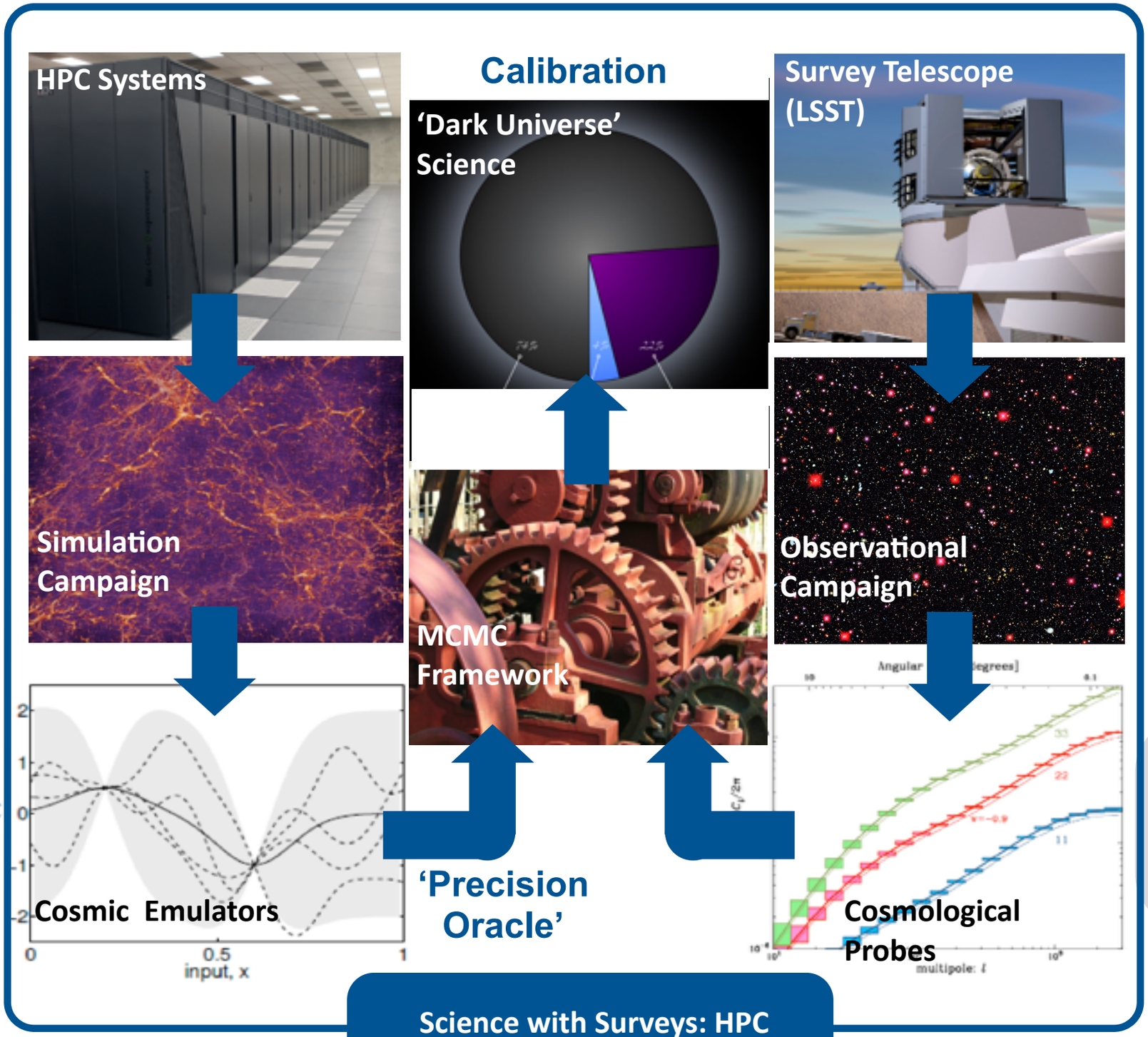
Precision Cosmology: 'Big Data' Meets Supercomputing

Supercomputer Simulation Campaign

Major stats + ML+ sampling + optimization collaboration

Simulations + CCF

Emulator based on Gaussian Process Interpolation in High-Dimensional Spaces



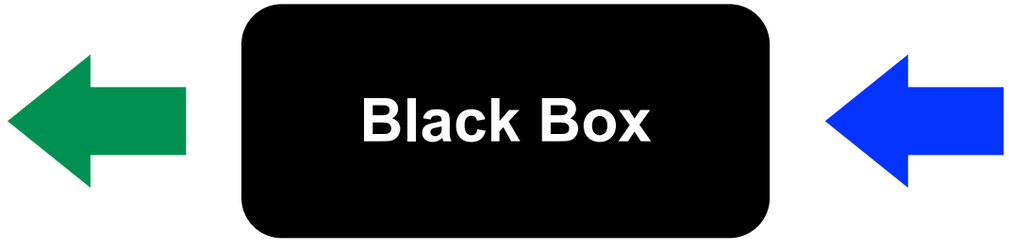
Mapping the Sky with Survey Instruments

Observations: Statistical error bars will 'disappear' soon!

