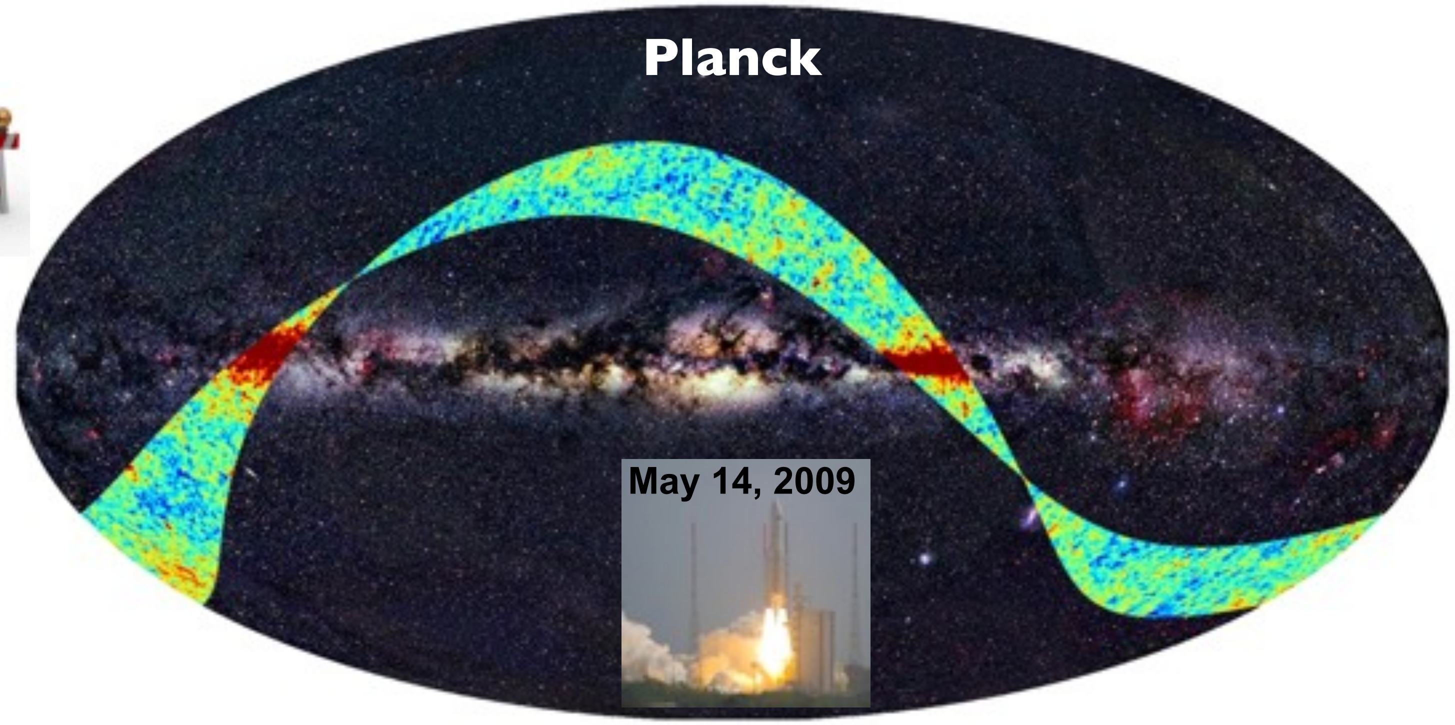
The Coyote Universe

Katrin Heitmann, ISR-1, LANL

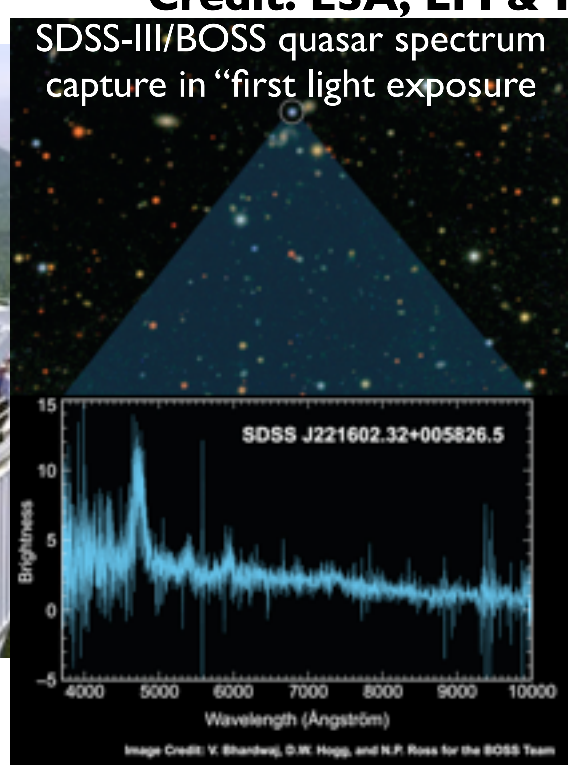
In collaboration with:
Jim Ahrens, Suman Bhattacharya, Salman Habib, David Higdon, Chung Hsu, Earl Lawrence, Zarija Lukic, Charlie Nakhleh, Paul Sathre, Christian Wagner, Martin White, Brian Williams

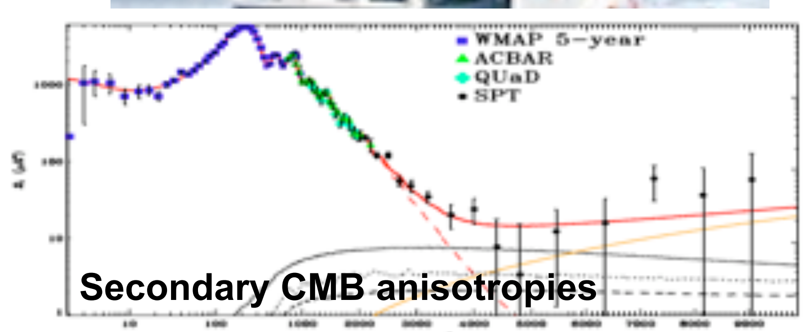
LA-URs: 09-06131, 08-07921, 08-05630

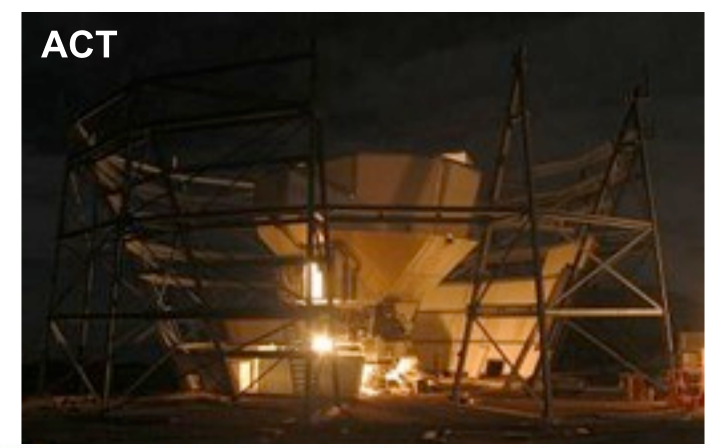
Exciting Times for Cosmology!

