LA-UR-08-7921

Approved for public release; distribution is unlimited.

Title: The Coyote Universe II: Cosmological Models and Precision Emulation of the Nonlinear Matter Power Spectrum

Author(s):
Katrin Heitmann, LANL/ISR-1
Martin White, UC Berkeley (T-2)
Christian Wagner, AI Potsdam (ISR-1)
Slman Habib, LANL/T-2
David Higdon, LANL/CCS-6
Brian J. Williams, LANL/CCS

Intended for: Astrophysical Journal

Los Alamos
NATIONAL LABORATORY
EST. 1943

Form 836 (7/06)

DRAFT VERSION DECEMBER 11, 2008
Preprint typeset using LaTeX style emulateapj v. 26/01/00

LA-UR-08-xxxx

THE COYOTE UNIVERSE II: COSMOLOGICAL MODELS AND PRECISION EMULATION OF THE NONLINEAR MATTER POWER SPECTRUM

KATRIN HEITMANN[1], MARTIN WHITE[2], CHRISTIAN WAGNER[3], SALMAN HABIB[4], DAVID HIGDON[5], AND BRIAN J. WILLIAMS[5]

[1] ISR-1, ISR Division, Los Alamos National Laboratory, Los Alamos, NM 87545
[2] Departments of Physics and Astronomy, University of California, Berkeley, CA 94720
[3] Astrophysikalisches Institut Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam
[4] T-8, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545
[5] CCS-6, CCS Division, Los Alamos National Laboratory, Los Alamos, NM 87545

Draft version December 11, 2008

ABSTRACT

The nonlinear power spectrum contains a large amount of cosmological information. Especially, weak lensing has been established as a powerful probe of dark energy, if theoretical uncertainties can be controlled – or even better avoided. For large field weak lensing surveys the nonlinear power spectrum has to be known at the one percent level in order to allow constraints on dark energy at the same level of accuracy. It is therefore of utmost importance to provide predictions for the nonlinear power spectrum at very high accuracy. The major obstacle for doing this is the cost of high-precision N-body simulations. In this paper we demonstrate that a set of only 38 cosmological models – the "Coyote Universe" – will allow us to predict the nonlinear dark matter power spectrum at the required high accuracy over a parameter range set by upcoming cosmic microwave background observations. This paper is the second in a series of three, with the final aim to provide a high-accuracy prediction scheme for the nonlinear dark matter power spectrum.

Subject headings: methods: N-body simulations — cosmology: dark matter power spectrum

1. INTRODUCTION

Since the discovery of dark energy a decade ago by Riess et al. (1998) and Perlmutter et al. (1999), our understanding of the nature of dark energy has made little progress. The main reason for this is that the dark energy equation of state w so far is consistence with a cosmological constant $w = -1$ at the ten percent accuracy level and we have no constraints on any possible time dependence yet. In order to be able to distinguish different models of dark energy, such as a cosmological constant or quintessence models from a possible break-down of general relativity on very large scales, our measurements of w and its time-dependence have to improve by an order of magnitude to percent-level accuracy. To achieve this ambitious goal, different dark energy probes which measure the growth of structure and the expansion history of the Universe have to be employed.

To date, the four most promising probes are: (i) Supernovae Type Ia, to measure the expansion history of the Universe, (ii) clusters of galaxies, to measure the expansion history and growth of structure, (iii) baryon acoustic oscillations, to measure the expansion history, and (iv) weak lensing, to measure the expansion history and the growth of structure. The last two probes, baryon acoustic oscillations and weak lensing, rely strongly on precise predictions of the nonlinear matter power spectrum. In the case of baryon acoustic oscillations, the measurements are carried out on very large scales. Therefore, higher order perturbation theory might offer a promising way to obtain precise predictions for the nonlinear matter power spectrum (see, e.g., Crocce & Scoccimarro 2006; Matsubara 2008; Pietroni 2008; Carlson et al. 2009 and references therein). Weak lensing measurements cover much smaller scales, out to $k \sim 1-10\ h\mathrm{Mpc}^{-1}$ and even smaller in the future. On these scales, we have to rely on numerical simulations to obtain the required level of accuracy. The simulations have to be at least as accurate, optimally even more accurate, as the observations. As was shown by e.g. Huterer & Takada (2003) a wide-field weak lensing survey such as the Supernova/Acceleration Probe (SNAP [1]) requires power spectrum predictions at the 1% level accurary, and a survey such as the Large Synoptic Survey Telescope (LSST [2]) requires predictions at the 0.5% level accuracy to avoid biasing of cosmological parameter estimations.