CCF= Cosmic Calibration Framework (2006)

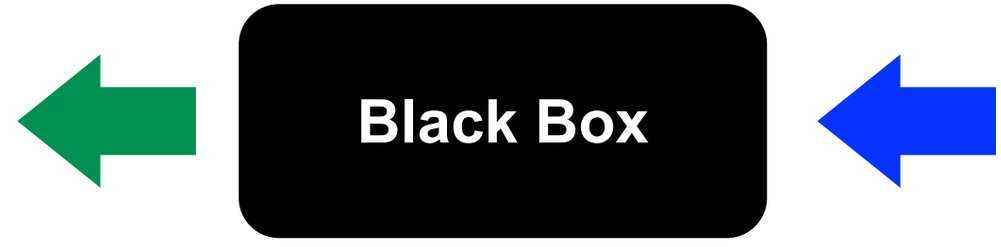
Precision Cosmology: Statistics + Machine Learning

Old School Stats: Limited Data
('benign' black box,
controlled environment)



Find the (simple)
'data model'

Machine Learning: 'Big Data'
('vicious' black box,
potentially uncontrolled environment)



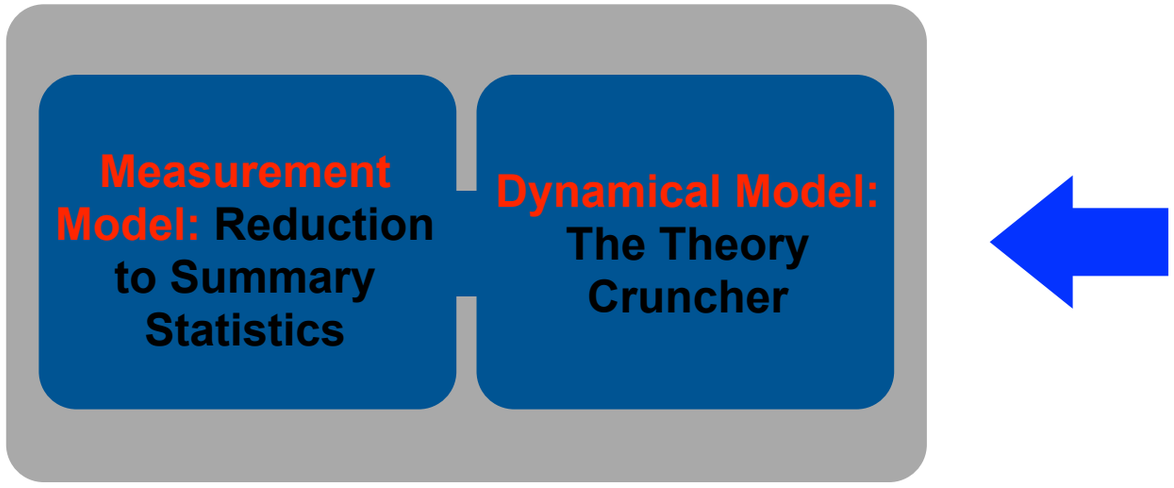
Find the (arbitrary)
'dynamical map'

Both (cartoon) approaches useful but by no means exhaustive in terms of scientific value--
with a good theory (forward model) and sufficient computing the box is not black!
We are not 'model-free' --

Statistics meets Machine Learning in the Supercomputing/Big Data Environment

Modeling cycle: Predict, (Cross) Validate, Optimize

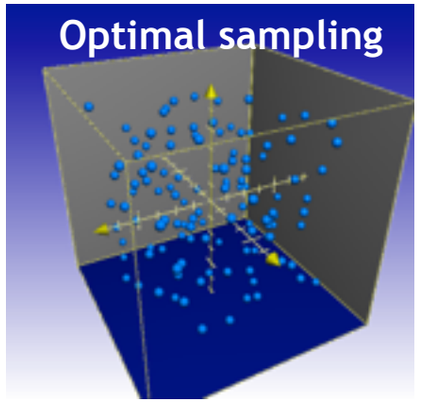
Summary Statistics
(restricted to
what can be
measured)

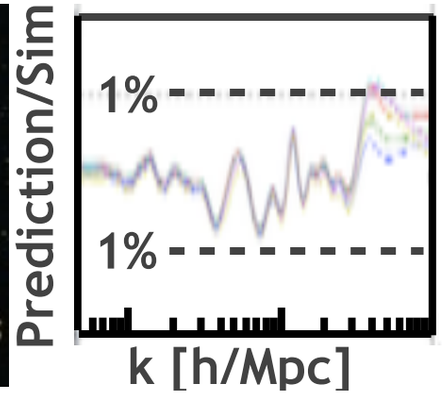


Theory and Modeling Inputs
(prior ranges)