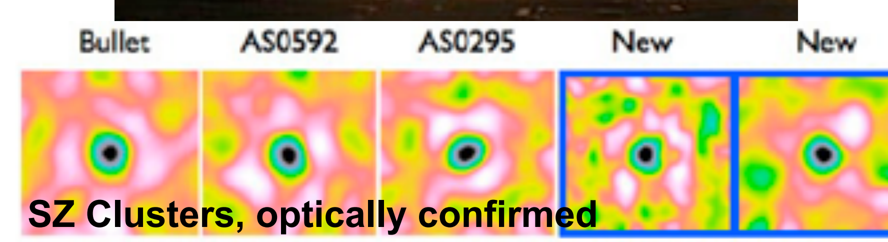


Credit: ESA, LFI & HFI Consortia. Background optical image: Axel Mellinger



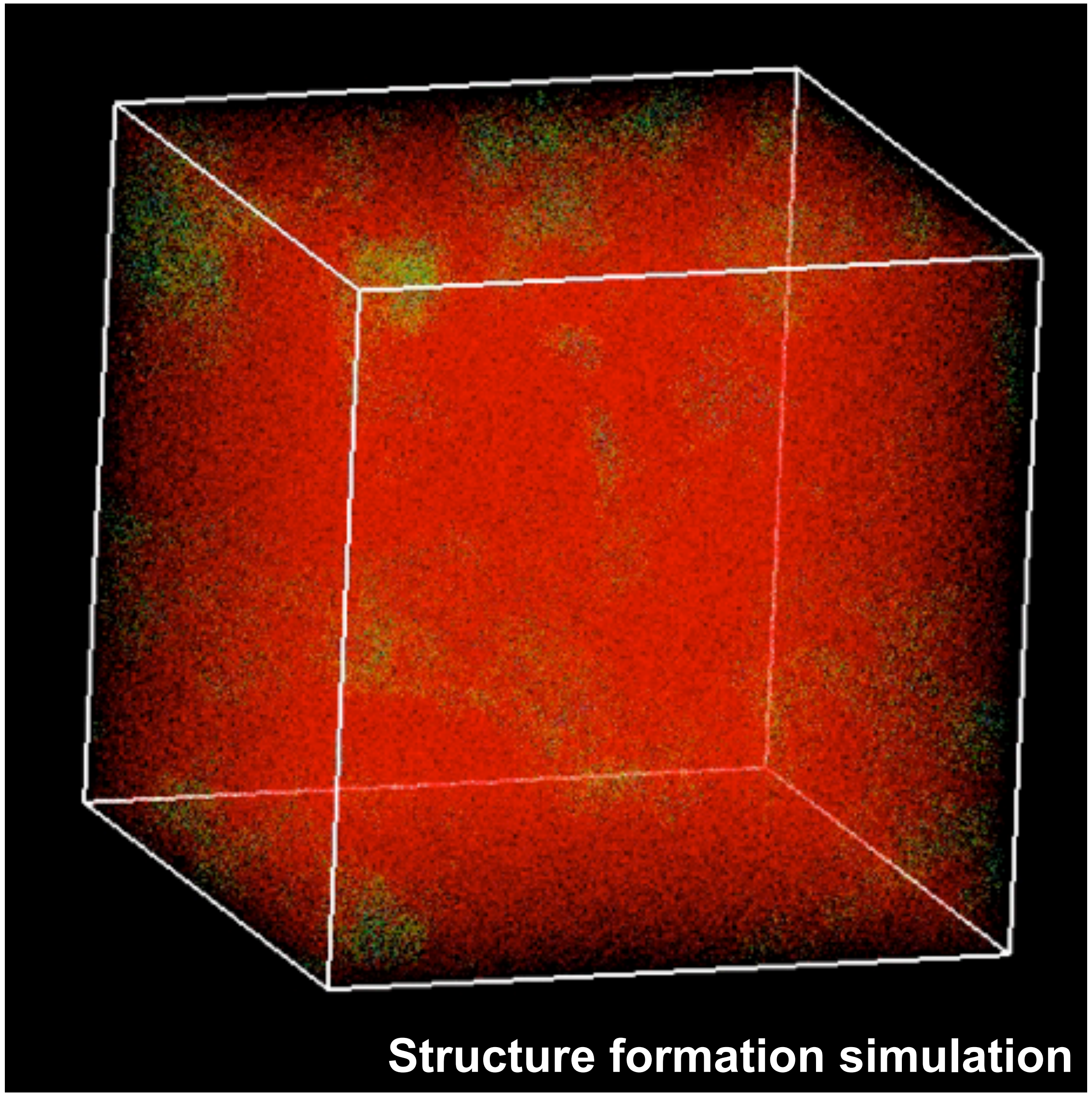




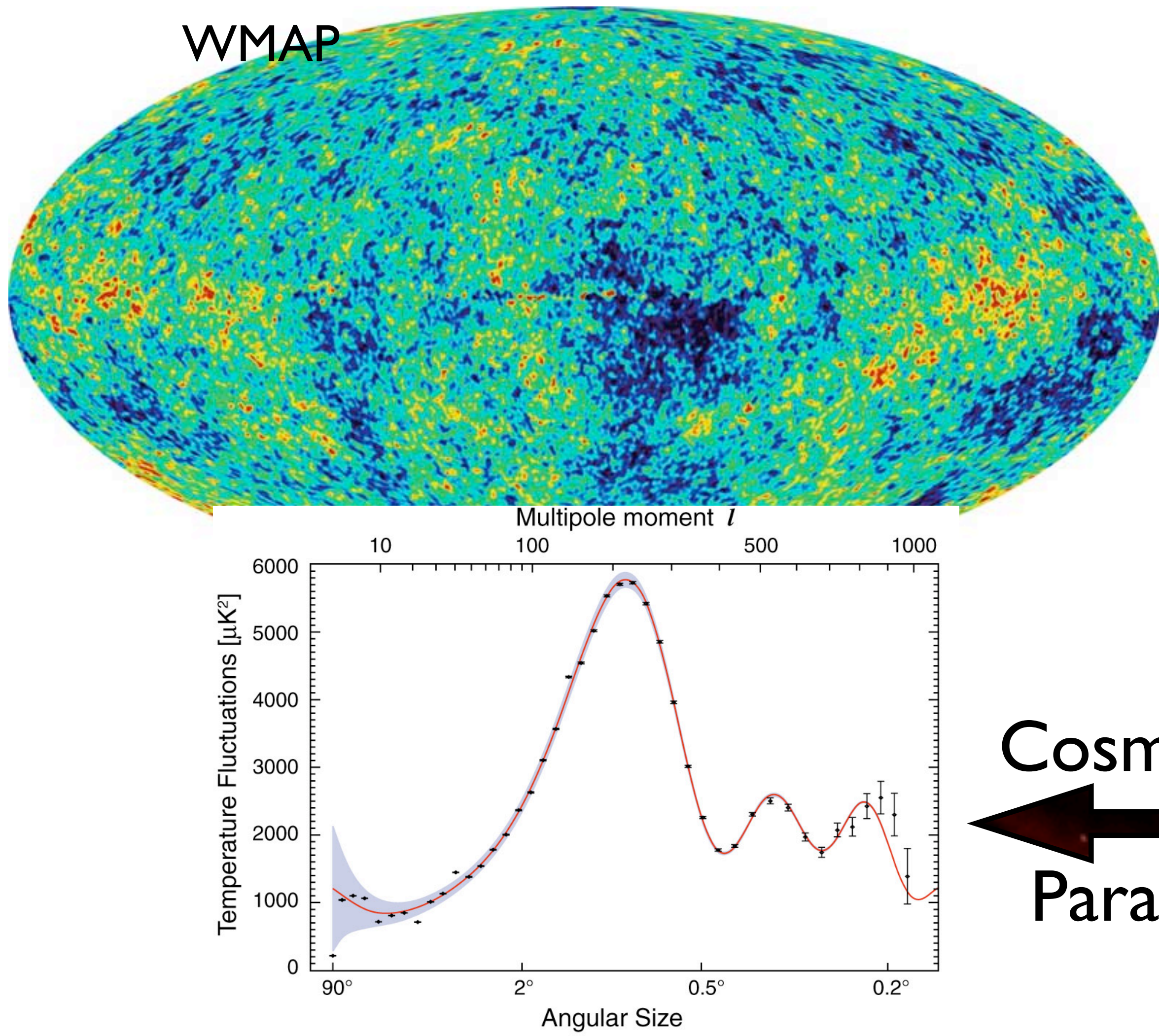


Physics from the Matter Power Spectrum

- **Nonlinear regime of structure formation as measured by galaxy surveys holds wealth of new information on cosmology**
  - Matter power spectrum P(k), 3-point function, mass function...
- **Example of physics beyond the Standard Model of Particle Physics and Cosmology**
  - Nature of dark matter
  - Nature of dark energy
  - Properties of neutrinos
- **Probes of P(k) at small scales**
  - Galaxy power spectrum, difficulty: bias, how to connect the light to the mass
  - Weak lensing, direct measurement of the mass distribution
  - Lyman-alpha forest, difficulty: gas physics

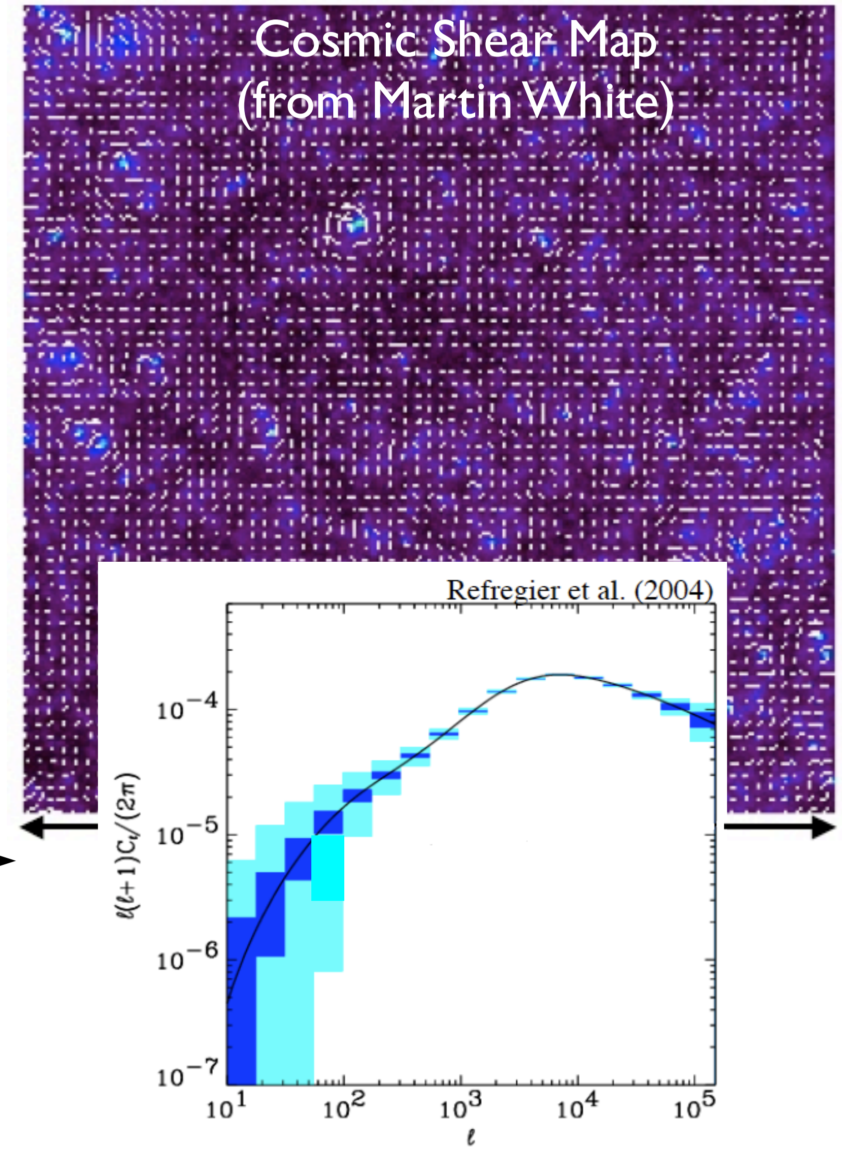


The Challenge



Cosmological

Parameters



- **For cosmological constraints**
  - Run your favorite MCMC code, e.g. CosmoMC
  - Need to calculate different statistics, e.g. P(k), ~10,000 - 100,000 times for different models
  - Current fitting functions for these statistics (tuned to simulations) accurate at the 10% level for different cosmologies, not good enough!
  - Brute force simulations: ~30 years on 2000 processor Beowulf Cluster...

The Challenge

Huterer & Takada (2005) on requirements for future weak lensing surveys: " While the power spectrum on relevant scales (0.1 < k [h/Mpc] < 10) is currently calibrated with N-body simulations to about 5-10%, in the future it will have to be calibrated to about 1-2% accuracy ..... These goals require a suite of high resolution N-body simulations on a relatively fine grid in cosmological parameter space, and should be achievable in the near future."

doubtful theorist

observer

(from S. Furlanetto)

J. Annis et al: Dark Energy Studies: Challenges to Computational Cosmology (2005): Dark energy studies will challenge the computational cosmology community to critically assess current techniques, develop new approaches to maximize accuracy, and establish new tools and practices to efficiently employ globally networked computing resources.......Code comparison projects should be more aggressively pursued and the sensitivity of key non-linear statistics to code control parameters deserves more careful systematic study......... Highly accurate dark matter evolution is only a first step

Precision Predictions for P(k)