These requirements bear two major challenges: first, we have to be able to proof that the simulations have reached the desired level of accuracy. In a recent paper (Heitmann et al. 2008) we have shown that at scales out to $k \sim 1\ h\mathrm{Mpc}^{-1}$ we can determine the nonlinear matter power spectrum at sub-percent accuracy. At smaller scales, baryonic physics becomes important at the few to ten percent level and has to be taken into account (White 2004; Zhan & Knox 2004; Jing et al. 2006; Rudd et al. 2008), a task which has to be tackled accurately in the near future. Second, in order to constrain cosmological parameters, we have to be able to cover a range of different cosmologies. Markov Chain Monte Carlo methods, which are commonly used for parameter determination, rely on tenth of thousands to hundred thousand models. Since an accurate N-body simulation on the scales of interest takes currently of the order of $\sim 30,000$ Cpu-hours, it is not feasible to run such simulations for each model. Running 10,000 N-body simulations with the required resolution on a 2048 processor Beowulf Cluster today, would take 30 years! We therefore need a method that allows us to obtain very accurate predictions for the matter power spectrum from a restricted number of simulations which then can be used for constraining cosmological parameters. In the following, we will refer to such a prediction scheme as emulator. The emulator will replace the results from N-body simulations for the matter power spectrum for a pre-defined set and range of cosmological parameters.

[1] http://snap.lbnl.gov
[2] http://www.lsst.org

In the cosmic microwave background (CMB) community several different paths have been suggested to speed up the calculation of the temperature anisotropy power spectrum. These include purely analytic (Tegmark & Zaldarriaga 2000; Jimenez et al. 2004) and combinations of analytic and semi-analytic fits (Kaplinghat et al. 2002). More recently neural network methods and machine learning techniques have been successfully used to generate very accurate temperature anisotropy power spectra (Fendt & Wandelt 2007a; Auld et al. 2007; Fendt & Wandelt 2007b; Auld et al. 2008). These methods are based on a large number of training sets, up to several tens of thousands. While this does not constitute a problem for anisotropy power spectra – codes such as CAMB and CMBFast are taking not much time – this approach is not feasible for the matter power spectrum.

As for the temperature anisotropy power spectrum, several attempts have been made in the past to find good approximations for the nonlinear matter power spectrum to avoid costly simulations. These range also from purely analytic derivations (e.g., Hamilton et al. 1991; Peacock & Dodds 1994 to semi-analytic fits which are calibrated against simulation results (e.g., Peacock & Dodds 1996; Smith et al. 2003). These approximations are accurate at best at the 5-10% level (see e.g. Heitmann et al. 2008 for a recent comparison of simulations with HALOFIT), hence are far from the requirements.

We have recently introduced the "Cosmic Calibration Framework" (Heitmann et al. 2006; Habib et al. 2007; Schneider et al. 2008) which combines sophisticated simulation designs with a Gaussian Process (GP) model to create a very accurate emulator from a very restricted set of simulations. By simulation design we mean the determination of parameter settings at which to carry out the simulations. One of the main reasons why the cosmic calibration framework can provide very accurate results from only a small number of training sets compared to the neural network approach is the optimization of the simulation design strategy to work well with the interpolation scheme. In this paper we will demonstrate that with only a small number of 38 simulations we can obtain an emulator for the nonlinear matter power spectrum at the level of one percent accuracy. We focus on the regime of $k \lesssim 1\ h\mathrm{Mpc}^{-1}$ and a redshift range between $z = 0$ and $z = 1$ covering the current space of interest for weak lensing measurements. Such an emulator will eliminate a major source of bias in determining cosmological parameters from weak lensing data.

