1998 → 2015

The Coyote Universe: Precision Simulations of the Large Scale Structure of the Universe

Katrin Heitmann, ISR-1, LANL

LBNL Seminar, March 12, 2009

In collaboration with:

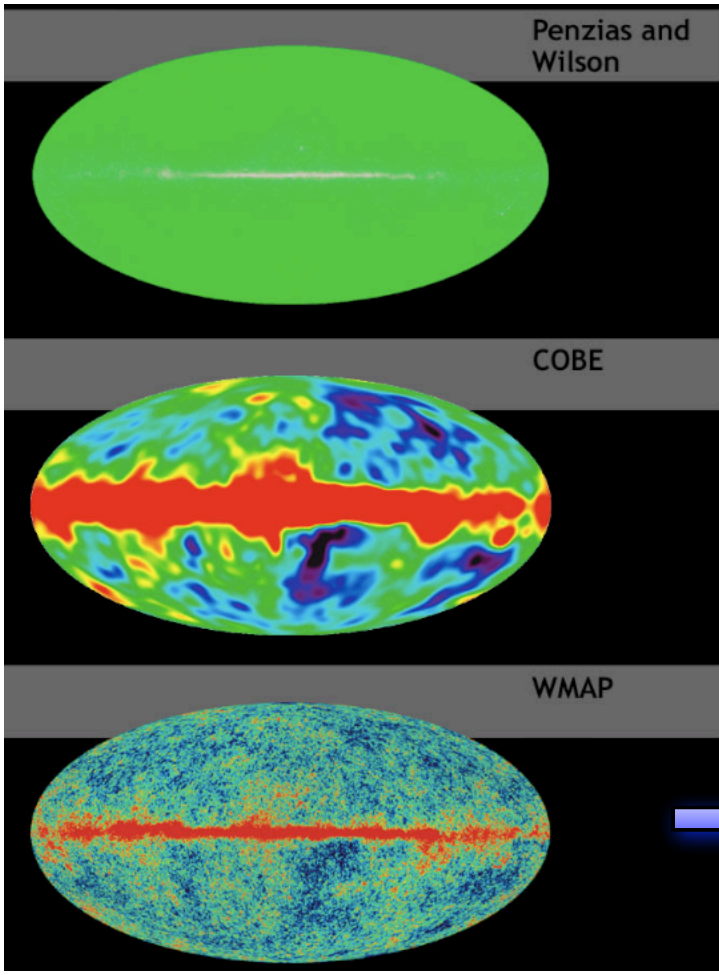
Suman Bhattacharya, Salman Habib, David Higdon, Zarija Lukic, Earl Lawrence, Charlie Nakhleh, Christian Wagner, Martin White, Brian Williams

SDSS, First Light in 1998

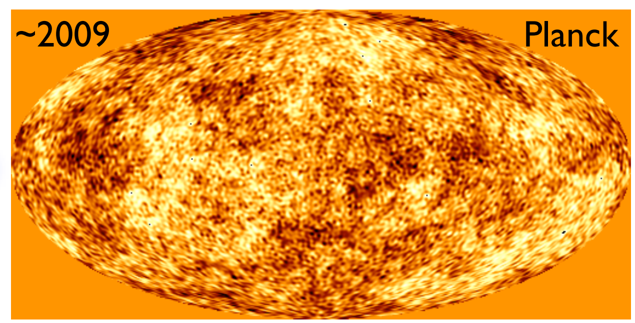
Visualization: Pat McCormick, CCS-1, LANL

Deep Lens Survey/LSST

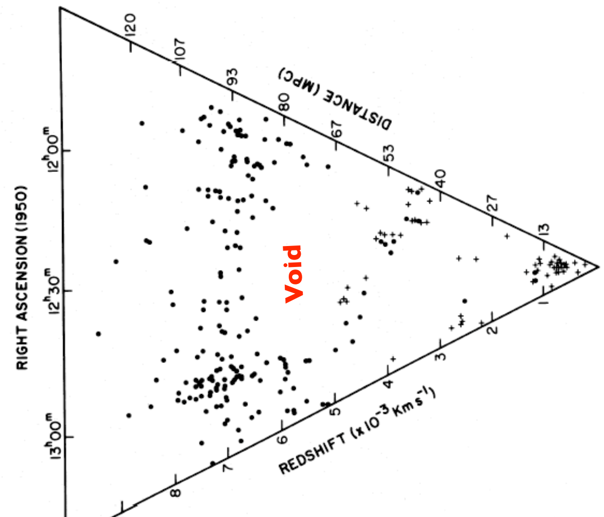
Progress in Cosmology I: CMB



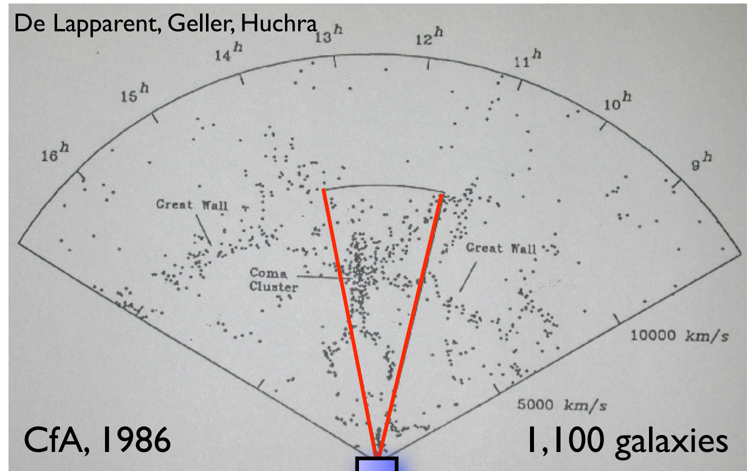
- **Cosmic microwave background measurements started the era of “precision cosmology”**
- **What made it “precision”?**
  - Physics “easy” to understand
  - At its wavelength the CMB dominates the sky

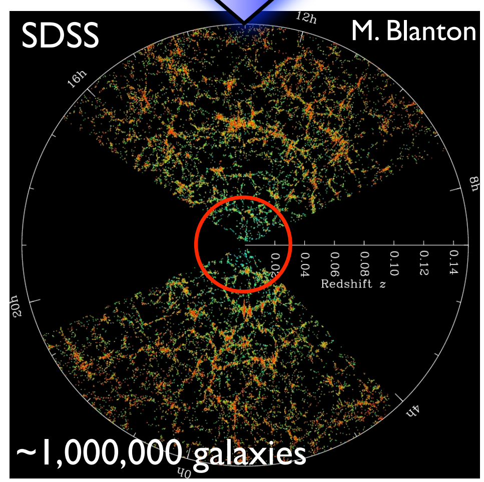


Progress in Cosmology II: LSS



Gregory & Thompson, 1978

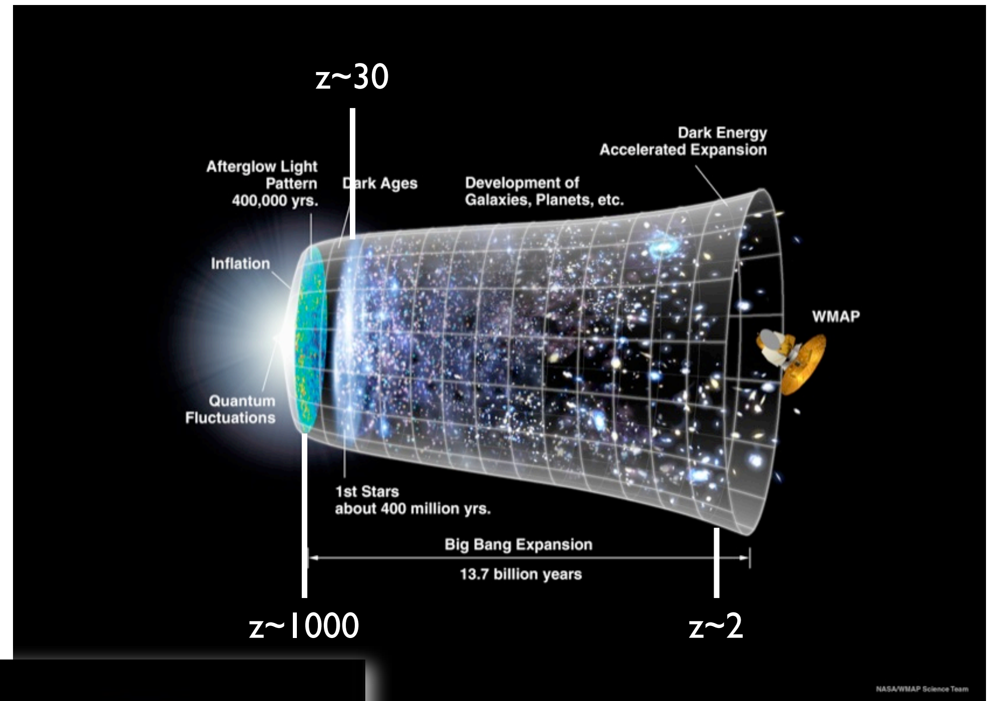


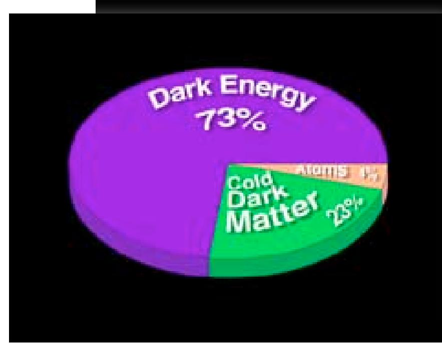


- **1978: Discovery of voids and superclusters, theory of hierarchical structure formation via gravitational instability emerges**
- **2006: SDSS has measured more than 1,000,000 galaxies, important discoveries such as the baryon oscillations by Eisenstein et al. cementing our picture of structure formation**

Standard Model of Cosmology

- **Good idea of the history of the Universe**
- **Good idea of the composition:**
  - ~73% of a mysterious dark energy
  - ~23% of an unkown dark matter component
  - ~4% of baryons
- **Constraints on ~20 cosmological parameters, including optical depth, spectral index, hubble constant,...**
- **Values are known to ~10%**
- **For comparison: the parameters of the Standard Model of Particle Physics are known with 0.1% accuracy!**





Why do we need higher accuracy?

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

~~w~-1+/-0.1 ... r ... onomers~~
~~this means th... sion history~~
~~is already well e... measured that~~
~~further refine... roduce at most~~
~~minor shifts ... e in... d history of~~
~~cosmi... ucture ... mation.~~

Simon D.M. White,
astro-ph/07043391

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

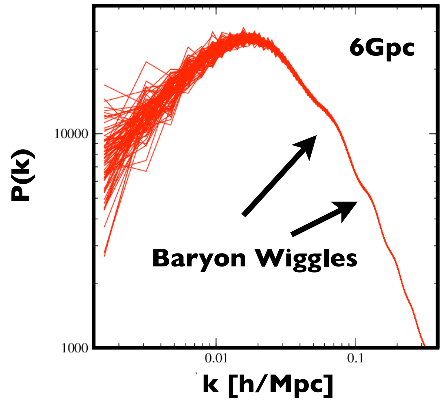
Why do we need higher accuracy?
-- An example: The spectral index and inflation

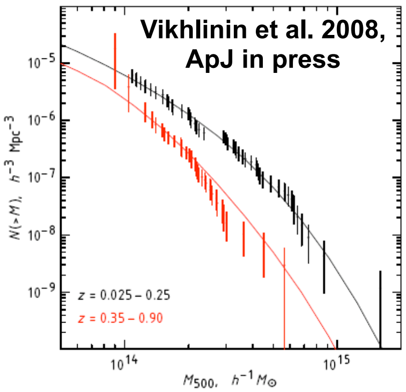
- Simple scaling arguments predict slope of the primoridal power spectrum to be n=1, constant, the Harrison-Zel'dovich power spectrum
- "Generic" inflationary models predict a slight deviation from n=1, usually smaller n < 1
- In addition: weak scale dependence, n(k), running
- If we could measure the spectral index and its k-dependence with high precision, we would have a smoking gun for inflation!

