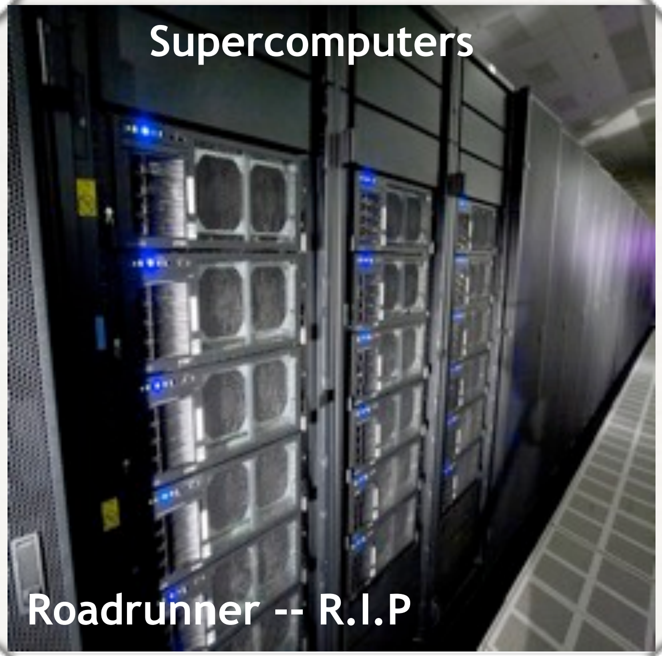
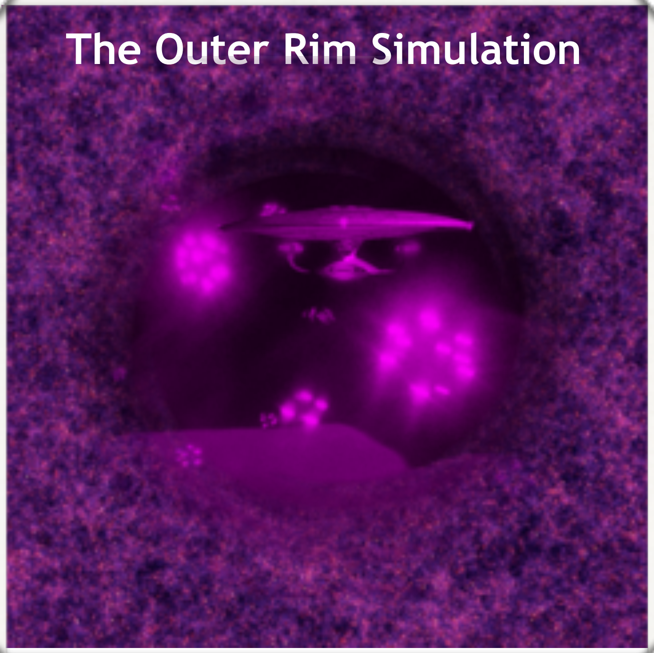
Cosmic Calibration -- Or How I Learned to Stop Worrying and Love Supercomputers

Katrin Heitmann

KICP Colloquium, November 6, 2013

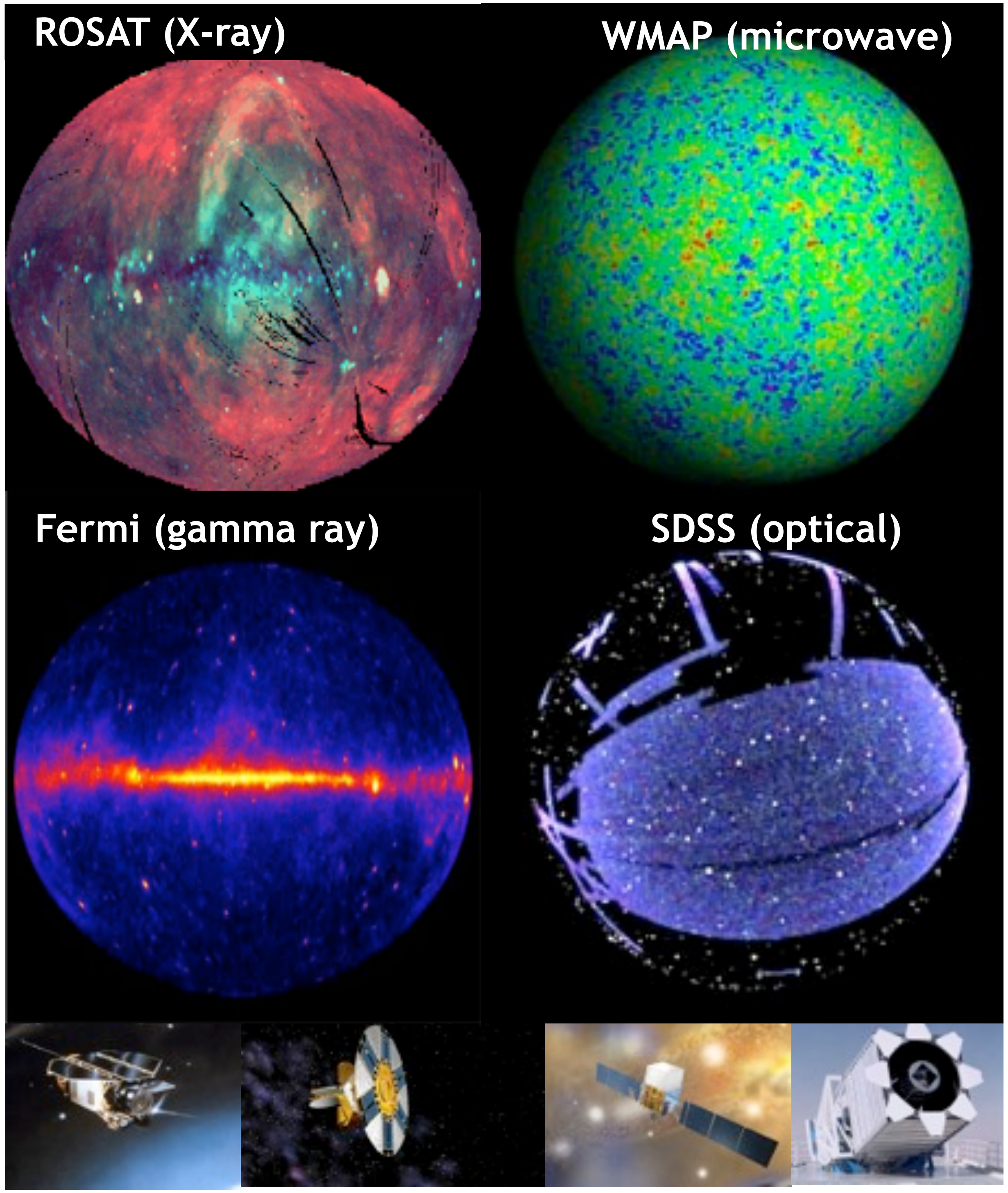




Thanks to many collaborators!

Modern Cosmology and Sky Maps

- Modern cosmology is the story of mapping the sky in multiple wavebands
- Maps cover measurements of objects (stars, galaxies) and fields (temperature)
- Maps can be large (Sloan Digital Sky Survey has~200 million galaxies, many billions for planned surveys)
- Statistical analysis of sky maps
- All precision cosmological analyses constitute a statistical inverse problem: from sky maps to scientific inference
- Therefore: *No* cosmology without (large-scale) computing



Outline

- **What have we learned so far from these observations?**
  - Content of the Universe
  - Evolution of the Universe
- **How do we extract this knowledge from the data?**
  - Focus here on optical surveys and simulations
- **Precision Cosmology:** Where do we want to go next and how do we get there?
- **Simulating cosmological surveys with HACC (Hardware/ Hybrid Accelerated Cosmology Code):** An N-body code designed for extreme scaling
- **Cosmic Calibration:** Building efficient prediction tools from expensive simulations

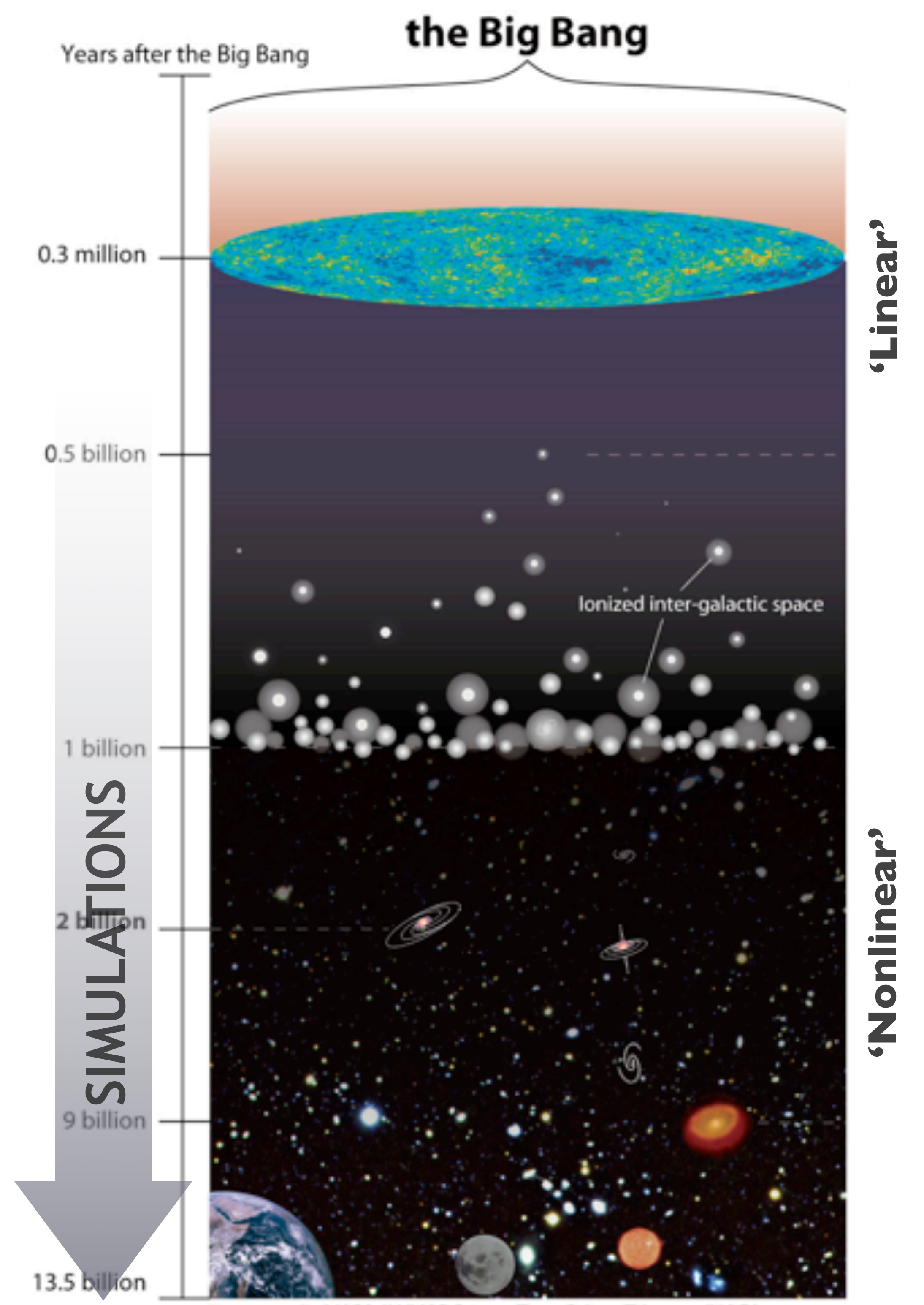
The Content of the Universe: It's dark!

- **Dark Energy:** Multiple observations show that the expansion of the Universe is accelerating (first in 1998, Nobel prize 2011)
- 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
- Independent of what we find, we will learn new, fundamental physics, this is not just the hunt for a couple numbers!
- **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.

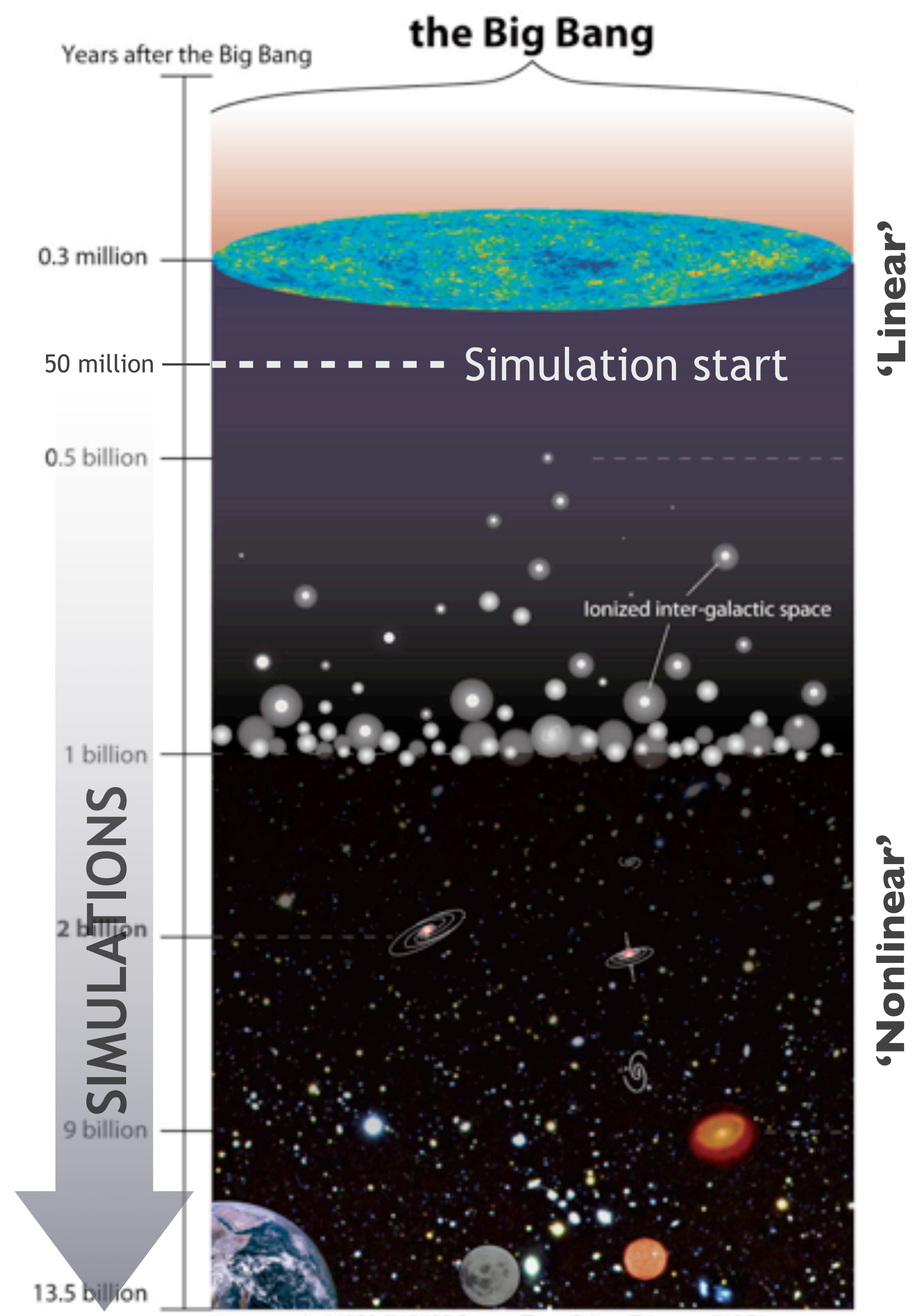
Structure Formation: The Basic Paradigm

- **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

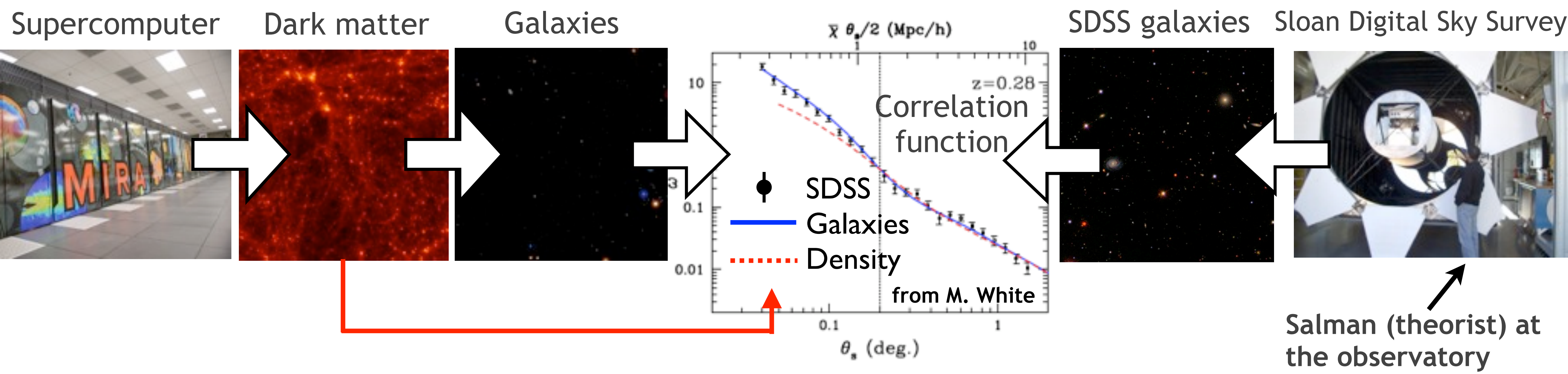


Structure Formation: The Basic Paradigm

- **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

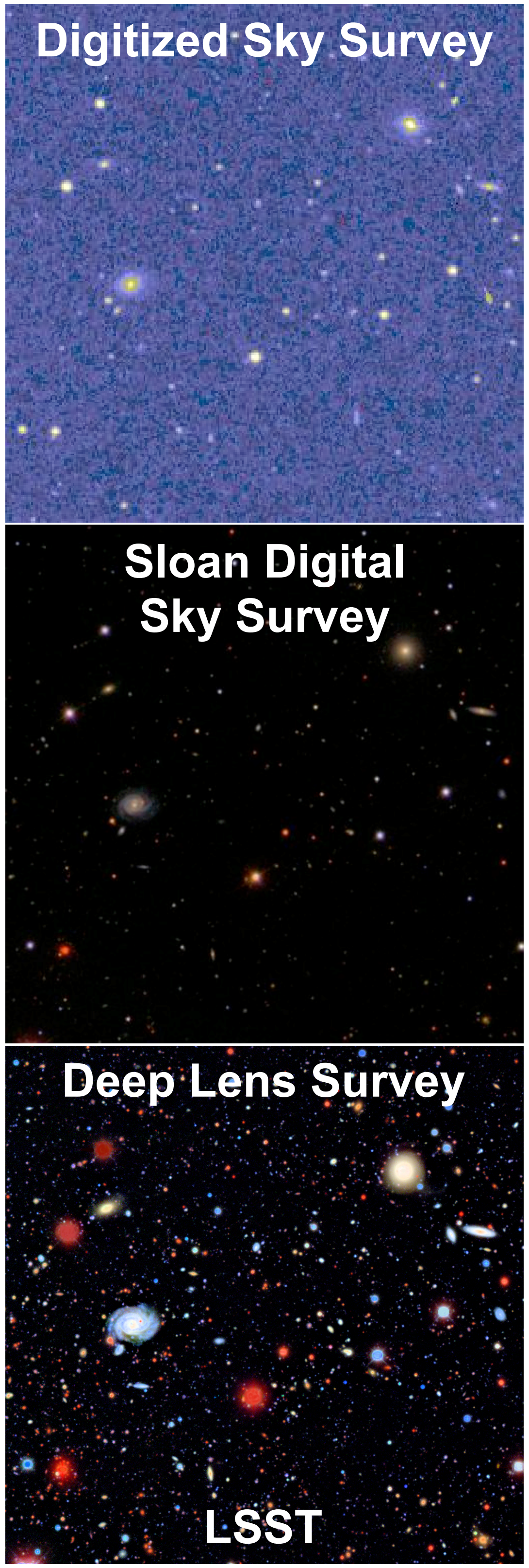


Connecting Theory and Observations

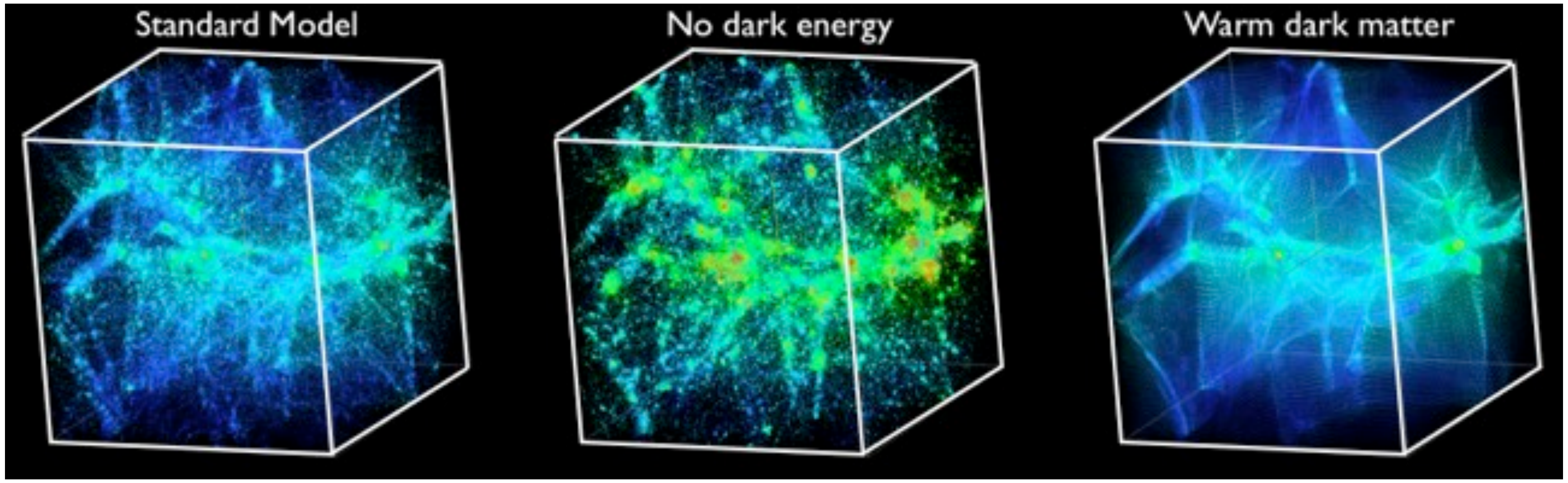


- Simulate the formation of the large scale structure of the Universe via dark matter tracer particles
- Take dark energy into account in the expansion history
- Measure the high-density peaks (dark matter halos) in the mass distribution
- "Light traces mass" to first approximation, therefore populate the halos with galaxies, number of galaxies depends on mass of halo (constraints from observations)
- Galaxy population prescription (hopefully) independent of cosmological model

Precision Cosmology: "Inverting" the 3-D Sky



- **Standard Model of Cosmology:** Verified at the 5-10% level across multiple observations, describes make-up and evolution of the Universe
- **Next generation observatories:** aim to push the current boundaries by orders of magnitude
- Scales that are resolved by future surveys become smaller and smaller, demanding (i) ever larger simulations with increased mass and force resolution; (ii) more detailed physics
- **Next frontier:** Nonlinear regime of structure formation
- **Future Targets:** Aim to control survey measurements to the ~1% level, **can theory and simulation keep up?**



Why do we need higher accuracy measurements?

Why do we need hig nts?

It's the f...... Universe, guys!
It deserves at least two
decimal places!

**Douglas Scott, UBC
at the Santa Fe Cosmology
Workshop in 2005**

Why do we need hig… …nts?