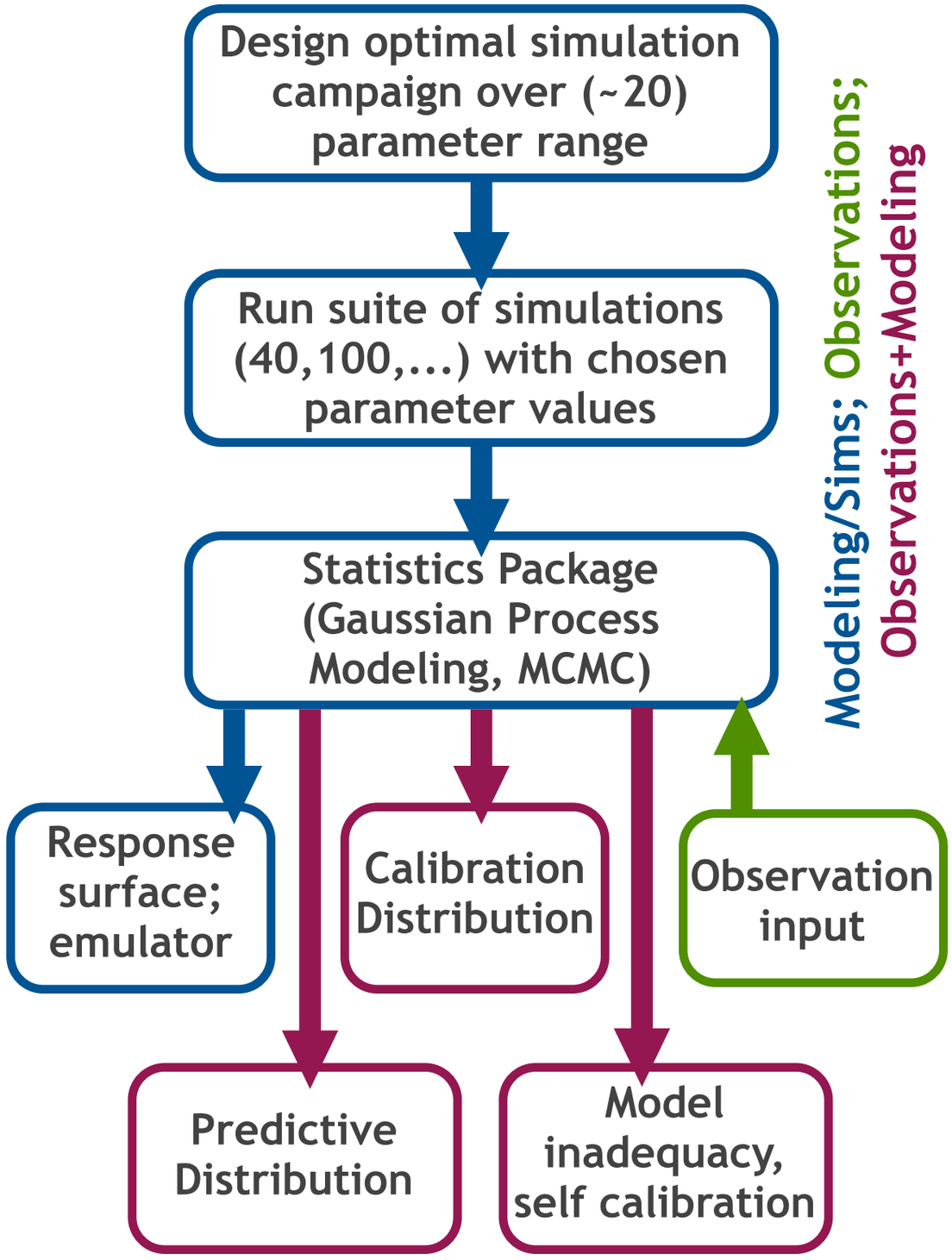
Cosmic Calibration: Solving the Inverse Problem

- **Challenge:** To extract cosmological constraints from observations in non-linear regime, need to run Markov Chain Monte Carlo code; input: 10,000 - 100,000 different models
- **Current strategy:** Fitting functions for e.g. P(k), accurate at 10% level, not good enough!
- **Brute force:** Simulations, ~30 years on 2000 processor cluster...
- **Only alternative:** emulators