Why do we need higher accuracy?
-- Another example: Dark energy

- **What is the nature of dark energy?**
  - Cosmological constant
  - Scalar field
  - Or none of this, but gravity is different on large scales..
- **In the absence of a good idea: try to characterize dark energy**
- **We have to determine the dark energy equation of state, w and its time variation**
- **At the moment: w=-1+/-0.1 from different data sources, dw/dt consistent with zero**
- **Promising probes: baryon acoustic oscillations (power spectrum), clusters (mass function), supernovae, weak lensing (power spectrum)**

Large Scale Structure Probes of Dark Energy

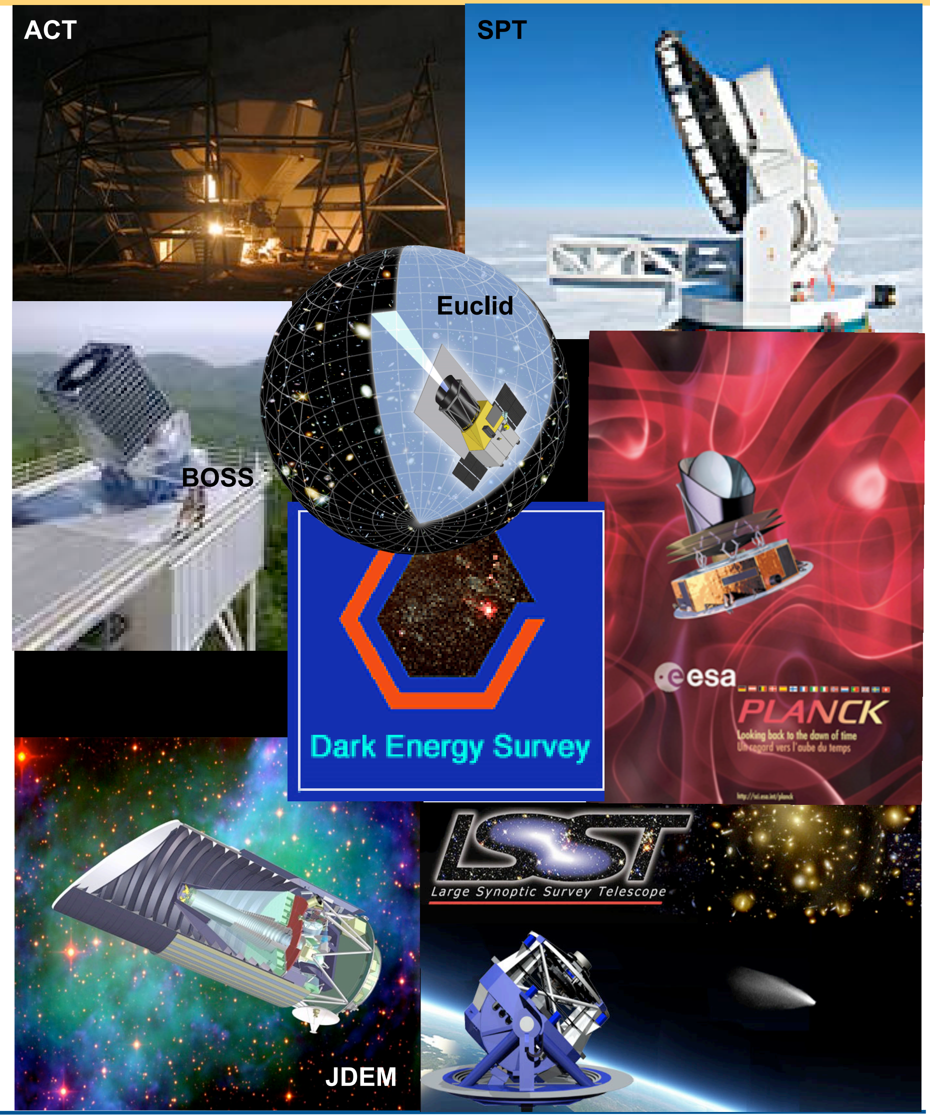




- **Baryon Accoustic Oscillations**
  - Precision requirement: 0.1% measurement of distance scale
  - Very large box sizes (~3 Gpc) to reduce sampling variance and systematics from nonlinear mode coupling
  - Gravity-only simulations largely adequate
- **Weak Lensing**
  - Precision requirement: 1% accuracy at k~1-10 h/Mpc
  - Large box sizes (~1Gpc) to reduce nonlinear mode coupling
  - At scales k > 1 h/Mpc: baryonic physics start to becomes important
- **Clusters**
  - Large box sizes (~1Gpc) for good statistics (~40,000 clusters)
  - Gas physics and feedback effects important
  - Well calibrated mass-observable relations

Precision Cosmology: Observations

- **JDEM**
  - 2000 supernovae, 300-1000 square degree lensing survey, w: ~4%, dw/dt:~10%
- **SPT (Southpole Telescope)**
  - 10 meter diameter telescope, thousand clusters, strong constraints on w
- **LSST (Large Synoptic Survey Telescope)**
  - 8.4 meter, digital imaging across the sky, supernovae, etc.
- **DES (Dark Energy Survey)**
  - Galaxy cluster study, weak lensing, 2000 SNe Ia, constraints on w at the one percent level
- **Planck**
  - High precision measurements of the microwave background out to l~2500



What about theory?

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."

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

Great Survey Size Simulations

- **"Billion/Billion" simulation:**
  - Gigaparsec box, billion particles
  - Smallest halos: ~10^{13} $M_{\odot}$ (100 particles)
  - 10 time snapshots: ~250GB of data
  - ~30,000 Cpu hours with e.g. Gadget-2, ~5 days on 256 processors (no waiting time in the queue included...)
  - Accuracy at k~1h/Mpc: ~1%
- **3 Gigaparsec, 300 billion particles**
  - Smallest halos: ~10^{12} $M_{\odot}$
  - 10 time snapshots: ~75TB
- **Physics:**
  - Gravitational physics
  - Gas physics
  - Subgrid models

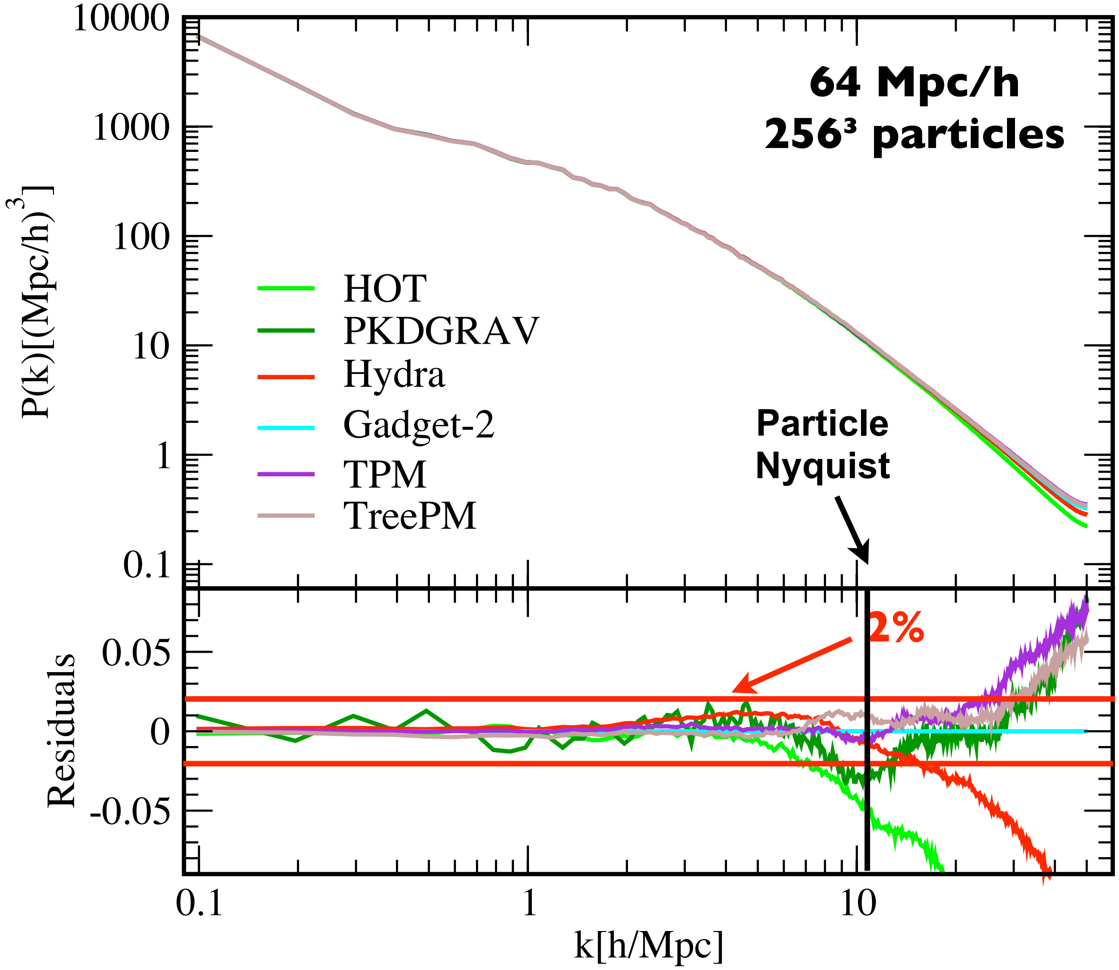
The Coyote Universe: Precision Predictions at the 1% level

Coyote-I: arXiv:0812.1052, Coyote-II: arXiv:0902.0429 (submitted to ApJ), Coyote-III, IV: in preparation

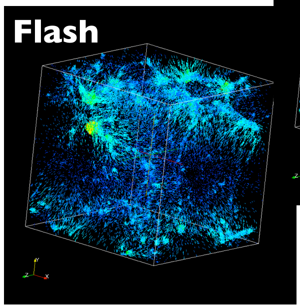
- **Large simulation suite run on LANL supercomputer "Coyote"**
  - 38 cosmological models with different dark energy equations of state
  - 1.3 Gpc cubed comoving volume, 1 billion particles each
  - 16 medium resolution, 4 higher resolution, and 1 very high resolution simulation for each model = 798 simulations, ~60Tb of data
- **Aim: precision predictions at the 1% accuracy level for different cosmological statistics**
  - dark matter power spectrum out to k~1h/Mpc; on smaller scales: hydrodynamics effects become important! (White 2004, Zhang & Knox 2004, Jing et al. 2006, Rudd et al. 2008)
  - shear power spectrum
  - mass function
- **Three parts to the project:**
  - Demonstrate 1% accuracy of the dark mattter simulations out to k=1h/Mpc ✓ (arXiv:0812.1052)
  - Develop framework which can predict these statistics from a minimal number of simulations ✓ (arXiv:09.02.0429)
  - Build prediction tools from simulation suite (Coyote III, IV, in progress)

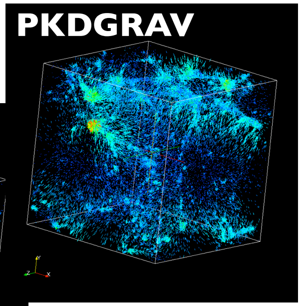
Code Comparison

Heitmann et al., ApJS (2005); Heitmann et al., Comp. Science and Discovery (2008)