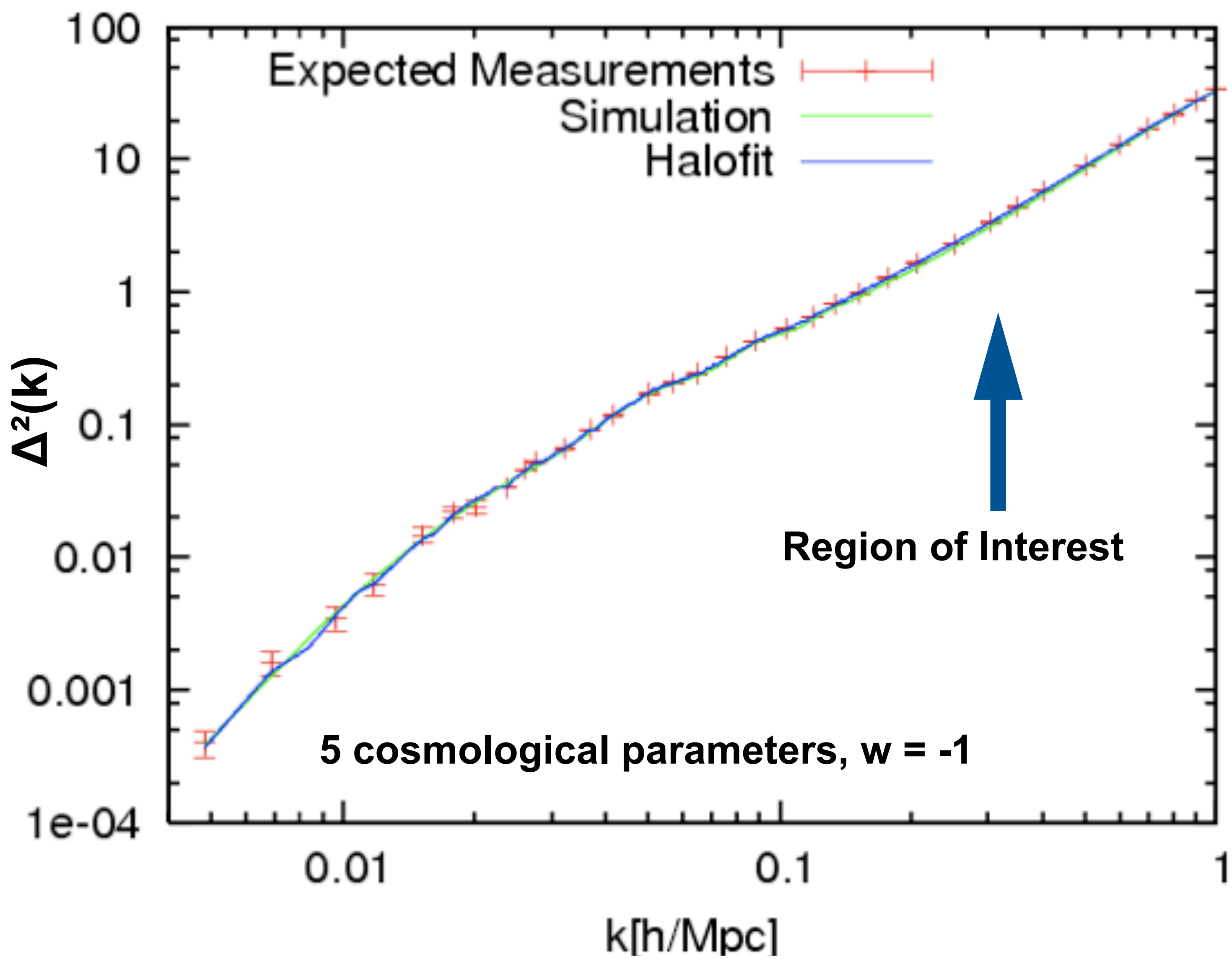
It's the f...... Universe, guys!
It deserves at least two
decimal places!

**Douglas Scott, UBC
at the Santa Fe Cosmology
Workshop in 2005**

- **Convincing argument! In addition:**
- **Fundamental physics questions await answers:**
  - Modified gravity or dark energy?
    - ★ Measure growth of structure and expansion history of the Universe
  - If dark energy: Cosmological constant or dynamical origin?
  - If modified gravity: How do structures form?

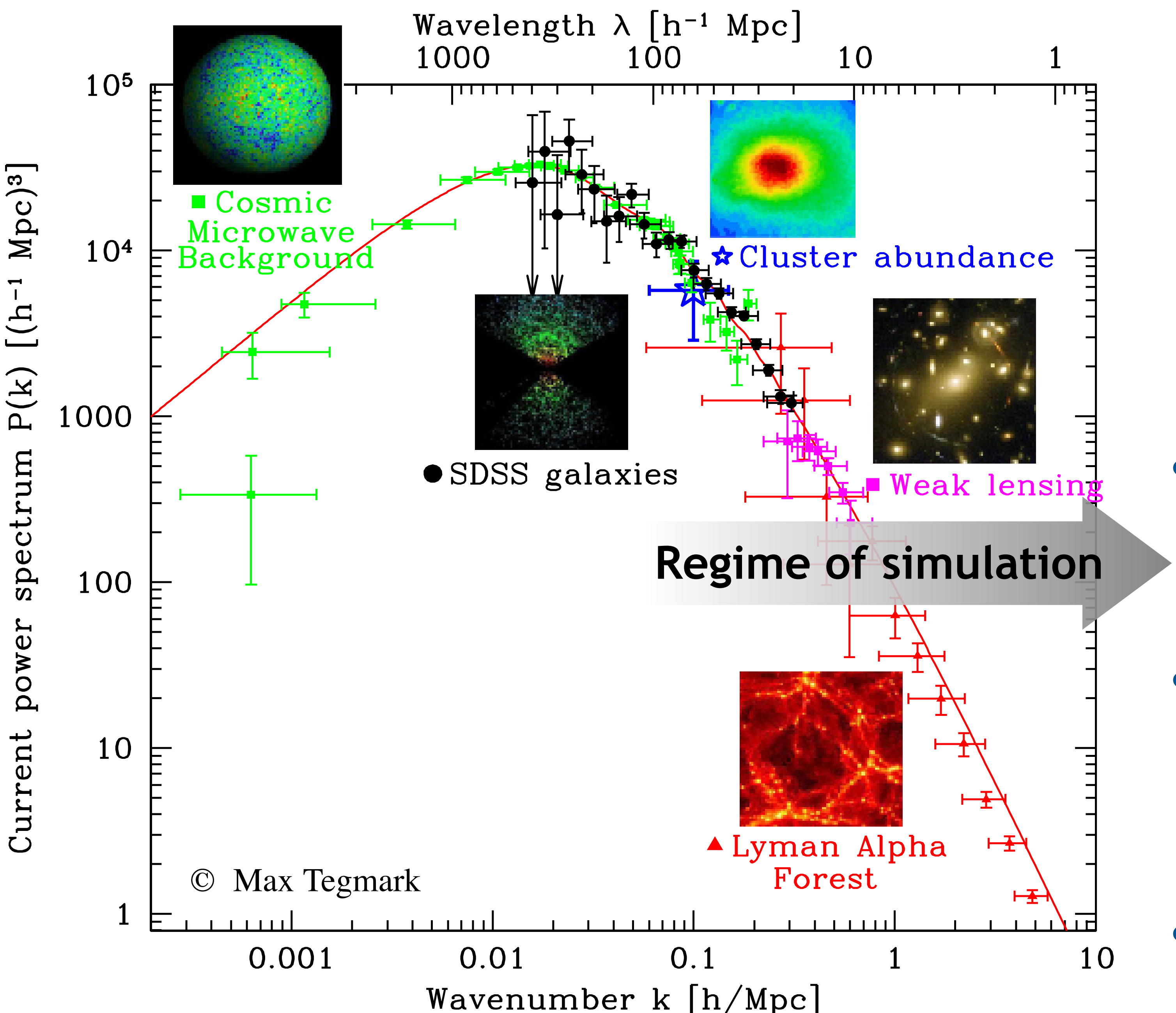
Why do we need higher accuracy in our predictions?

- Example in this talk: matter power spectrum
- Question: how badly will our constraints on dark energy be biased if we *do not* reach the same accuracy in our modeling as we might have in our data?
- Generate mock data set with the expected 1% error
- Analyze data with current method using HaloFit to model the matter power spectrum
  - HaloFit (Smith et al. 2003): semi-analytic fit for the power spectrum, based on modeling approach and tuned to simulations, accurate at the 5-10% level



$$\Delta^2(k) = \frac{k^3 P(k)}{2\pi^2};\ P(\vec{k}) = \langle \delta^2(\vec{k}) \rangle$$

The Matter Power Spectrum



Length scale of interest:
1 parsec (pc)=3.26 light years~$3\cdot10^{13}$km,
separation of stars in a galaxies
Mpc=10^6pc: ~ separation of bright galaxies

2-point correlation function:

$$\xi(\vec{x}) = \int \frac{d^3\vec{y}}{V}\delta(\vec{y}-\vec{x})\delta(\vec{y}) = \int \frac{d^3\vec{k}}{(2\pi)^3 V}\underbrace{|\delta_k|^2}_{\text{power spectrum}} e^{i\vec{k}\cdot\vec{x}}$$

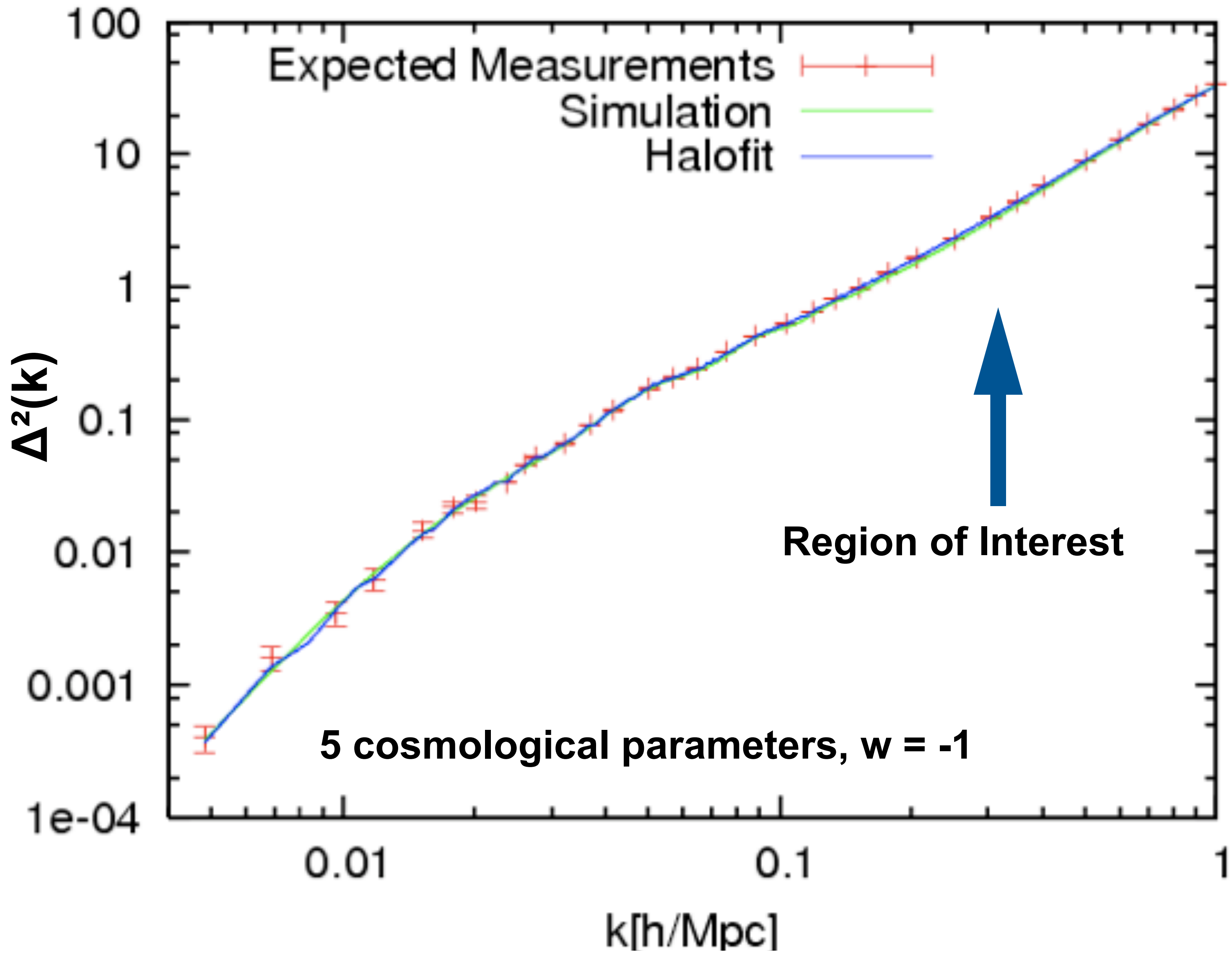
- 2-point correlation function: excess probability of finding an object pair separated by a distance r_{12} compared to that of a random distribution
- P(k): power spectrum, Fourier transform of correlation function

$$\Delta^2(k) = \frac{k^3 P(k)}{2\pi^2}$$

- Power spectrum very sensitive to physics of interest: amount and properties of dark matter, dark energy, neutrino mass, ...
- Many different probes for measuring P(k)

Why do we need higher accuracy in our predictions?

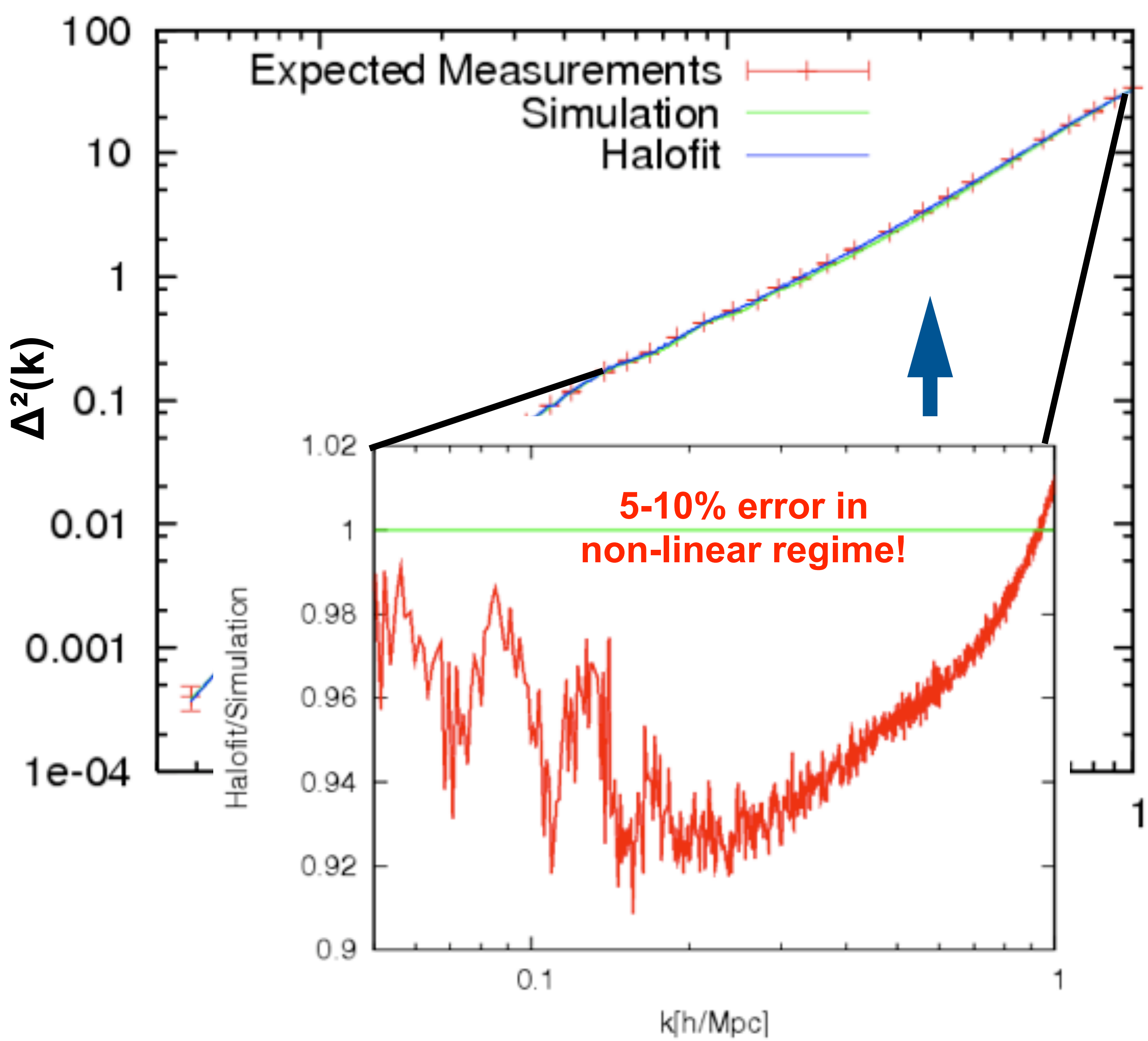
- Example in this talk: matter power spectrum
- Question: how badly will our constraints on dark energy be biased if we *do not* reach the same accuracy in our modeling as we might have in our data?
- Generate mock data set with the expected 1% error
- Analyze data with current method using HaloFit to model the matter power spectrum
  - HaloFit (Smith et al. 2003): semi-analytic fit for the power spectrum, based on modeling approach and tuned to simulations, accurate at the 5-10% level



$$\Delta^2(k) = \frac{k^3 P(k)}{2\pi^2};\ P(\vec{k}) = \langle \delta^2(\vec{k}) \rangle$$

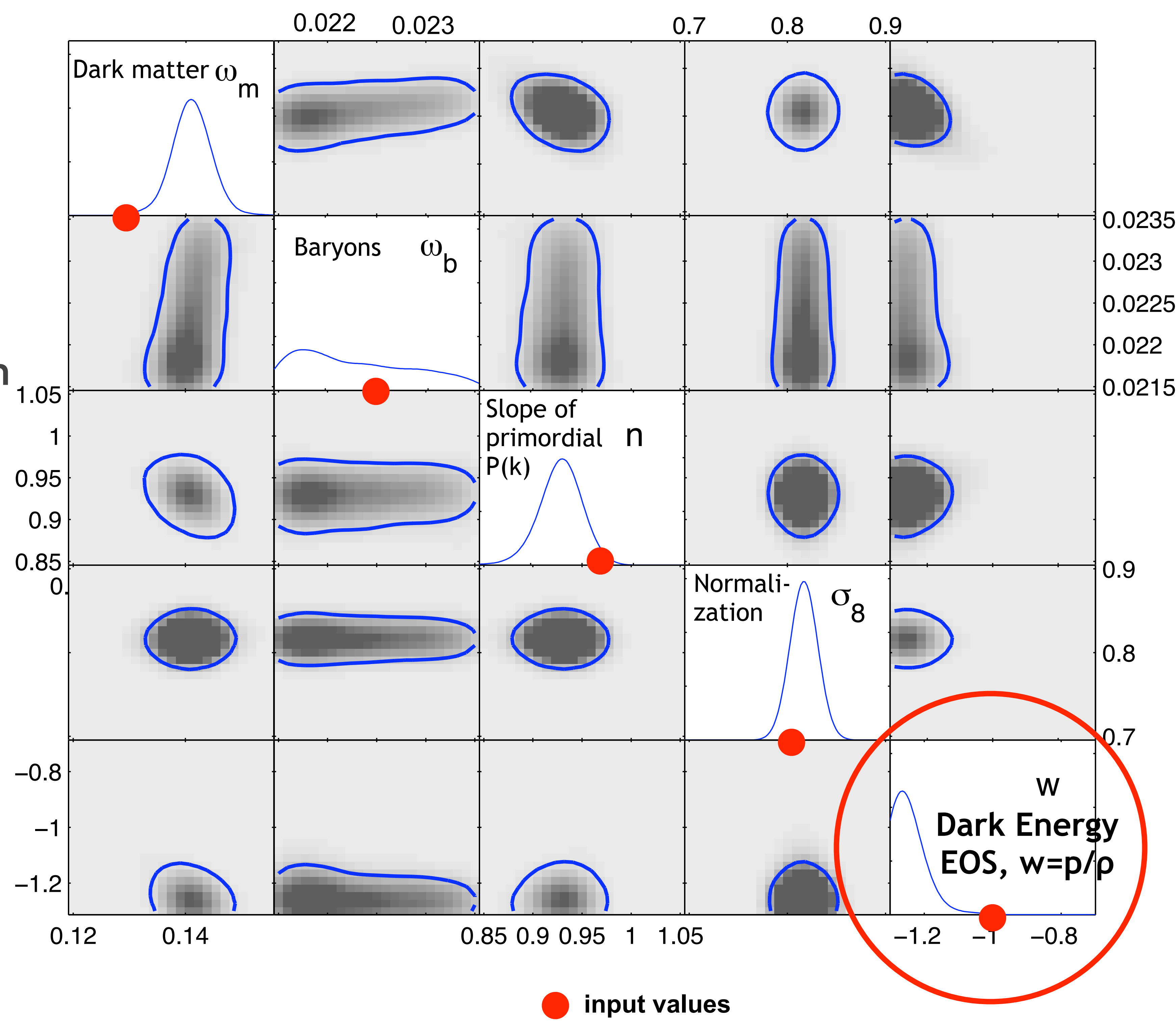
Why do we need higher accuracy in our predictions?

- Example in this talk: matter power spectrum
- Question: how badly will our constraints on dark energy be biased if we *do not* reach the same accuracy in our modeling as we might have in our data?
- Generate mock data set with the expected 1% error
- Analyze data with current method using HaloFit to model the matter power spectrum
  - HaloFit (Smith et al. 2003): semi-analytic fit for the power spectrum, based on modeling approach and tuned to simulations, accurate at the 5-10% level



Analysis of the “True data”

- Generate mock data from high-resolution simulation
- Use Halofit for analysis; remember, halofit ~5-10% inaccurate on scales of interest
- Parameters are up to 20% wrong! (We checked that with more accurate predictions the answer is correct)
- Only solution: **precision simulations**
- Analysis takes at least 10,000 input power spectra for MCMC, each simulation takes ~20,000 CPU hours
- With a 2000 node cluster running 24/7, our analysis will take ~30 years, hmmm...