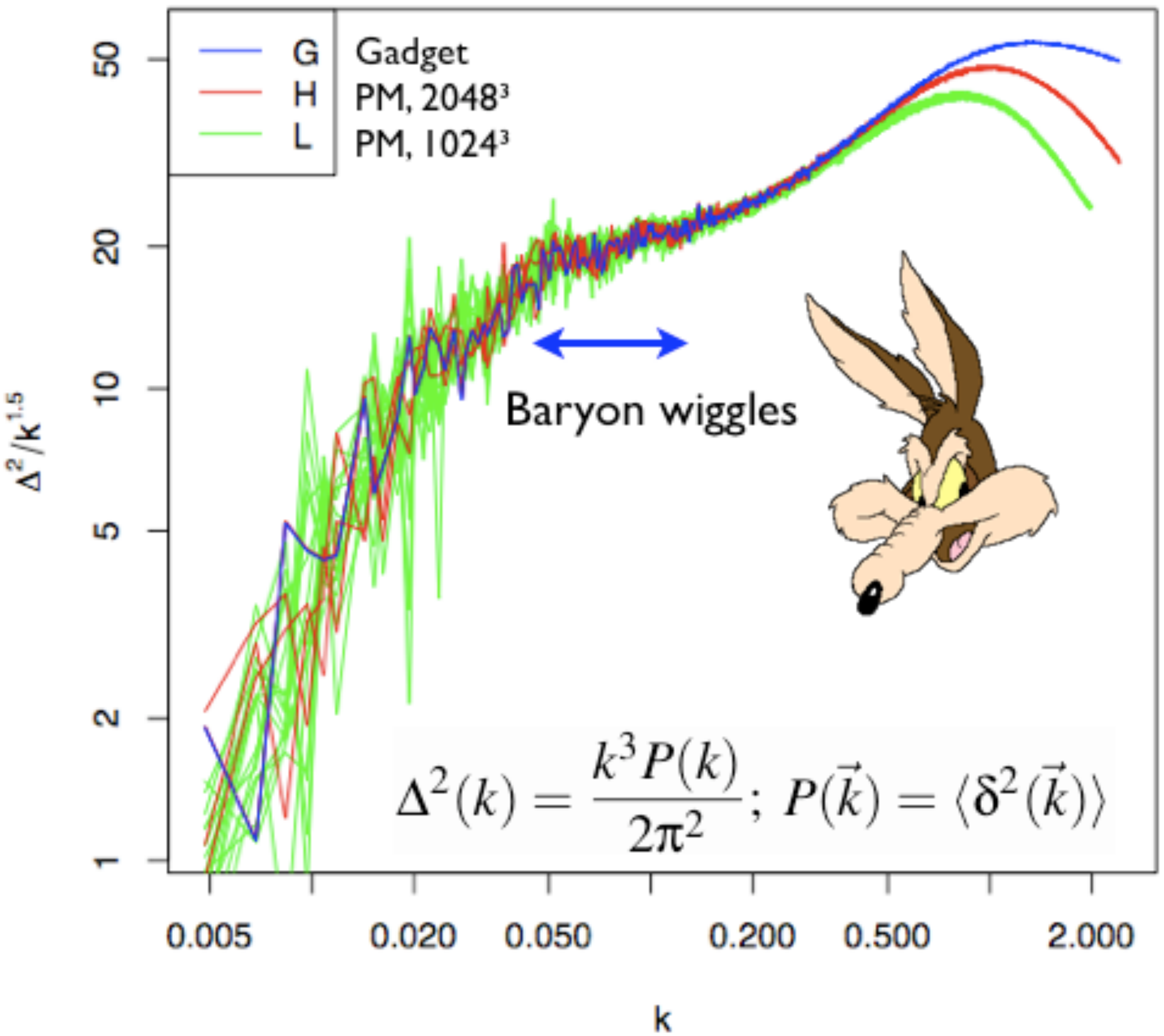
- **Aim: predict P(k) out to scales of k~1 h/Mpc at 1% accuracy between z=0 and z=1**
  - Regime of interest for current weak lensing surveys
  - Baryon physics at these scales sub-dominant
  - Dynamic range for simulations manageable
- **Step 1: Show that simulations can be run at the required accuracy (arXiv:0812.1052, Heitmann et al., ApJ 2010)**
  - Initial conditions, force and mass resolution, ...
  - Minimal requirement: 1 billion particles, 1.3 Gpc volume, 50 kpc force resolution, ~ 20,000 CPU hours, few days on 250 processors + wait time in queue ~ 1 week per simulation on “Coyote”, LANL cluster
- **Step 2: Cosmic Calibration Framework (arXiv:0902.0429, Heitmann et al., ApJ 2009)**
  - Build with a small number of high-precision simulations a prediction scheme (“emulator”) which provides the power spectrum for any cosmology within the parameter space under consideration
  - ~ 40 cosmological models sufficient
- **Step 3: Cosmic Emulator (arXiv:0912.4490, Lawrence et al., ApJ 2010)**
  - Carry out large number of simulations (~1,000) at varying resolution for 38 cosmologies, one high-resolution run per cosmology, emulator is “look-up” table
  - Emulator available at: www.lanl.gov/projects/cosmology/CosmicEmu

The Coyote Universe

Priors:

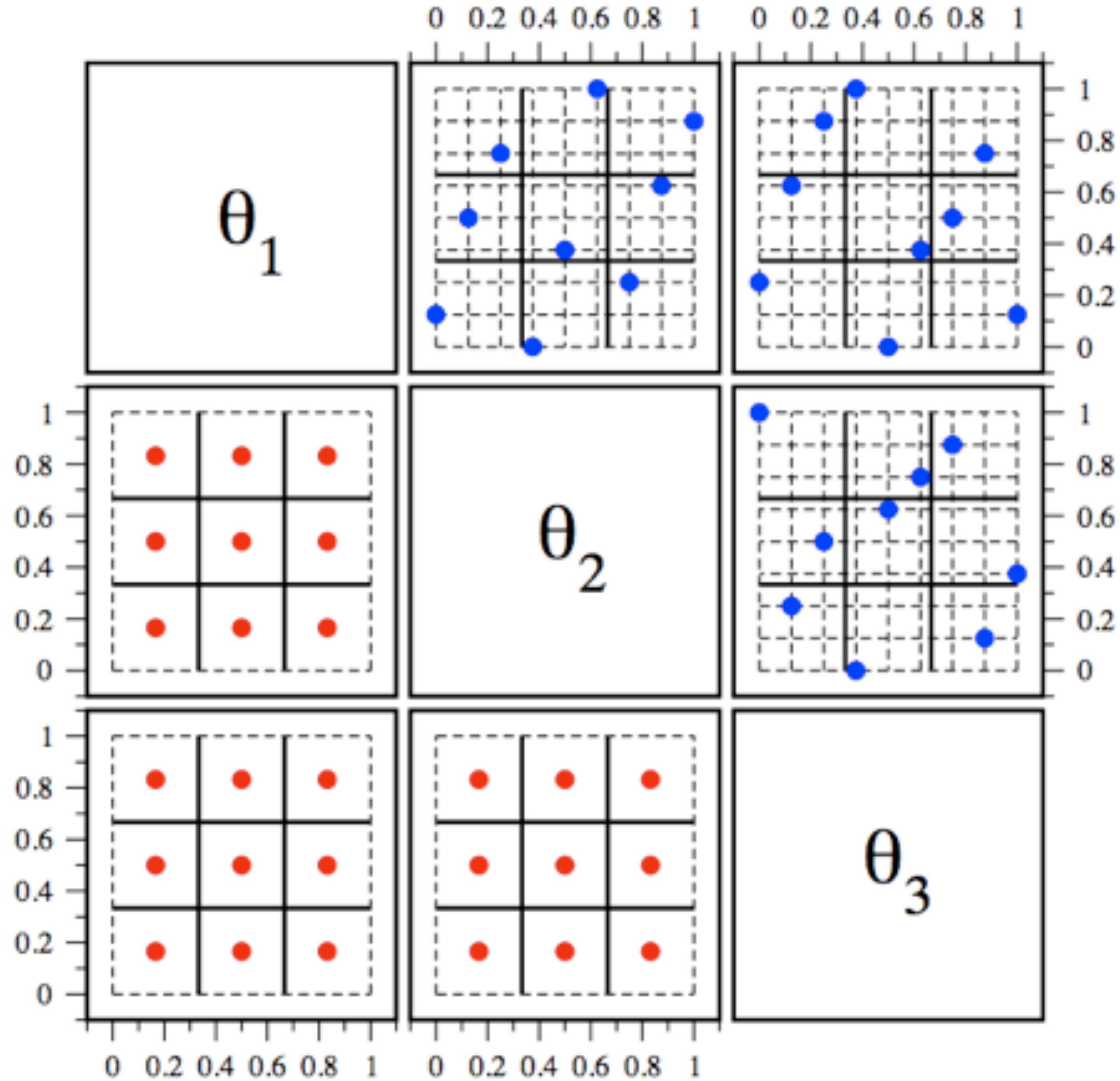
$0.020 \leq \omega_b \leq 0.025$
$0.11 \leq \omega_m \leq 0.15$
$0.85 \leq n_s \leq 1.05$
$-1.3 \leq w \leq -0.7$
$0.6 \leq \sigma_8 \leq 0.9$

- **37 model runs + ΛCDM**
  - 16 low resolution realizations (green)
  - 4 medium resolution realizations (red)
  - 1 high resolution realization (blue)
  - 11 outputs per run between z = 0 - 3
- **Restricted priors to minimize necessary number of runs**
- **1.3 Gpc boxes, $m_p \sim 10^{11} M_\odot$**
- **~1000 simulations, 60TB**



The Simulation Design

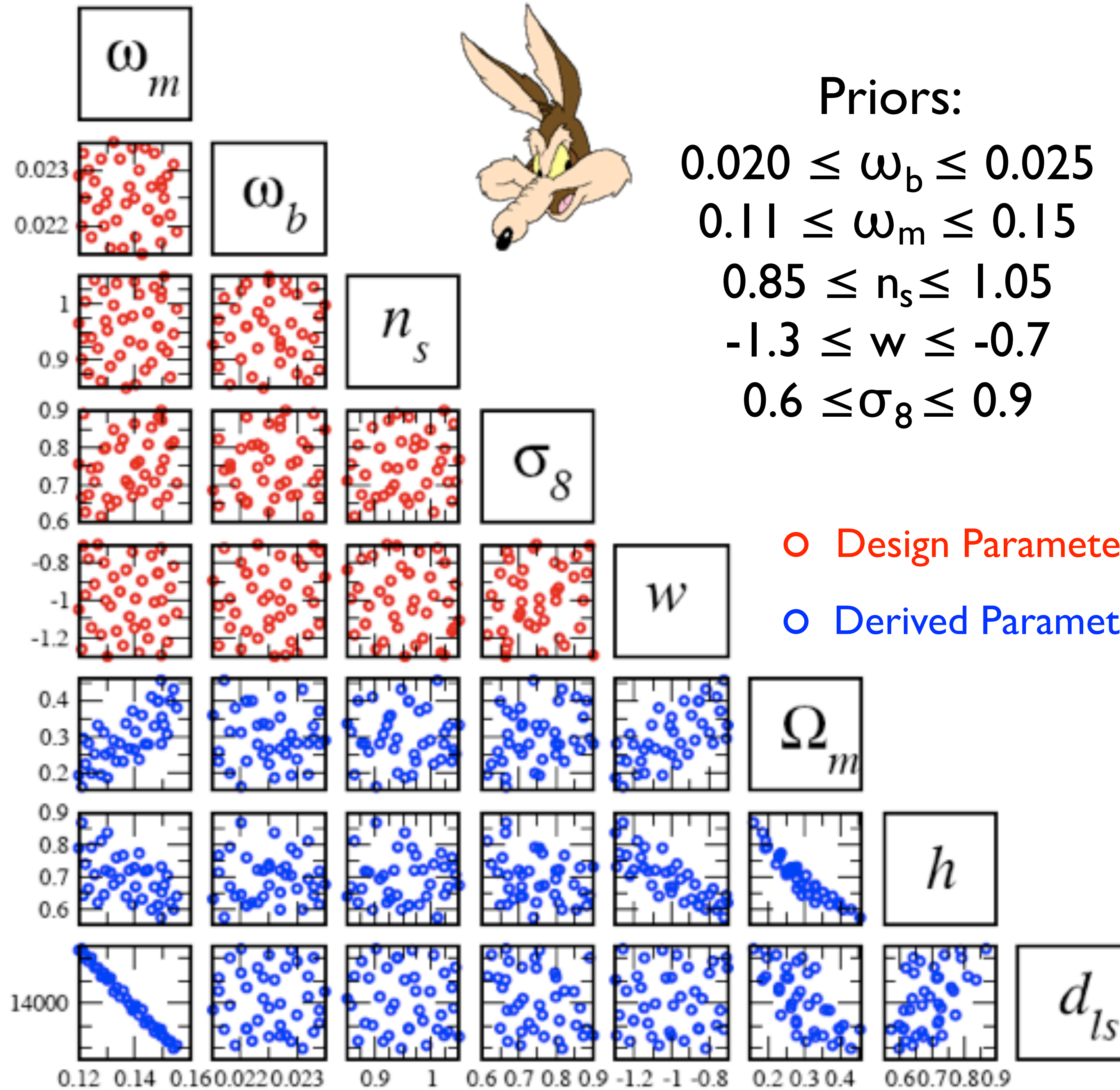
- **"Simulation design": for a given set of parameters to be varied and certain numbers of runs, at what settings should the simulations be performed?**
- **In our case: five cosmological parameters, tens of high-resolution runs are affordable**
- **First idea: grid**
  - Assume 5 parameters and each parameter should be sampled 3 times: 3^5=243 runs, not a small number, covarage of parameter space poor, allows only for estimating quadratic models ☹
- **Second idea: random sampling**
  - Good if we can perform many runs -- if not, most likely insufficient sampling of some of the parameter space due to clustering
- **Our approach: orthogonal-array Latin hypercubes (OA-LH) design**
  - Good coverage of parameter space
  - Good coverage in projected dimensions