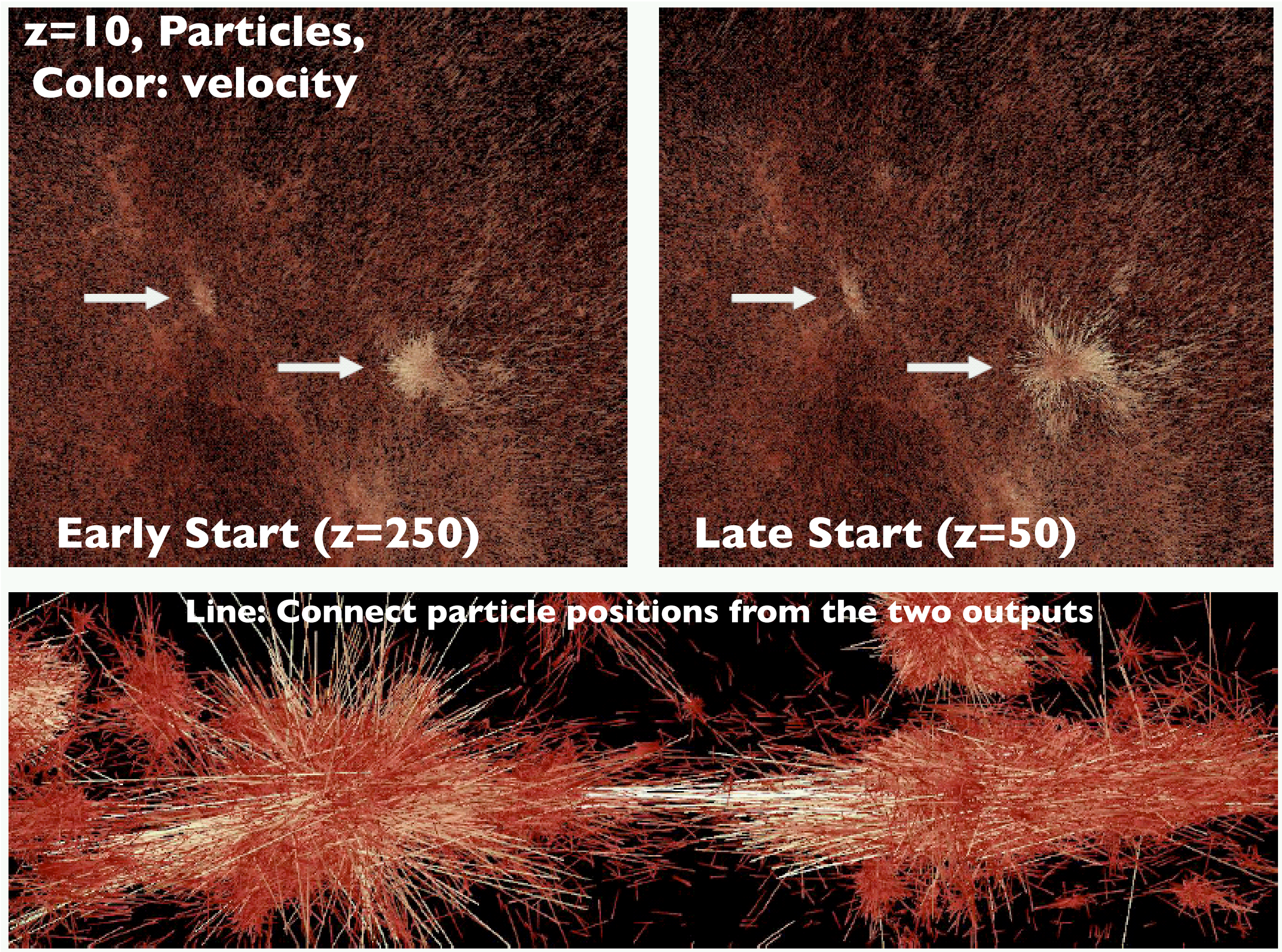
- Comparison of ten major codes (subset is shown)
- Each code starts from same initial conditions
- Each simulation is analysed in exactly the same way
- Overall, good agreement between codes for different statistics at the 5-10% level





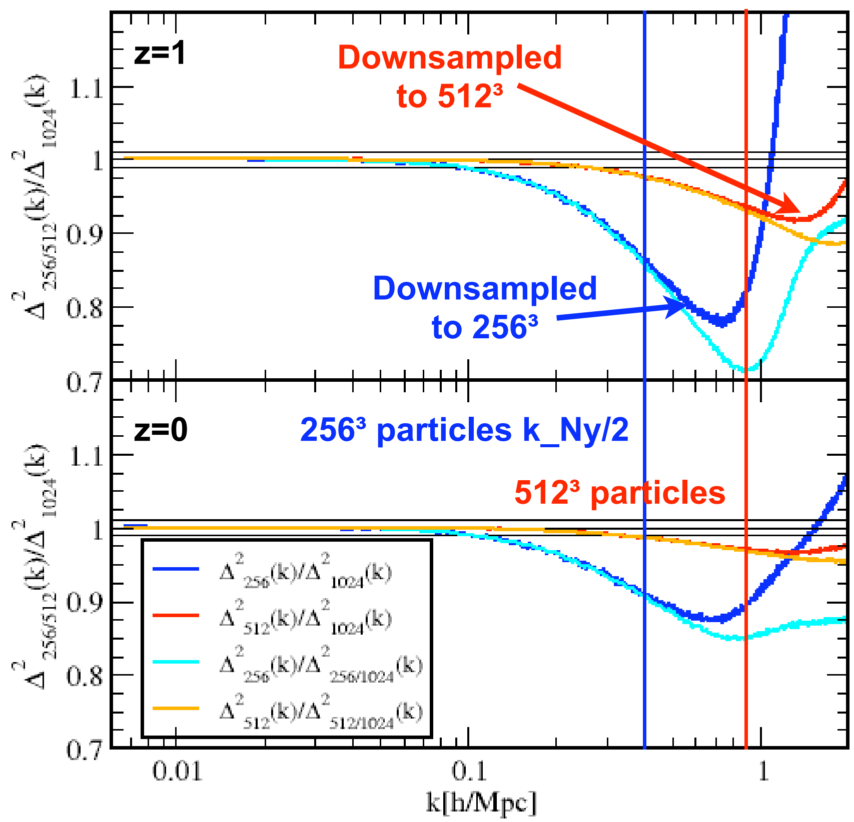
Initial Redshift

Haroz, Ma & Heitmann (2008); Haroz & Heitmann (2008)



Mass Resolution

- **Test with different particle loading in 1Gpc box**
  - Run 1024^3 particles as reference
  - Downsample to 512^3 and 256^3 particles and run forward
  - In addition: downsample z=0,1 1024^3 results to characterize shot noise problem
- **For precision answers: interparticle spacing has to be small!**
- **Requirement: k < k_Ny/2**
- **Gigaparsec box requires billion particle minimum**
- **Force resolution is not the limiting factor, but mass resolution is**



The Cosmic Calibration Framework

Heitmann et al., ApJL (2006); Habib et al., PRD (2007); Schneider et al., PRD (2008), Heitmann et al., arXiv:0902.0429

- **We have simulation accuracy under control at the 1% level out to k~1h/Mpc**
  - Mass resolution, box size, initial start, force resolution, and time step criteria exist!
- **For cosmological constrains from e.g. SDSS:**
  - Run your favorite Markov Chain Monte Carlo code, eg. CosmoMC
    - MCMC: directed random-walk in parameter space
  - Need to calculate P(k) ~ 10,000 - 100,000 times for different models
  - 30 years of Coyote time (2048 processor Beowulf Cluster), impossible!
- **What we need: framework that allows us to provide, e.g., P(k) for a range of cosmological parameters**
- **The Cosmic Calibration Framework provides:**
  - Simulation design, an optimal strategy to choose parameter settings
  - Emulation, smart interpolation scheme that will replace the simulator and will generate power spectra, mass functions... with controlled errors
  - Uncertainty and sensitivity analysis
  - Calibration -- combining simulations with observations to determine best-fit cosmology

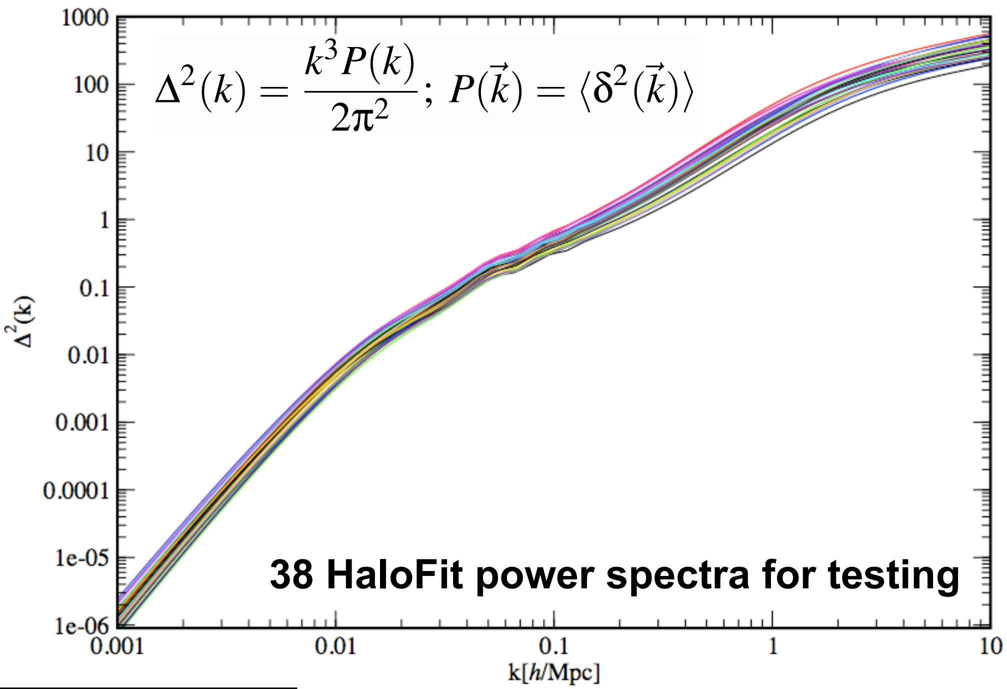
The Coyote Universe in Numbers

Priors:

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

Best Fit CIs:

$0 \leq \tau \leq 0.34$

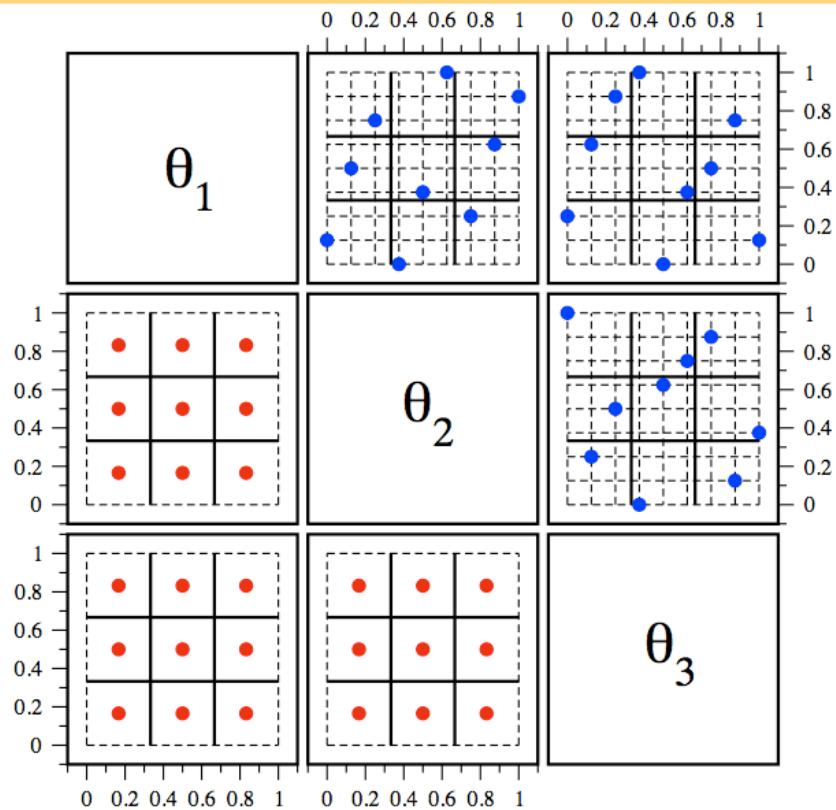


* 16 realization PM, 512^3 on 1024^3
* 4 realization PM, 1024^3 on 2048^3
* 1 realization Gadet-2, 1024^3 with 5kpc
* 11outputs per run between z=0,3