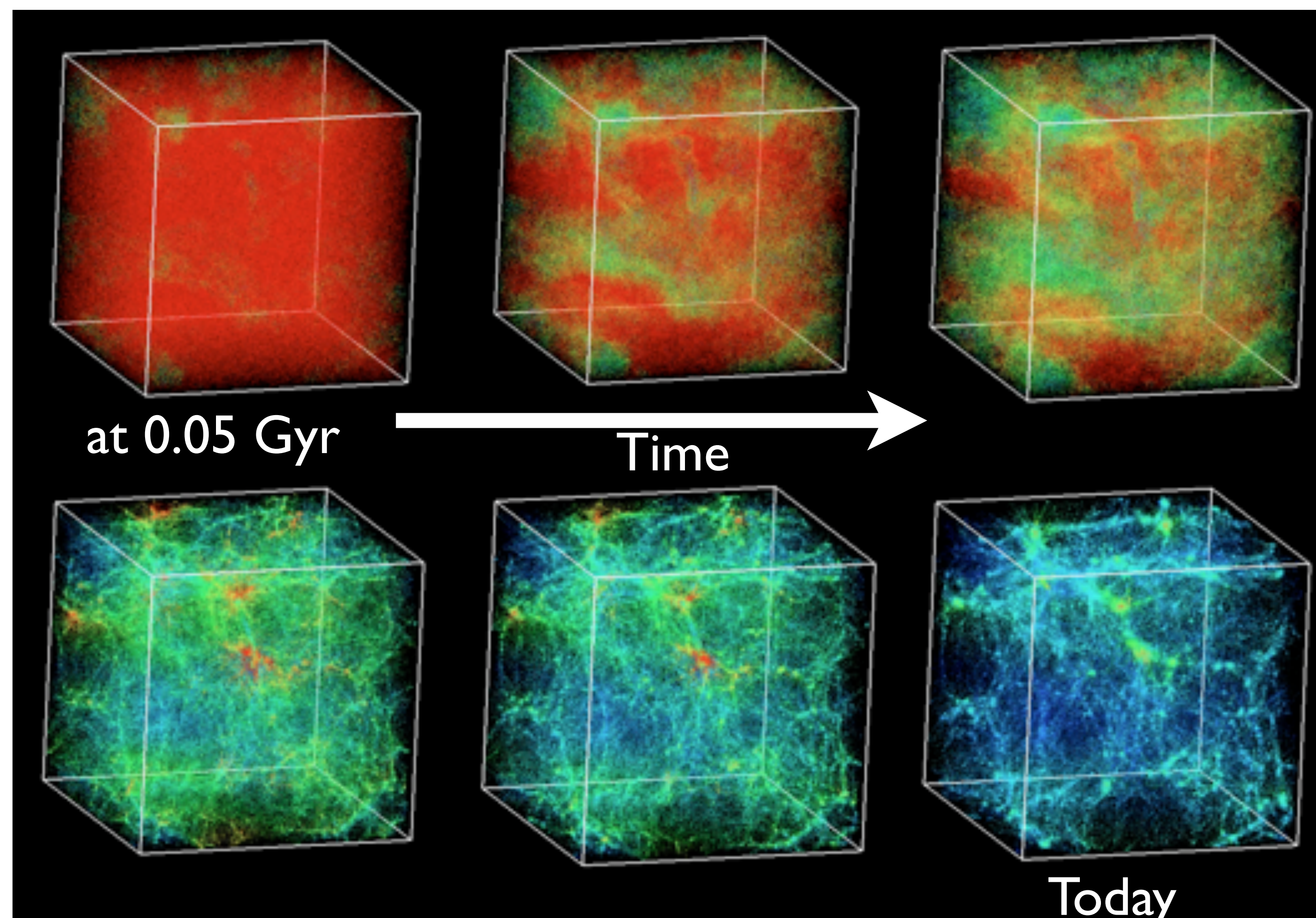
What do we need --

- **Very accurate predictions for the cosmological measurements of interest (power spectra/correlation functions, mass functions, galaxy power spectra ... [YOUR FAVORITE STATISTICS HERE]**
  - For gravity-only and galaxies in post-processing: HACC (Hardware/Hybrid Accelerated Cosmology Code, coming next)
  - For full treatment of baryonic effects: another talk at another time ...
- **Very fast predictions for your favorite statistics**
  - While HACC is fast, we need something much faster, sub-seconds preferably!
  - Cosmic Calibration Framework: Prediction tools build from a small but very accurate number of simulations (last part of the talk)

HACC's Task: Computing the Universe

- Gravity dominates at large scales, key task: solve the Vlasov-Poison equation (VPE)
- VPE is 6-D and cannot be solved as PDE, therefore N-body methods
- Particles are tracers of the dark matter in the Universe, mass typically at least ~10^9 $M_\odot$
- At smaller scales, add gas physics, feedback etc., sub-grid modeling inevitable

$$
\begin{aligned}
\frac{\partial f_i}{\partial t} + \dot{\mathbf{x}}\frac{\partial f_i}{\partial \mathbf{x}} - \nabla\phi\frac{\partial f_i}{\partial \mathbf{p}} &= 0, \qquad \mathbf{p} = a^2\dot{\mathbf{x}}, \\
\nabla^2\phi &= 4\pi G a^2(\rho(\mathbf{x},t) - \langle\rho_{\mathrm{dm}}(t)\rangle) = 4\pi G a^2 \Omega_{\mathrm{dm}}\delta_{\mathrm{dm}}\rho_{\mathrm{cr}}, \\
\delta_{\mathrm{dm}}(\mathbf{x},t) &= (\rho_{\mathrm{dm}} - \langle\rho_{\mathrm{dm}}\rangle)/\langle\rho_{\mathrm{dm}}\rangle), \\
\rho_{\mathrm{dm}}(\mathbf{x},t) &= a^{-3}\sum_i m_i \int d^3\mathbf{p}\, f_i(\mathbf{x},\dot{\mathbf{x}},t).
\end{aligned}
$$

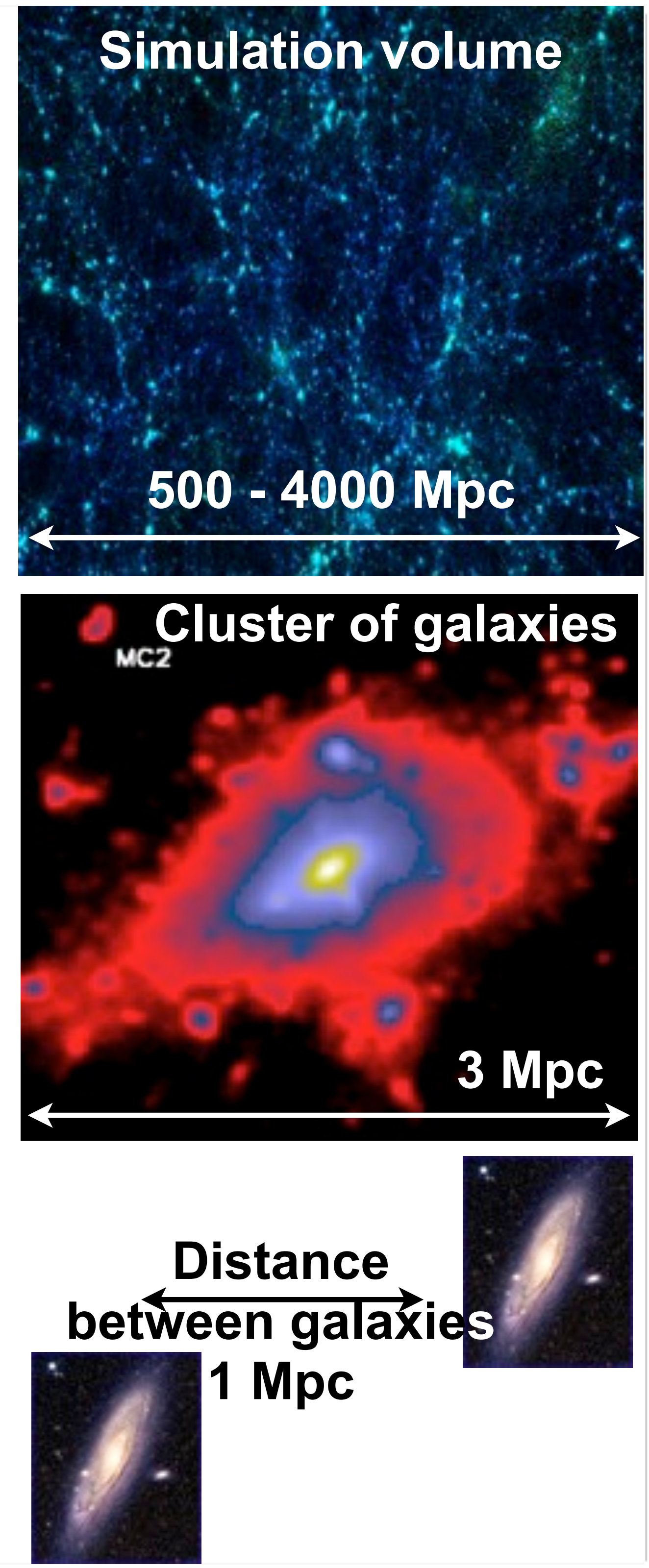


"The Universe is far too complicated a structure to be studied deductively, starting from initial conditions and solving the equations of motion."

Robert Dicke (Jayne Lectures, 1969)

Computing the Universe: Simulating Surveys

- **Simulation Volume:** Large survey sizes impose simulation volumes ~(**4** Gpc)^3, memory required ~**100**TB **-** **1**PB
- **Number of Particles:** Mass resolution depends on ultimate object to be resolved, ~$10^8 \mathrm{M}_\odot - 10^{10} \mathrm{M}_\odot$, N~$10^{11} - 10^{12}$
- **Force Resolution:** ~kpc, yields a (global) spatial dynamic range of **10**^{6}
- **Throughput:** Large numbers of simulations required (**100**'s **--1000**'s), development of analysis suites, and emulators; peta-exascale computing exploits
- **Computationally very challenging!** HACC is aimed to meet these requirements



The HACC Story Begins ...

Andrew White   Dec 7, 2007   + What if you had a petaflop/s

- **... with an email:** Los Alamos National Lab offers the opportunity to run open science projects on the fastest supercomputer in the world for the first six months of the machine's existence: **Roadrunner**
- **Roadrunner:** First machine to achieve Petaflop performance via Cell-acceleration, CPU/Cell hybrid architecture (more details later) (equivalent to ~200,000 laptops)
- **The Challenges:**
  - The machine has a "crazy" architecture, requiring major code re-designs and rewrites (we ended up writing a brand **new code**)
  - Roadrunner probably **one of a kind**, code-design needs to be flexible and **portable** to other future architectures
- **Cosmologists are poor -- so we took on the challenge!**
- **Outcome: MC^3 (Mesh-based Cosmology Code on the Cell) which later morphed into HACC, N-body code to simulate large-scale structure formation in the Universe**

The HACC Story Begins ...

Andrew White | Dec 7, 2007 | + What if you had a petaflop/s

- **... with an email:** Los Alamos National Lab offers the opportunity to run open science projects on the fastest supercomputer in the world for the first six months of the machine's existence: **Roadrunner**
- **Roadrunner:** First machine to achieve Petaflop performance via Cell-acceleration, CPU/Cell hybrid architecture (more details later) (equivalent to ~200,000 laptops)
- **The Challenges:**

  Andy White: "forward-looking"

  - The machine has a "~~crazy~~" architecture, requiring major code re-designs and rewrites (we ended up writing a brand **new code**)
  - Roadrunner probably **one of a kind**, code-design needs to be flexible and **portable** to other future architectures
- **Cosmologists are poor -- so we took on the challenge!**
- **Outcome: MC³ (Mesh-based Cosmology Code on the Cell) which later morphed into HACC, N-body code to simulate large-scale structure formation in the Universe**

The Story Continues … and makes it into "Die Süddeutsche"

Newspaper I read every morning

sueddeutsche.de

Politik | Wirtschaft | Geld | Kultur | Sport | Leben | Karriere | München & Region | Bayern | M

Home > Digital   Supercomputer - Rasend schnell

Supercomputer

Rasend schnell

"He is the fastest calculator in the world:"

Er ist der schnellste Rechner der Welt: der amerikanische Supercomputer "Roadrunner" hat die Petaflop-Grenze geknackt. Sein Job: die Simulation von Atombombenexplosionen.

Twittern 0 | Empfehlen | Senden | +1 0

Ein Rechner der US-Regierung schafft erstmals mehr als eine Billiarde Operationen in der Sekunde (Petaflops) und ist damit nun der schnellste Computer der Welt. Das berichten das US-Energieministerium und der Hersteller IBM am Montag.

Der Computer namens Roadrunner wurde am Los Alamos National Laboratory (LANL) in New Mexico installiert. Er wird zuvorderst für die Forschung an US-Atomwaffen rechnen. Der neu konstruierte Roadrunner ist auf einen Schlag mehr als doppelt so schnell wie der bisherige Spitzenreiter der "Top 500"-Liste der Supercomputer.

Anfangs soll der Roadrunner aber vor allem wissenschaftliche Probleme lösen. Beispielsweise sind Tests von Klimamodellen vorgesehen, doch rechnet das LANL mit Anwendungen in diversen Bereichen, darunter die Kosmologie, die Entwicklung von Antibiotika oder die Astrophysik. Danach wird der Supercomputer laut LANL militärischen Aufgaben zugeteilt und unter Geheimhaltung Explosionen nuklearer Waffen simulieren, um physikalische Modelle zu verbessern und das Vertrauen in das nukleare Arsenal der USA ohne tatsächliche Atomtests zu erhalten.

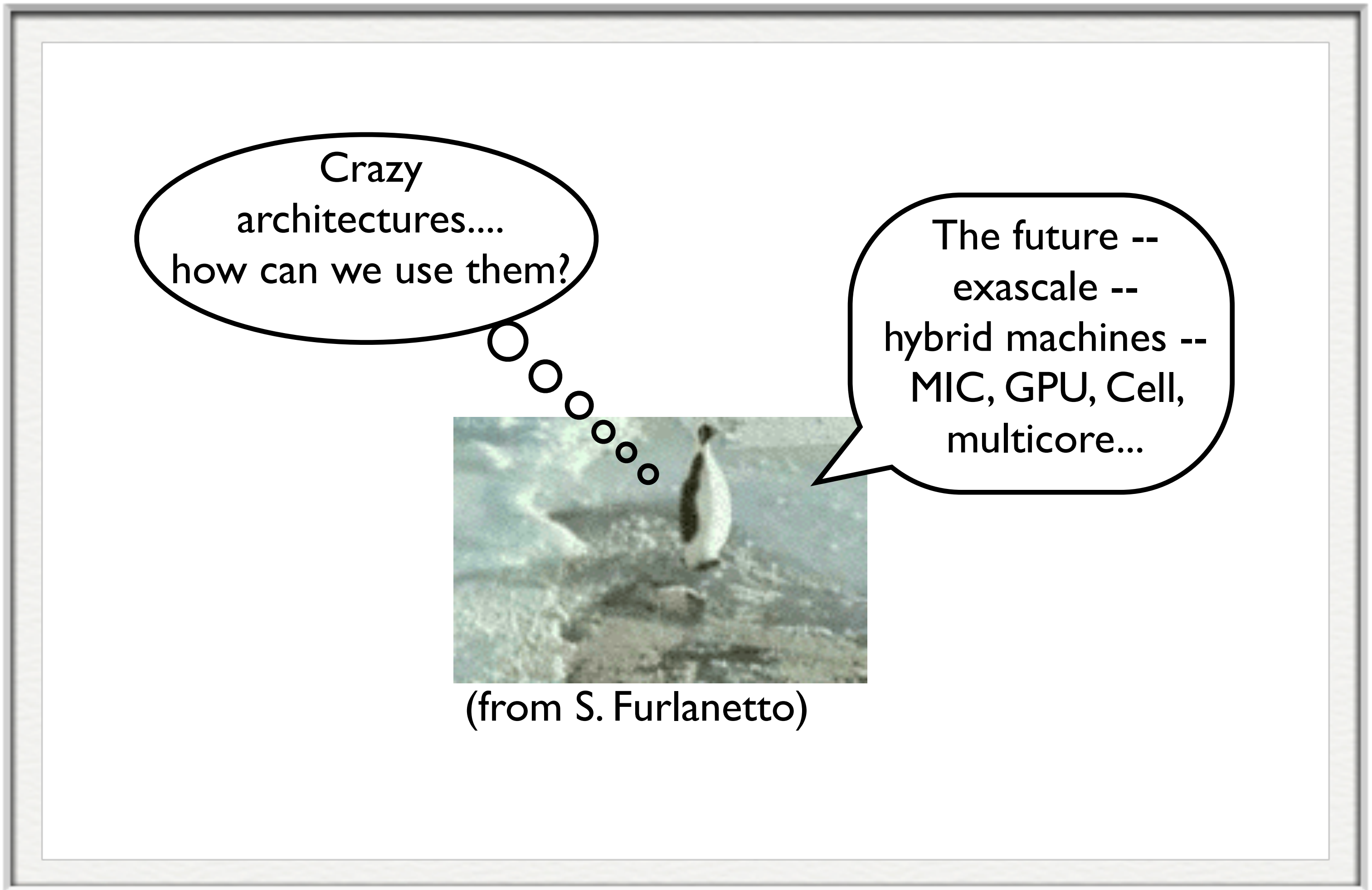
Supercomputer mit Vorbildfunktion "leads by example"

"Für uns und die HPC-Community ist es hoch erfreulich, dass es ein System gibt, das diese Marke geknackt hat", sagt Thomas Lippert, Leiter des Jülich Supercomputing Centre. Dadurch werde dem Supercomputing berechtigte Aufmerksamkeit zuteil.

Technologisch dürfte Roadrunner Vorbildwirkung haben. "Es zeichnet sich ab, dass Hybrid-Technologie auf jeden Fall Zukunft haben", meint Lippert. Damit sind Systeme gemeint, die klassische CPUs mit Beschleunigern wie beispielsweise den Cell-Chips oder Grafikprozessoren kombinieren.

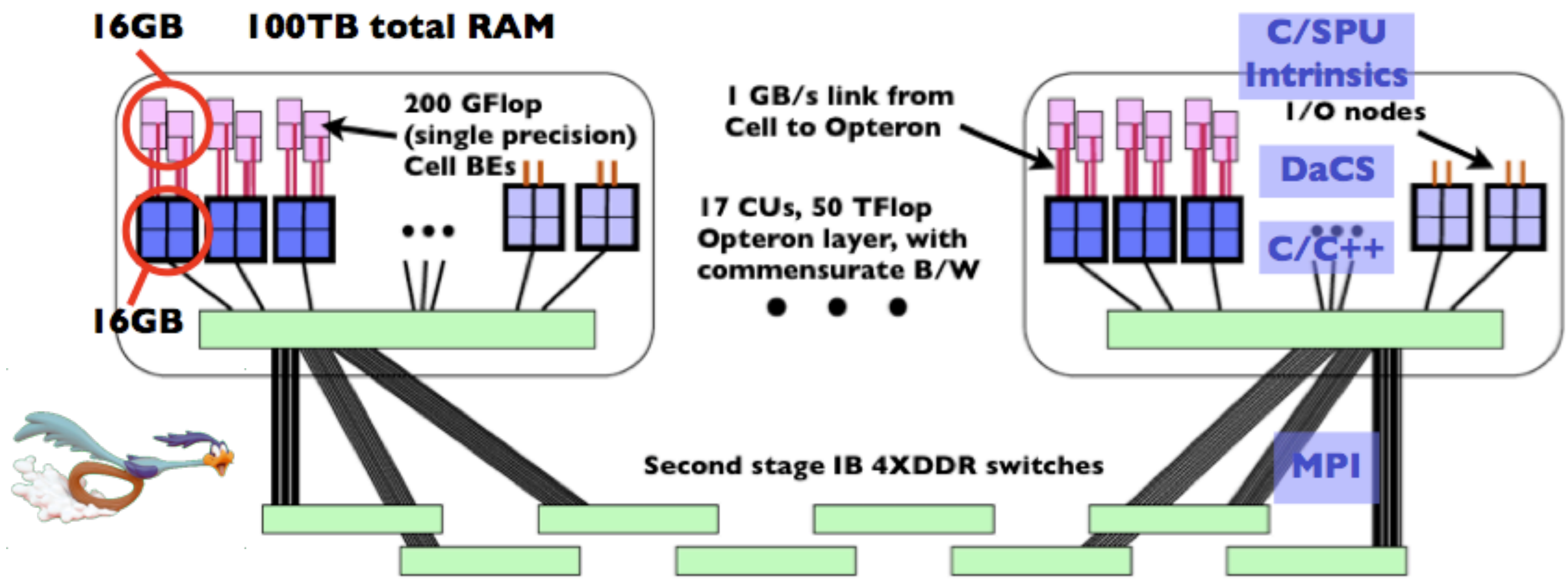
"technology of the future"

So we started thinking --



(from S. Furlanetto)

The Roadrunner Architecture

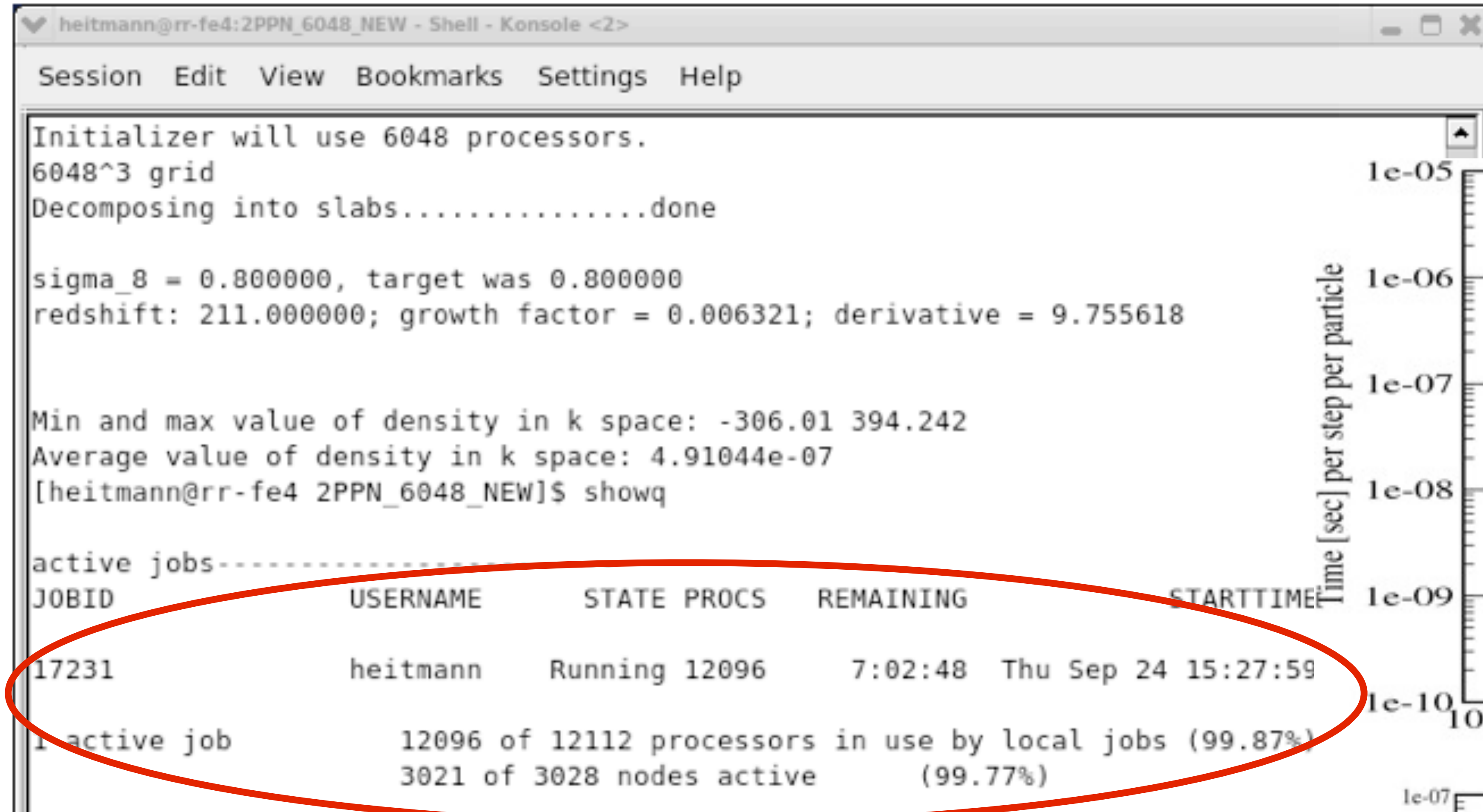


- **Opterons** have little compute (5% of total compute) but half the memory and balanced communication, for N-body codes, memory is limiting factor, so want to make best use of CPU layer
- Cells **dominate** the compute but communication is poor, 50-100 times out of balance (also true for CPU/GPU hybrid systems)
- **Multi-layer programming model:** C/C++/MPI (Message Passing Interface) for Opterons, C/Cell-intrinsics for Cells

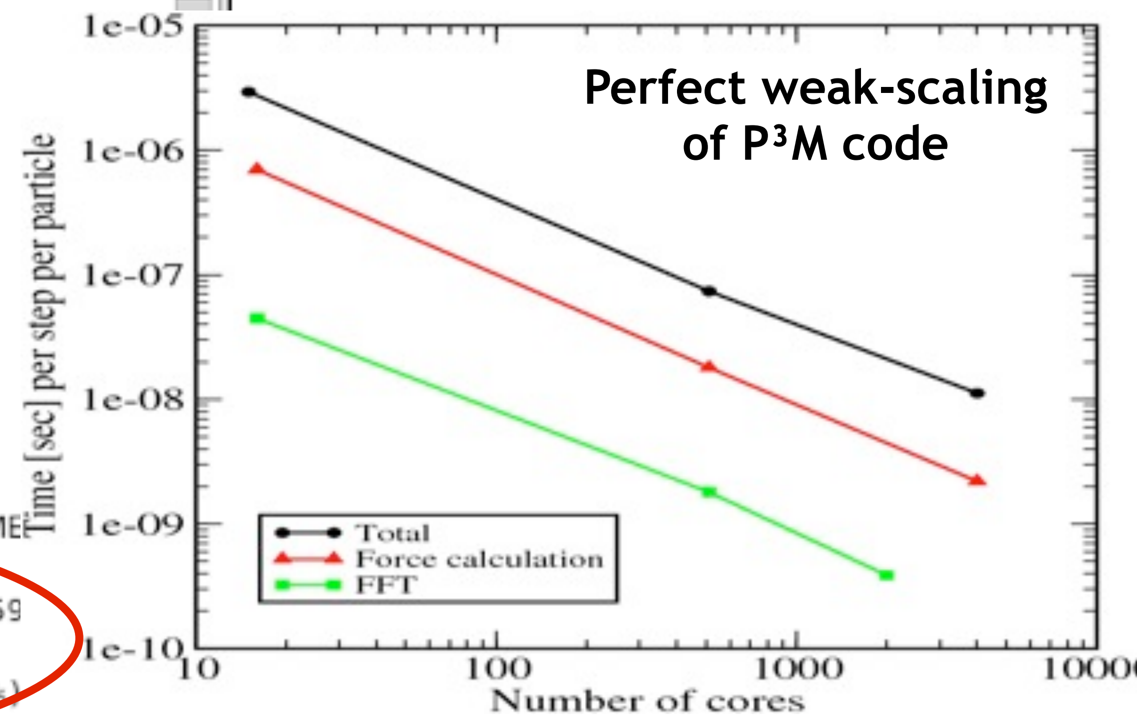
Design Challenges and Solutions for MC^3 CHANGE

- **Challenges (summarized from last slide):**
  - Opterons have half of the machine's memory, balanced communication, but not much compute, standard programming paradigm, C/C++/MPI
  - Cells have other half of machine's memory, slow communication, **lots** of compute (95% of machine's compute power), new language required
- **Design desiderata:**
  - Distribute memory requirements on both parts of the machine
  - Give the Cell lots of (communication limited) work to do, make sure that Cell part is easy to code and later on easy to replace by different programming paradigm
- **Our Solution:** P^3M algorithm (long range - short range split)
  - Particle Mesh (PM) solver for long-range force, FFT based, grid lives on the Opterons, all coarse-grained parallelism here, base grid is maximized
  - Direct Particle-Particle solver for short range force, particles live on the Cells, lots of compute, simple data structure, easy to implement, can be replaced by tree for different architecture
  - Overloading trick to minimize communication needs, only simple grid information flows between Cells and Opterons; enables node level short-range force plug-ins