Example: 3 parameters to vary, 9 runs we can do
First step: OA design -- an OA distributes runs uniformly in certain projections of the full parameter space, here: 2 D
Second step: LH design -- perturbe each position of the runs in such a way, that they do not overlapp when projected
Third step: optimization of the distances of the points

The Simulation Design

- **Observational considerations**
  - Planck will provide very accurate measurements of “vanilla parameters”
  - Right now from WMAP-5, BAO: ω_m, ω_b, n_s known at 2-3%
  - w, σ_8 less well known
  - Determine best-fit value for *h* for each model from distance to surface of last scattering, know at 0.3%
- **For good emulator performance from very small number of runs**
  - Not too broad priors
  - Not too many parameters



Priors:

$0.020 \leq \omega_b \leq 0.025$

$0.11 \leq \omega_m \leq 0.15$

$0.85 \leq n_s \leq 1.05$

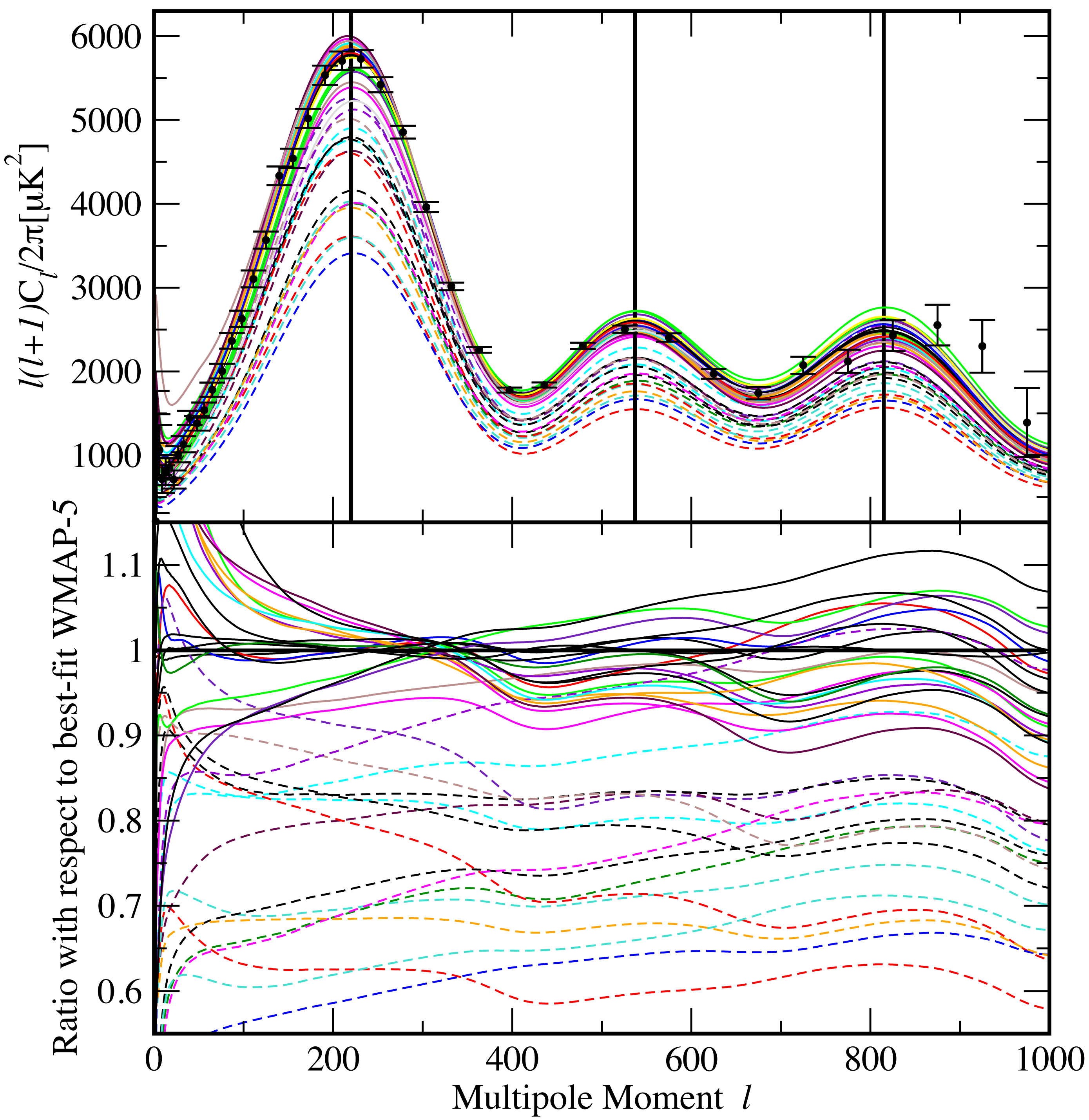
$-1.3 \leq w \leq -0.7$