- **37 model runs + ΛCDM**
- **Restricted priors, to minimize necessary number of runs**
- **1.3 Gpc boxes, m_p~10^{11}M_solar**
- **~800 simulations, 60Tb of data**

The Simulation Design

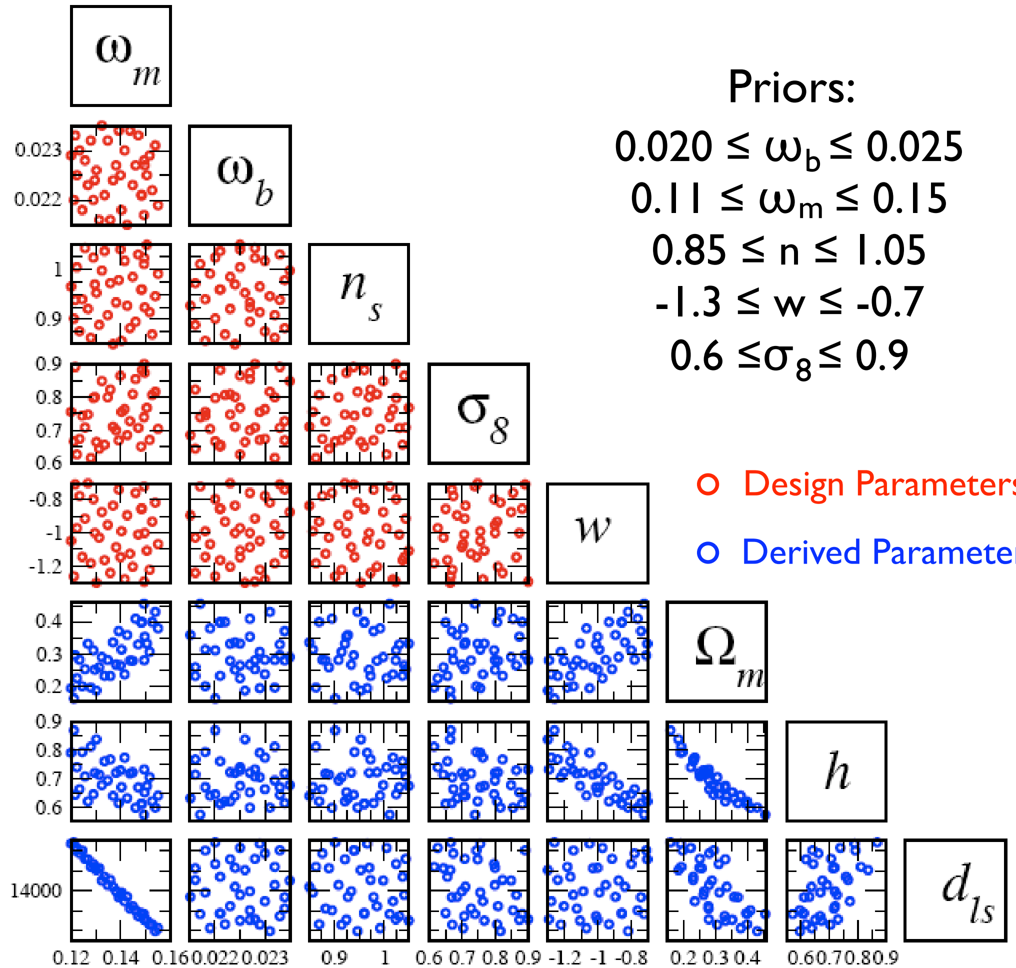
- **"Simulation design": for a given set of parameters to be varied and certain numbers of runs that can be done, at what settings should the simulations be performed?**
- **In our case: five cosmological parameters, tens of high-resolution runs are affordable**
- **First idea: grid**
  - Assume 5 parameters 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 projects 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 known at 2-3%
  - w, σ_8 less well known
- **For good emulator performance from very small number of runs**
  - Not too broad priors
  - Not too many parameters



Priors:

$0.020 \leq \omega_b \leq 0.025$

$0.11 \leq \omega_m \leq 0.15$

$0.85 \leq n \leq 1.05$

$-1.3 \leq w \leq -0.7$

$0.6 \leq \sigma_8 \leq 0.9$

o Design Parameters

o Derived Parameters

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 componet analysis
  - 5 PC bases needed, fifth PC basis pretty flat
  - next step: modeling the weights
  - Here: Gaussian Process modeling

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

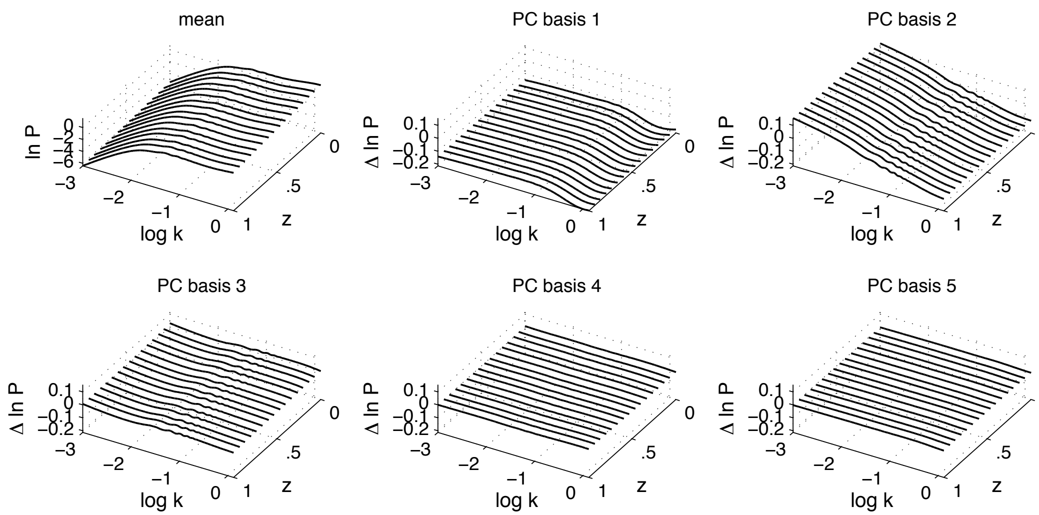
Number of basis function, here: 5

Basis functions, here: PC basis

Weights, here: GP model

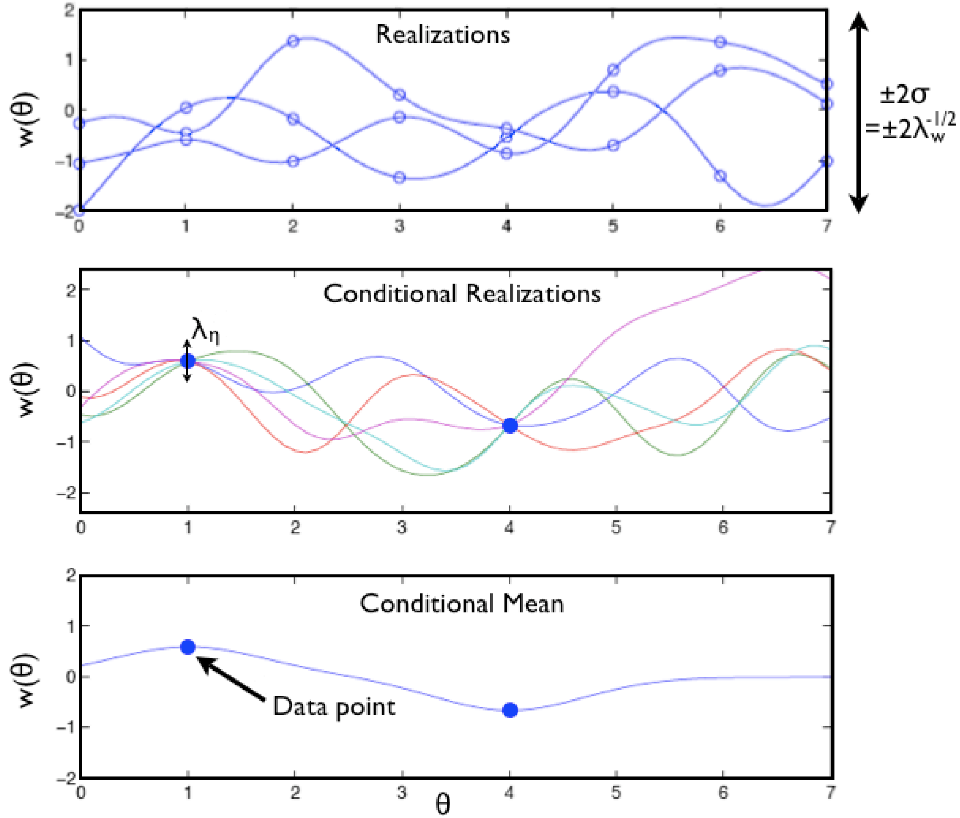
Cosmological parameters

Number parameters, 5



Gaussian Process Models

- **Nonparametric regression scheme, particularly well suited for interpolation of smooth functions**
- **Local interpolator, works well with space-filling sampling techniques**
- **Extending the notion of a Gaussian distribution over scalar or vector random variables into function space**
- **Gaussian distribution is specified by a scalar mean μ and a covariance matrix, GP specified by a mean function and a covariance function**



Unconditioned GP:

$$w \sim N(0, \lambda_w^{-1} R)$$

$$R_{ij} = \exp\{-||\theta_i - \theta_j||^2\}$$

Conditioned GP:

$$\begin{pmatrix} \theta \\ \theta^* \end{pmatrix} \sim N\left(0, \begin{pmatrix} K & K_*^T \\ K_* & K_{**} \end{pmatrix}\right)$$

$$\theta \;|\theta_* \sim N(K_* K^{-1}\theta, K_{**} K^{-1} K_*^T)$$

Gaussian Process Models

What is the probability to find a pair of points y1 and y2 in the plane?