MC³ Performance

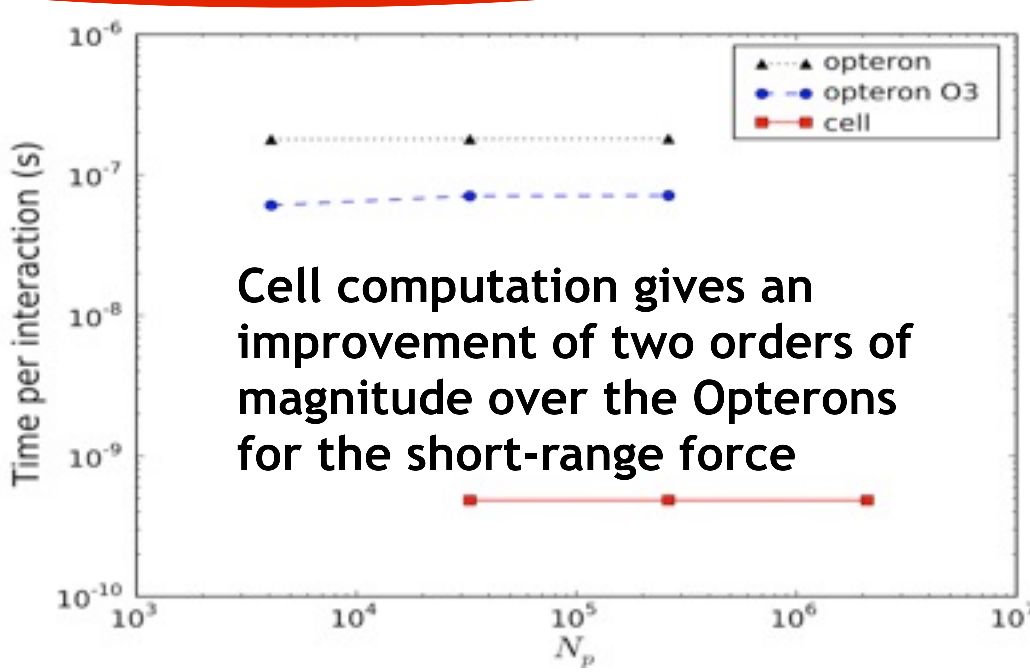



```
Initializer will use 6048 processors.
6048^3 grid
Decomposing into slabs...............done

sigma_8 = 0.800000, target was 0.800000
redshift: 211.000000; growth factor = 0.006321; derivative = 9.755618


Min and max value of density in k space: -306.01 394.242
Average value of density in k space: 4.91044e-07
[heitmann@rr-fe4 2PPN_6048_NEW]$ showq

active jobs------------------------
JOBID               USERNAME      STATE PROCS   REMAINING            STARTTIME

17231               heitmann    Running 12096     7:02:48  Thu Sep 24 15:27:59

1 active job          12096 of 12112 processors in use by local jobs (99.87%)
                       3021 of 3028 nodes active      (99.77%)
```

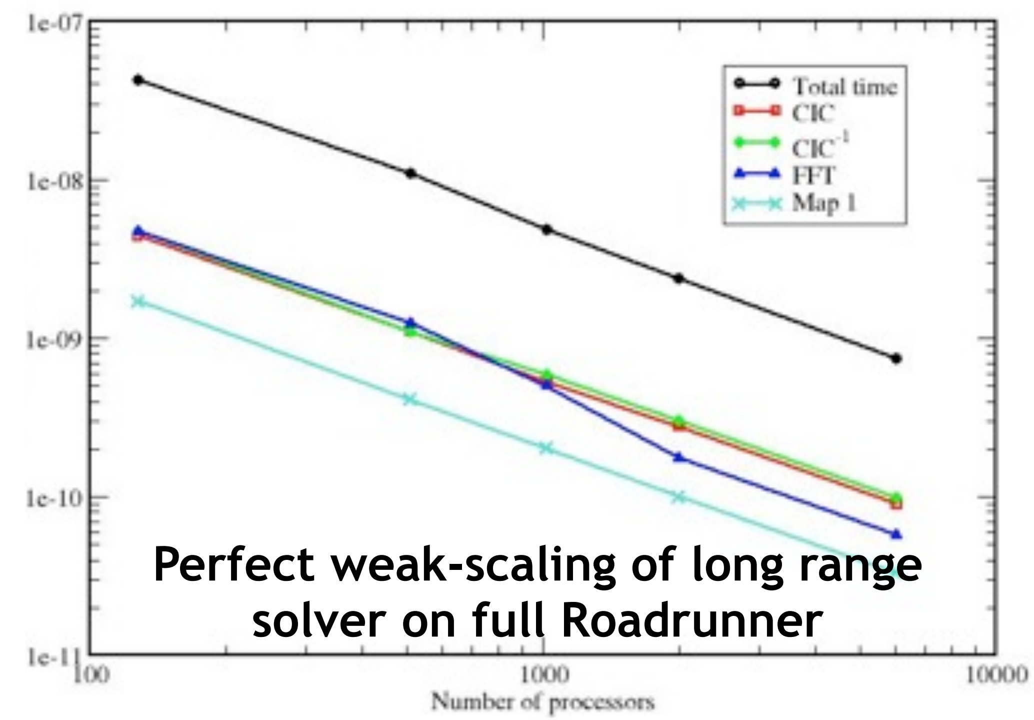


Perfect weak-scaling of P^3M code

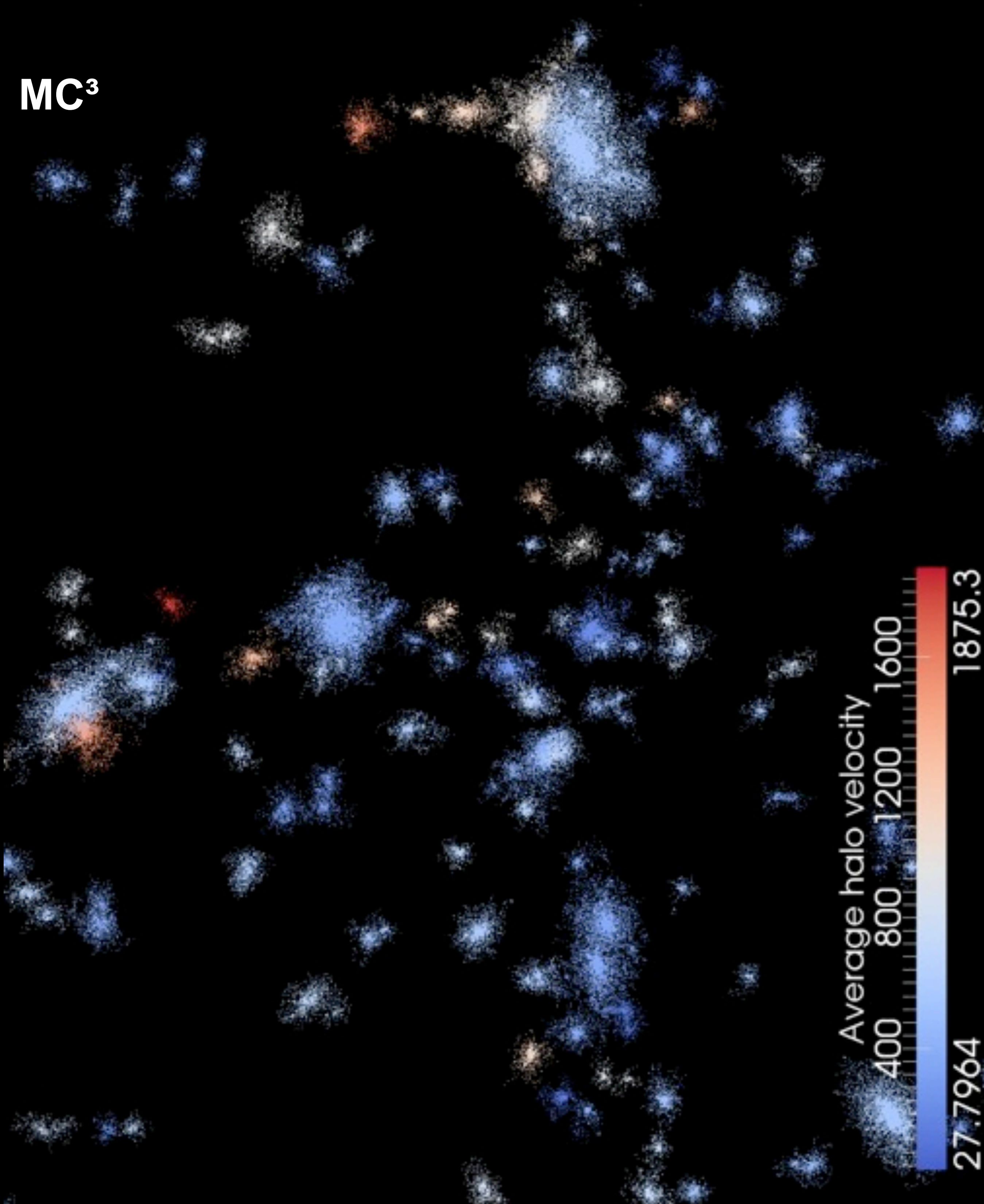
Weak scaling: problem size grows while core count increases
Strong scaling: problem size fixed while core count increases

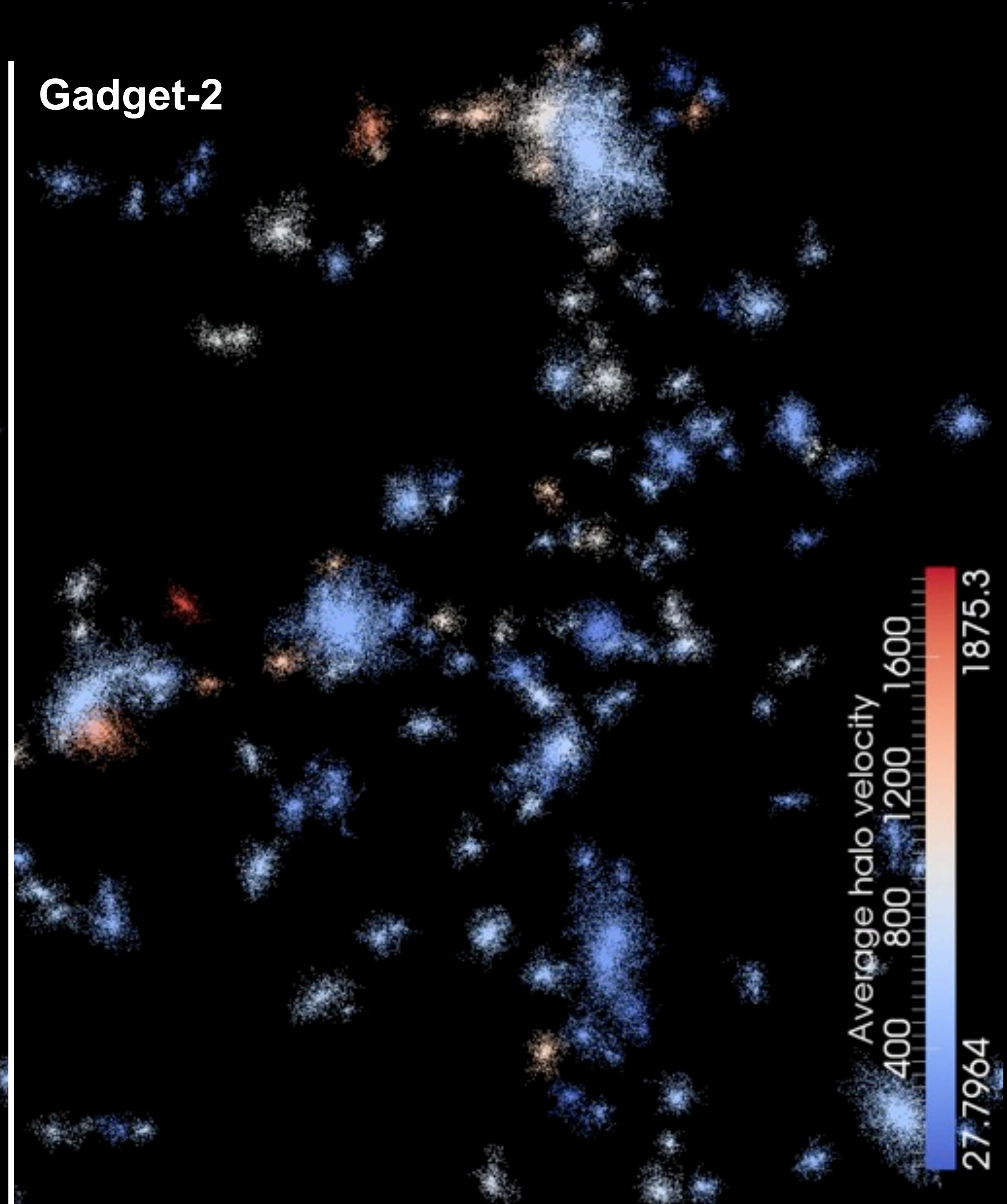


Cell computation gives an improvement of two orders of magnitude over the Opterons for the short-range force

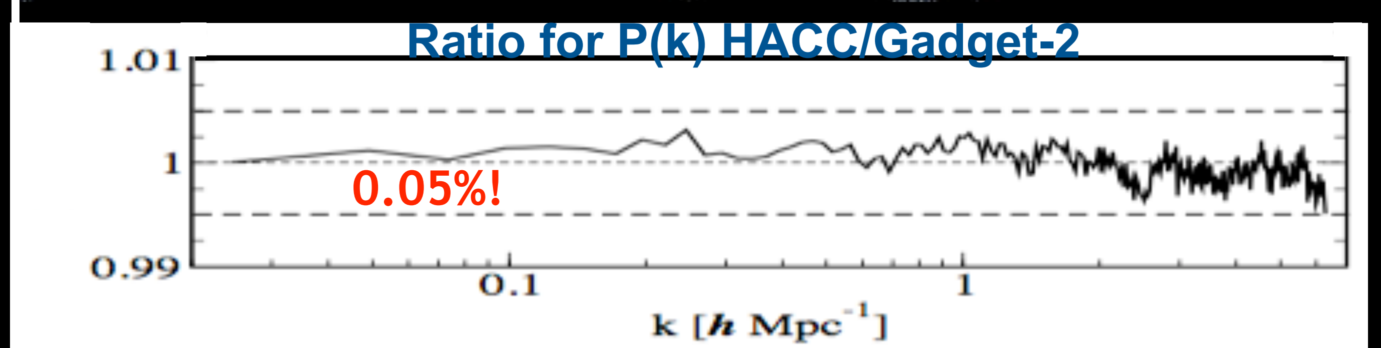


Perfect weak-scaling of long range solver on full Roadrunner





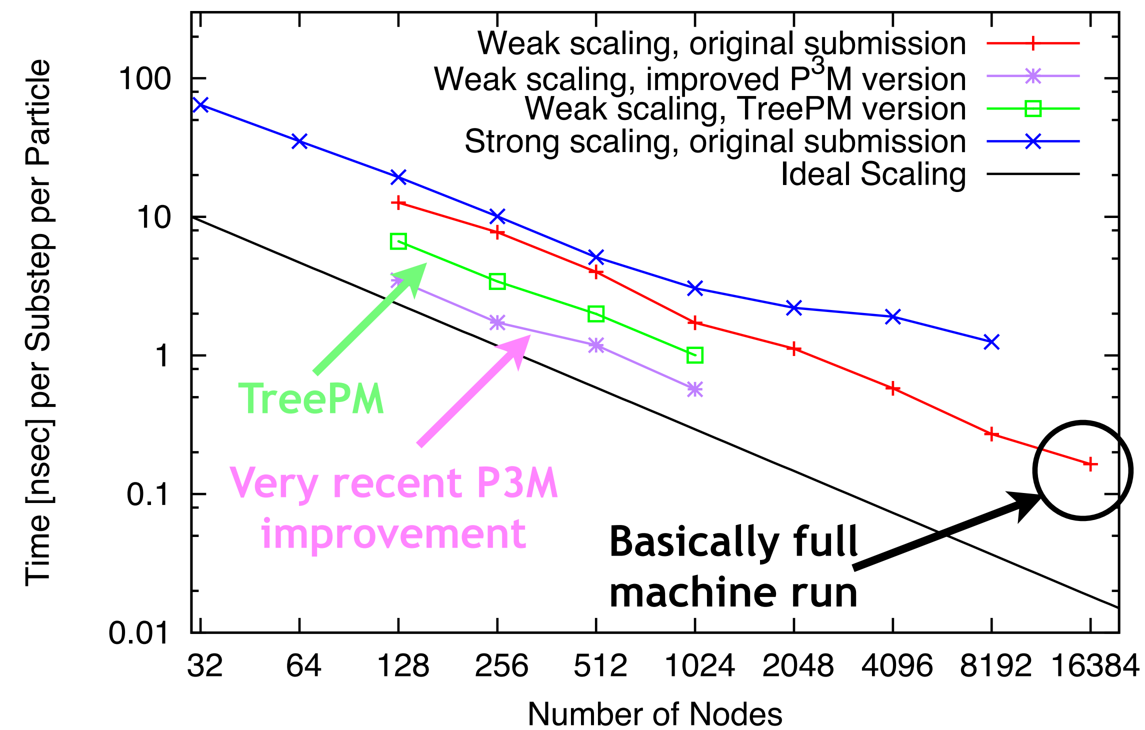
Snapshot from Code Comparison simulation, ~25 Mpc region; halos with > 200 particles, b=0.15
Differences in runs: P³M vs. TPM, force kernels, time stepper: MC³: a; Gadget-2: log(a)
Power spectra agree at sub-percent level



The Story continues, MC³ becomes HACC: CPU+GPU

- **Proof of concept for easy portability:** replace Cell part by GPU implementation
- Paul Sathre (CS undergraduate at the time) in summer 2010 successfully ports code within weeks (with guidance from Adrian Pope)
- **New challenges:**
  - CPU/GPU performance and communication out of balance **AND** unbalanced memory (CPU/main memory dominates)
  - New programming language on GPU, OpenCL
- **With the arrival of Titan in 2013 (GPU accelerated supercomputer at Oak Ridge National Lab):**
  - Nick Frontiere completely rewrote and optimized HACC-P3M version for GPUs
  - NVIDIA's Justin Luitjens and Argonne's Vitali Morozov wrote TreePM version

Weak and Strong Scaling Results



S. Habib et al. 2013: SuperComputing13, Gordon Bell Finalist (decision: Nov. 21)

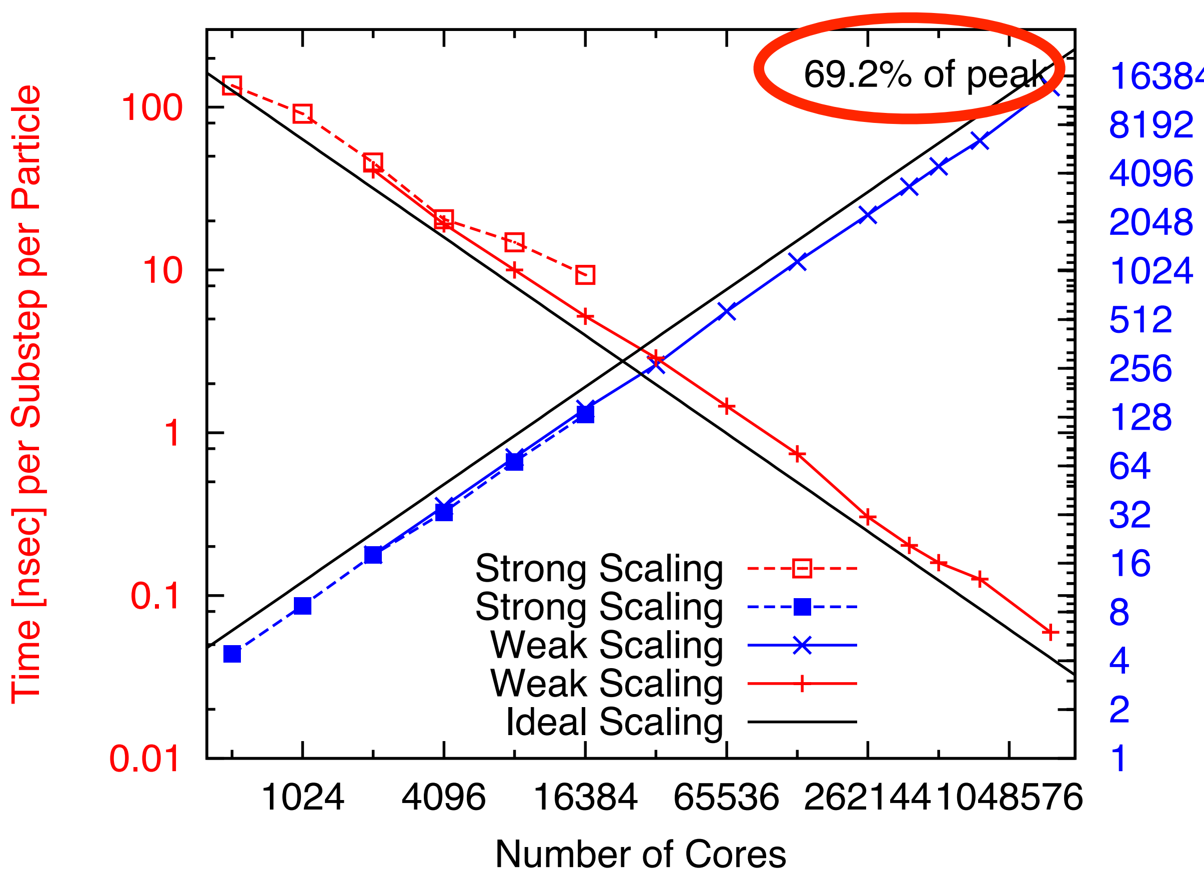
"The Gordon Bell Award recognizes outstanding achievement in high-performance computing applications."

- Achieved 20.54 Pflops peak performance evolving 1.23 trillion particles in test run on ~75% of machine (full machine currently not available)
- Why is this important? Qualify for more computing time!