This paper – establishing that with a small set of high-precision simulations a power spectrum emulator with sub-percent accuracy can be constructed – is the second in a series of three communications. In the first, we have demonstrate that it is possible to obtain dark matter power spectra at the sub-percent level accuracy out to $k \sim 1h\mathrm{Mpc}^{-1}$ and derived a set of requirements for such simulations. The third paper of the series will present results from the complete simulation suite based on the cosmologies presented in the current paper and publicly release a precision power spectrum emulator. The simulation suite is called the "Coyote Universe" after the cluster it has been carried out on.

The paper is organized as follows. In Section 2 we describe in detail the cosmic calibration framework with special emphasis on building a matter power spectrum emulator from a very small set of simulations. We explain the design strategy for generating the training sets and discuss the emulation step and demonstrate the emulator accuracy. Next we provide a sensitivity analysis which basically shows which cosmological parameter changes the power spectrum on what scales. Finally, we test the complete framework on a simulated data set using the emulator for the data analysis. We conclude in Section 3.

2. THE COSMIC CALIBRATION FRAMEWORK

The Cosmic Calibration Framework consists of four interlocking steps: (i) the simulation design, which determines at what parameter settings to generate the training sets, (ii) the emulation which replaces the simulator, (iii) the uncertainty and sensitivity analysis, and (iv) the calibration against data to find parameter constraints.

In the following we will discuss each of these steps in detail with special emphasis on generating a very accurate emulator to determine the nonlinear matter power spectrum.

2.1. *Design of the model grid*

As discussed in the Introduction, one of the major challenges in building an accurate emulator for the nonlinear matter power spectrum is due to the very high cost of N-body simulations. We have to find a scheme which builds upon an as small number of training sets as possible. An important step therefore is to decide on the model grid which will comprise the training sets. This decision is guided by the interpolation scheme used to build the emulator and current constraints from observations on the parameter space of interest. Space-filling Latin hypercube (LH) designs have been proven to be well suited for the GP model based approach (Sacks et al. 1989a; Currin et al. 1991).

The input on the parameter choices and ranges from observations are based on the most recent CMB observations from WMAP-5 (Komatsu et al. 2008). In the following, we will discuss these two components for our design – the statistical and the observational inputs.

2.1.1. *Latin Hypercube Sampling*

The aim is to find a distribution of the parameter settings – the design – which covers the parameter space optimally with a small number of sampling points. The first design that might come to mind, is a simple grid. Suppose we want to vary 5 cosmological parameters and we can sample each parameter only three times – at the maximum, at the minimum, and in the middle of each parameter range. This would require $3^5 = 243$ simulations. First, 243 is not a very small number, second, the parameter space is not well covered at all with only three sample points per parameter. While some interpolation schemes

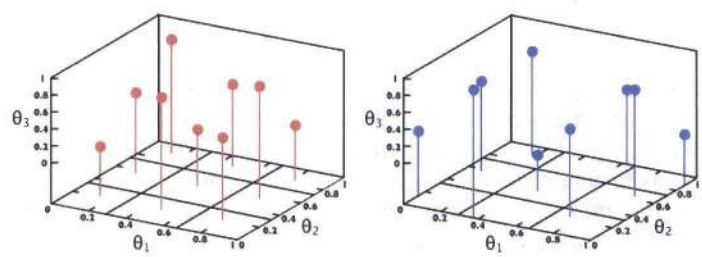

FIG. 1.— Left panel: an orthogonal array (OA) based design for 3 parameters, θ_1, θ_2, θ_3 and nine sampling points. Right panel: the OA based design has been perturbed in such a way that the one-dimensional projection onto any parameter leads to an equally spaced distribution of sample points. The projection into any two dimension leads to a space filling design.