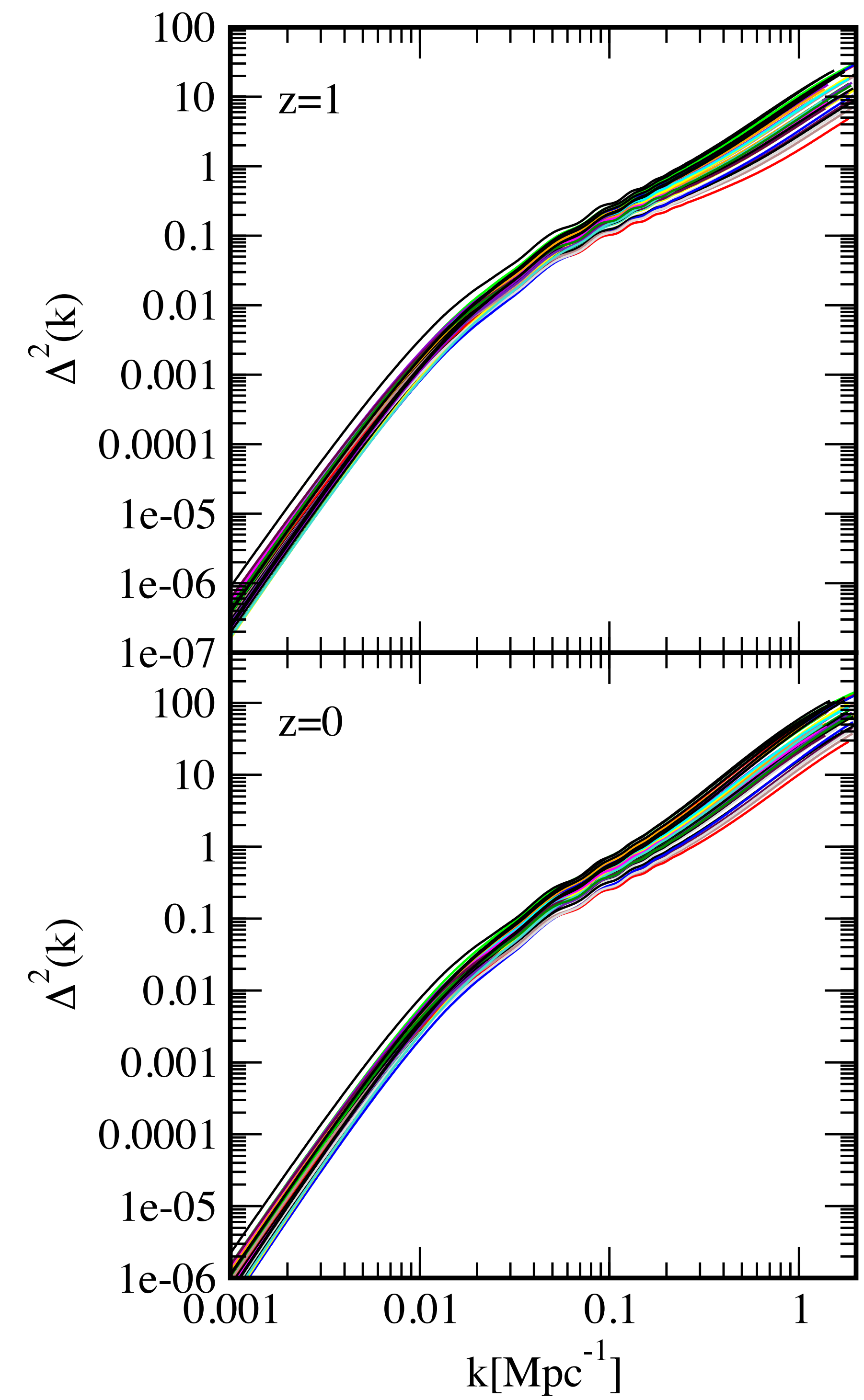
$0.6 \leq \sigma_8 \leq 0.9$

o Design Parameters

o Derived Parameters

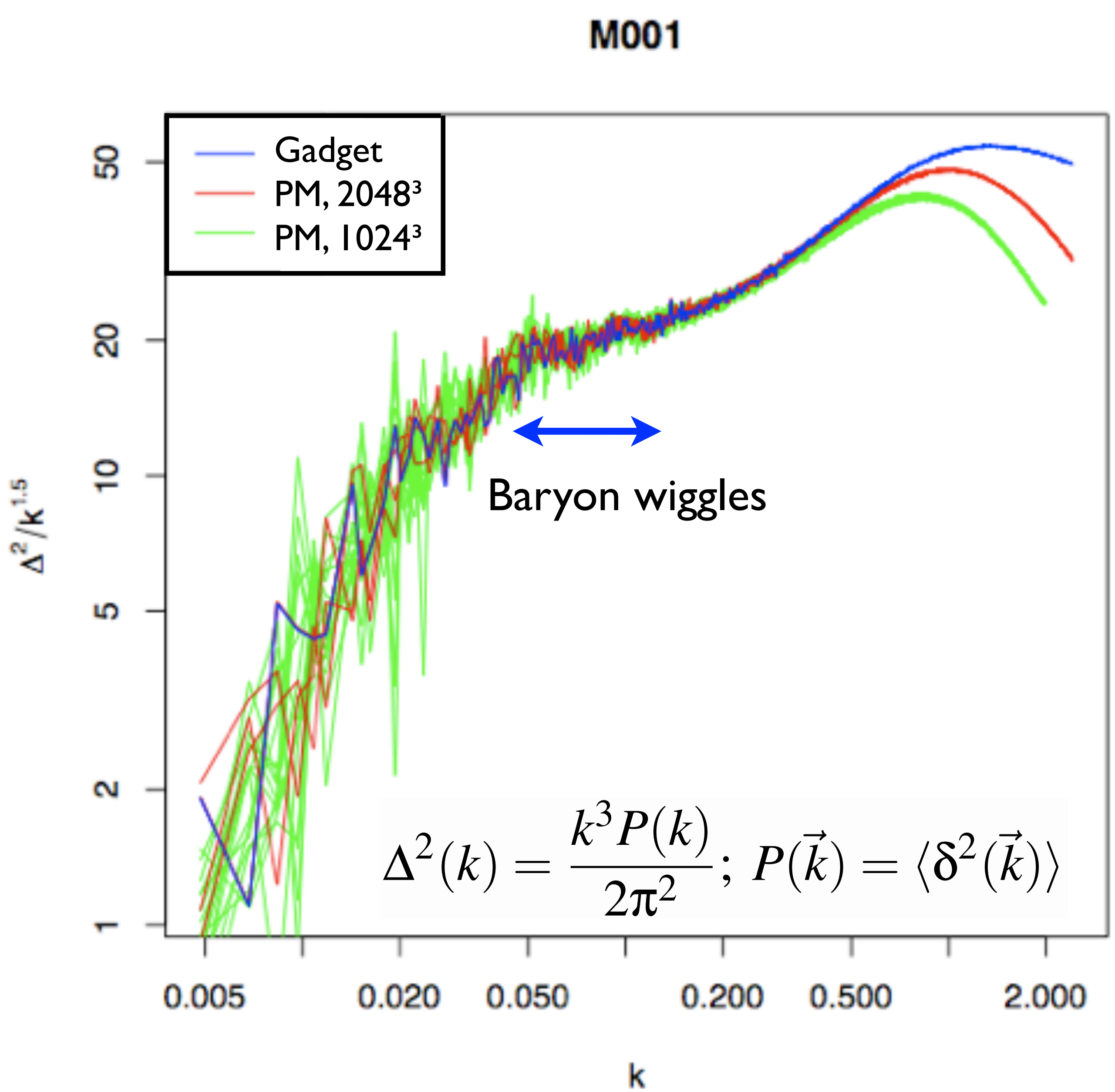
Model Selection





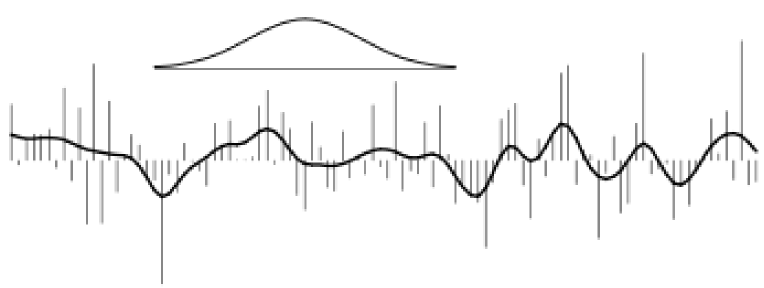
Next step: Smooth Power Spectrum

- Three different resolutions: 16 realizations low resolution PM, 4 realization medium resolution PM, one high-resolution Gadget run
- Pick scale on which baryon oscillations are most prominent
- Major challenge: Make sure that baryon features are not washed out or enhanced due to realization scatter
- For very low k: Sparse sampling and large scatter, difficult to handle
- Solution: Perturbation theory

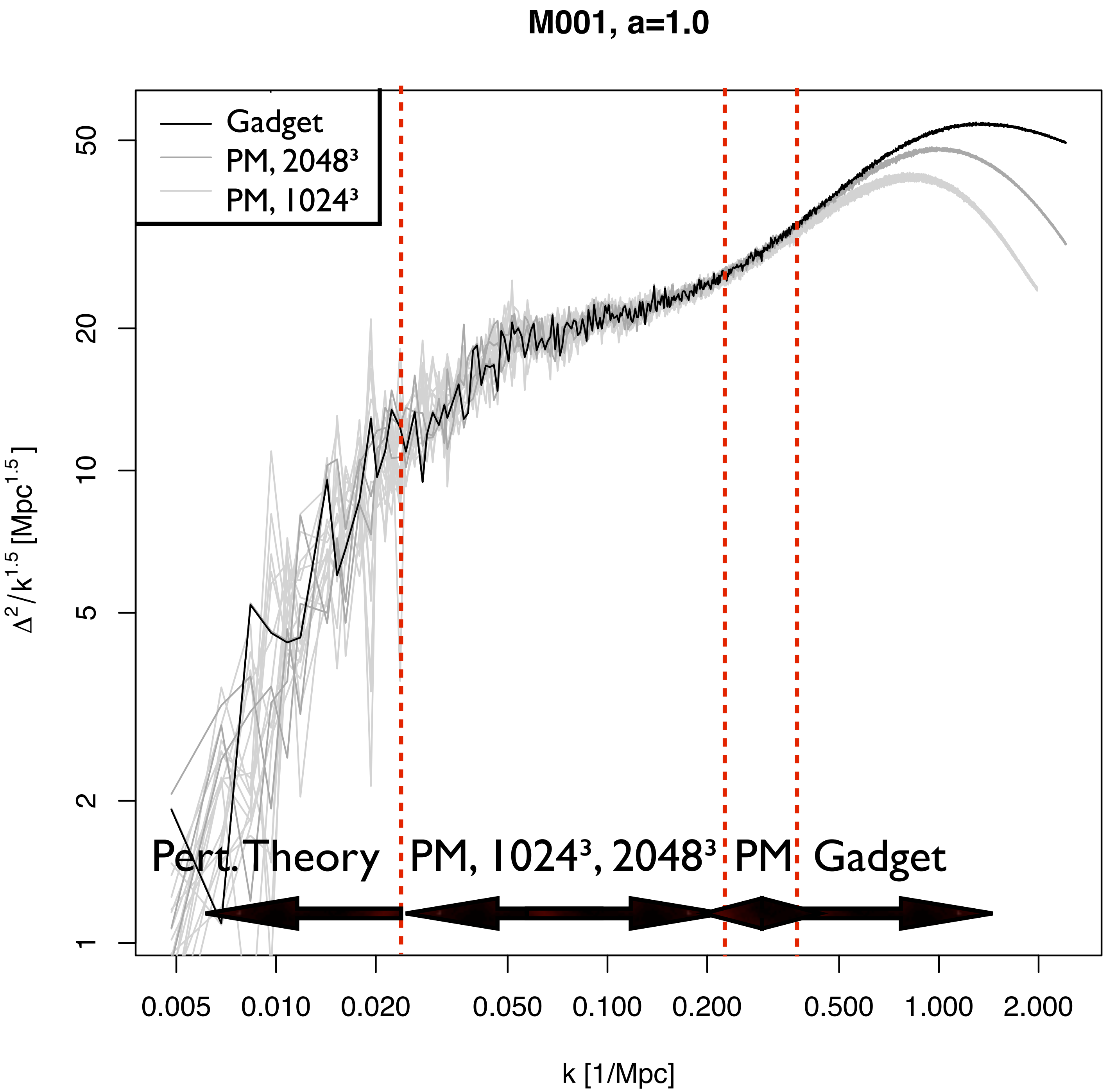


Smoothing Procedure

- **Match different resolution runs and perturbation theory**
- **Construct smooth power spectra using a process convolution model (Higdon 2002)**
- **Basic idea: calculate moving average using a kernel whose width is allowed to change to account for nonstationarity**