K-1

$$P(y_2, y_1, K) = N \exp\left(-\frac{1}{2}(y_1, y_2)\begin{pmatrix} a & b \\ b & c\end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right)$$

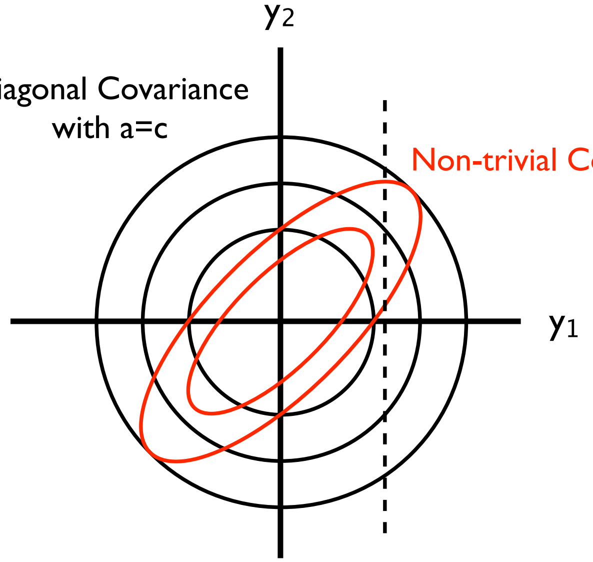
For a given y1, the distribution for y2 is:

$$= N' \exp\left(-\frac{1}{2}(c y_2^2 + 2b y_1 y_2)\right)$$

$$= N'' \exp\left(\frac{(y_2 - (-y_1 \frac{b}{c}))^2}{1/c}\right)$$

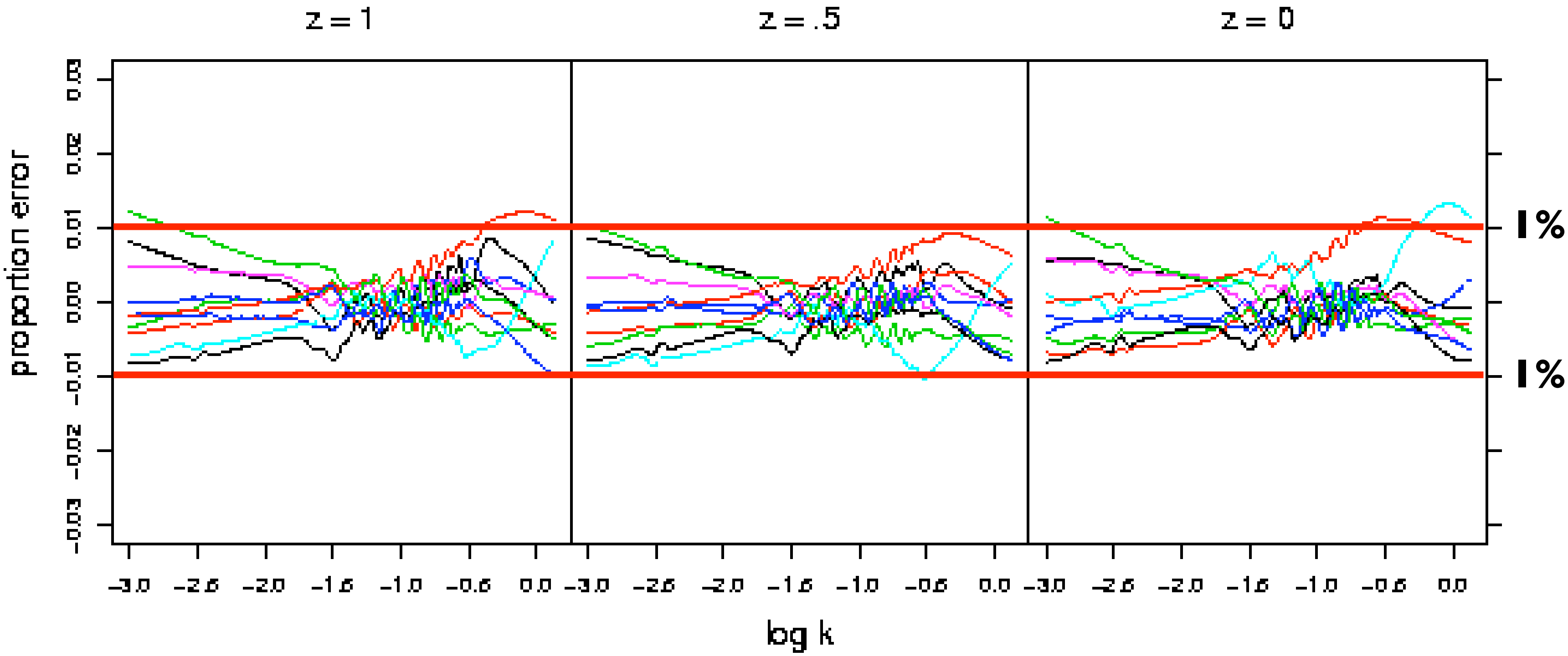
Conditioned mean of $y_2 = -y_1 \frac{b}{c}$

Variance of $y_2 = 1/c$



Even though the joint distribution of y1 and y2 is mean-zero, the conditioned distribution of y2 is not mean-zero, if the covariance matrix is not diagonal

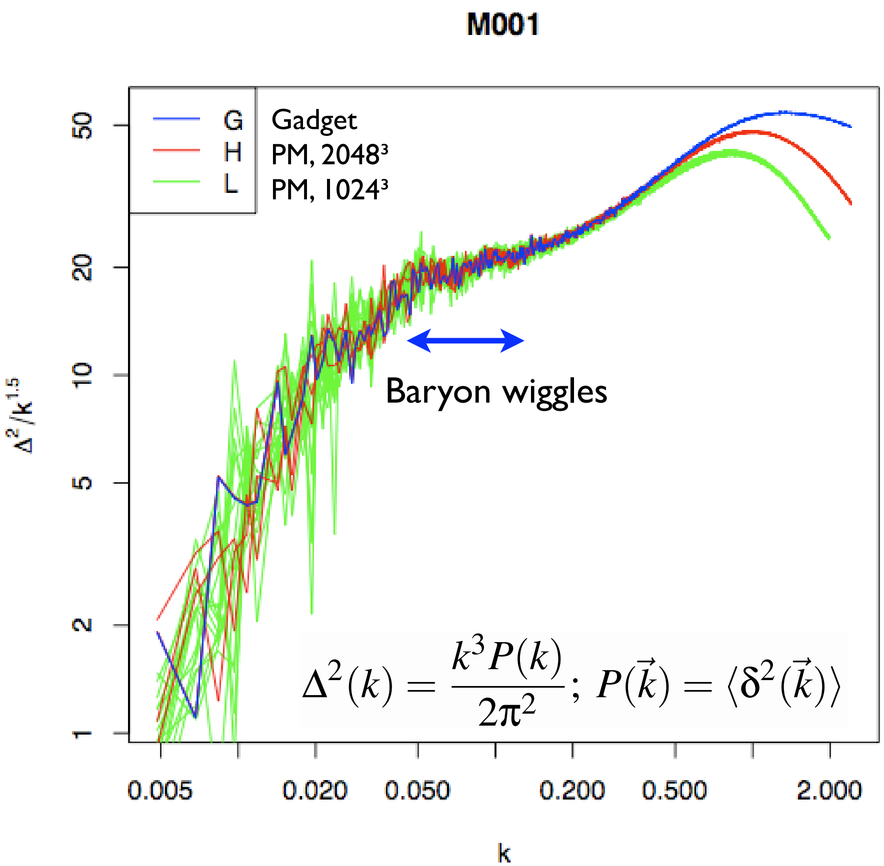
Emulator Performance



- Emulator: interpolation scheme, which allows us to predict the power spectrum at non-simulated settings in the parameter space under consideration
- Build emulator from 37 HaloFit runs according to our design
- Generate 10 additional power spectra within the priors with Halofit and the emulator
- Emulator predictions are accurate at the sub-percent level!

The Smoothing Procedure

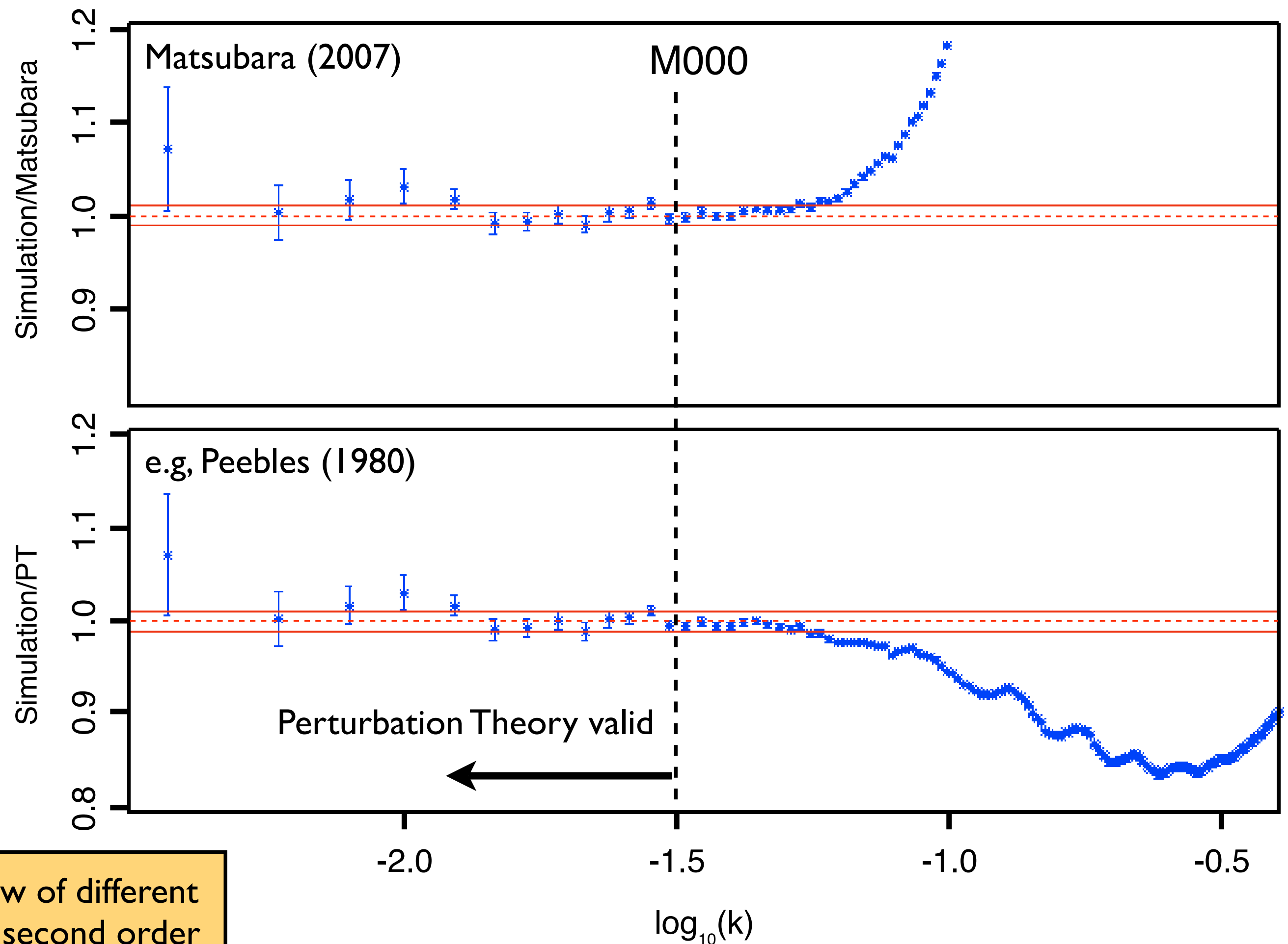
- Three different resolutions: 16 realizations low resolution PM, 4 realization medium resolution PM, one high-resolution Gadget run
- Make sure that features are not washed out
- Construct smooth power spectra using a process convolution model (Higdon 2002)
- Basic idea: calculate moving average using a kernel whose width is allowed to change over to account for nonstationarity
- For very low k: sparse sampling and large scale scatter, difficult to handle
- Maybe: perturbation theory