rely on grid designs, this is not a feasible approach for us. Fortunately, the GP model interpolation scheme does not require a grid design. Another approach would be just to randomly sample the parameter space. As for the grid design, this approach might work well if a large number of sample points is available. For a small number, however, the points could be clustered in some parts of the parameter space and not cover others. Well-suited designs for GP model emulators are LH based designs. An LH design is an $n \times m$ matrix in which each column is a random permutation of $\{1,....,m\}$. The use of LH designs in application where the aim is to predict the outcome of some quantity at untried parameter settings from a restricted set of simulations was first proposed by McKay et al. (1979). As we discuss in some more detail later, GP models rely on local information for their interpolation strategy. Therefore it is important, that the parameter settings for the simulation runs on which the interpolation will be based, provide a good coverage in the whole parameter space. Space-filling LH designs and variants therefore provide an optimal approach for this.

Very often, the LH designs are combined with other design strategies such as orthogonal array (OA)-based designs or they are optimized in other ways, e.g., by symmetrizing them (more details are given below). By combining different design strategies, different attributes of the sampling strategies will lead to even better designs and short-comings of a specific design will be eliminated. As a last step, optimization schemes for spreading out the points evenly in a projected space are often applied. One such optimization scheme is based on minimizing the maximal distance between points in the parameter space, which will lead to an even coverage. For a discussion on different design approaches and their specific advantages see, e.g., Santner et al. (2003).

Let us now briefly discuss two design strategies well suited for cosmological applications in which the number of parameters is still much less than the number of simulations that can be performed. We first explain how to set up an optimal OA-LH design and then give a description of an optimal symmetric LH design. The former has been used in previous work in cosmology (Heitmann et al. 2006; Habib et al. 2007), the latter will be used in this paper to construct the design for the Coyote Universe. For illustration purposes, we will use a very simple, three-dimensional case with three parameters, θ_1, θ_2, θ_3 and nine sampling points.

In order to create an OA-LH design, we break up our strategy into two steps: the orthogonal arrays and the construction of the orthogonal-array based Latin hypercube on top of that. We follow very closely the description by Tang (1993) and Leary et al. (2003). An orthogonal array distributes runs uniformly in certain projections of the full parameter space. A more mathematically rigorous description is given by the following definition: An n by m matrix $\mathbf{A}$ with entries from a set $1,2,...,s$ is called orthogonal array of strength r, size n with m constraints and s levels if each $n \times r$ submatrix of $\mathbf{A}$ contains all possible $1 \times r$ rows with the same frequency λ. Here λ is termed the index of the array, and $n = \lambda s^r$. The array is denoted OA(n,m,s,r) (Tang 1993).

For our application, n denotes the number of simulation runs which we can perform and m specifies the number of parameters we want to vary (these can be cosmological parameters as well as numerical input parameters). Therefore, m also reflects the number of dimensions in the parameter hypercube. s defines how many levels of stratification each column in the