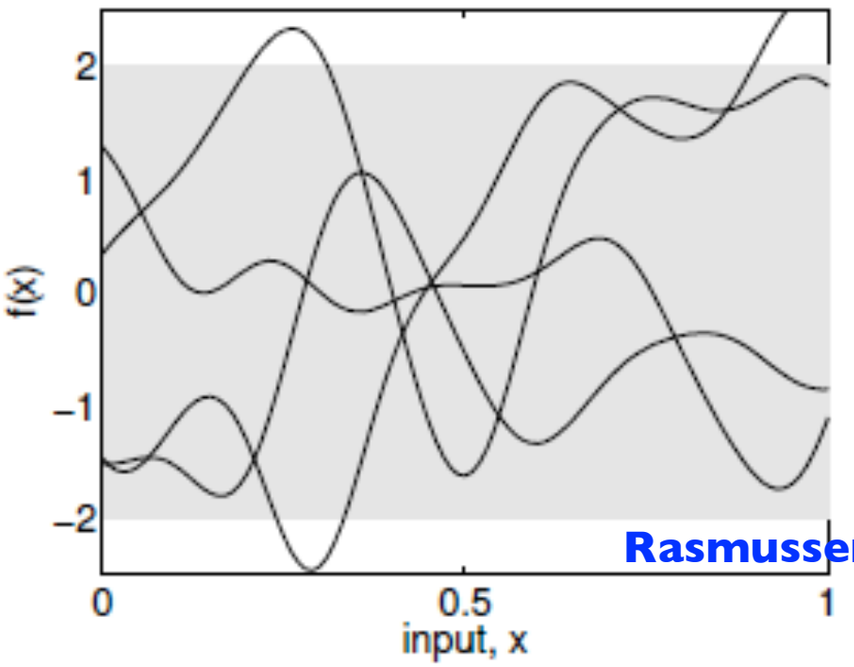




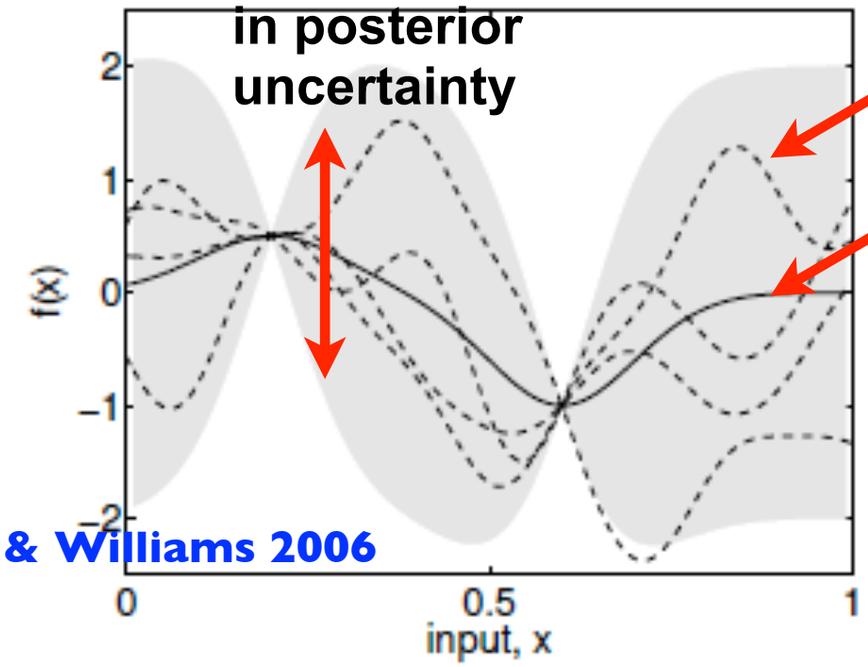
Heitmann et al. 2006, Habib et al. 2007



Gaussian Process Modeling



Prior **distribution over random functions: global mean zero (although individual choices clearly are not mean-zero), variance assumed to be independent of x, 2-SD band in gray**



Rasmussen & Williams 2006

Posterior **distribution conditioned on exact information at two x points, consider only those functions from the prior distribution that pass through these points**

Acceptable functions

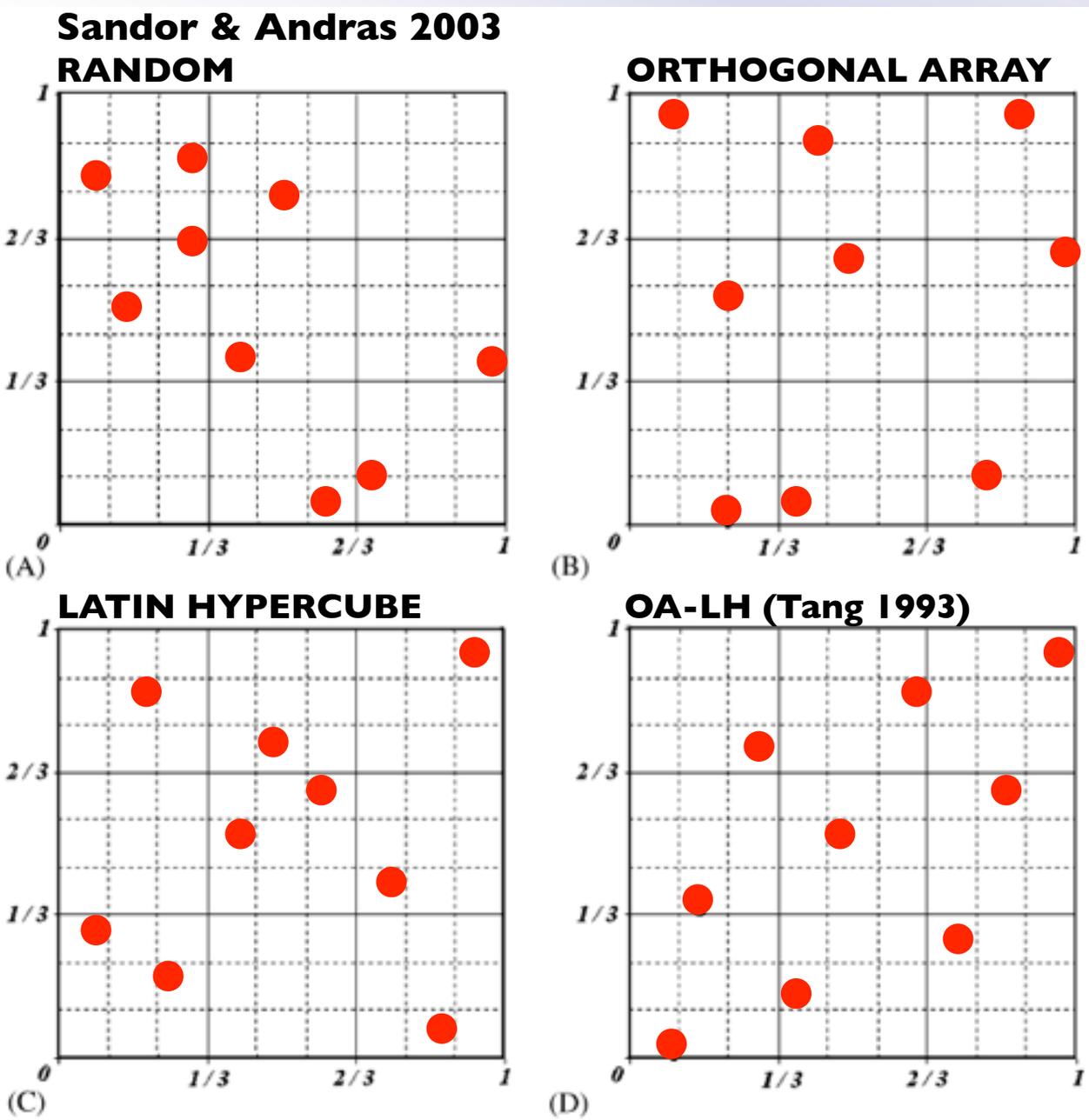
Mean value function (not mean zero!)