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

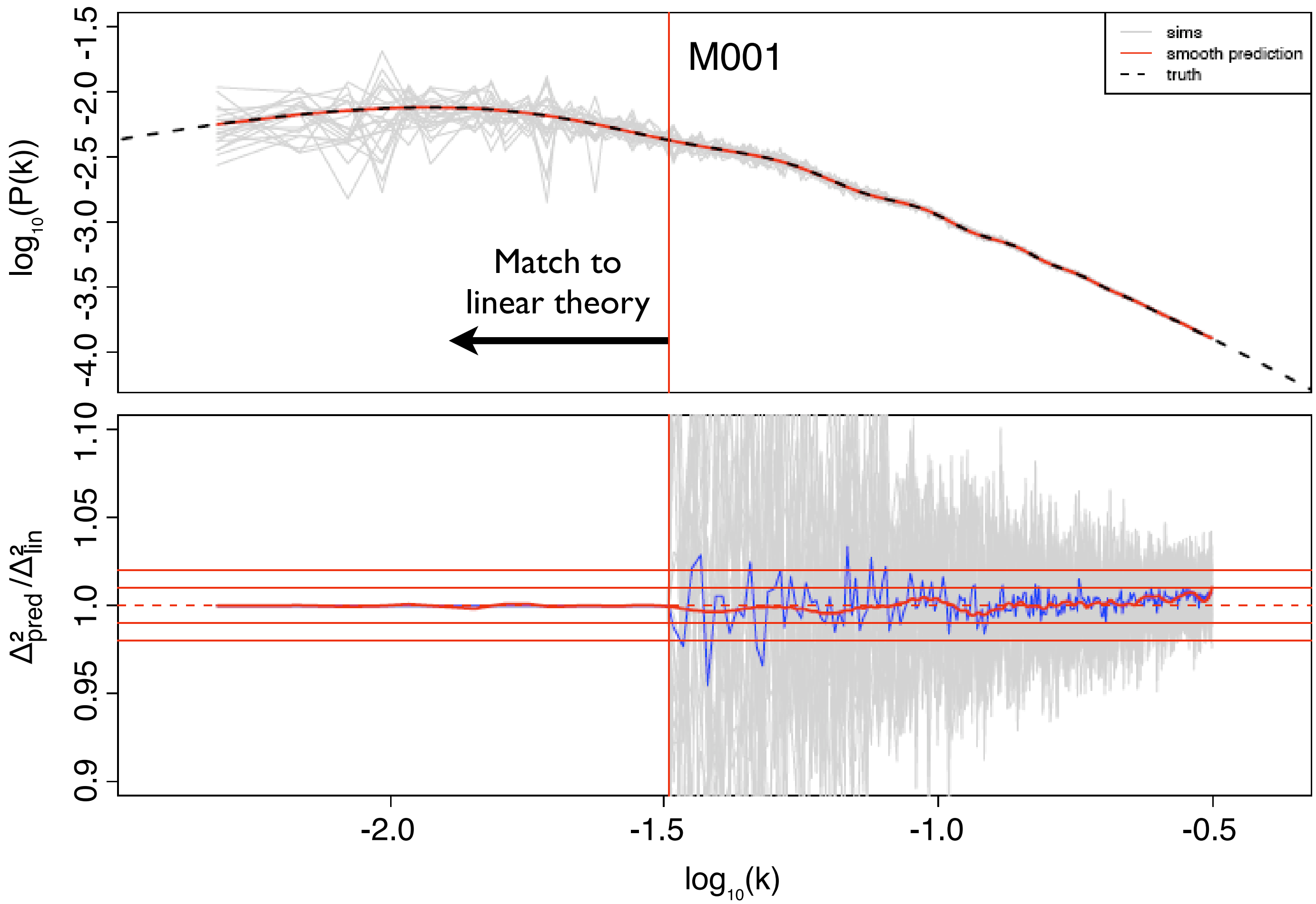
Perturbation Theory for low k

- **To reduce run-to-run scatter: 30 realization of 2.78Gpc boxes with PM code**
- **Compare different perturbation theory ideas**
- **PT works well below 1% accuracy out to at least k=0.03/Mpc**

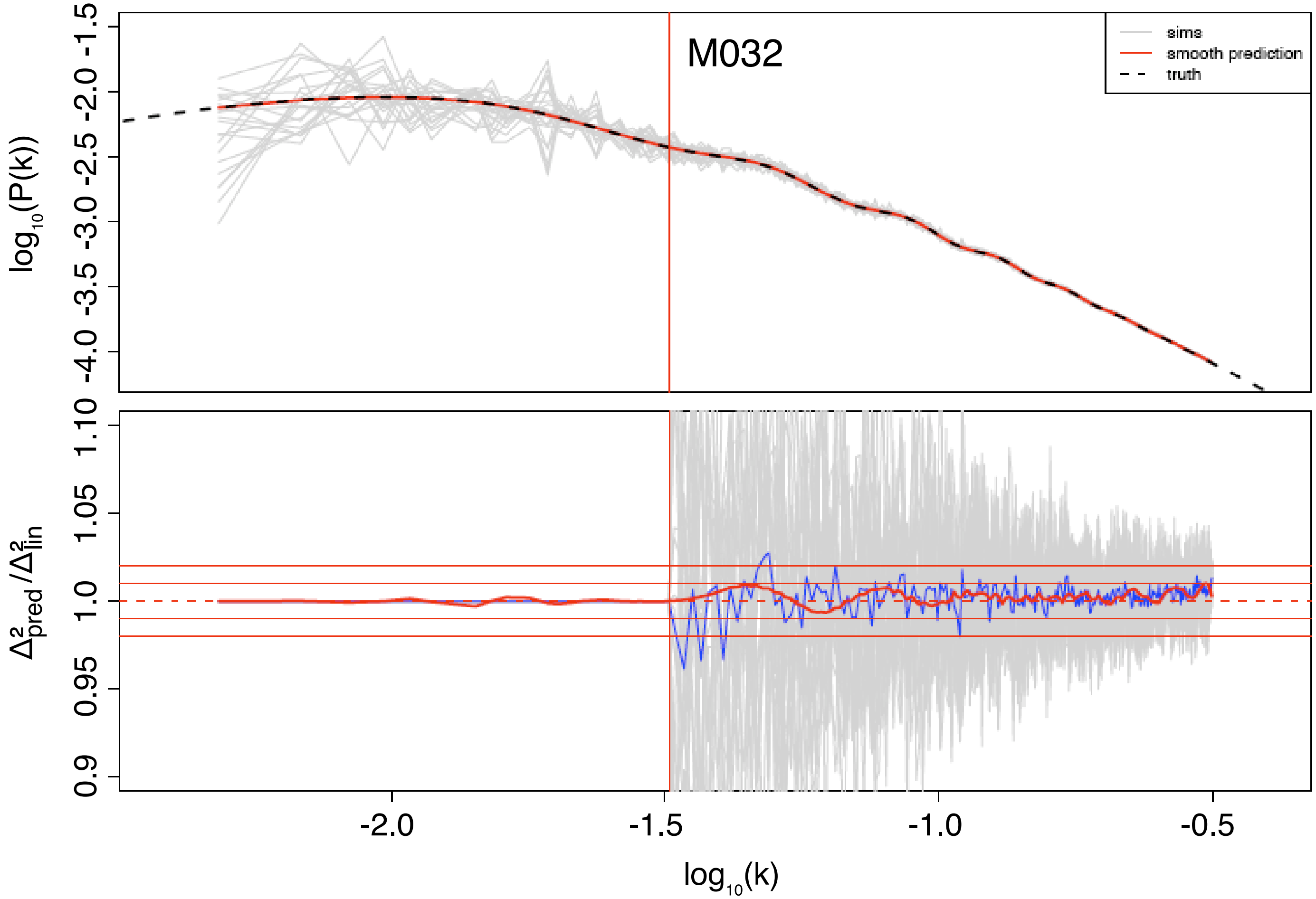


For an excellent review of different methods and first full second order calculation, see: J. Carlson et al. 2009