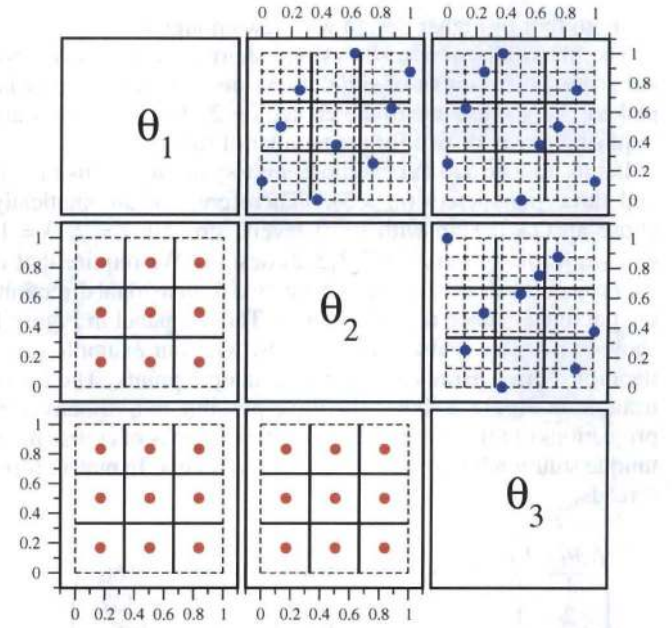

FIG. 2.— Projections of the design shown in Figure 1 into two dimensions. The lower triangle shows the projection of the OA design, the upper triangle of the OA-LH design. Note that when projected into one dimension, the OA-LH design leads to an even coverage and no points are on top of each other.

matrix $\mathbf{A}$ has. In order to fill the parameter space well, it is often not enough to fill it well globally. Especially if two or more parameters interact strongly with each other, one would like to have a good space-filling design in the subspace of these parameters. In other words, if one projects the design down into, e.g, two dimensions, such projection should have then space-filling properties in those two dimensions too. The parameter r, the strength of OA designs, indicates the projections for which the LH design based on that OA are guaranteed to be space-filling. For example, if $r = 3$, then all 1, 2 and 3 dimensional projections will be space-filling. Obviously, r cannot be larger than m, the number of parameters varied. The strength r is usually restricted to two or three for several reasons: (i) Creating designs which have space-filling properties (besides the global space filling property) in projected subspace higher than three dimensions is algorithmically extremely demanding. (ii) In most applications, only a small number of parameters influence the response heavily. Statisticians call this the "20-80 rule" – 20% of the parameters are responsible for 80% of the outcome variation. Therefore, one would like to capture these parameters well. Furthermore, outcome variation is often dominated by a small number of single parameter and two-factor interaction effects, which are adequately covered by OA-LH designs based on $r = 2$ or 3. (iii) The number of simulations often has to be kept small, therefore r can not be chosen too large, since the number of simulations n is connected to r via $n = \lambda s^r$. As for r, s is also restricted by the number of runs one can possibly perform. It is very often set to $s = 2$ which then requires the number of runs to be a power of two. The frequency λ with which a permutation repeats, is kept small too. To create a design, the strategy is to fix strength first, and try to find an OA design that has approximately the right number of runs and at least as many parameters as one needs. If not, then the strength is reduced by one and one tries again. Usually, this strategy is started with OA designs of strength 3, though many more designs of strength 2 exists. It is rarely possible to find a

strength four or higher design with few enough runs.

The above discussion shows one short-coming of orthogonal arrays: the number of simulation runs cannot arbitrarily be picked. Very often due to the choice $s = 2$, the OA designs are requiring power of two for the number of runs. [3]

Let us now turn to our example: nine sampling points ($n = 9$) and three parameters ($m = 3$). Therefore, we automatically chose an OA design with $s = 3$ levels, strength $r = 2$, $\lambda = 1$, and we leading to an OA(9,3,3,2) design. We require that if we project our design down into any two dimensional direction, the parameter space is well covered. The left panel in Figure 1 shows a possible realization of an OA with our example specifications, three parameters and nine sample points. The lower triangle in Figure 2 shows the three possible two dimensional projections of this design. This specific design is of course not a unique solution but fulfills all our requirements. In matrix form it reads:

$$\begin{pmatrix} \theta_1 & \theta_2 & \theta_3 \\ \hline 1 & 1 & 1 \\ 2 & 1 & 3 \\ 3 & 1 & 2 \\ 1 & 2 & 2 \\ 2 & 2 & 1 \\ 3 & 2 & 3 \\ 1 & 3 & 3 \\ 2 & 3 & 2 \\ 3 & 3 & 1 \end{pmatrix} \overset{[0...1]}{\Leftrightarrow} \begin{pmatrix} 0.166 & 0.166 & 0.166 \\ 0.5 & 0.166 & 0.832 \\ 0.832 & 0.166 & 0.5 \\ 0.166 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.166 \\ 0.832 & 0.5 & 0.832 \\ 0.166 & 0.832 & 0.832 \\ 0.5 & 0.832 & 0.5 \\ 0.832 & 0.832 & 0.166 \end{pmatrix} \quad (1)$$

From this 9×3 matrix we can now verify that each of the three 9×2 sub-matrices indeed contains all possible 1×2 rows with the same frequency $\lambda = 1$. On the right hand side of Equ. (1) we simply rescaled the design points into the [0...1] space which is shown in Figures 1 and 2.