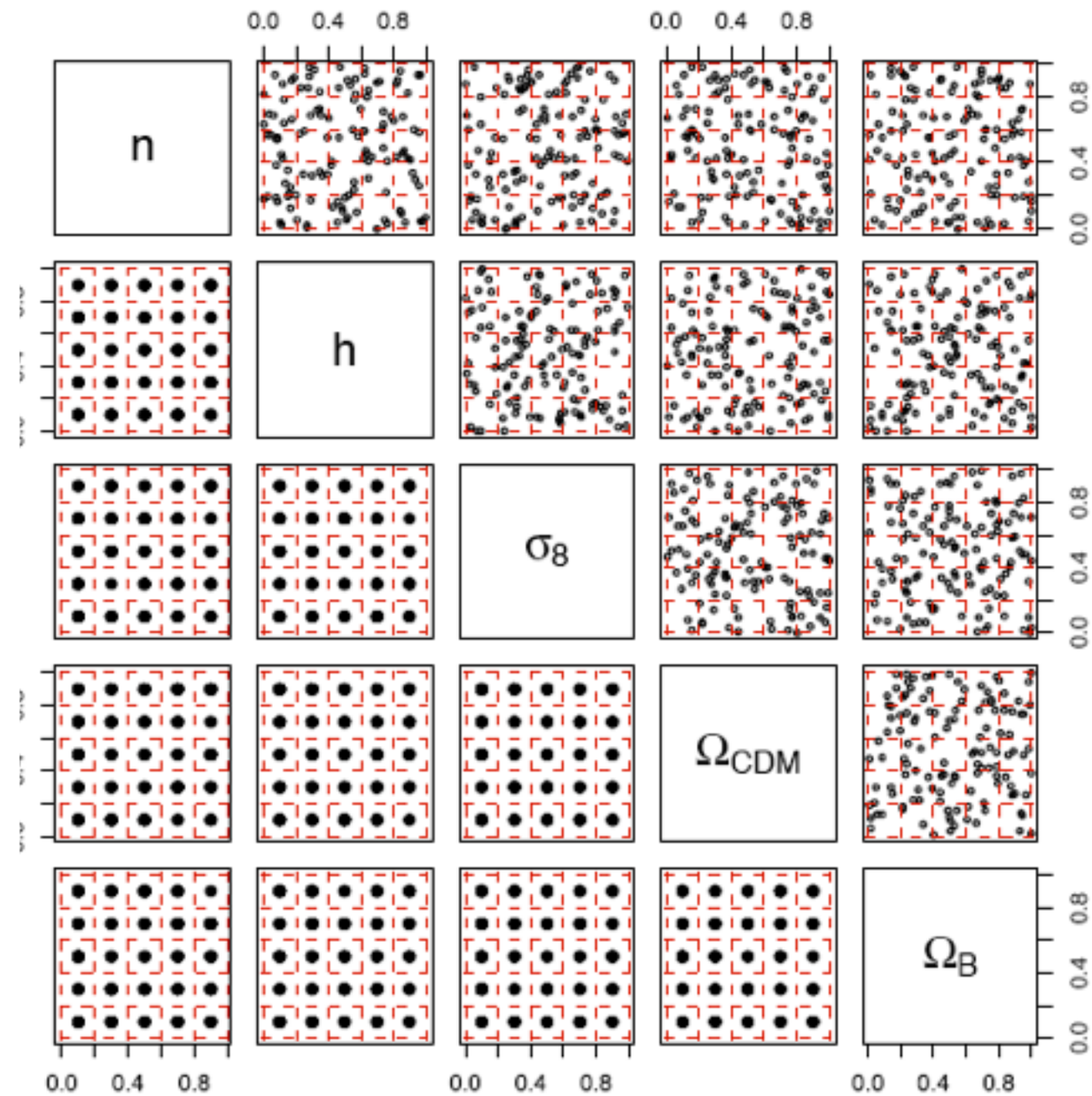
Avoid overfitting by using priors on hyperparameters and by controlling the learning process

- GPs are *nonparametric*, so there is no need to worry if the functions can fit the data (e.g., linear functions against nonlinear data), even with many observations still have plenty of candidate functions

- With GP models, the choice of prior distribution over random functions is essentially a statement of the properties of the initial covariance function, these properties can be specified in terms of a set of *hyperparameters*, using data to determine these defines the *learning* problem for the GP approach