Test on Linear Power Spectrum

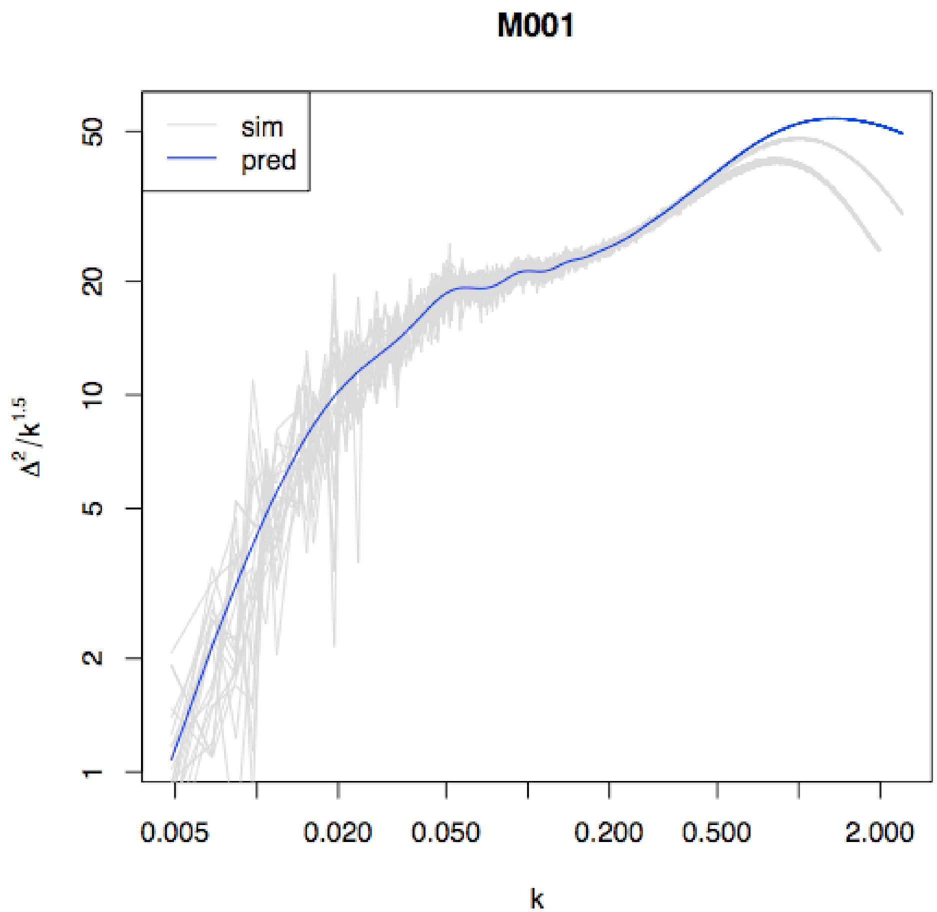


Test on Linear Power Spectrum



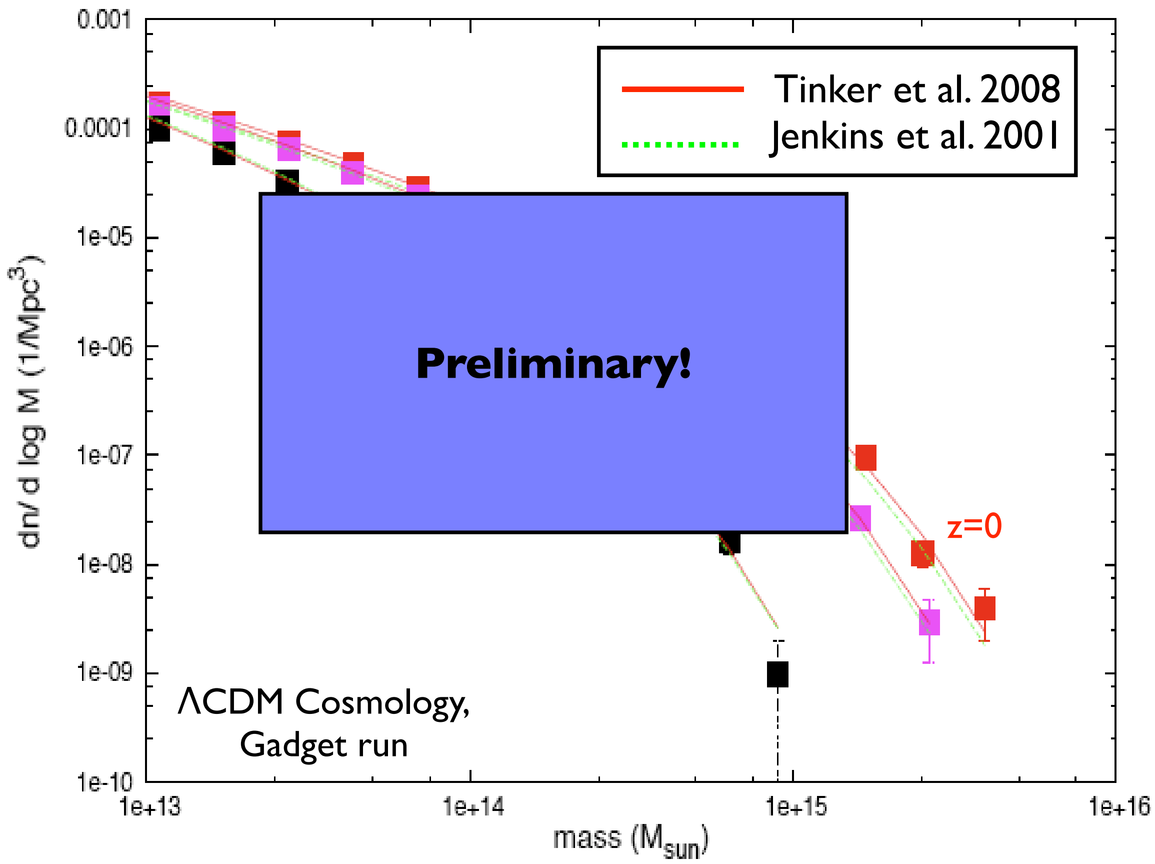
Results and Tests

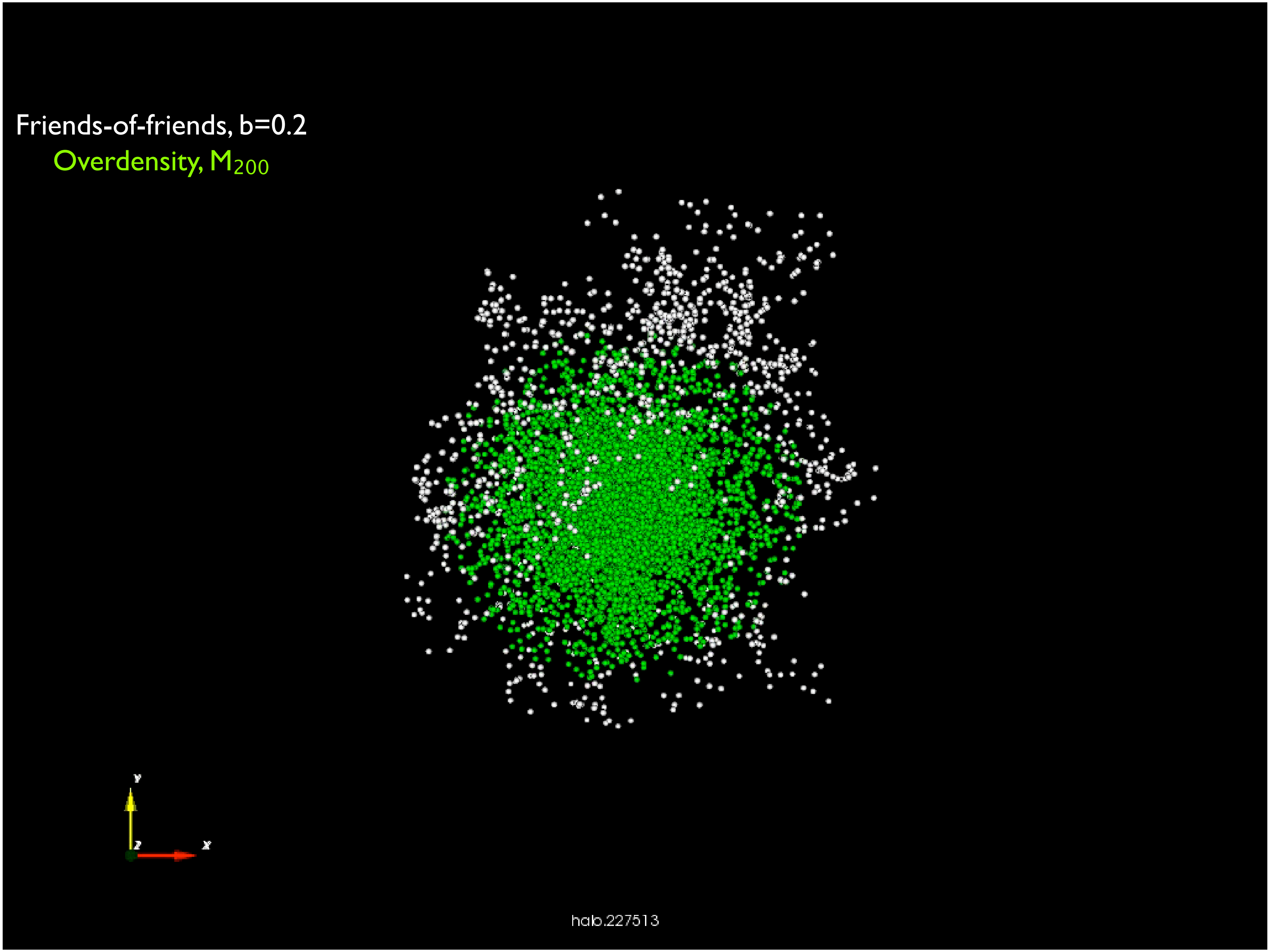
- Smoothed result from combination of PT and simualtions
- Recap: 37+1 models, 20 realizations at different resolution to cover the complete k-range of interest, 37+1 smooth power spectra
- Last step: construct emulator
- Tests: change parameters in smoothing procedure, predict power spectrum for M000 etc
- So far: everything works at the 1% level
- Emulator written in C with additional Fortran interface



The Halo Mass Function

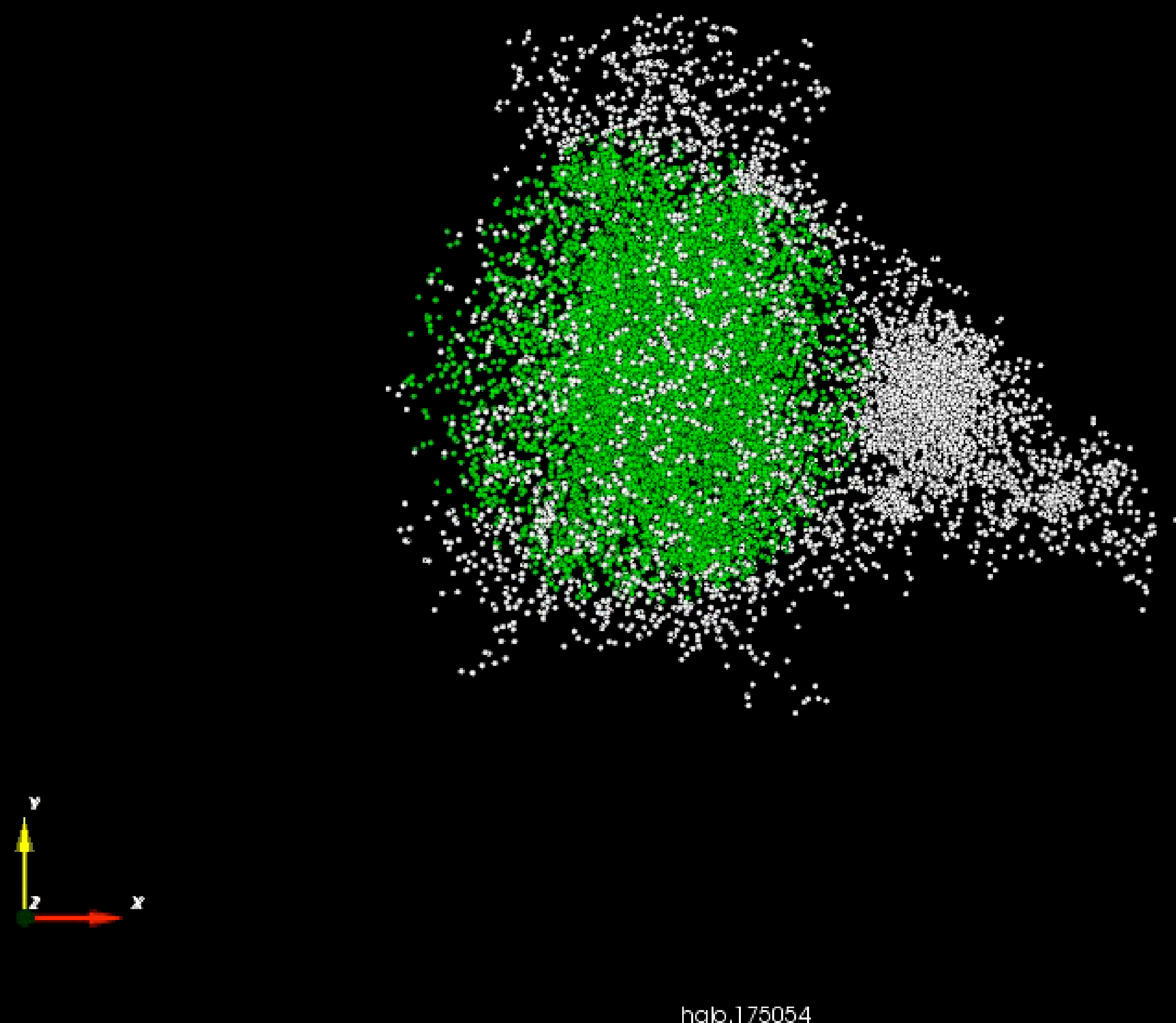
- Statistics describing the halo mass distribution in the Universe
- n(M): number density of halos with mass > M in a comoving volume element
- Evolution of mass function is highly sensitive to cosmology because matter density controls rate at which structure grows
- After Press/Schechter: semi-analytic fits by Sheth & Tormen (1999), Jenkins et al . (2001), Warren et al. (2006), Tinker et al. (2008) and many more...
- Dependence on halo definition, here overdensity (SO_{180b})