In order to cover the parameter space even better, in the next step – the latin hypercube sampling – we will perturb the positions of each sampling points from $\mathbf{A}$ via the following prescription: for each column of $\mathbf{A}$, the λs^{r-1} positions with entry k are replaced by a permutation of

$$[(k-1)\lambda s^{r-1}+1, (k-1)\lambda s^{r-1}+2, \ldots, (k-1)\lambda s^{r-1}+\lambda s^{r-1} = k\lambda s^{r-1}. \quad (2)$$

This means in our example that the entries for $k = 1$ will be replaced by 1,2,3, the entries for $k = 2$ will be replaced by 4,5,6, and the entries for $k = 3$ by 7,8,9. Another way of describing this step is the following: The latin hypercube algorithm demands that no two sample points are in the same column and due to the OA base no two sample points will be in the same row either. This requirement is equivalent to the following: if the two dimensional projection of each two parameters is project down to one dimension, the sampling in this one dimension is again evenly covered with points. The right panel in Figure 1 shows a possible realization of this in three dimensions, derived from perturbing the orthogonal array in the left panel. The upper right triangle in Figure 2 shows the two-dimensional projection of this design. The solid lines show the division for the orthogonal array while the dashed lines show the additional sub-division. Note that each sample point lies on a unique horizontal and vertical dashed line – if we would project the sample points down into any one direction, the one-dimensional space would be evenly covered. In matrix form, our OA-LH design looks like follows:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 1 & 9 \\ 7 & 3 & 5 \\ 2 & 5 & 6 \\ 5 & 4 & 1 \\ 8 & 6 & 7 \\ 3 & 7 & 8 \\ 6 & 9 & 4 \\ 9 & 8 & 2 \end{pmatrix} \overset{[0...1]}{\Leftrightarrow} \begin{pmatrix} 0 & 0.125 & 0.25 \\ 0.375 & 0 & 1 \\ 0.75 & 0.25 & 0.5 \\ 0.125 & 0.5 & 0.625 \\ 0.5 & 0.375 & 0 \\ 0.875 & 0.625 & 0.75 \\ 0.25 & 0.75 & 0.875 \\ 0.625 & 1 & 0.375 \\ 1 & 0.875 & 0.125 \end{pmatrix} \quad (3)$$

Note that we have replaced the entries randomly - we convinced ourselves "by eye" that we have good coverage in 2-D projection. Leary et al. (2003) suggest to choose an optimal strategy to ensure even better coverage of the parameter space. These optimization strategies are often used for the projected parameter space. In order for the points to spread out, one has to determine the closeness of points. This can be defined as the smallest distance between any two points. A design that maximizes this measure is said to be a maximin distance design. For more details, see Santner et al. (2003). The designs in Heitmann et al. (2006) and Habib et al. (2007) combine the OA-LH based design with a maximum distance design in each two-dimensional projection. Other optimization methods are based on an entropy criterion which is based on the minimization of -log $|R|$, where R is the covariance matrix of the design (Shewry & Wynn 1987) or Integrated Mean Squared Error (Sacks et al. 1989b).

Our example shows just one way to realize an OA-LH design. It can be implemented straightforwardly and leads to the desired good coverage of the parameter space. After the design has been determined in the [0,1] parameter space, it then can be easily translated into the physical parameter space of interest. At this point, when projected down to one dimension, the even coverage of the parameter space in one dimension is of course lost, but still no two sample points will fall on top of each other in projection.