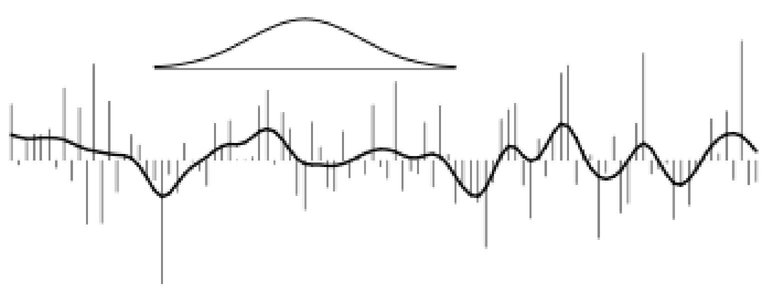


- **Kernel width determined by the variation in the data, free variable**

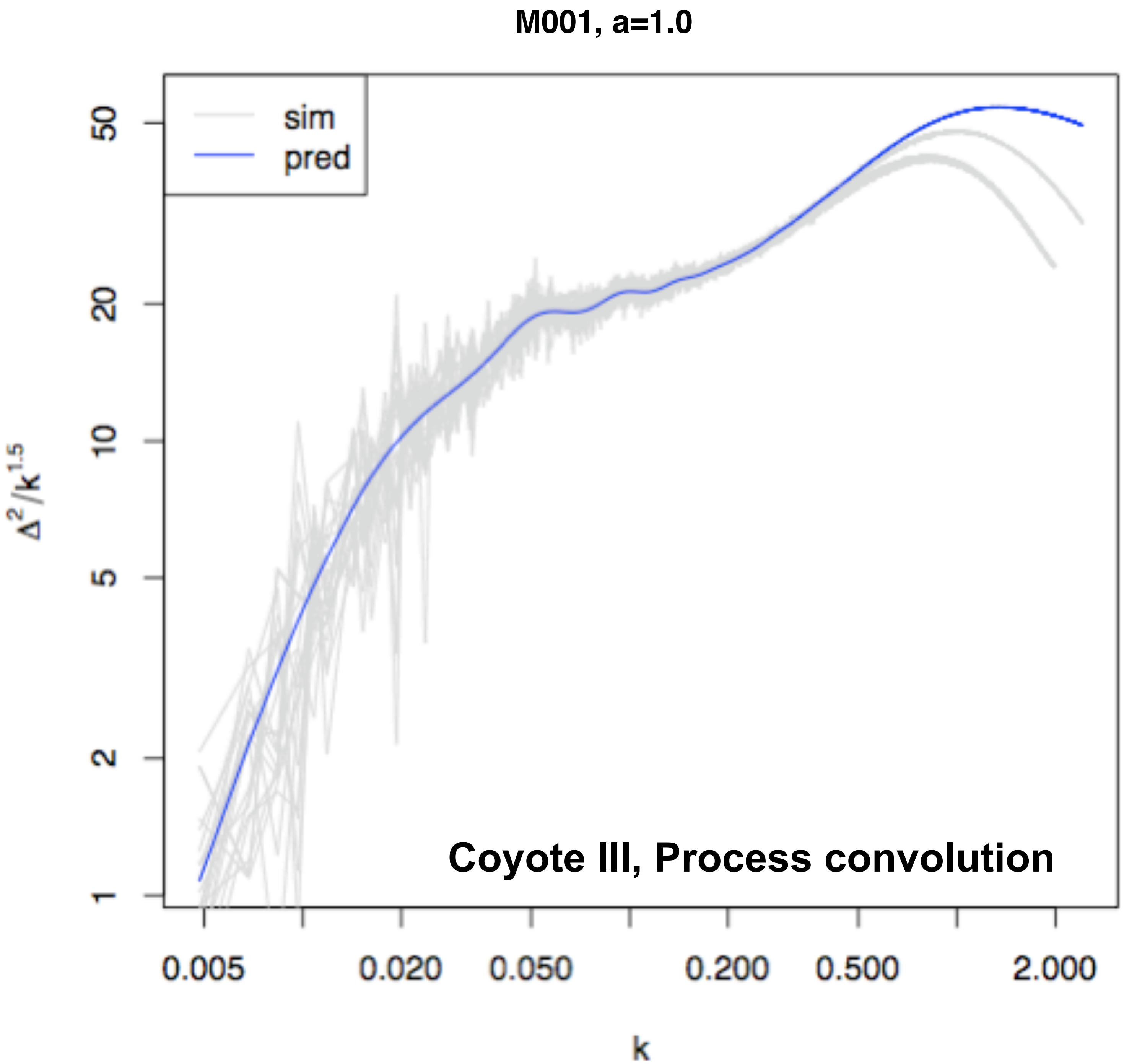


Smoothing Procedure

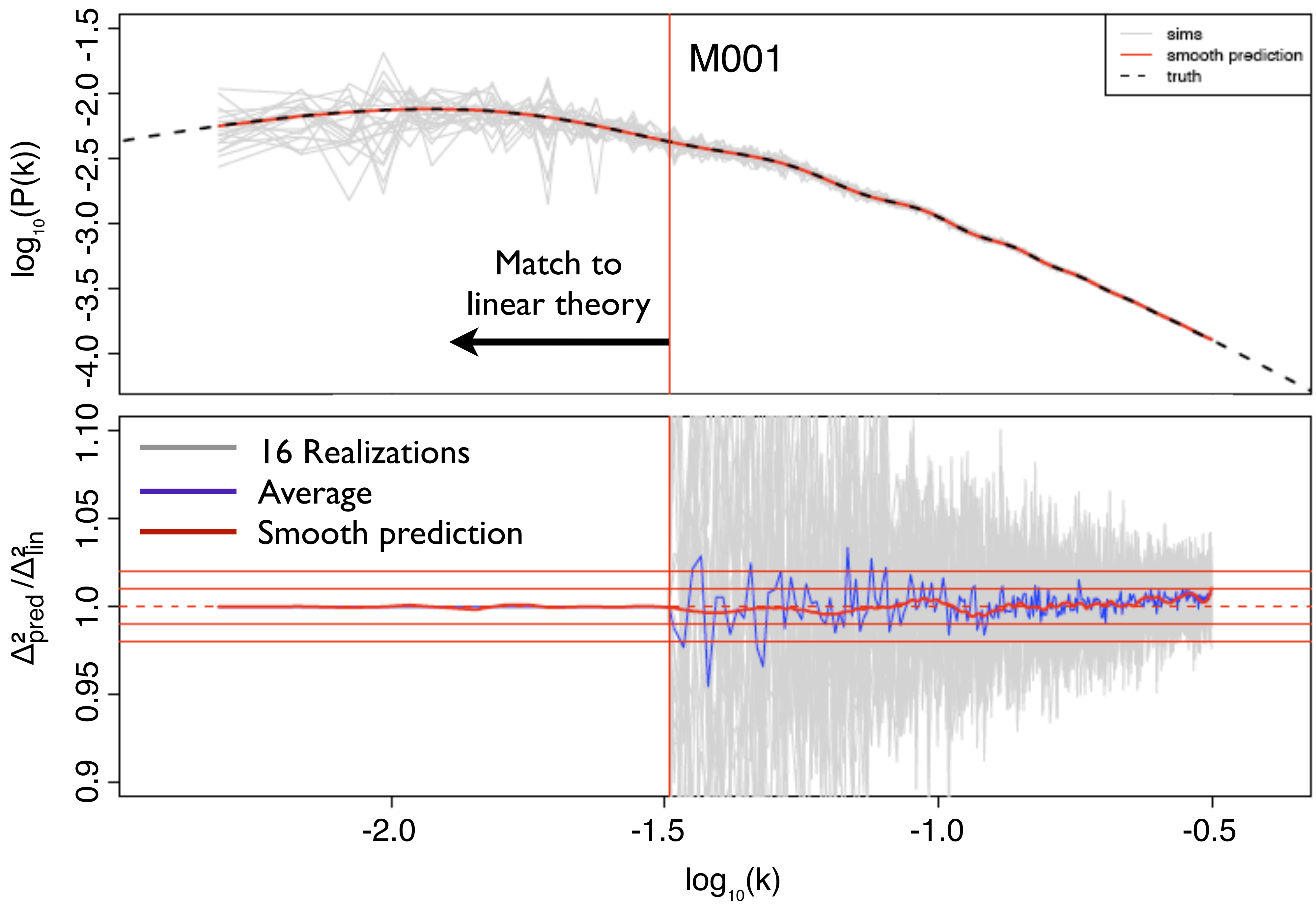
- **Match different resolution runs and perturbation theory**
- **Construct smooth power spectra using a process convolution model (Higdon 2002)**
- **Basic idea: calculate moving average using a kernel whose width is allowed to change to account for nonstationarity**



- **Kernel width determined by the variation in the data, free variable**



Test on Linear Power Spectrum



The Interpolation Scheme

- **After having specified the simulation design: build interpolation scheme that allows for predictions for any cosmology within the priors**
- **Model simulation outputs using a p_η - dimensional basis representation**
  - Find suitable set of orthogonal basis vectors $\phi_i(k,z)$, here: principal component analysis
  - 5 PC bases needed, fifth PC basis pretty flat
  - next step: modeling the weights
  - Here: Gaussian Process modeling (non-parametric regression approach, local interpolator; specified by mean function and covariance function)

$$\ln\left\{\frac{\Delta^2(k,z)}{2\pi k^{3/2}}\right\} = \sum_{i=1}^{p_\eta} \phi_i(k,z)\, w_i(\theta) + \epsilon \qquad \theta \in [0,1]^{p_\theta}$$