Sampling

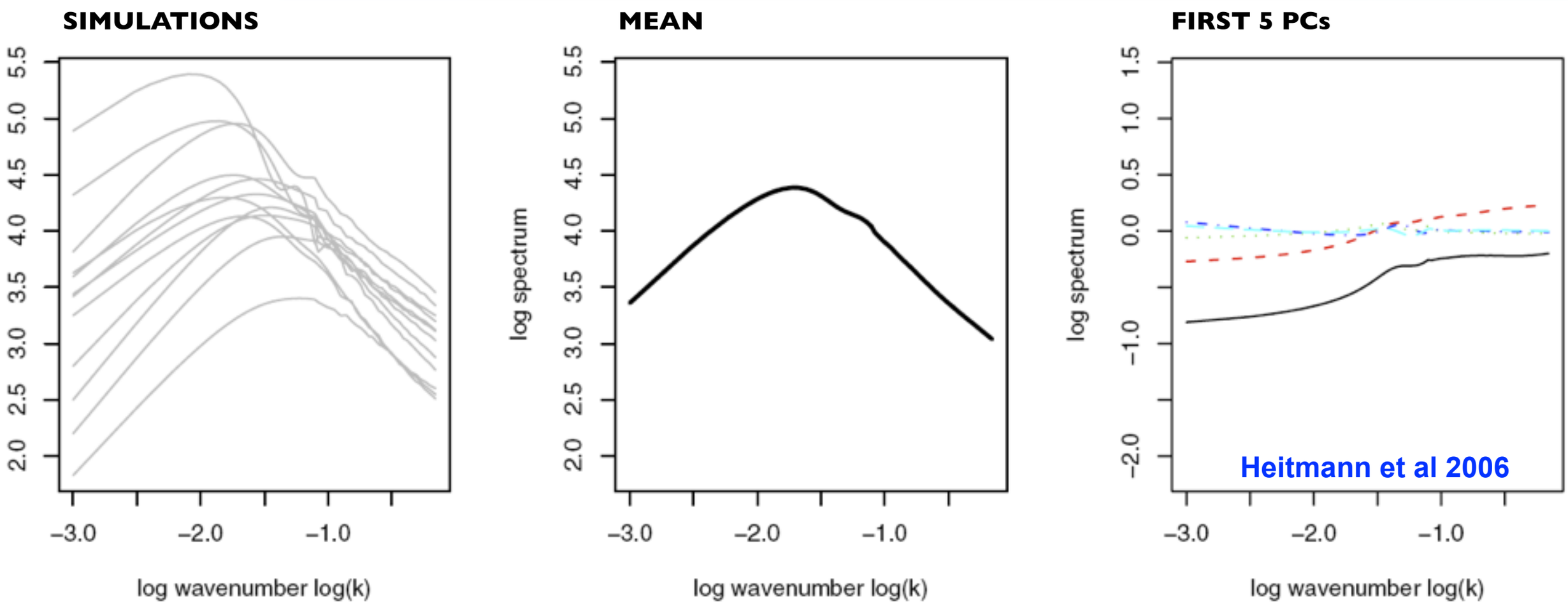




Strive for “equidistribution” property over the sampling space, best approach when ignorant of functional variation, well-suited to GPs.

Practical 128 point, 5-level, strength-2-based design [level=#variable slices, strength=(lower) dimension to be sampled, #columns=#variables, #rows=#trials]

Data Compression: PCA Basis



Mean-adjusted Principal Component Representation

$$\eta(k;\theta) = \sum_{i=1}^{p_\eta} \phi_i(k) w_i(\theta) + \epsilon, \qquad \theta \in [0,1]^{p_\theta}$$

COSMOLOGICAL/MODELING PARAMETERS

PC BASIS FUNCTIONS

GP WEIGHTS

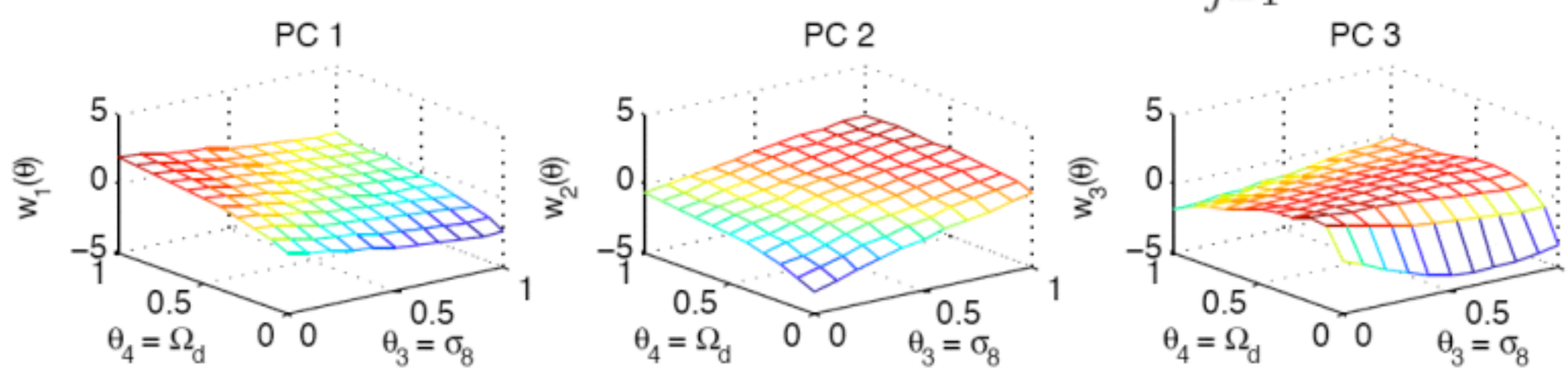
STANDARDIZED PARAMETER DOMAIN

GP Predictions at new Settings

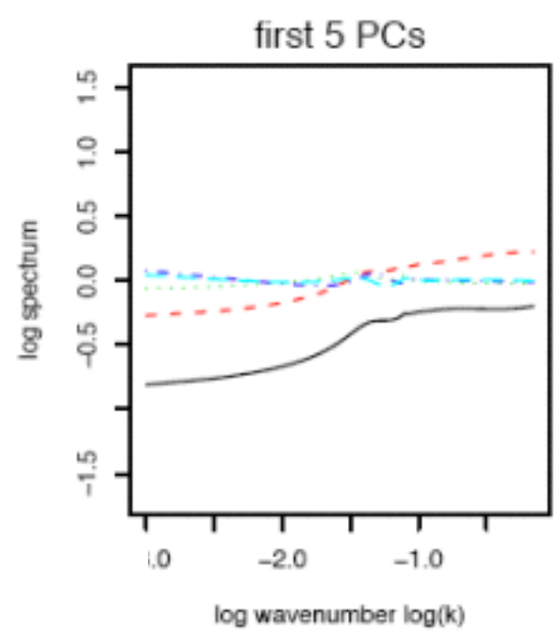
Gaussian process (GP) models are used to estimate the weights $w_j(\theta)$ at untried settings