As mentioned above, the major restriction of OA-LH based designs is that they cannot be setup for an arbitrary number of simulation runs. This is specifically a concern, if one can only run a very restricted number of simulations. In addition, the setup of an OA-LH can be non-trivial. Very often, one has to rely on OA libraries which are restricted in their parameters and also not always easily available. Li & Ye (2000) propose an alternative, space-filling design strategy, which offers a compromise between the computing efforts and the design optimality - optimal symmetric latin hypercube designs. Following their definition, a LH design is called a symmetric LH (SLH) design if it has the following property: For any row i of a LH design, there

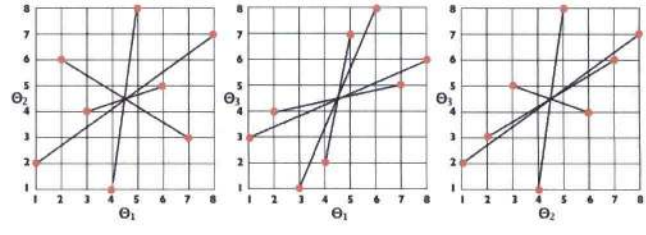

FIG. 3.— Two-dimensional projections of the SLH design given in Eqn. (4). We have connected the symmetric design points to show the reflection through the center.

[3] The other "short-coming" of orthogonal arrays is that they are not easy to construct. A library with OAs can be found here: http://www.research.att.com/ njas/oadir/index.html

exists another row in the design which is the ith row's reflection through the center. For example, in an $n \times m$ Latin hypercube with levels from 1 to n, if $(a_1, a_2, ..., a_m)$ is one of the rows, the vector $(n+1-a_1, n+1-a_2, ..., n+1-a_m)$ should be another row in the design. The symmetry imposes a space-filling requirement on the designs considered up front, which carries through to all projections. An example for an SLH design is given by:

$$\begin{pmatrix} 1 & 2 & 3 \\ 8 & 7 & 6 \\ 4 & 1 & 2 \\ 5 & 8 & 7 \\ 7 & 3 & 5 \\ 2 & 6 & 4 \\ 3 & 4 & 1 \\ 6 & 5 & 8 \end{pmatrix}. \quad (4)$$

In this design, the ith+1 row is symmetric to row i. We start the construction of this design from our original OA-LH design. Row by row we derive the symmetric ith+1 row until in the 7th row we have to ensure by hand that the LH nature of the design is preserved. Obviously, the number of rows has to be even for this kind of design. As before, we do not attempt to optimize the resulting design. Its two-dimensional projection is shown in Figure 3. Li & Ye (2000) provide an excellent discussion of optimal SLH designs, including a description of possible algorithmic implementations and comparison with traditional optimal LH designs. As an example, they show that the computational effort to find an optimal LH design by starting with an SLH design reduces roughly by a factor of ten for a 25×4 design from 48026 seconds to 3574 seconds on a workstation. As before, the SLH design is usually optimized in the last step, often with respect to a distance based criterion which spreads out the points in two-dimensional projections. Two standard search algorithms for optimal SLH design are the columnwise-pairwise (CP) algorithm by Ye (1998) and the simulated annealing (SA) algorithm Morris & Mitchell (1995). Simply put, these algorithms are based on columnwise exchanges of entries which will keep the symmetry properties (since the corresponding symmetric pairs are always switched at the same time). They are iterative procedures, which will stop after the certain pre-set optimization criterion is fulfilled or the process is interrupted by time limitations. Very often, several designs are produced at the same time and the most optimal will be kept in the end.

In the following, we will use a LH design which is optimized via a distance criterion. In detail, 20 optimizations with CP and 20 with SA were carried out, and the best was chosen in the end where the quality was measured by a distance criterion. For CP, 10 of the designs had a symmetry requirement and the other 10 did not. For SA, 10 of the designs had a symmetry requirement and were optimized with a local optimization criterion, and the other 10 did not have a symmetry requirement and were optimized with a more global optimization criterion. The best design from all of these came from one of the optimizations using SA, a global optimization criterion, and no symmetry requirement.