Another Challenge: Multi-core systems, BG/Q

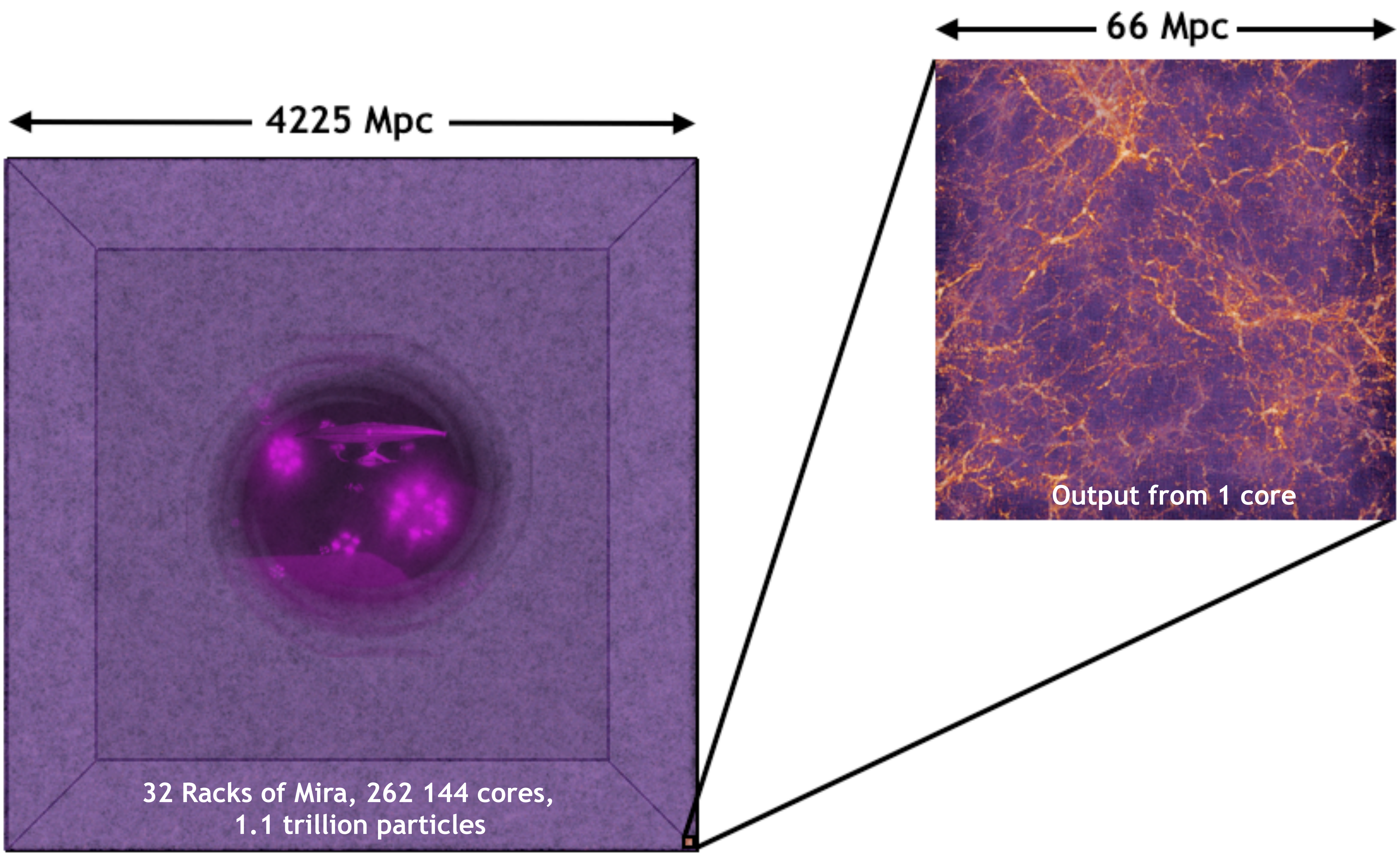
- **Proof of concept for "easy" portability II: IBM Blue Gene (BG) systems**
  - BG/Q Mira at Argonne: 10 PFlops, arrived in 2012, 750,000 cores, 16GB per node
  - BG/Q Sequoia at Livermore: twice as large
- **New challenges:**
  - BG/Q systems have many cores but no accelerators
  - Slab-decomposed FFT does not scale well on very large number of cores
- **Solutions:**
  - Particle-particle interaction now replaced by tree, OpenMP node parallel
  - Pencil decomposed FFT
  - Adaptive time stepping
- **Achieved 13.94PFlops on Sequoia, 90% parallel efficiency on 1,572,864 cores**
- **3.6 trillion particle benchmark run**

Weak Scaling up to 96 Racks; Strong Scaling, 1024^3 Particles

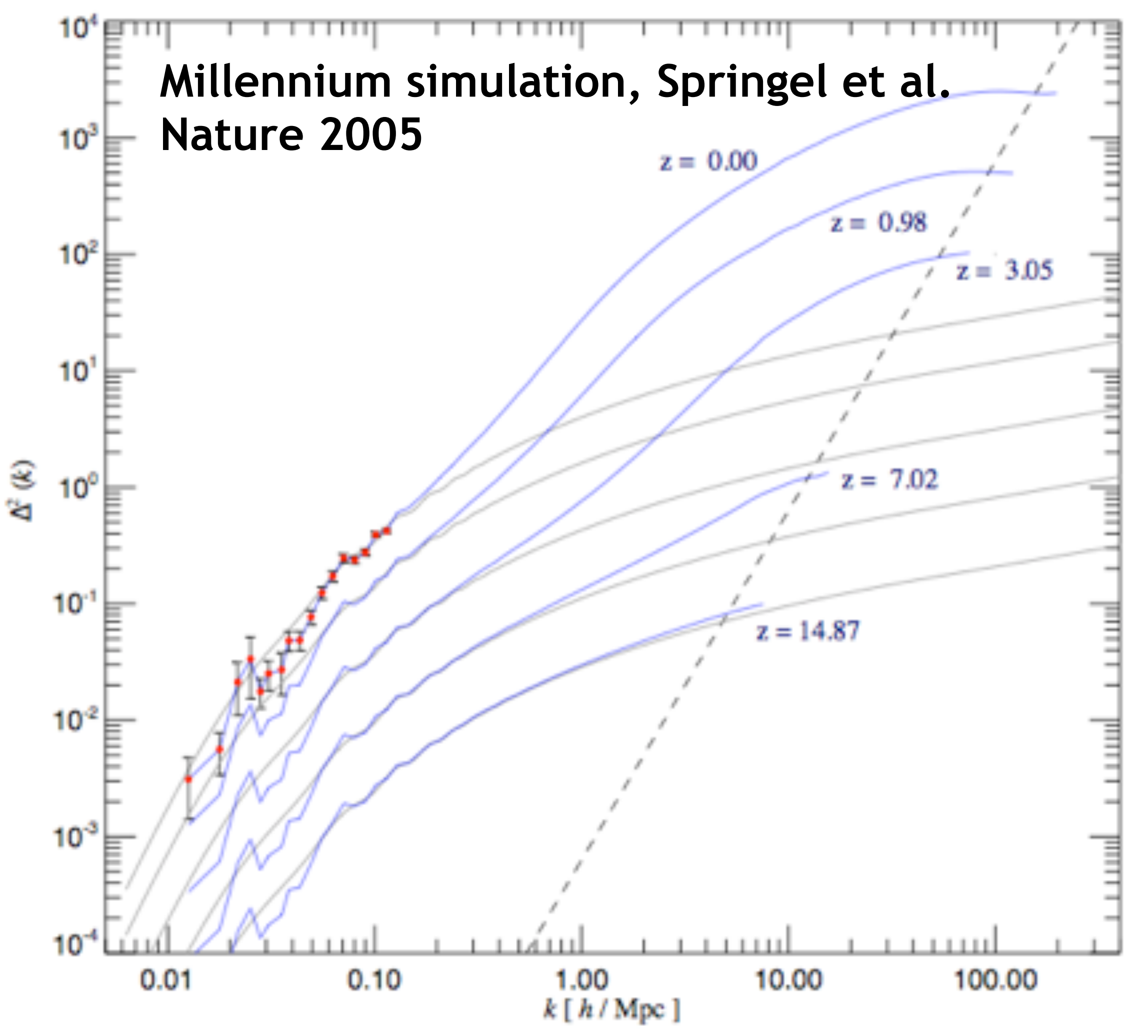


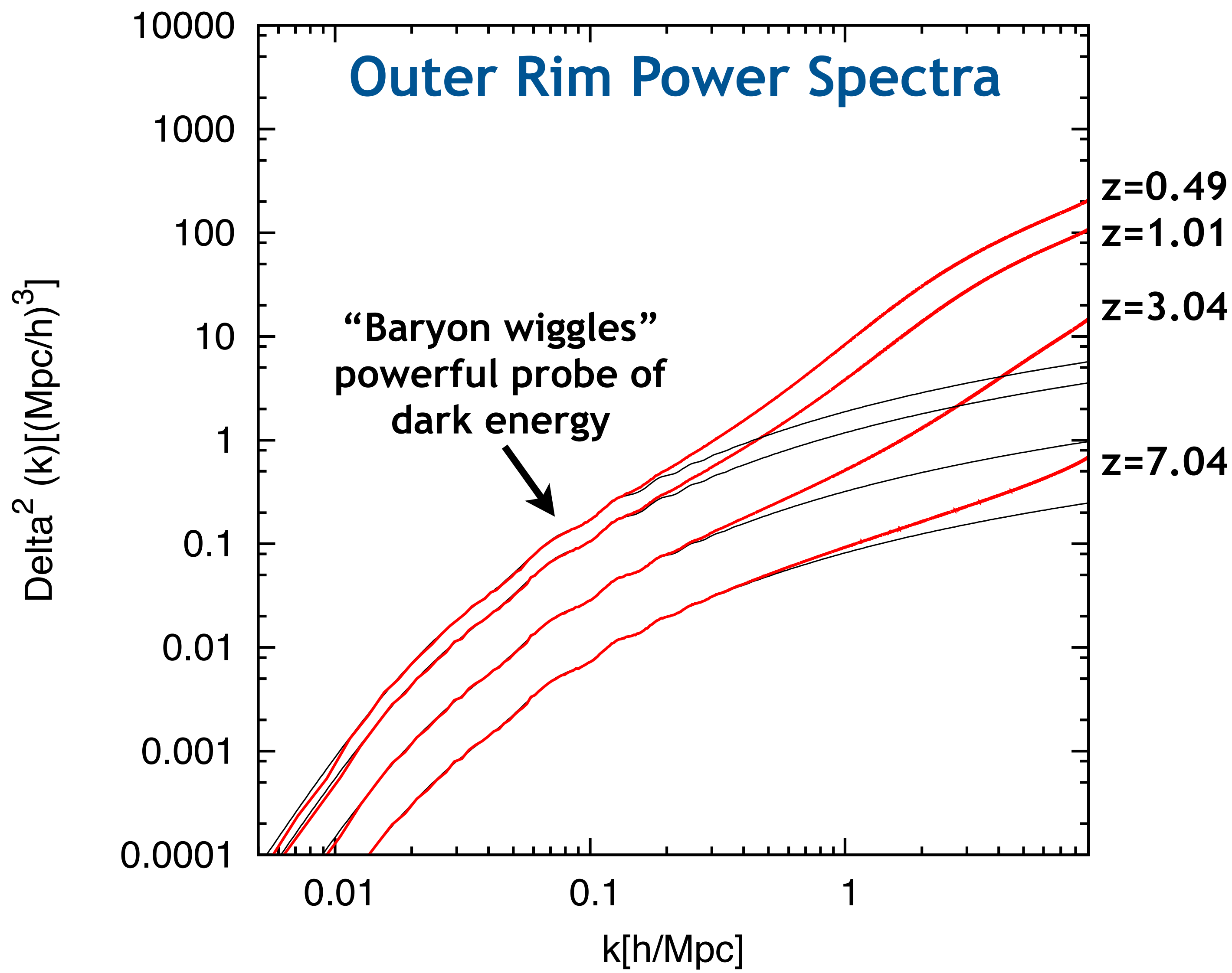
Habib et al. SC12, Gordon Bell Finalist

Mira Science: The Outer Rim Simulation



Some HACC Science Results --

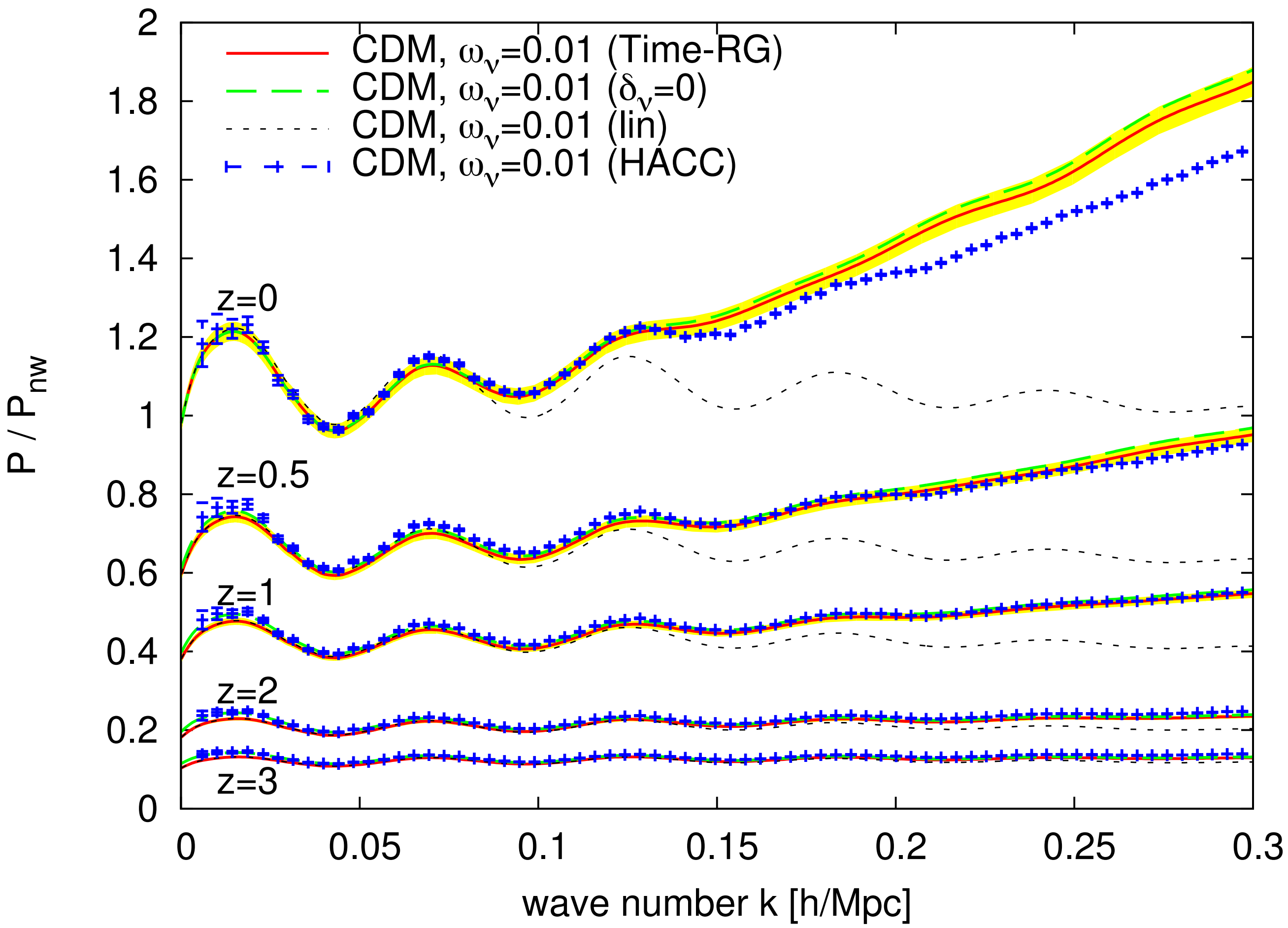




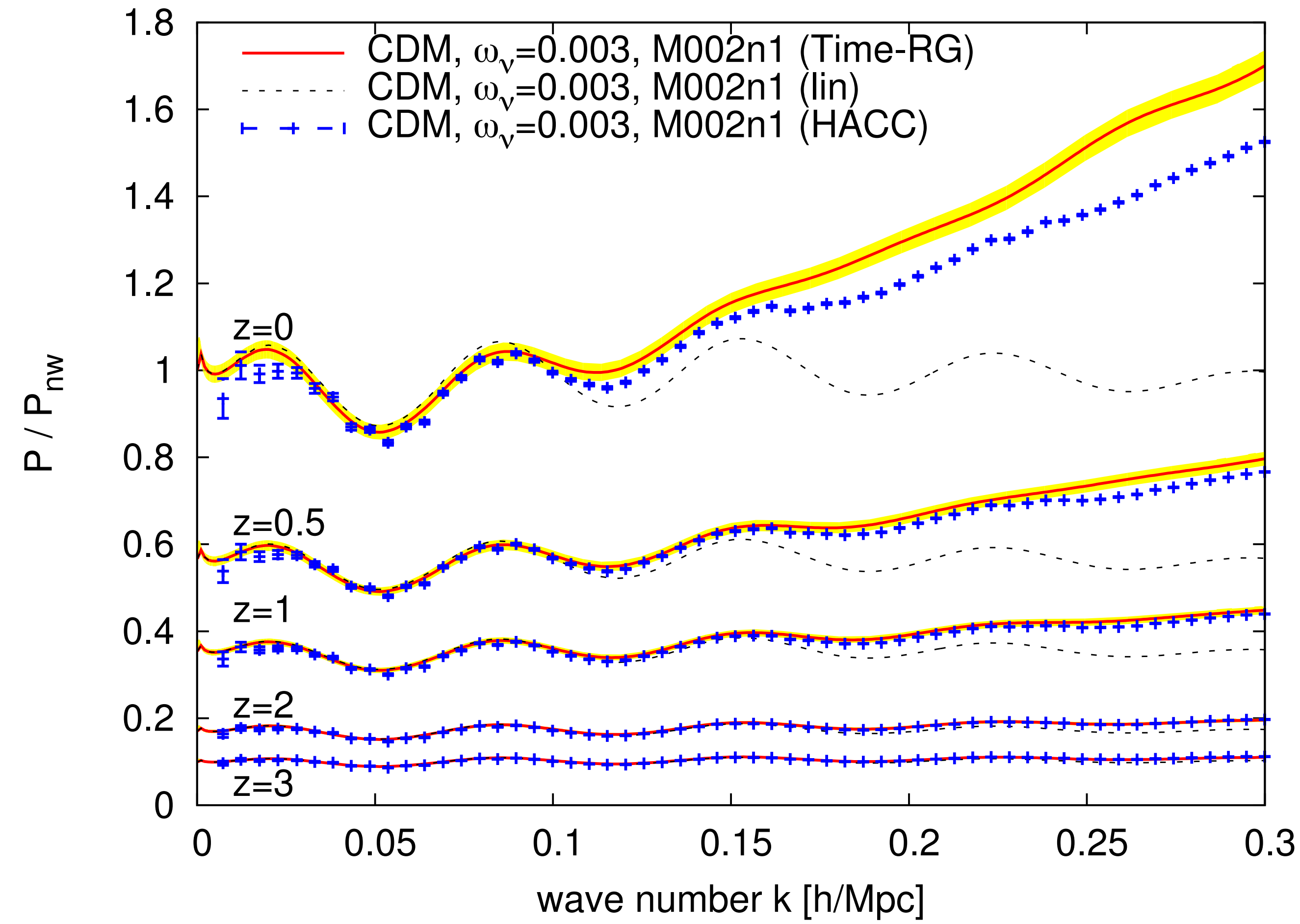
- Mass resolution of Millennium simulation and Outer Rim run very similar (~ $10^9 M_\odot$ particle mass), but volume different by a factor of 216 (Outer Rim volume = Millennium XXL)
- Exceptional statistics at high resolution enable many science projects

Some HACC Science Results --

LCDM + massive neutrinos



Early dark energy + massive neutrinos

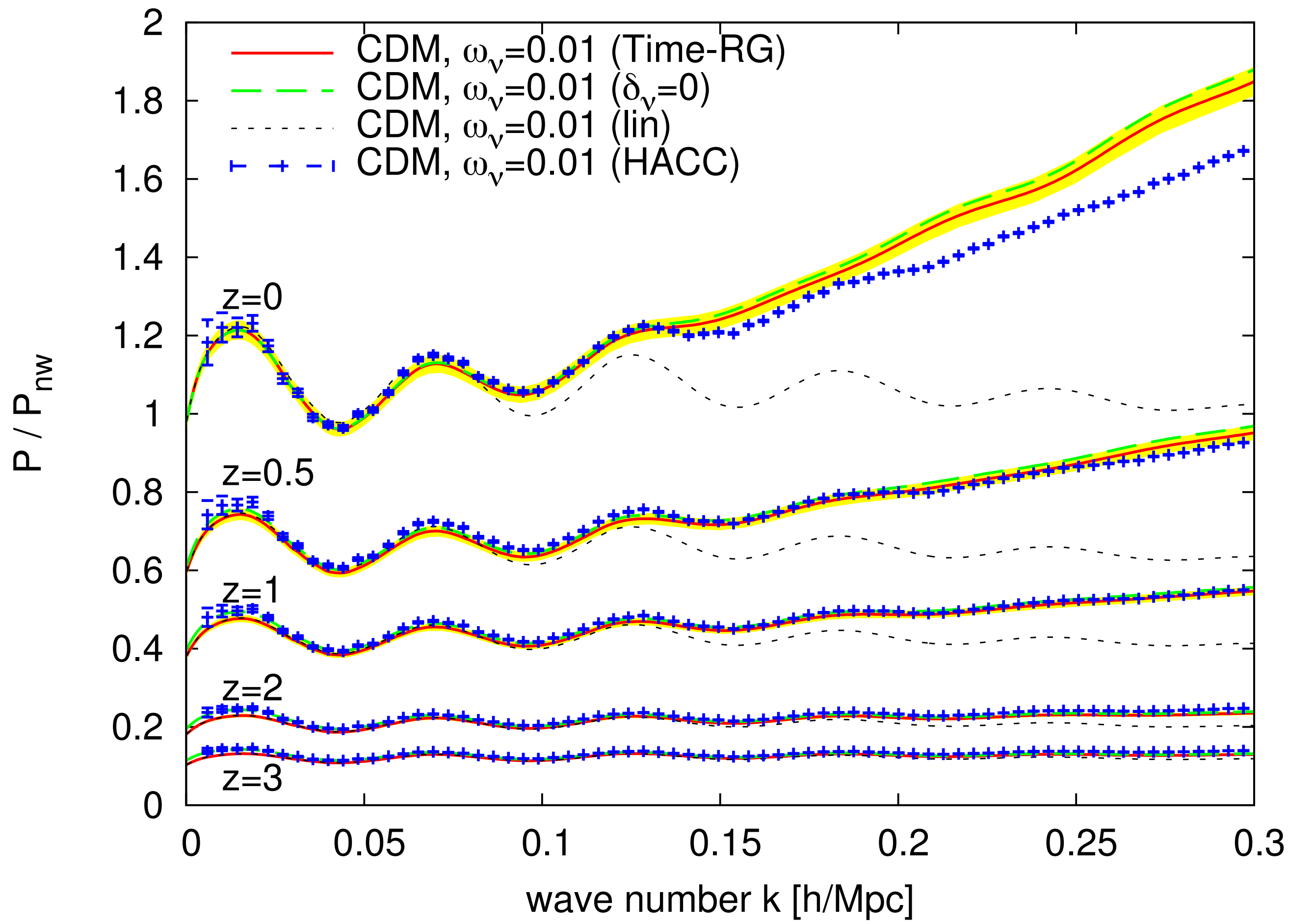


- Investigate different physical effects on e.g. the matter power spectrum at high accuracy
- Here: 2100 Mpc simulations with ~32.8 billion particles, $10^{10}\mathrm{M}_\odot$ particle mass to study massive neutrinos as well as time varying dark energy EOS
- Develop "higher order" perturbation theory approach