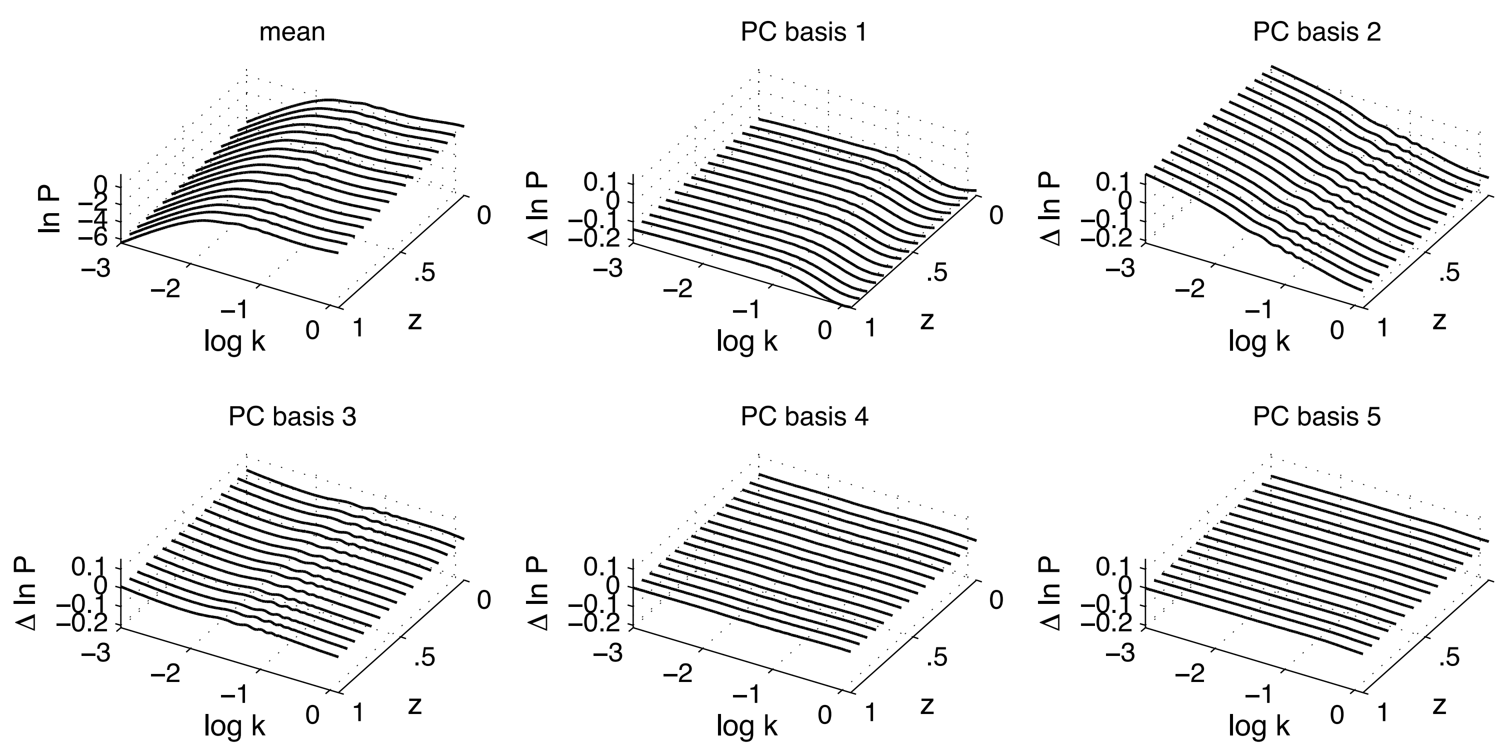
Cosmological parameters

Number of parameters, 5

Number of basis function, here: 5

Basis functions, here: PC basis

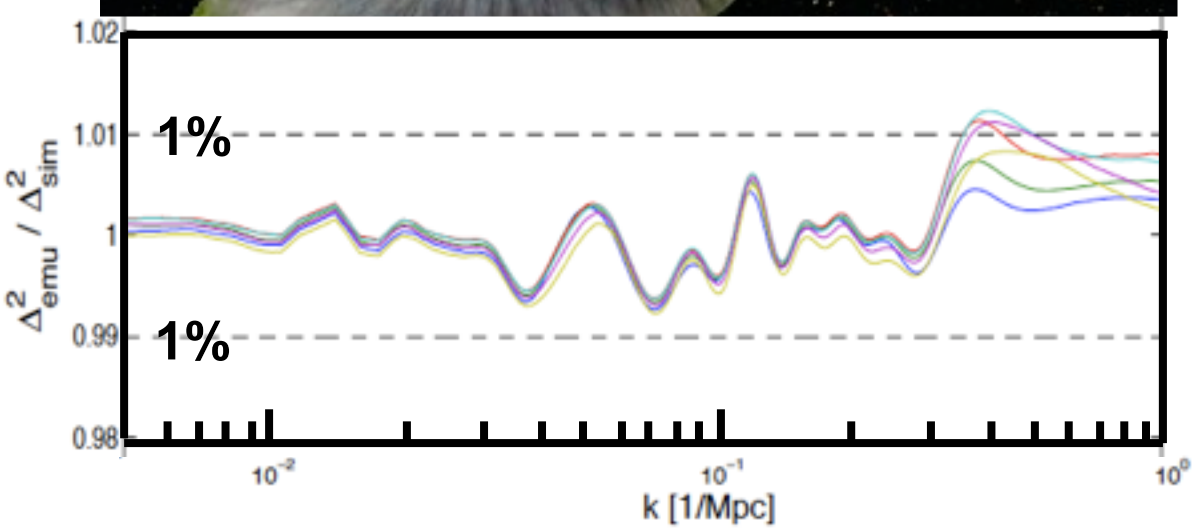
Weights, here: GP model



The Cosmic Emu(lator)

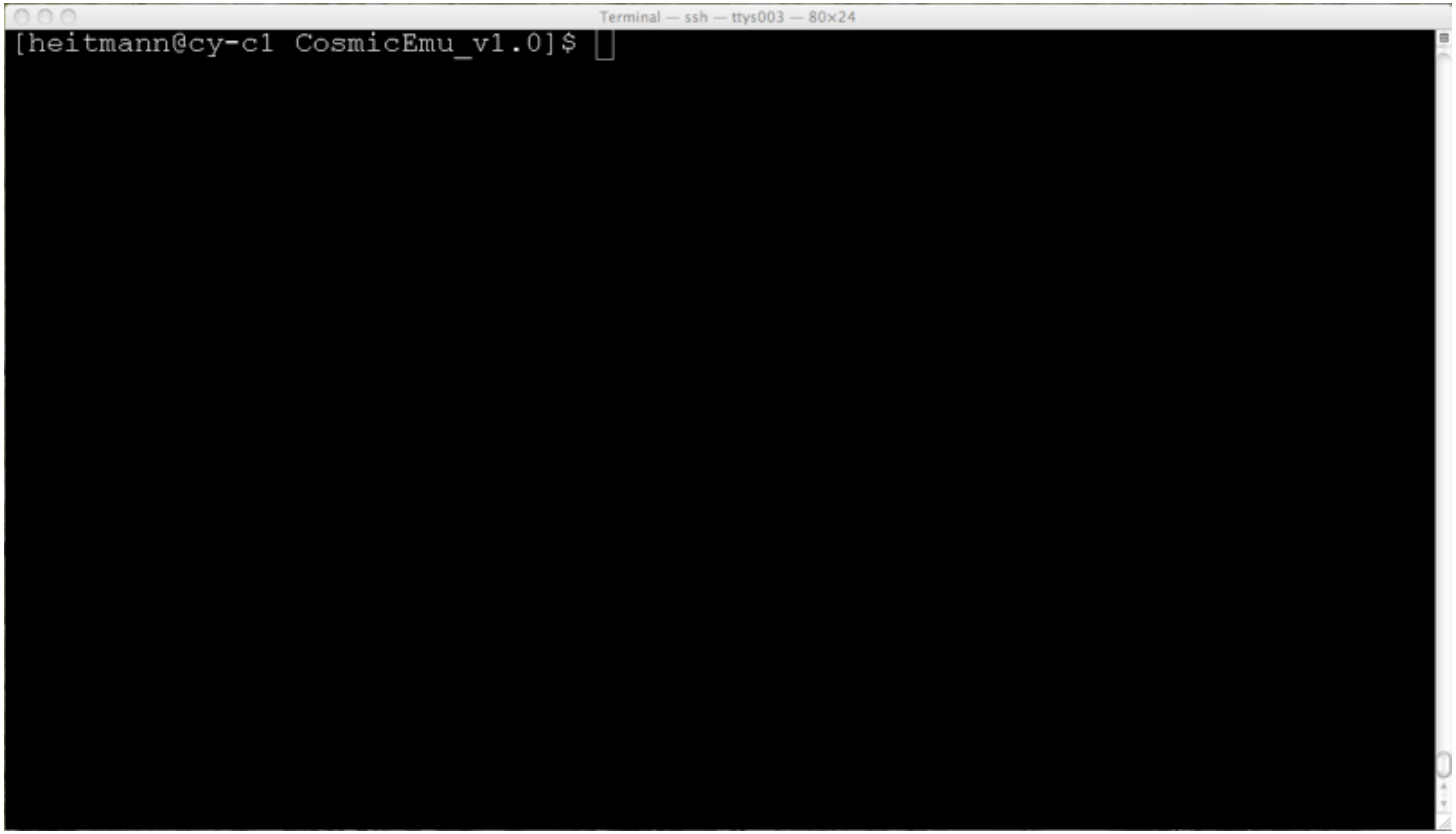
- **Prediction tool for matter power spectrum has been built**
- **Accuracy within specified priors between z=0 and z=1 out to k=1 h/Mpc at the 1% level reached**
- **Emulator has been publicly released, C code**
- **Next steps**
  - Extend k-range
  - Include more physics, e.g. neutrinos
  - Other statistics, e.g. shear spectrum

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



Emulator Efficiency

- **Success of CMBFast: cut down computational time from hours to seconds/minutes, while retaining the accuracy**
- **Cosmic emulators: basically instantaneously, from weeks to sub-seconds**

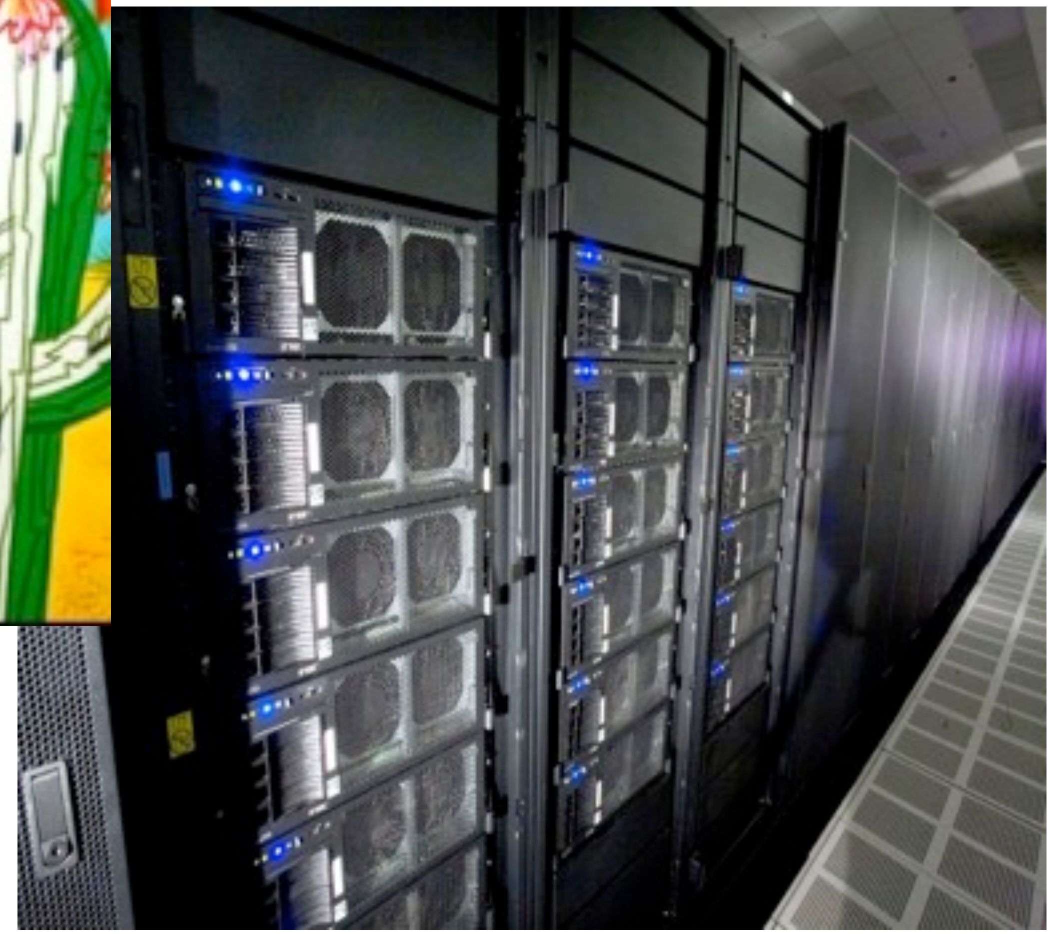


Lessons Learned

- **Coyote Universe idea was born in Sept. 2006, final emulator was released in Dec. 2009**
  - January 10, 2008, after a year of testing we think we understand the simulation errors at the 1% level and start the first runs; February 2, email from Martin White: ”Sanity checks. I've been thinking about these runs some more and worrying about lots of different things. It's so hard to know we're doing things right to 1%!”
- **First end-to-end calculation of “simplest”, but non-trivial problem (power spectrum at 1% out to k~1 h/Mpc) to provide precision prediction tool**
- **Collaboration of three different communities: cosmology, computer sciences, statistics, applied mathematics**
  - Many tools already exist, we do not want to reinvent them! Communication with other communities therefore very important, but will be slow at the start
- **Simulation infrastructure: running and analyzing 1000 simulations is not easy...**
  - Need for integrated analysis tools
  - Automatization of running the code and checking the outputs
- **Serving the data**
  - Data can be used for many other projects we have not thought about
  - Currently, ship the data by “hand”
  - Dedicated database would be very desirable
- **Computational and storage capacities demanding (one Coyote run ~250GB output for 12 snapshots, 300 billion particle run would be ~75TB)**

After the Coyote Universe...

... the Roadrunner Universe!