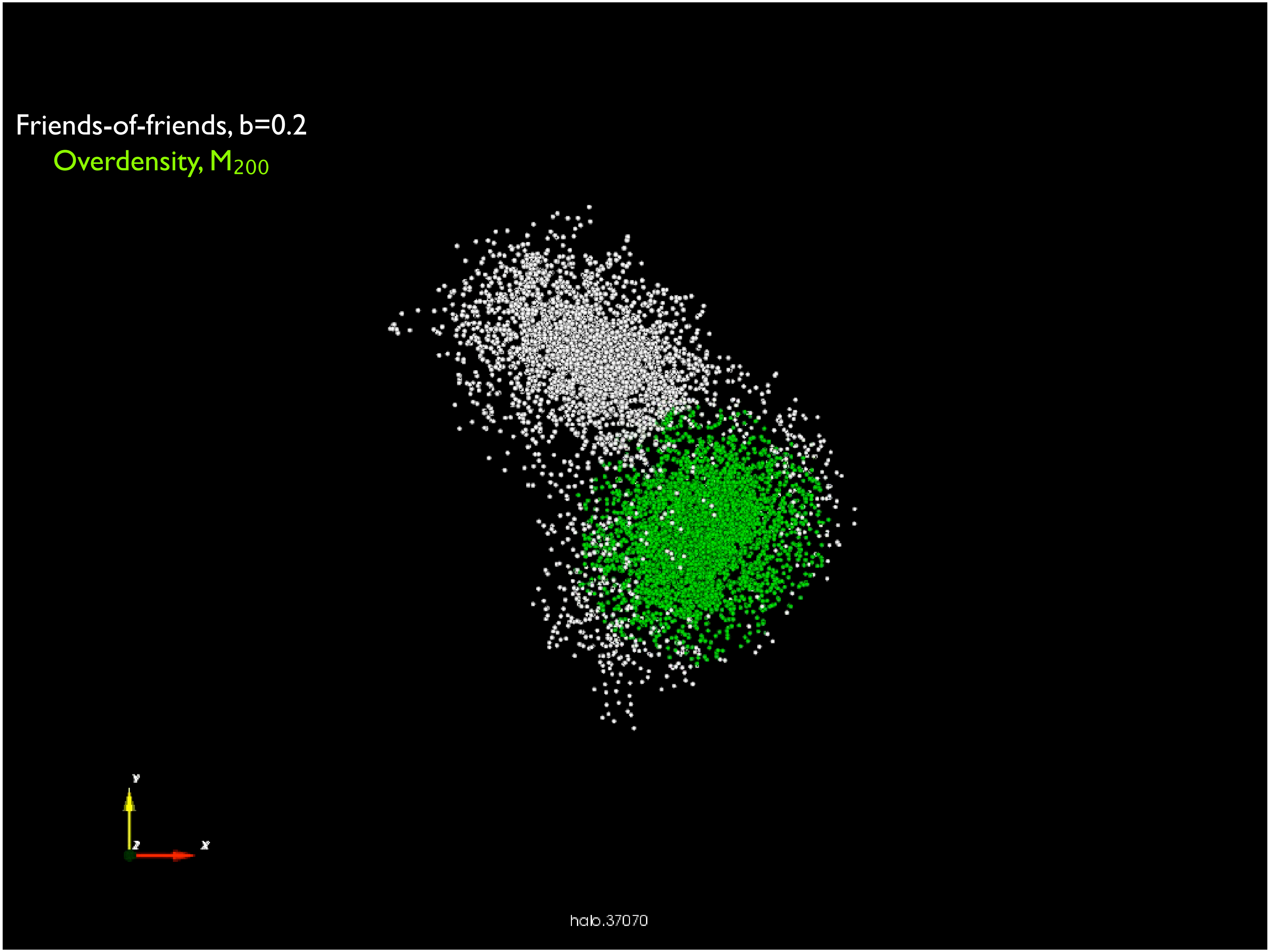



Friends-of-friends, b=0.2
Overdensity, M200
y
z
x
hab.227513

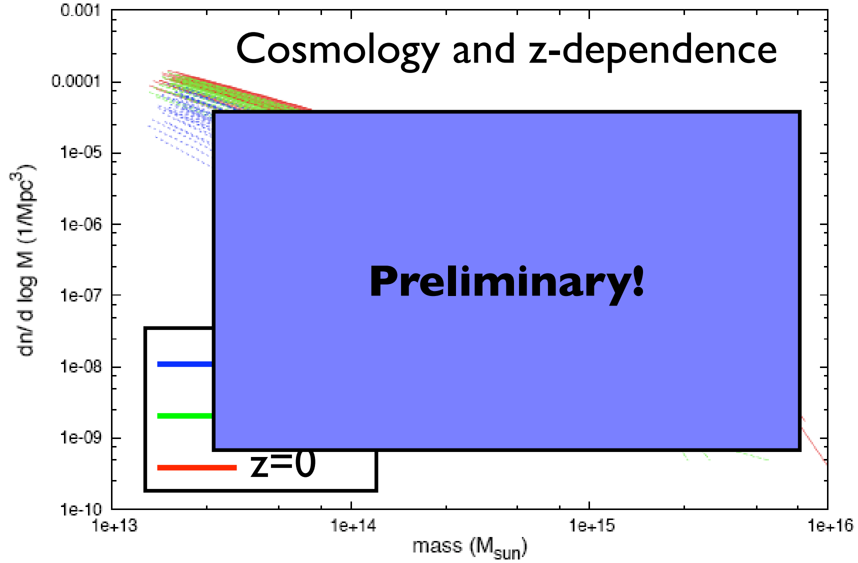
Friends-of-friends, b=0.2

Overdensity, M_{200}



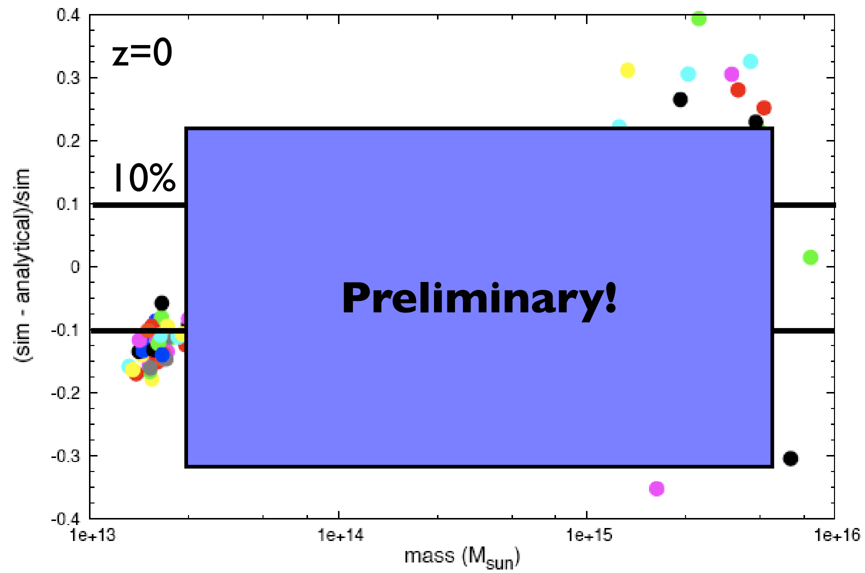

Friends-of-friends, b=0.2
Overdensity, M_{200}
y
x
z
halo.37070

The Halo Mass Function

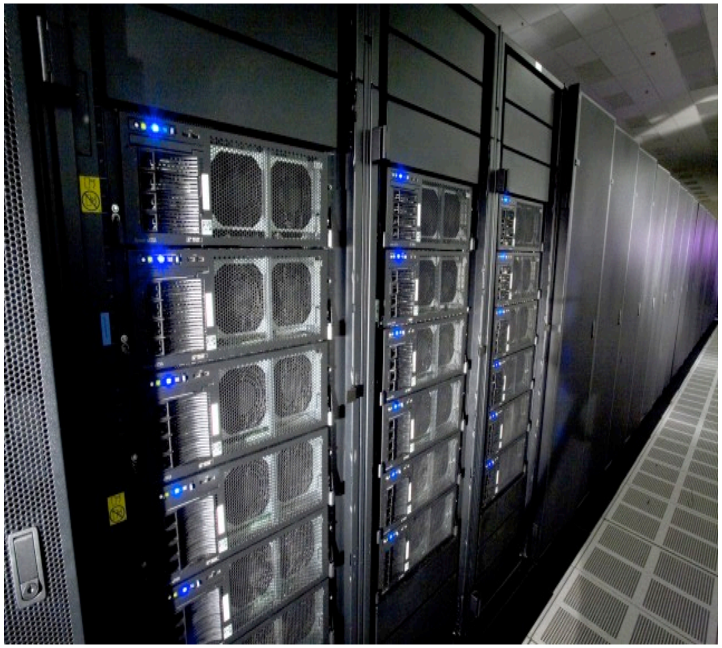


Idea: build an emulator for mass function at different redshifts, different cosmologies, and different halo defintions (linking length, overdensity)

Comparison to Tinker et al. 2008, which was derived for w=-1, agreement at z=0 at ~10%, slightly worse at higher z



The Next Step: The Roadrunner Universe



The Roadrunner Universe

inside
Los Alamos

ABOUT LANL · NEWS · LIBRARY · JOBS

Search

News Media Home

News Center

News Releases

- Archive

- Fact Sheets

Links - Daily News and Information

News Clips

Publications

Contacts

Los Alamos Supercomputer Remains Fastest in World

Contact: Kevin N. Roark, knroark@lanl.gov, (505) 665-9202 (04-388)

LOS ALAMOS, N.M., November 18, 2008 — ***New TOP500 list is announced IBM/LANL Roadrunner hybrid supercomputer still #1***

The latest list of the TOP500 computers in the world has been announced at the SC08 supercomputing conference in Austin, Texas, and continued to place the Roadrunner supercomputer at Los Alamos National Laboratory as fastest in the world running the LINPACK benchmark—the industry standard for measuring sustained performance.

Story Tools

Printer friendly version

Send article to a friend

123 View most sent articles

Roadrunner is currently housed at the Nicholas Metropolis Center for Modeling and Simulation at Los Alamos where it reached a sustained 1.105 petaflop/s on November 2, 2008.

"Petaflop/s" is computer jargon—peta signifying the number 1 followed by 15 zeros (sometimes called a quadrillion) and flop/s meaning "double-precision floating point operations per second."

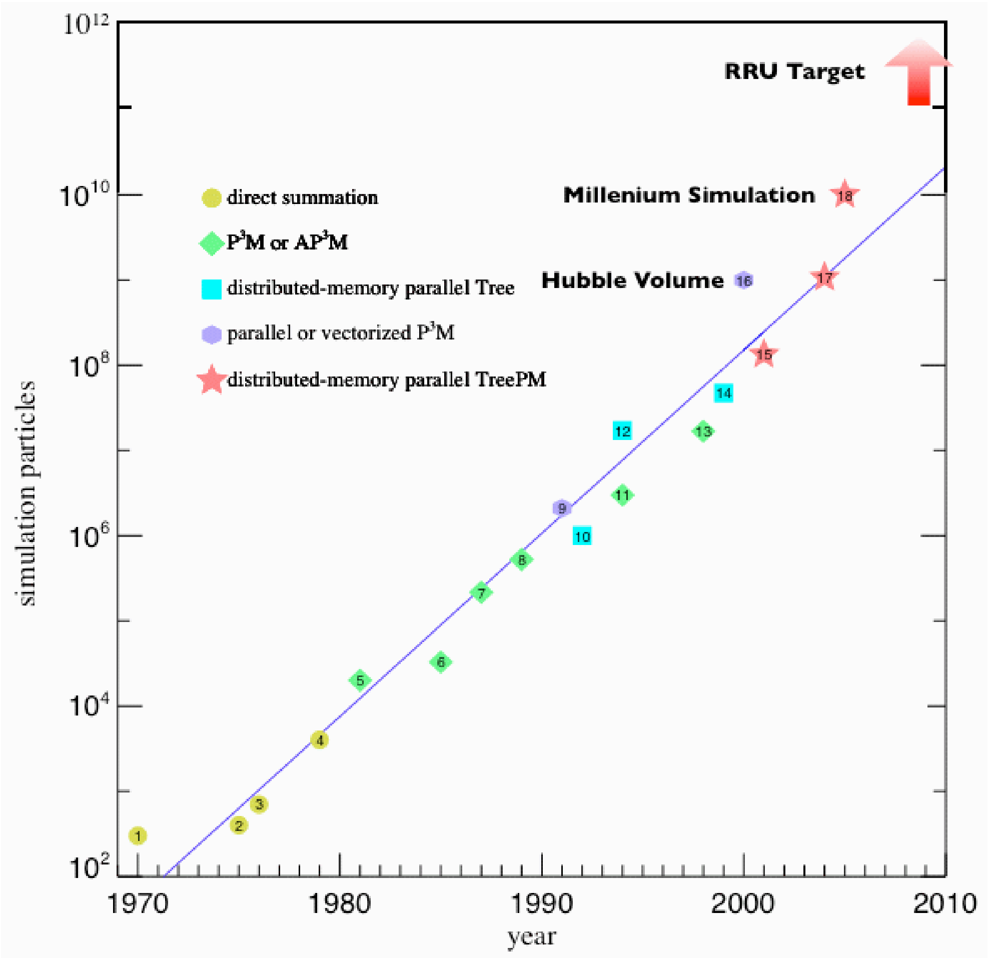
"The full Roadrunner system is now fully installed at Los Alamos and has entered its acceptance phase and is operating at or above designed performance," said Andrew White, Roadrunner project director. "We are looking forward to the integration phase where we use the machine to do some fascinating calculations in the unclassified realm, to see what it can really do."

The Roadrunner Universe is one of eight science projects selected for first six months of runtime! Equivalent to 100 Million Cpu hours on conventional hardware

The Roadrunner Universe

Collaboration: S. Habib, J. Ahrens, L. Ankeny, C.-H. Hsu, D. Daniel, N. Desai, P. Fasel, K.H., Z. Lukic, G. Mark, A. Pope

- New hybrid P^3M code
- Large suite of very large volume/large number of particle simulations with different cosmologies and realizations
- Lessons learnt from the Coyote Universe
- Large fraction of analysis needs to be done on the fly
- What information should be stored?



Conclusions

- **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
- **Future simulations**
  - Very large data sets
  - Emphasis on analysis, what should be done
  - How should data be made available to the community?