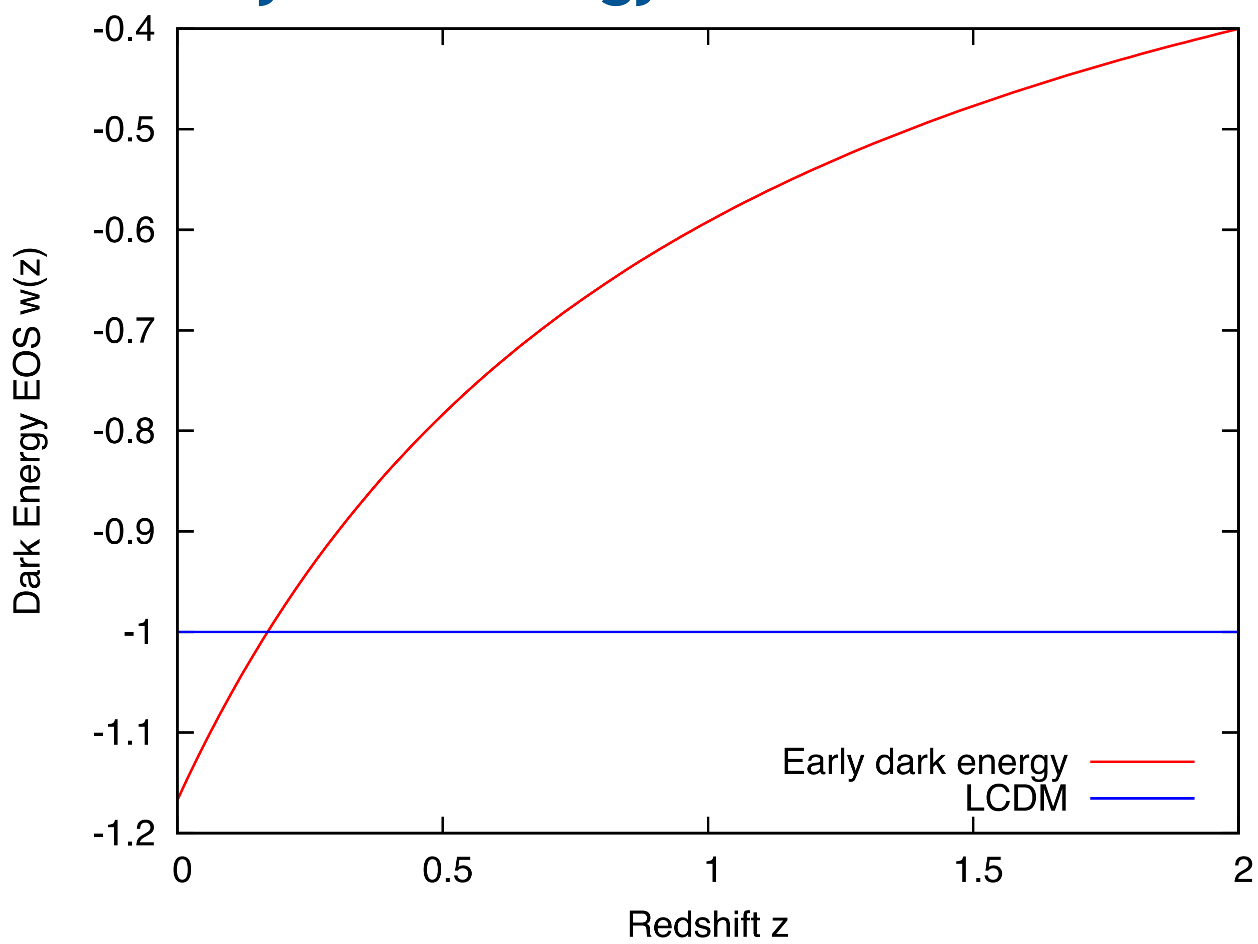
A. Upadhye et al. 2013

Some HACC Science Results --

LCDM + massive neutrinos



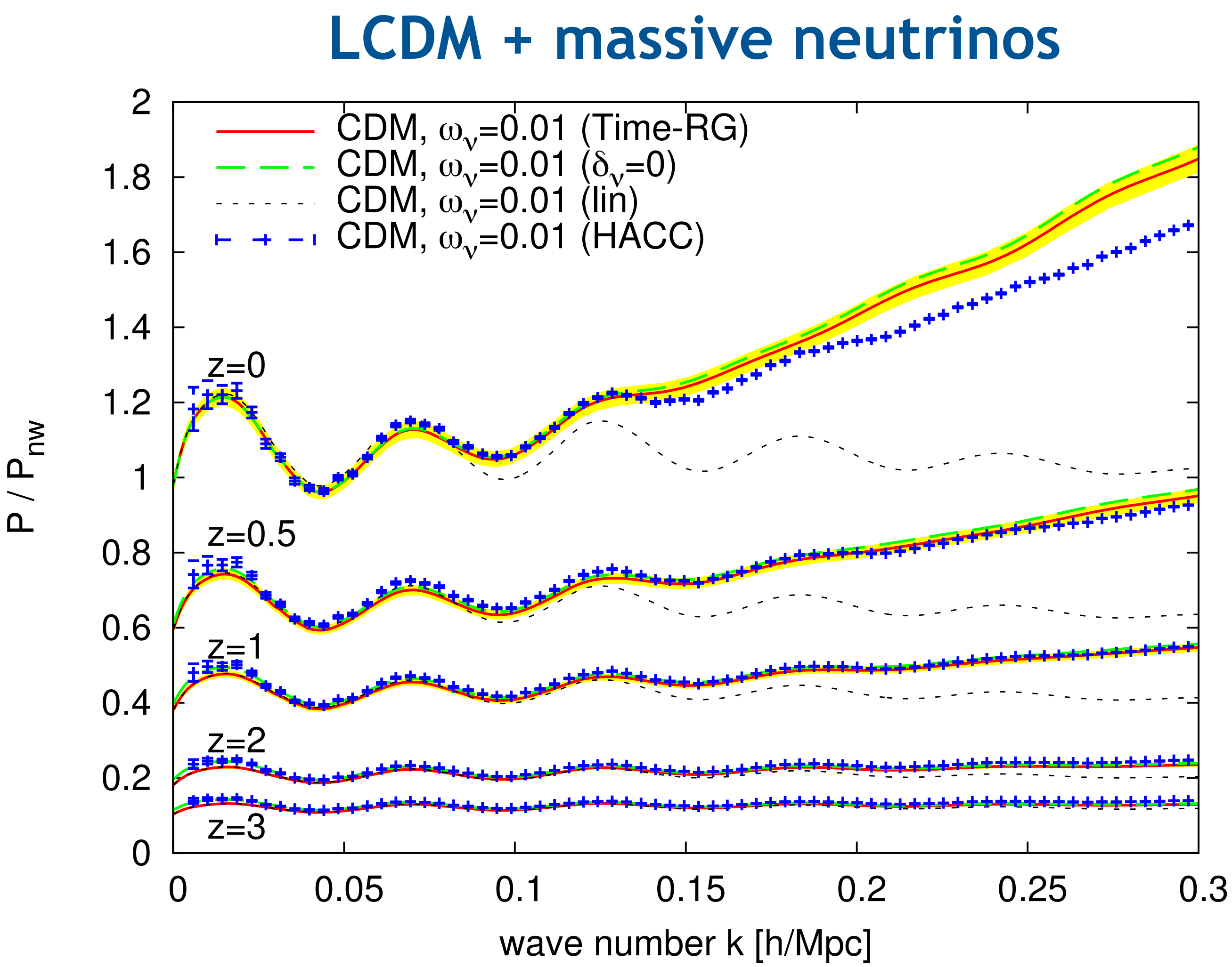
Early dark energy + massive neutrinos

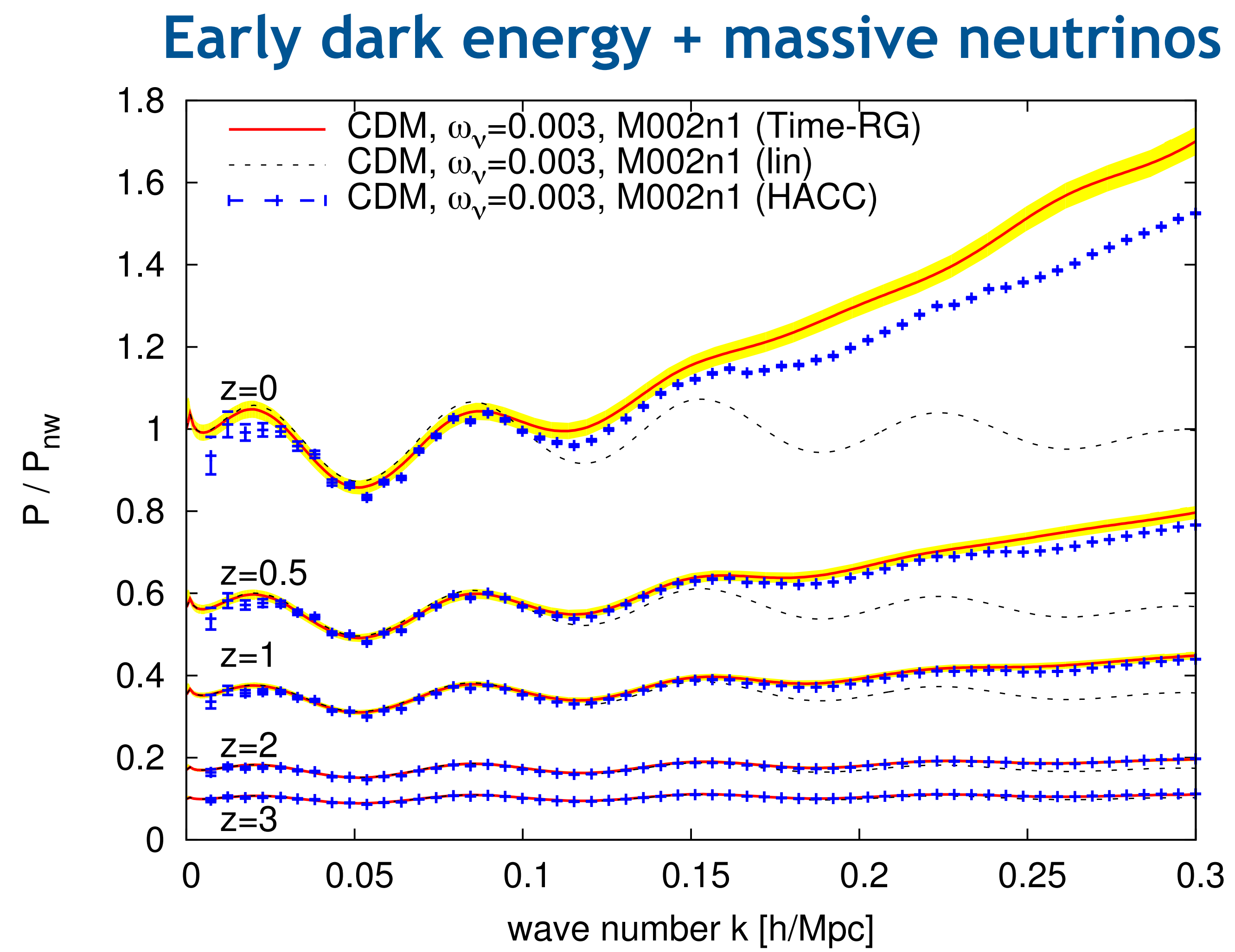


- Investigate different physical effects on e.g. the matter power spectrum at high accuracy
- Here: 2100 Mpc simulations with ~32.8 billion particles, $10^{10}\mathrm{M}_{\odot}$ particle mass to study massive neutrinos as well as time varying dark energy EOS
- Develop “higher order” perturbation theory approach

A. Upadhye et al. 2013

Some HACC Science Results --



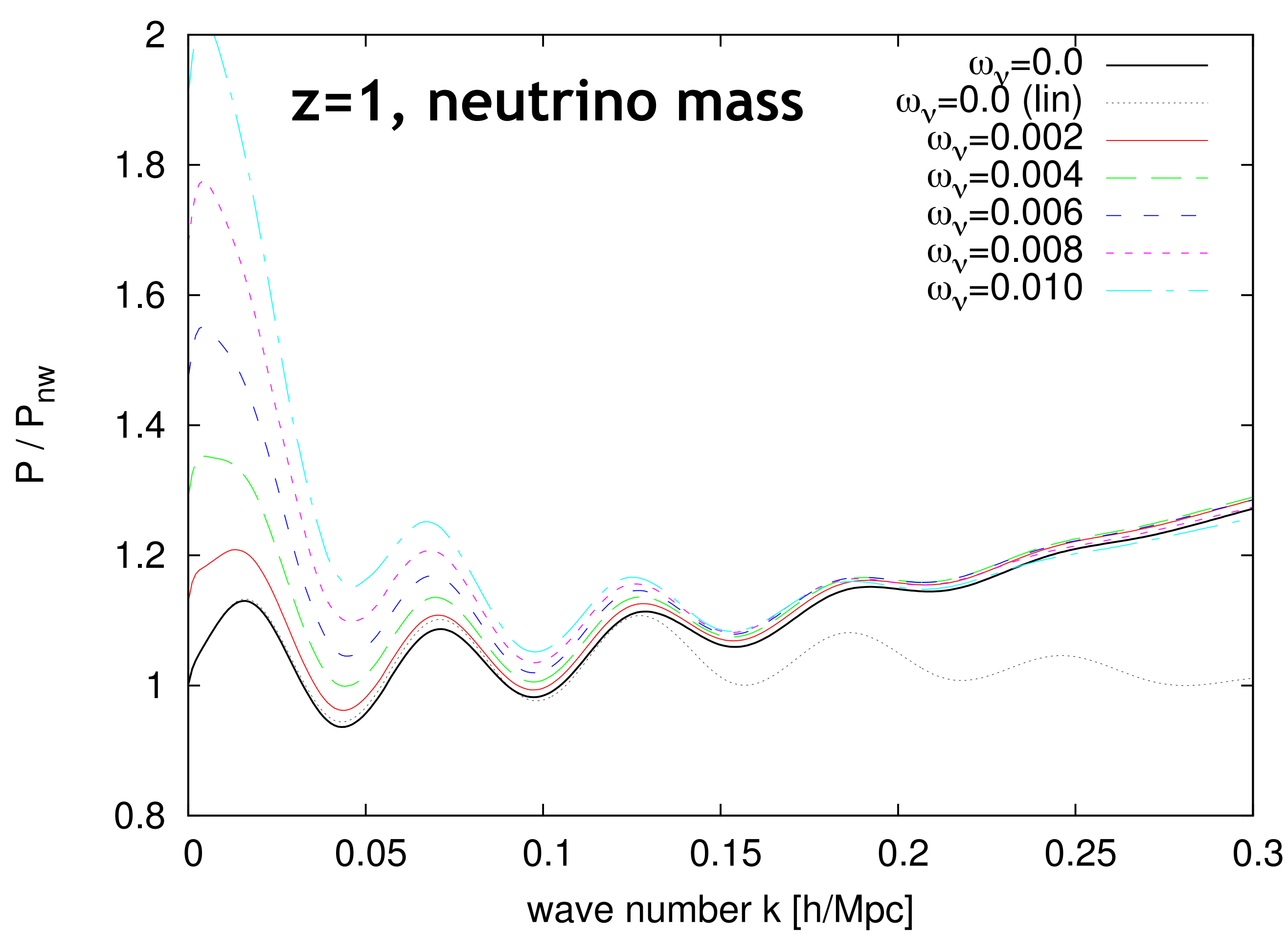


- Investigate different physical effects on e.g. the matter power spectrum at high accuracy
- Here: 2100 Mpc simulations with ~32.8 billion particles, $10^{10}\mathrm{M}_\odot$ particle mass to study massive neutrinos as well as time varying dark energy EOS
- Develop "higher order" perturbation theory approach

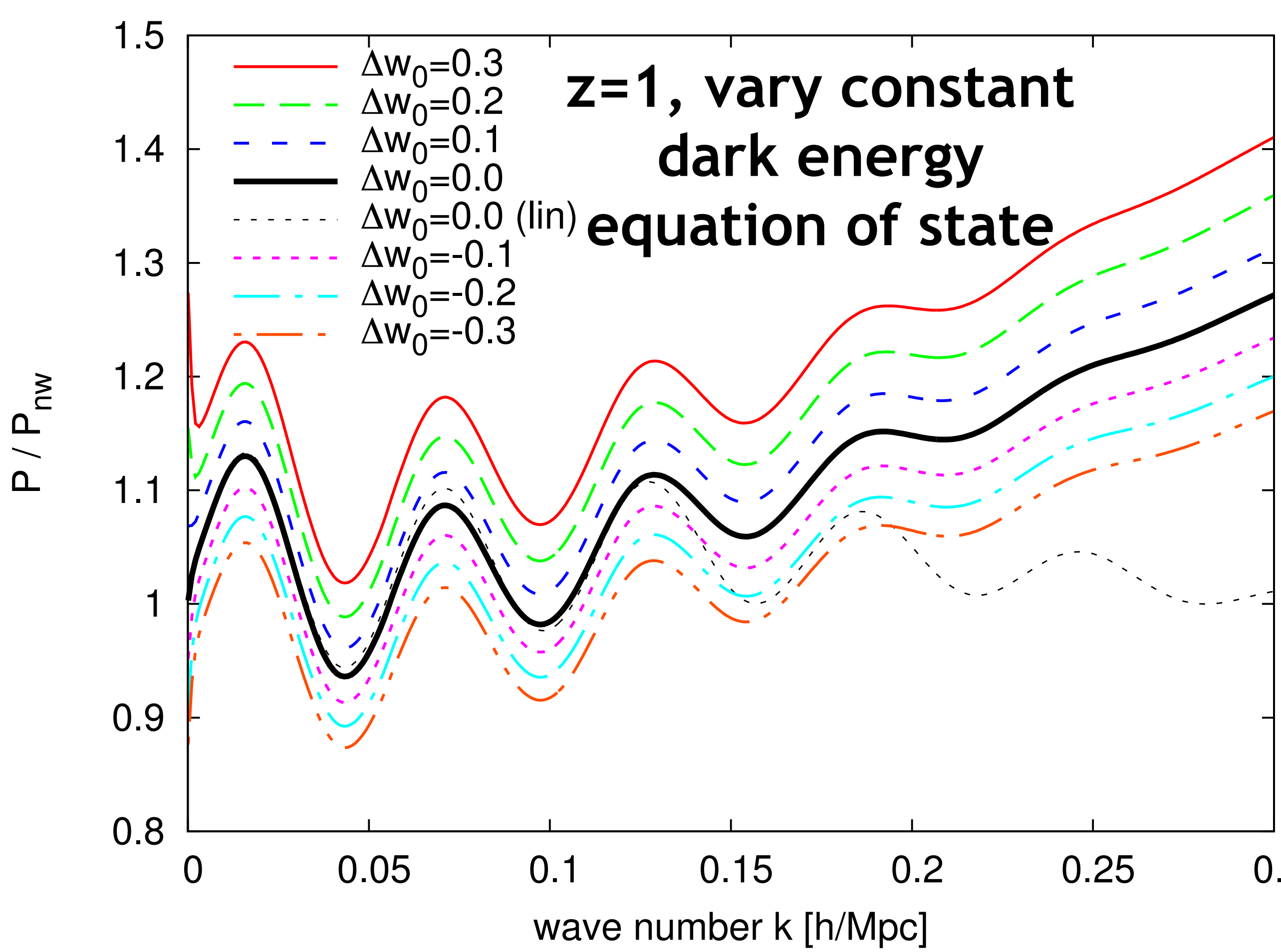
A. Upadhye et al. 2013

Some HACC Science Results --

LCDM + massive neutrinos



Early dark energy + massive neutrinos

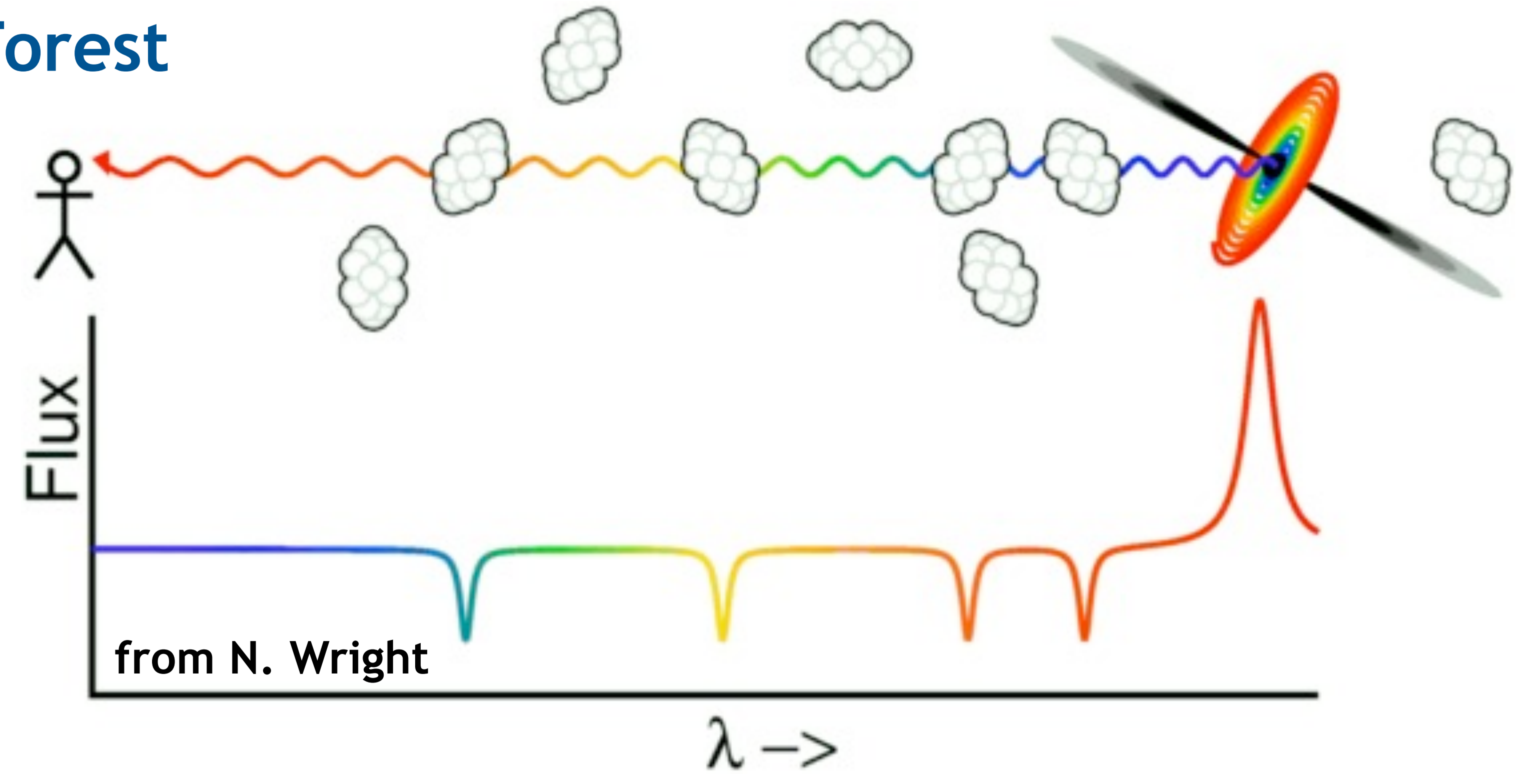


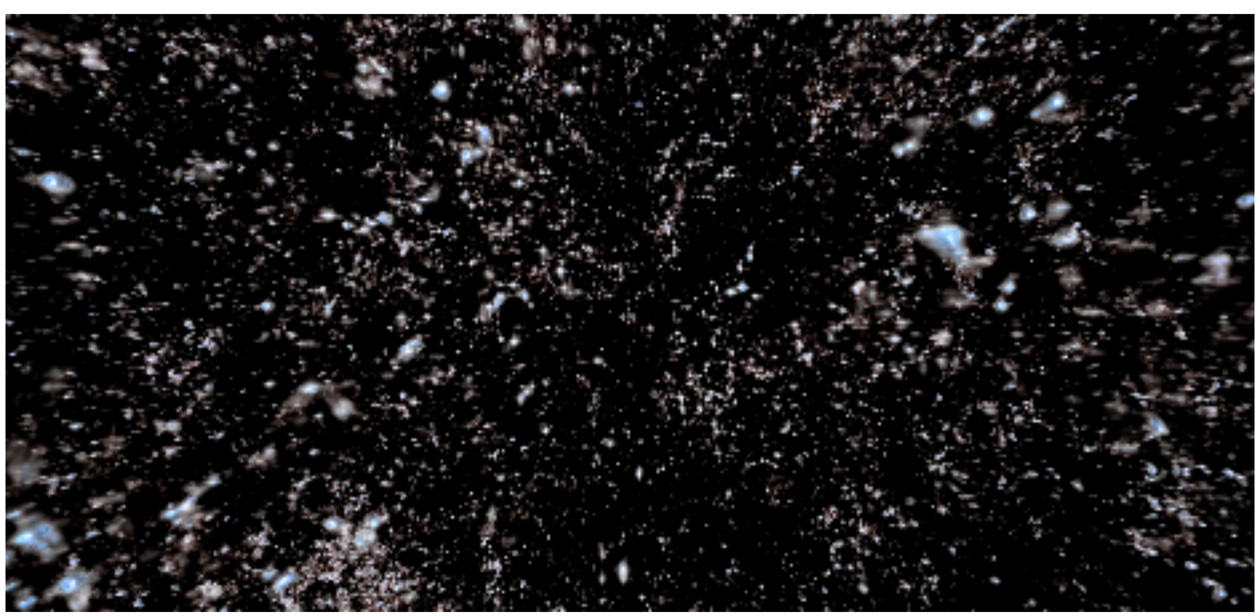
- Investigate different physical effects on e.g. the matter power spectrum at high accuracy
- Here: 2100 Mpc simulations with ~32.8 billion particles, $10^{10}\mathrm{M}_\odot$ particle mass to study massive neutrinos as well as time varying dark energy EOS
- Develop "higher order" perturbation theory approach

A. Upadhye et al. 2013

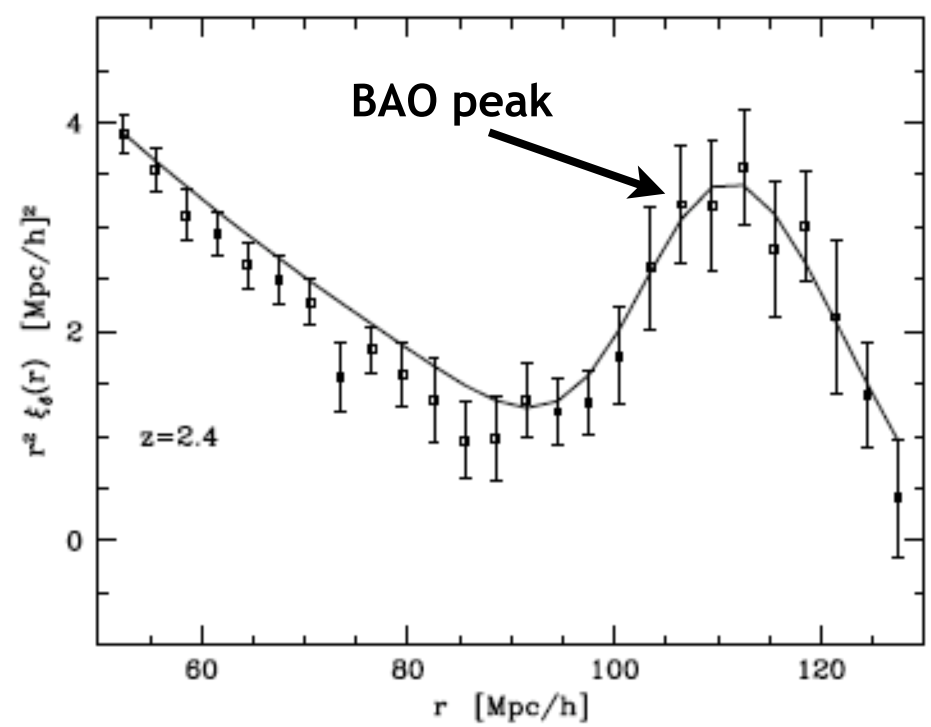
Some HACC Science Results --

Lyman-alpha Forest





Roadrunner view (halos) of the Universe at z=2 from a 64 billion particle run (9 runs on one weekend)