$$\widehat{\eta}(\theta; k) = \sum_{j=1}^{p_\eta} w_j(\theta)\phi_j(k)$$

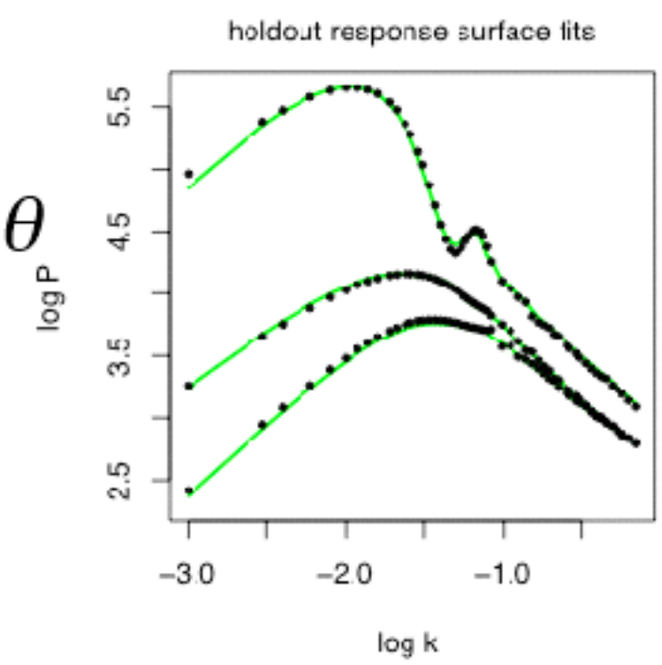
GP models $w_j(\theta)$



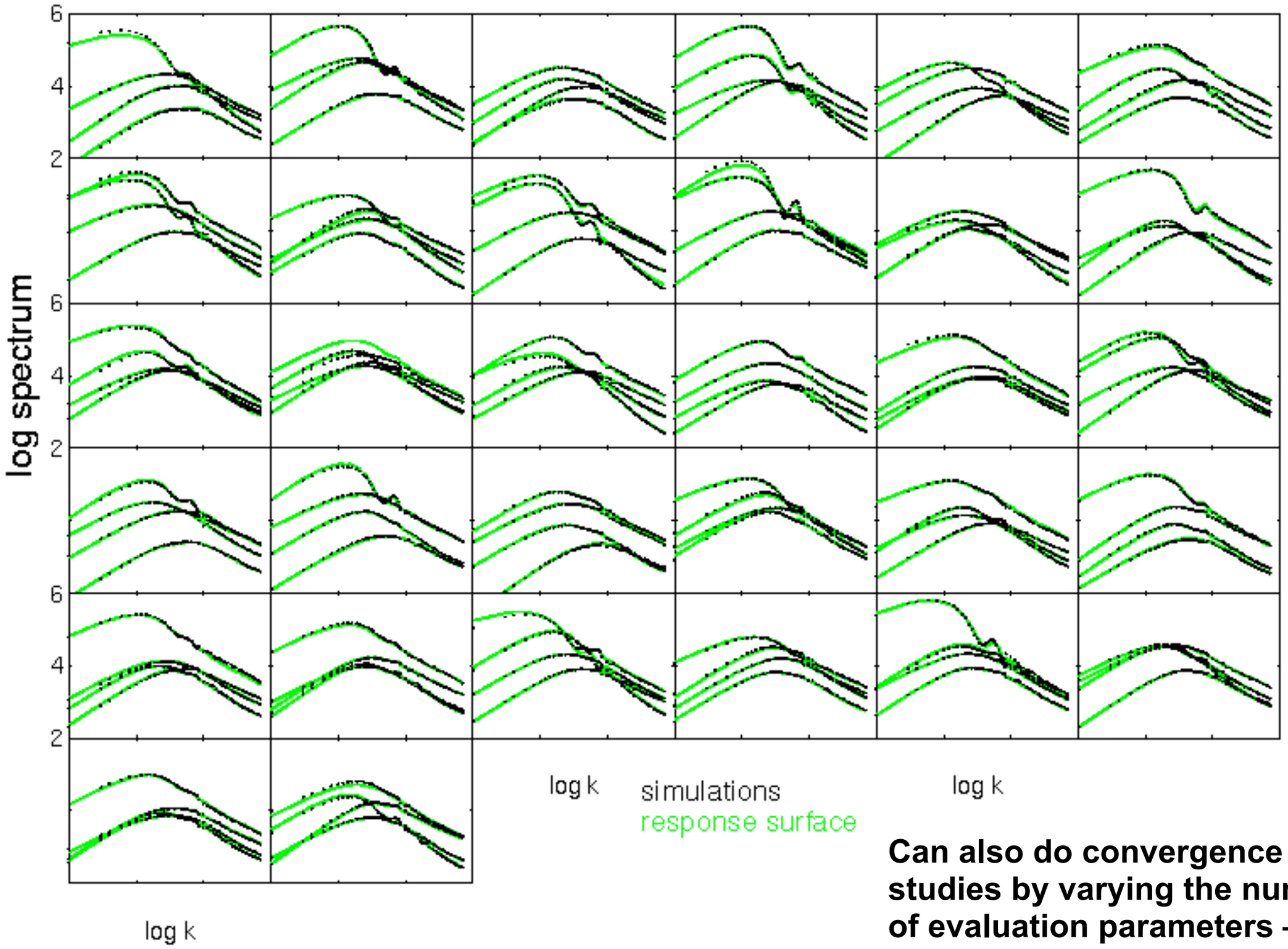
Bases $\phi_j(\theta)$



Prediction at new θ



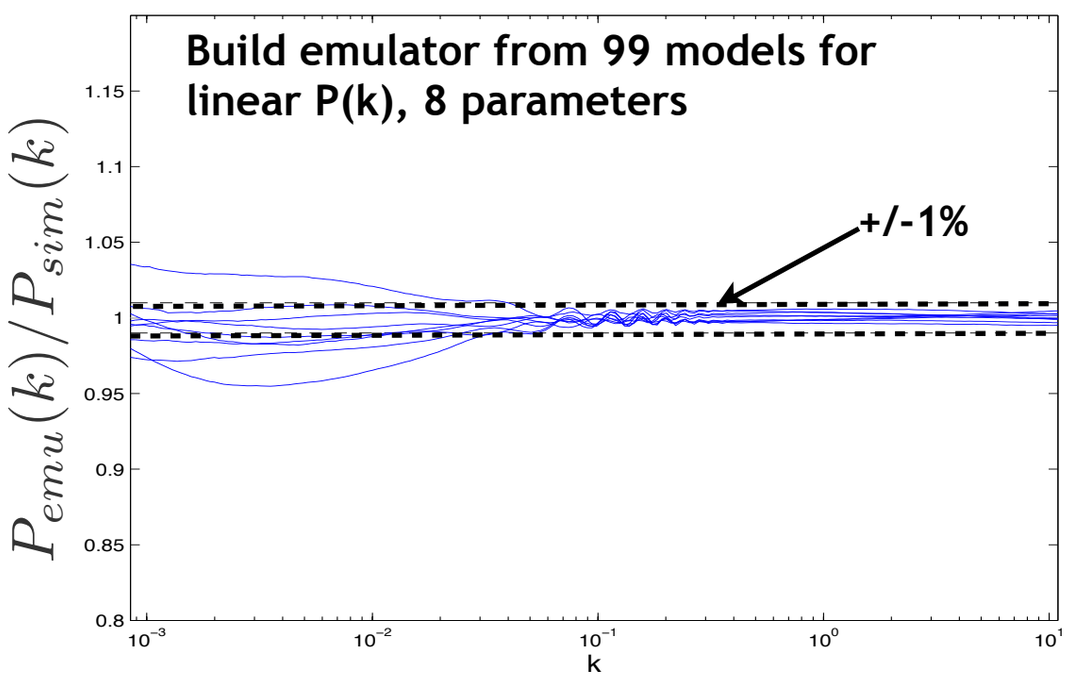
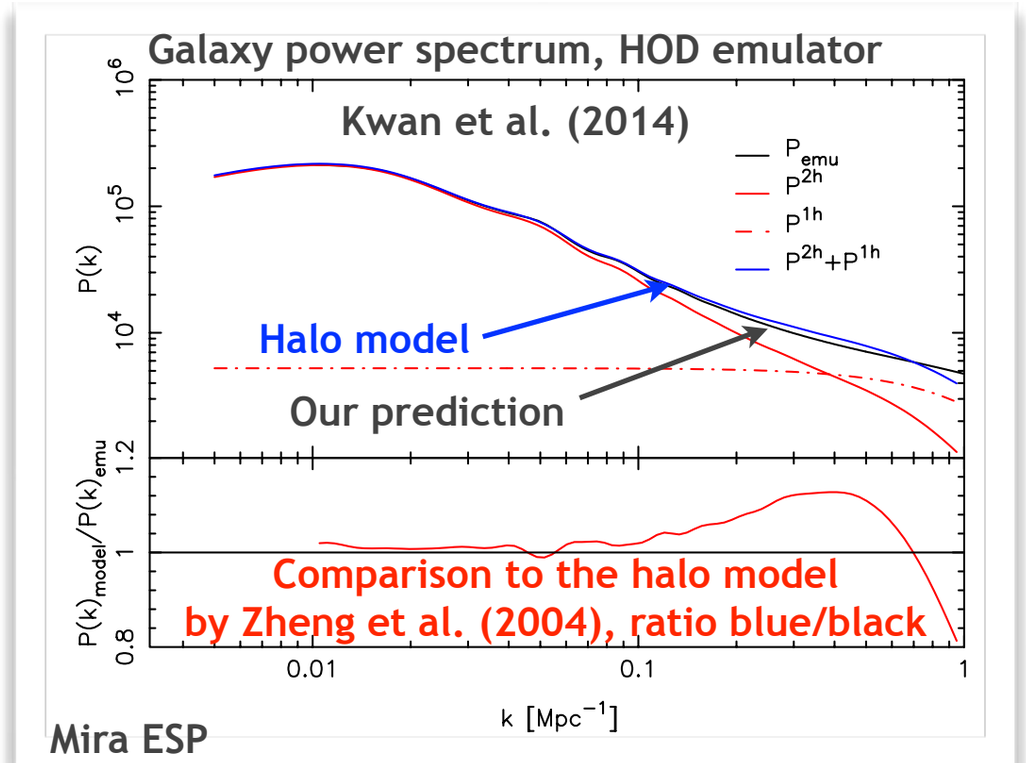
Testing: Hold-Outs



Can also do convergence studies by varying the number of evaluation parameters —

Cosmic Emulator in Action

- Instantaneous ‘oracle’ for nonlinear power spectrum, reduces compute time from weeks to sub-seconds, accurate at 1% out to k~1/Mpc, 3% out to k~10/Mpc
- Enables direct MCMC with results from full simulations for the first time

Next Steps —

- **Discrepancy Modeling:** What happens if your forward model isn't correct?
- **Nested/Adaptive Sampling:** Convergent/Learning approach to emulation
- **Covariance Emulation:** Emulate covariances rather than just the mean
- **Accuracy Limits:** Theory for convergence (a posteriori so far)
- **Limits of Dimensionality:** How high can we go?