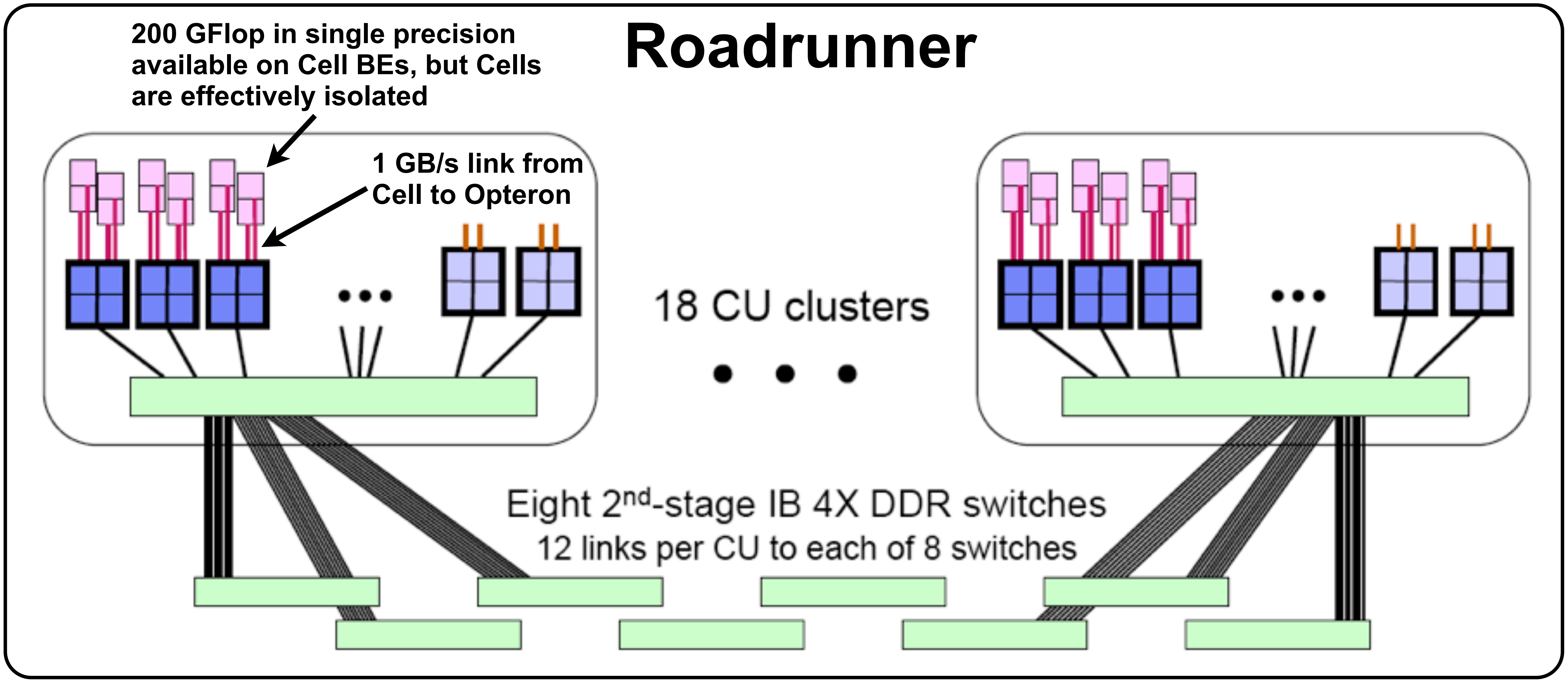


- **Development of new cosmology code to make use of HPC hybrid archictures**
- **Code up and running, will enable the next step in precision cosmology**

HACC: Hardware Accelerated Cosmology Codes

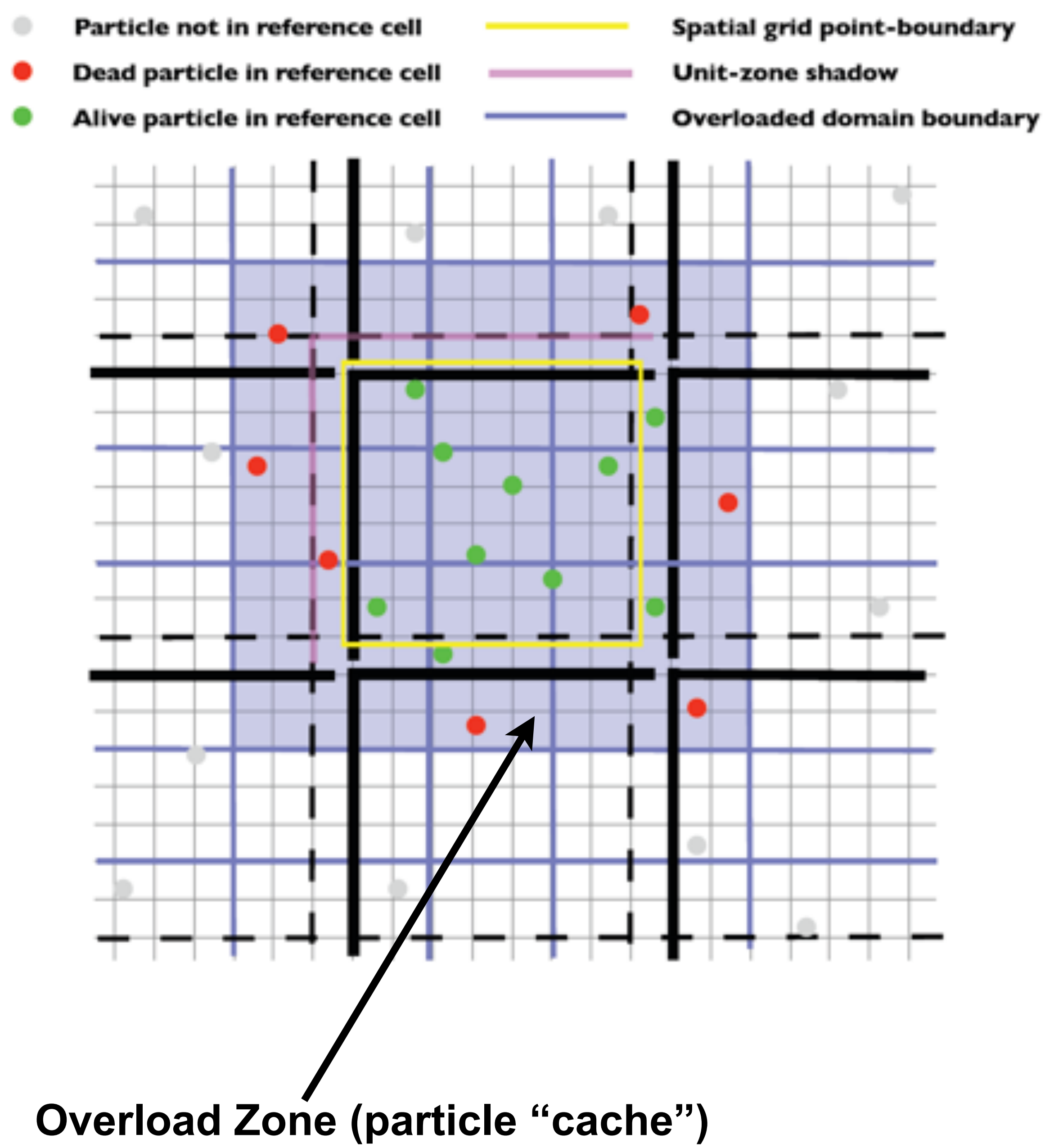
Habib et al Journal of Physics 2009; Pope et al. Computing in Science and Engineering 2010

- Challenge on future supercomputers: hybrid architectures, standard hardware is coupled to accelerating units (100 times speed up)
- Examples: Roadrunner at LANL (Cells), next version of Jaguar at ORNL (GPUs)
- Challenges: new programming paradigms, data movements slow between sub-units



HACC: Hardware Accelerated Cosmology Codes

- **Use P³M algorithm**
  - Grid lives on Opteron layer, large FFT Poisson solves
  - Particles live on Cell and interact on “subgrid scales” by fast hand-coded routines
  - Only simple grid info flows between Cells and Opterons
- **Avoid particle communication**
  - Particle communication between Cells at every short-range time step would be too slow, avoid this using **particle caching** (trade memory for speed)
  - Nearest neighbor refresh at every long timestep
- **Speed up: 32^3 particles, 200 timesteps on Opterons only (8 MPI ranks): 11 hours; on Operton/Cell hybrid: 2 minutes**
- **Cell part can easily be exchanged by different module, serial implementation due to overloading idea**
  - GPU implementation finished last week
  - For Opterons: tree instead of N^2 (under development)
- **On the fly analysis tools implemented**



Summary and Outlook

- **Nonlinear regime of structure formation requires simulations**
  - No error controlled theory
  - Simulated skies/mock catalogs essential for survey analysis
- **Simulation requirements are demanding, but can be met**
  - Only a finite number of simulations can be performed
- **Cosmic Calibration Framework**
  - Accurate emulation of several statistics matching code errors
  - Allows fast calibration of models vs. data
- **Challenges for the future**
  - More physics needs to be taken into account, uncertainties from not knowing
  - Computational and storage capacities will be demanding
  - Simulation infrastructure, running very large number of simulation need automation
  - Serving the data, simulation results should be available to broader community
  - Communication with other communities: statisticians, computer scientists, applied mathematicians, it takes some time to learn each other's language!
  - New supercomputer architectures pose challenges to cosmology codes but will also allow us to run simulations in days which now take weeks

References

- **Gaussian Process Modeling Introduction:**
  - ***Gaussian Processes for Machine Learning,*** C.E. Rasmussen, K.I. Williams, MIT Press, www.GaussianProcess.org/gpml
- **Cosmic Calibration Framework (www.lanl.gov/projects/cosmology/CosmicEmu)**
  - ***The Coyote Universe III: Simulation Suite and Precision Emulator for the Nonlinear Matter Power Spectrum,*** E. Lawrence et al., ApJ submitted, arXiv:0912.4490
  - ***The Coyote Universe II: Cosmological Models and Precision Emulation of the Nonlinear Matter Power Spectrum,*** K. Heitmann et al., ApJ 705, 156 (2009), arXiv:0902.0429
  - ***Constraining Cosmological Parameters by Combining Simulations and Physical Observations,*** D. Higdon et al., in The Handbook of Applied Bayesian Analysis, Oxford University Press (2010)
  - ***Simulations and Cosmological Inference: A Statistical Model for Power Spectra Means and Covariances,*** M.D. Schneider et al., Phys. Rev. D78, 063529 (2008)
  - ***Cosmic Calibration: Constraints from the Matter Power Spectrum and the Cosmic Microwave Background,*** S. Habib et al., Phys. Rev. D76, 083503 (2007)
  - ***Cosmic Calibration,*** K. Heitmann et al., ApJ 646, L1 (2006)
- **Gaussian Process Modeling for other Astrophysics Problems**
  - ***Nonparametric Reconstruction of the Dark Energy Equation of State,*** T. Holsclaw et al., Phys. Rev. D.
  - ***Gaussian Process Modeling of Asteroseismic Data,*** B.J. Brewer, D. Stello, MNRAS 395, 22206 (2009)
  - ***Photometric Redshift Prediction Via Gaussian Process Regression,*** M. Way et al. ApJ 706, 623 (2009)