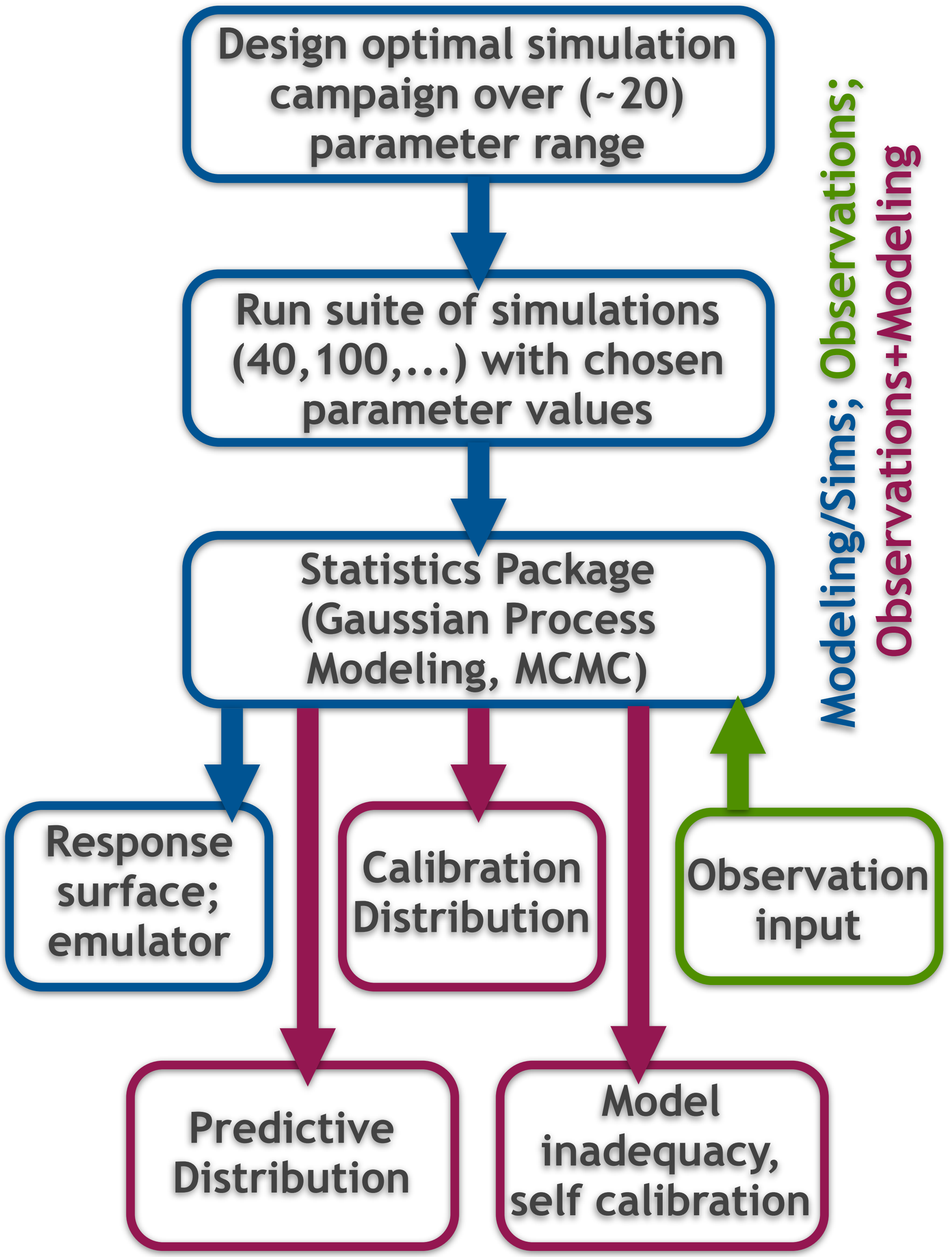
M. White et al. 2010

Cosmic Calibration: Solving the Inverse Problem

- **HACC works great!** But still cannot generate simulations in seconds ...
- **Challenge:** To extract cosmological constraints from observations in nonlinear regime, need to run Marko Chain Monte Carlo code; input: 10,000 - 100,000 different models
- **Current strategy:** Fitting functions for e.g. P(k), accurate at 10% level, as we saw this is not good enough!
- **Our alternative:** Emulators, fast prediction schemes built from a manageable set of simulations
- **Example here: Power spectrum emulator**
  - Step 1: Show simulations have required accuracy (Heitmann et al. 2005, 2008, 2010)
  - Step 2: Determine minimum number of simulations needed and develop sophisticated interpolation scheme that provides the power spectrum for any cosmology within a given parameter space prior (Heitmann et al. 2006, 2009; Habib et al. 2007)
  - Step 3: Carry out simulation and build final emulator (Lawrence et al. 2010, Heitmann et al. 2013)

Cosmic Calibration Framework

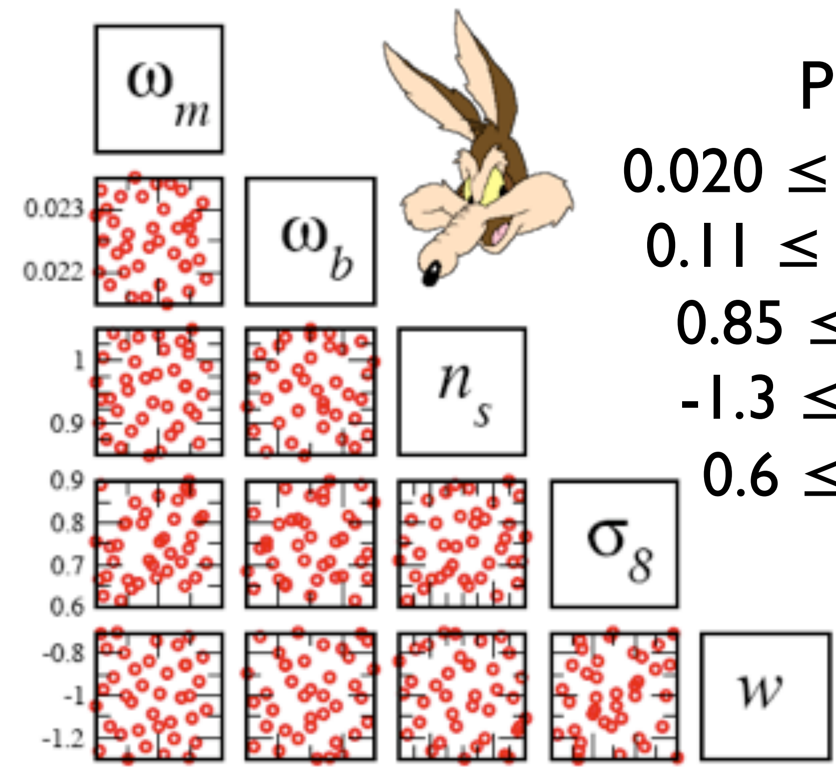
- Step 1: Design simulation campaign, rule of thumb: O(10) models for each parameter
- Step 2: Carry out simulation campaign and extract quantity of interest, in our case, power spectrum
- Step 3: Choose suitable interpolation scheme to interpolate between models, here Gaussian Processes
- Step 4: Build emulator
- Step 5: Use emulator to analyze data, determine model inadequacy, refine simulation and modeling strategy...



The Simulation Design for wCDM Cosmologies

- "Simulation design": For a given set of parameters to be varied and a fixed number of runs, at what settings should the simulations be performed?
- Example: Five cosmological parameters, tens of high-resolution runs are affordable
- First idea: Grid
  - Space filling but poor projection properties
- Second idea: Random sampling
  - Good projection properties but poor space coverage
- Our approach: Orthogonal-array Latin hypercubes (OA-LH) design
  - Stratified random sampling approach
  - Good projection properties AND space filling

The Coyote Universe



Priors:

$0.020 \leq \omega_b \leq 0.025$

$0.11 \leq \omega_m \leq 0.15$

$0.85 \leq n_s \leq 1.05$

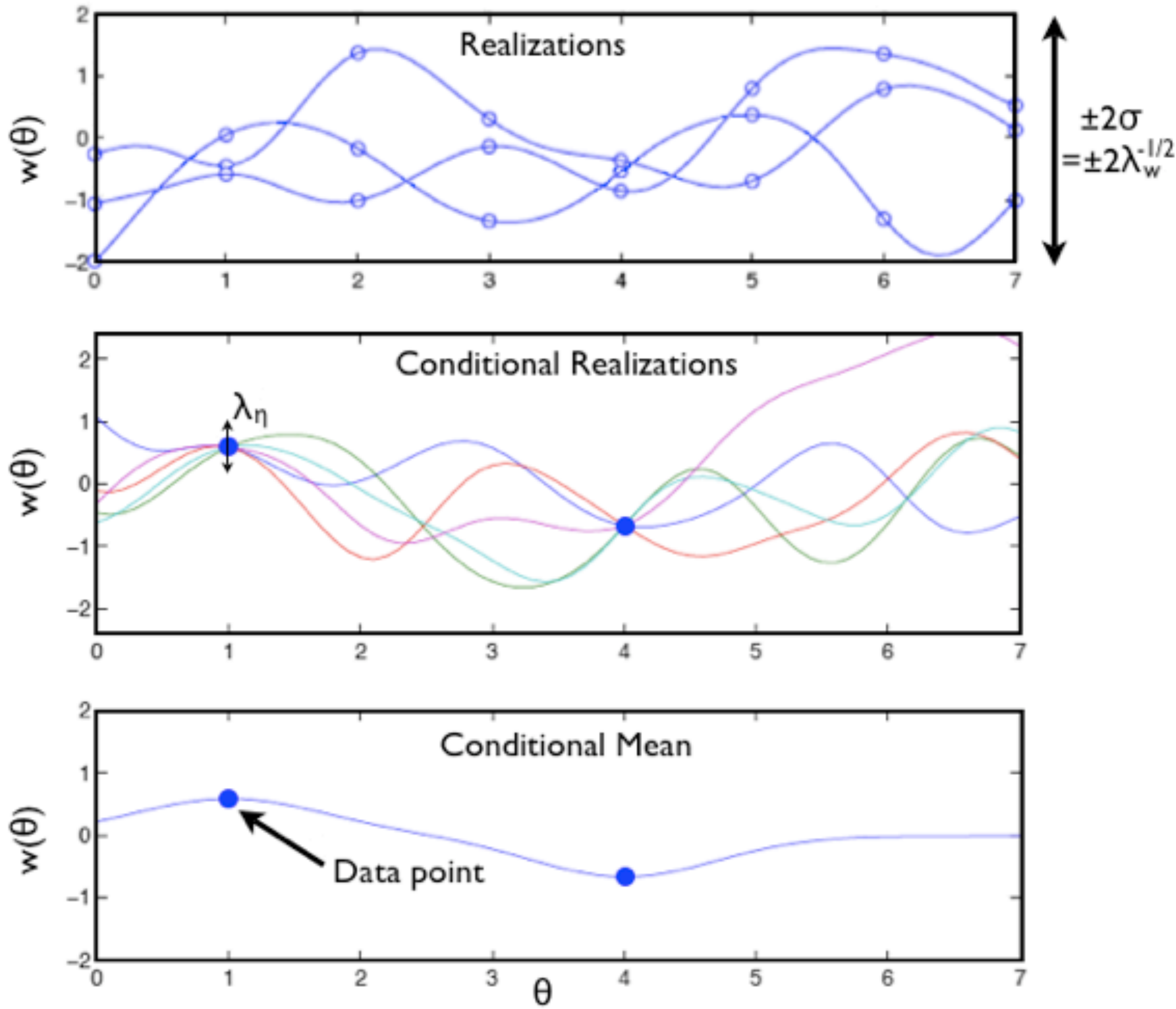
$-1.3 \leq w \leq -0.7$

$0.6 \leq \sigma_8 \leq 0.9$

Priors are informed by current cosmological constraints, the tighter the priors, the easier to build a prediction tool. Restriction in number of parameters also helps!

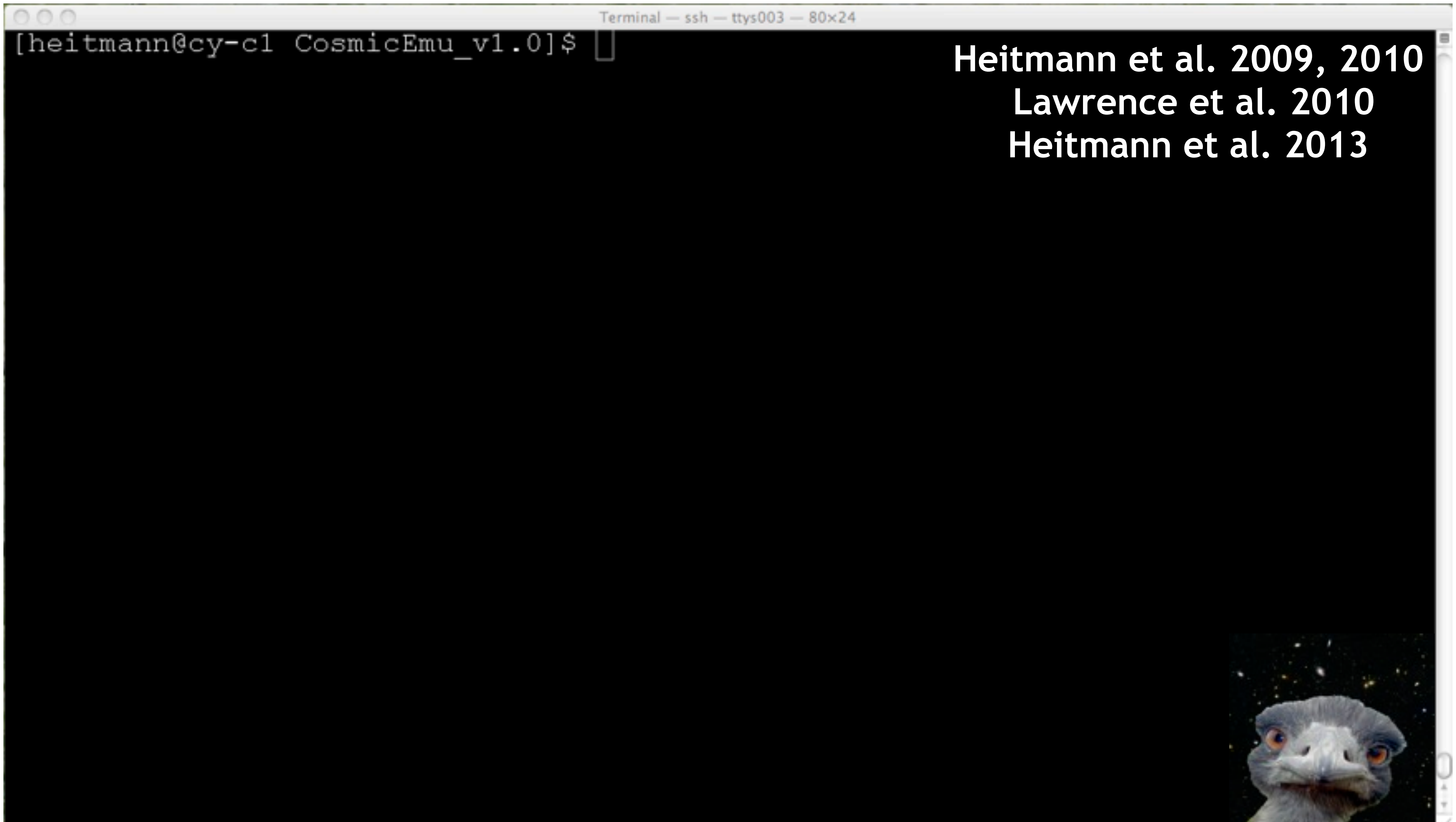
The Interpolation Scheme: Gaussian Processes + PCA

- After simulation design specification: Build non-parametric interpolation scheme
- Gaussian Process (GP): fits in function space
- GP involves matrix inversion in conditioning step ("curse of dimensionality")
- Data compression: Express power spectra in terms of principal component (PC) basis (can use other basis too)
- GP over over PC coefficients



Cosmic Emulator in Action

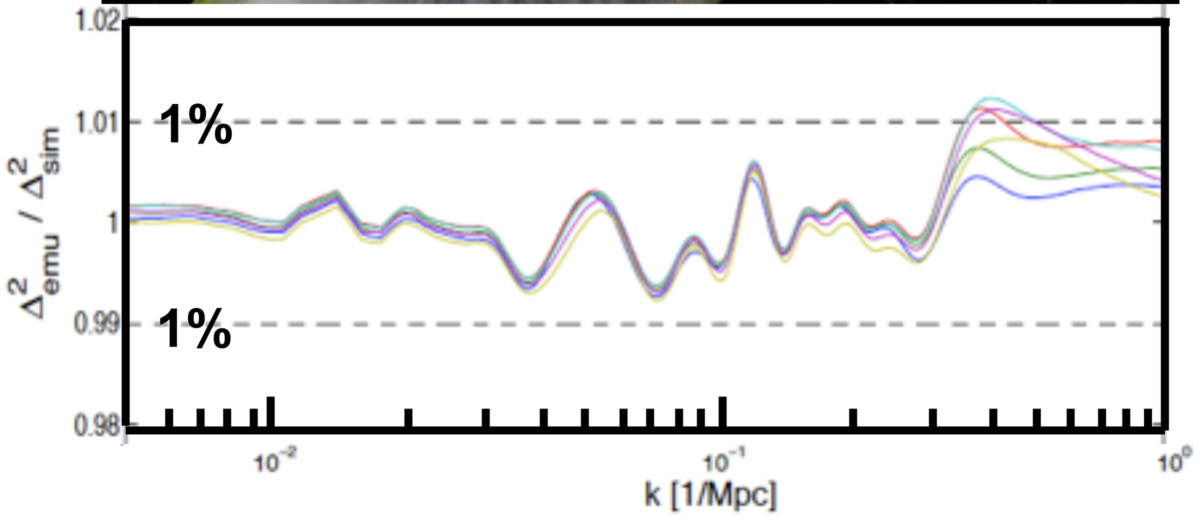
- Instantaneous 'oracle' for nonlinear power spectrum, reduces compute time from weeks to negligible, accurate at 1% out to k~1/Mpc for wCDM cosmologies
- Enables direct MCMC with results from full simulations



The Cosmic Emu(lator)

- Prediction tool for matter power spectrum has been constructed
- Accuracy within specified priors between z=0 and z=1 out to k=1 *h*/Mpc at the 1% level achieved
- Emulator has been publicly released, C code
- Extension: Include *h* as sixth parameter, out to k=10 *h*/Mpc and z=4
  - Nested simulations to cover large k-range
  - Approach degrades accuracy to ~3%

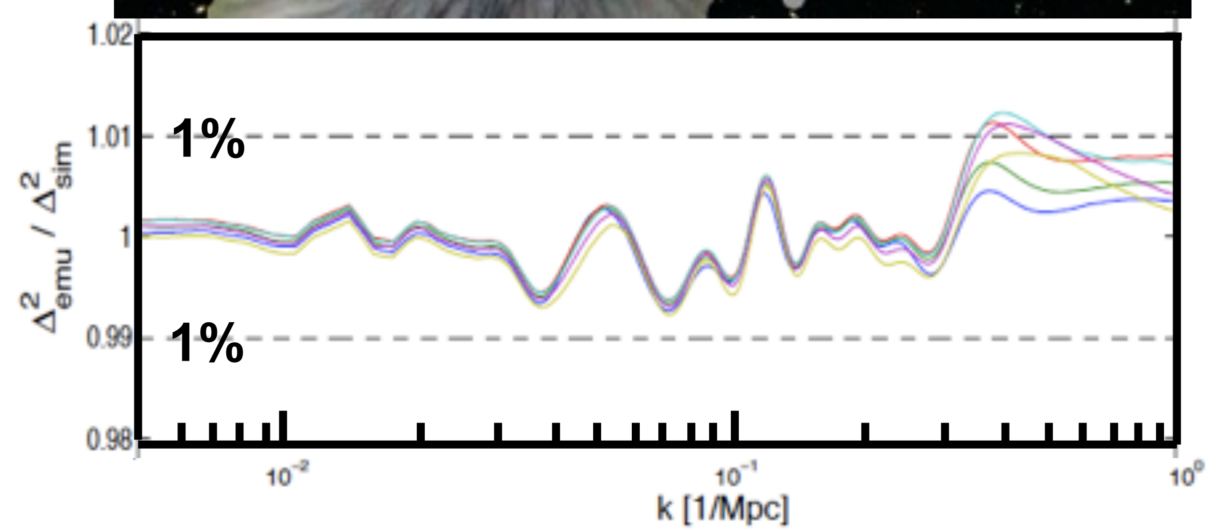
Emulator performance: Comparison of prediction and simulation output for a model not used to build emulator at 6 redshifts.



The Cosmic Emu(lator)

- Prediction tool for matter power spectrum has been constructed
- Accuracy within specified priors between z=0 and z=1 out to k=1 *h*/Mpc at the 1% level achieved
- Emulator has been publicly released, C code
- Extension: Include *h* as sixth parameter, out to k=10 *h*/Mpc and z=4
  - Nested simulations to cover large k-range
  - Approach degrades accuracy to ~3%

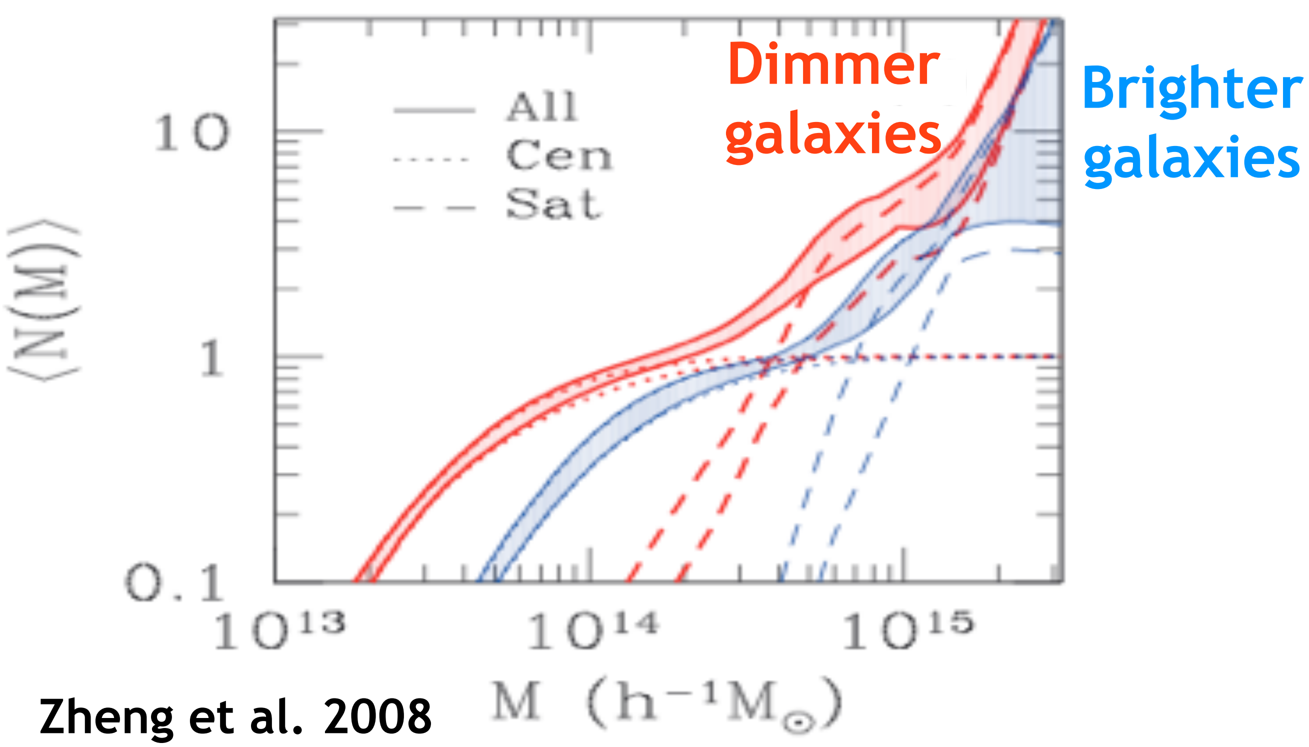
Emulator performance: Comparison of prediction and simulation output for a model not used to build emulator at 6 redshifts.



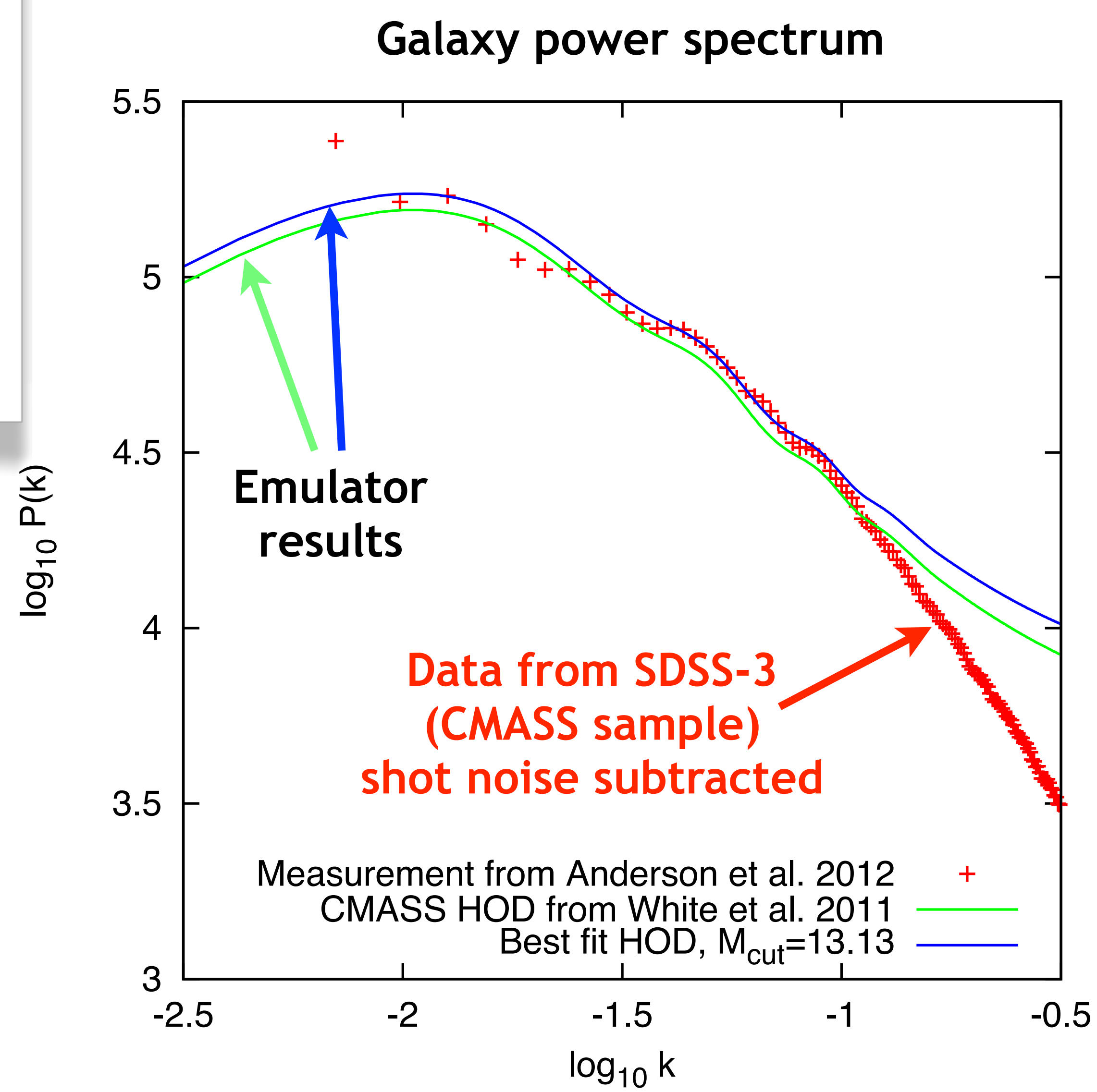
Emulating the Galaxy Power Spectrum

- Idea: find halos in the simulation, depending on their mass, populate them with galaxies
- Above mass threshold: each halo hosts central galaxy; the heavier the halo, the more satellites
- 5 parameter model, parameters adjusted by comparison to observations, different classes of galaxies = different models

$$N_{\rm cen}(M) = \frac{1}{2}{\rm erfc}\left[\frac{\ln(M_{\rm cut}/M)}{\sqrt{2}\sigma}\right] \qquad N_{\rm sat}(M) = \left(\frac{M-\kappa M_{\rm cut}}{M_1}\right)^{\alpha}$$



Zheng et al. 2008



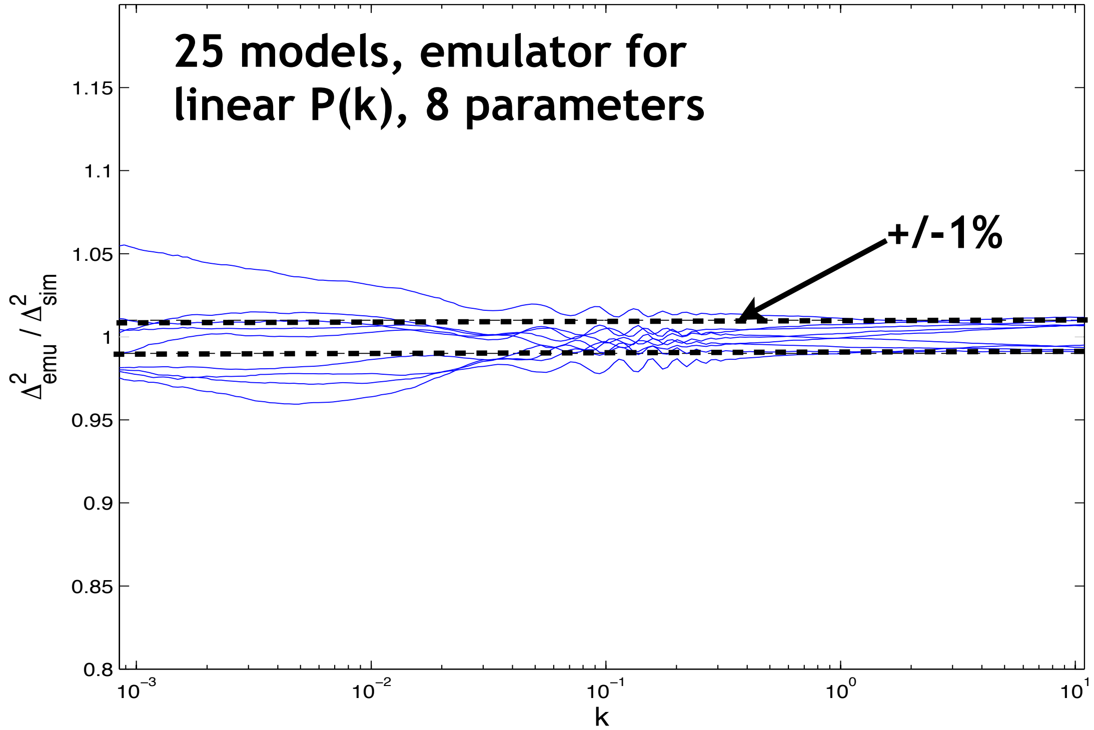
Kwan et al. 2013

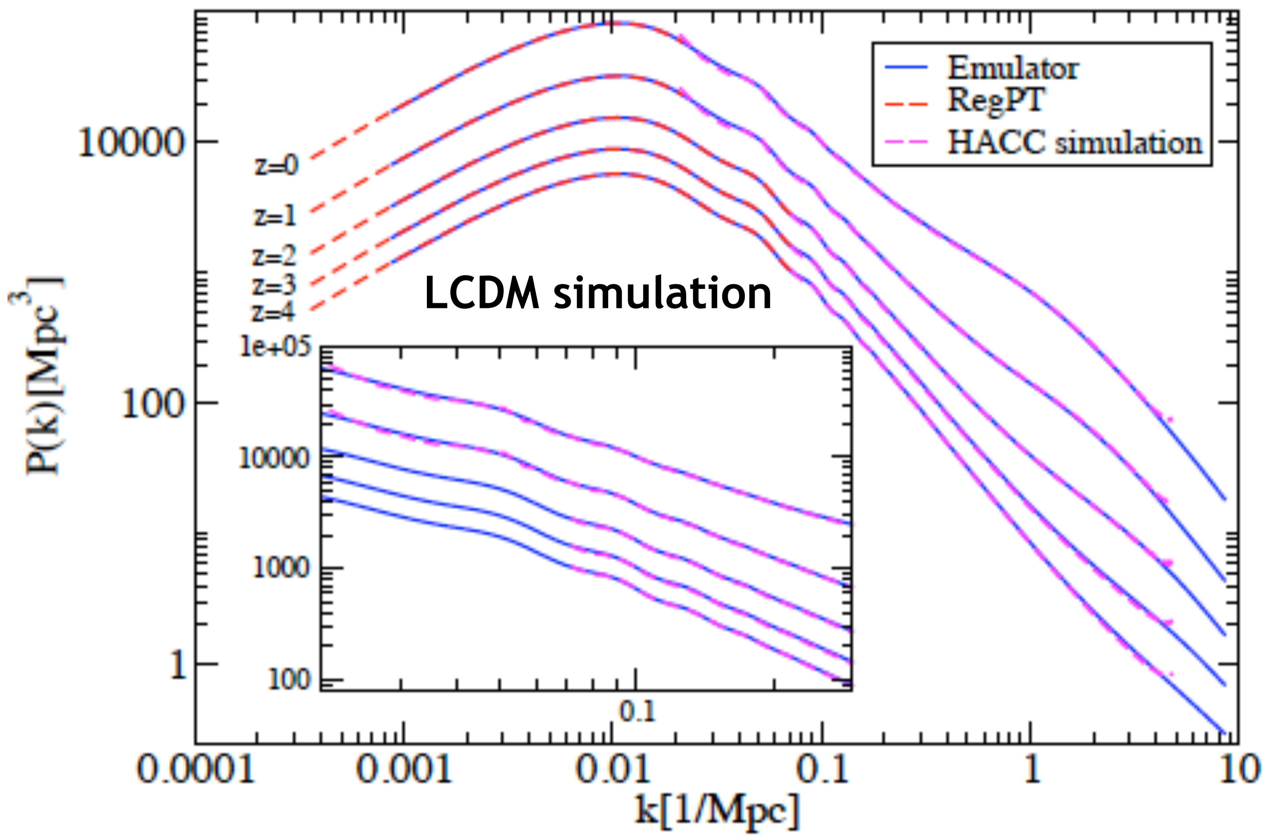
The Next Step: The Mira Universe

- Extend parameter space to include varying w(z) and massive neutrinos
- Build "nested designs": enable to build emulator from first set of 25 models, improve with additional 27 models, final precision with 99 models overall
- Various emulators for P(k), mass function, c-M relation, derived quantities...
- LCDM done, finalizing set-up based on this run

Parameters

$$\begin{aligned}
0.12 &\leq \omega_m \leq 0.155 \\
0.0215 &\leq \omega_b \leq 0.0235 \\
0.7 &\leq \sigma_8 \leq 0.9 \\
0.55 &\leq h \leq 0.85 \\
0.85 &\leq n_s \leq 1.05 \\
-1.3 &\leq w_0 \leq -0.7 \\
-1.5 &\leq w_a \leq 1.15 \\
0.0 &\leq \omega_\nu \leq 0.01.
\end{aligned}$$



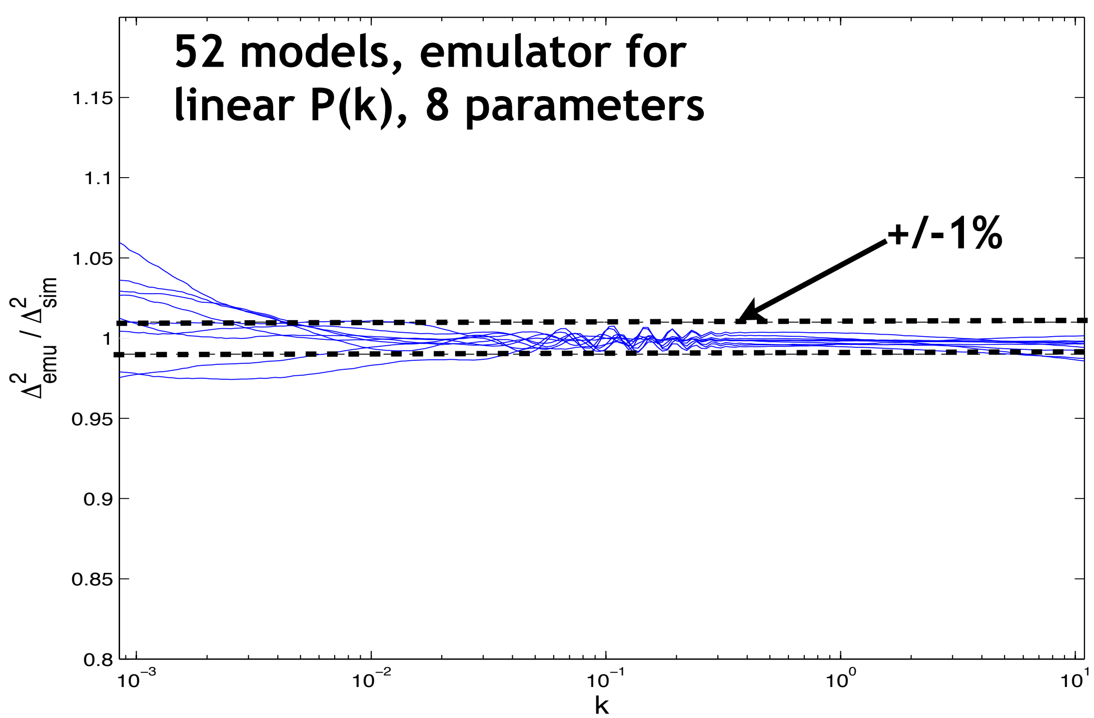


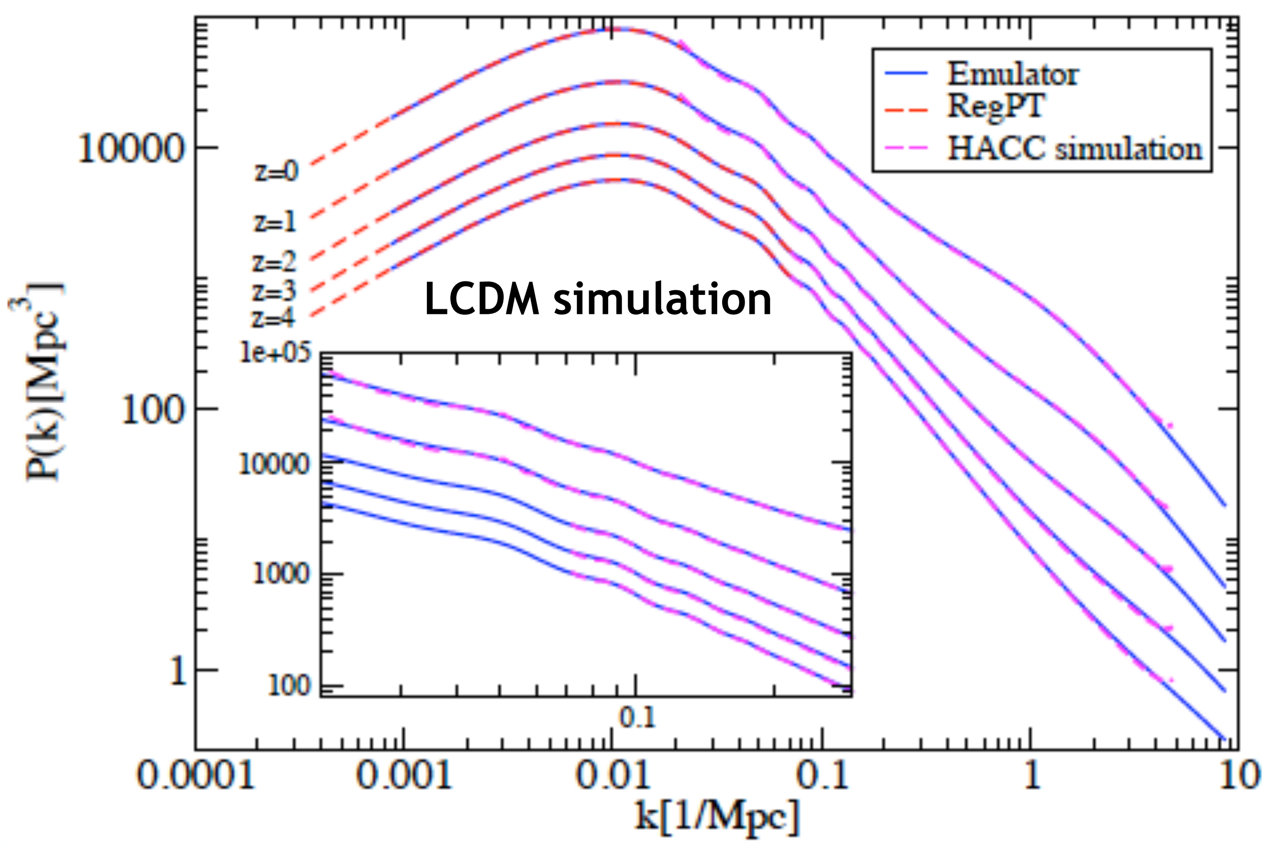
The Next Step: The Mira Universe

- Extend parameter space to include varying w(z) and massive neutrinos
- Build "nested designs": enable to build emulator from first set of 25 models, improve with additional 27 models, final precision with 99 models overall
- Various emulators for P(k), mass function, c-M relation, derived quantities...
- LCDM done, finalizing set-up based on this run

Parameters

$$
\begin{aligned}
0.12 \le &\ \omega_m \le 0.155\\
0.0215 \le &\ \omega_b \le 0.0235\\
0.7 \le &\ \sigma_8 \le 0.9\\
0.55 \le &\ h \le 0.85\\
0.85 \le &\ n_s \le 1.05\\
-1.3 \le &\ w_0 \le -0.7\\
-1.5 \le &\ w_a \le 1.15\\
0.0 \le &\ \omega_\nu \le 0.01.
\end{aligned}
$$



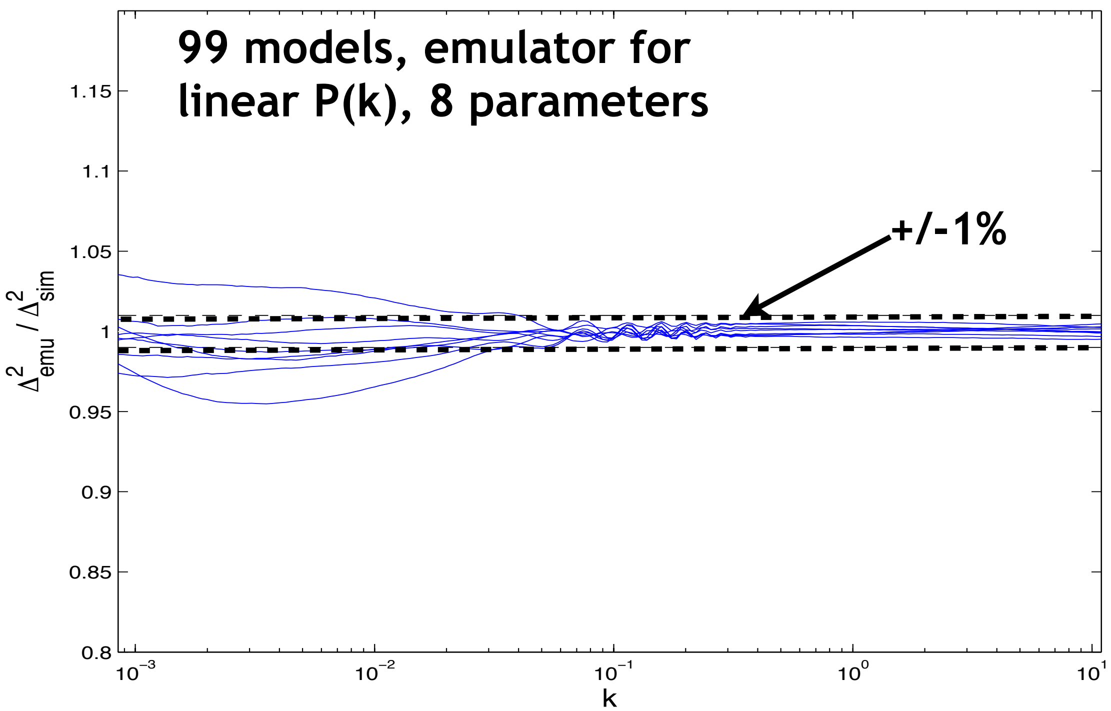


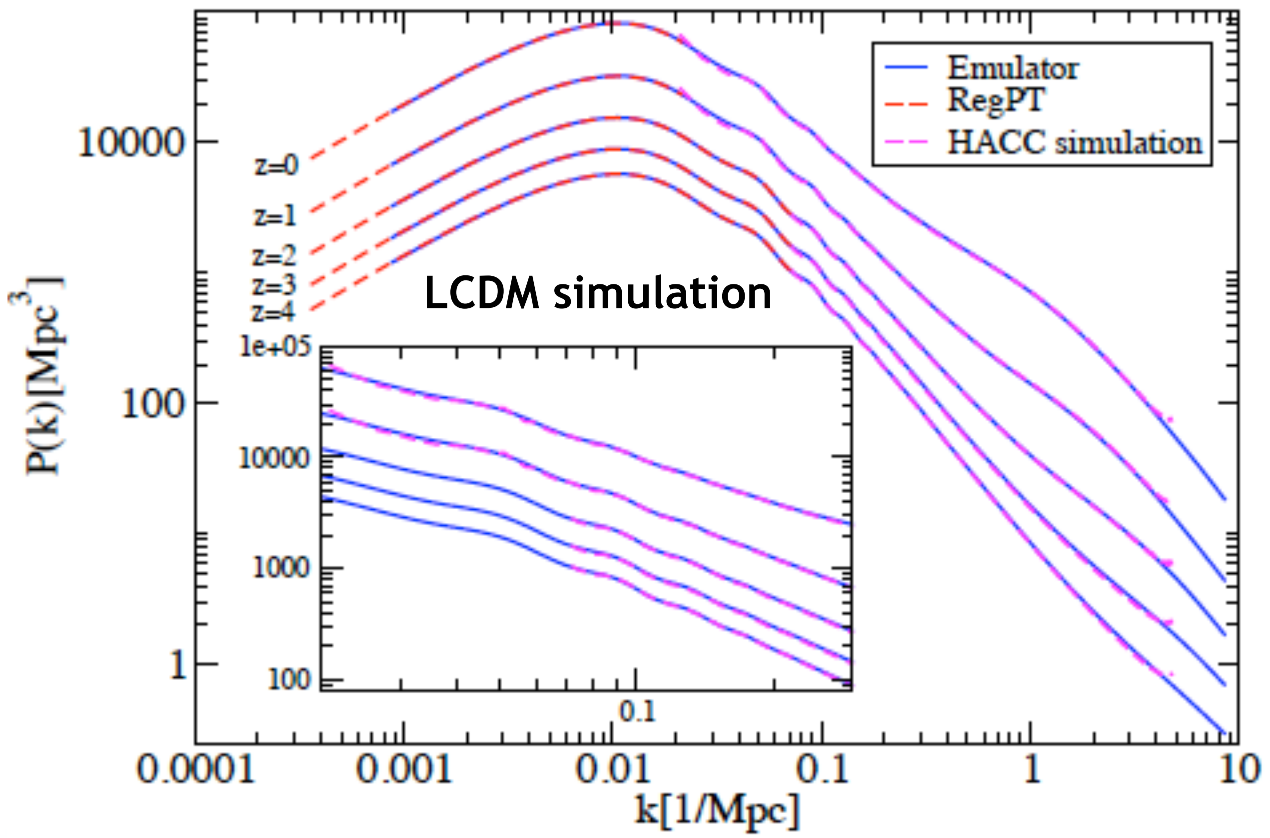
The Next Step: The Mira Universe

- Extend parameter space to include varying w(z) and massive neutrinos
- Build “nested designs”: enable to build emulator from first set of 25 models, improve with additional 27 models, final precision with 99 models overall
- Various emulators for P(k), mass function, c-M relation, derived quantities...
- LCDM done, finalizing set-up based on this run

Parameters

$$\begin{aligned}
0.12 \leq \omega_m &\leq 0.155 \\
0.0215 \leq \omega_b &\leq 0.0235 \\
0.7 \leq \sigma_8 &\leq 0.9 \\
0.55 \leq h &\leq 0.85 \\
0.85 \leq n_s &\leq 1.05 \\
-1.3 \leq w_0 &\leq -0.7 \\
-1.5 \leq w_a &\leq 1.15 \\
0.0 \leq \omega_\nu &\leq 0.01.
\end{aligned}$$





Summary and Outlook

- **Precision cosmology needs high accuracy predictions! Can we avoid being theory limited?**
- **HACC is a large scale computational tool to address this challenge**
- **Cosmic Calibration Framework allows us to build fast prediction tools for ongoing and future surveys**
- **Largest ever high-resolution run, *The Outer Rim Simulation*, currently running, analysis ongoing**
  - Many hurdles had to be overcome to make this happen, including memory management, I/O, adaptive time stepper, and analysis tools
  - Exciting science extracted, e.g. strong lensing predictions
- ***The Mira Universe* will lead to an unprecedented set of simulations, spanning 8 cosmological parameters, including different dark energy models and neutrinos**
- **In this talk: Focus on power spectrum science but many more science results have already been extracted from HACC simulations**