2.1.2. *Observational Considerations*

We take as our basic 5 parameters $\omega_m \equiv \Omega_m h^2$, $\omega_b \equiv \Omega_b h^2$, n_s, w, and σ_8 where Ω_m contains the contributions from the dark matter and the baryons. Thus we restrict ourselves to power-law models (no running of the spectral index), to spatially flat models without massive neutrinos and to dark energy models with constant equation of state. A sixth parameter, the redshift or time, simply requires us to dump data from each run at multiple epochs.

The effect of massive neutrinos can be roughly approximated by decreasing Ω_m (Brandbyge et al. 2008). We do not expect any realistic dark energy model to have a constant equation of state, but we wanted to begin with the most restrictive parameter space in order to validate our methods. The next generation of experiments will pose at best weak constraints on any time variation of w, and in this sense our constant w can be thought of as an appropriate average of $w(z)$. Using growth matching techniques (White & Vale 2004; Linder & White 2005; Francis et al. 2007) one can map the power spectrum from a complex $w(z)$ onto one with a constant w with reasonable accuracy.

The normalization is another area where choices need to be made. Historically the amplitude of the power spectrum was set by σ_8, the amplitude of the *linear theory* matter power spectrum smoothed with a top-hat on scales of $8\,h^{-1}$Mpc

$$\sigma_8^2 \equiv \int \frac{dk}{k}\, \Delta^2_{\rm lin}(k) \left[\frac{3 j_1(kR)}{kR}\right]^2 \Bigg|_{R=8\,h^{-1}{\rm Mpc}}, \quad (5)$$

with the linear power spectrum being defined as

$$\Delta^2_{\rm lin}(k) \equiv \frac{k^3 P_{\rm lin}(k)}{2\pi^2}. \quad (6)$$

This scale and normalization was chosen because the fluctuations of counts of L_* galaxies in cells of this size is close to unity. With the advent of the *COBE* data it became common to quote the normalization at horizon scales e.g. Bunn & White

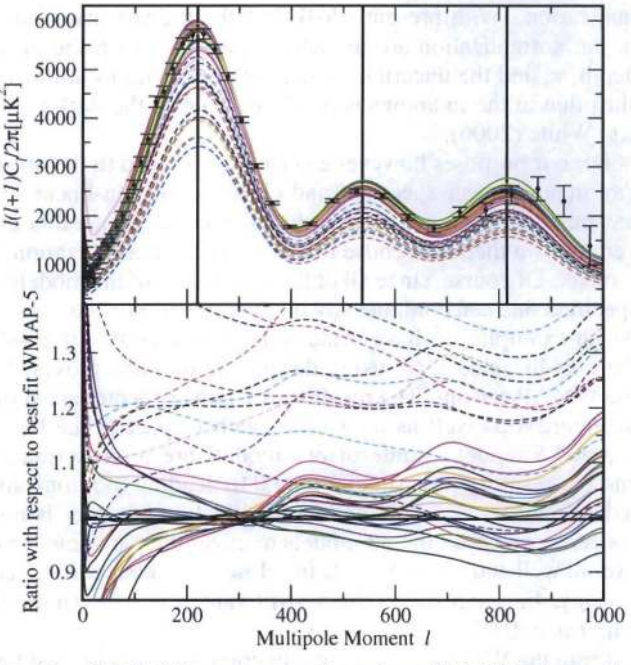

FIG. 4.— Best-fit TT power spectra for each model in Table 1 using the WMAP-5 results. The only parameter which has been optimized by minimizing χ^2 is the optical depth τ. The upper panel shows the resulting power spectra, the black points with error bars show WMAP-5 data points and the thick black line the best-fit WMAP-5 model. The lower panel shows the residuals for the each model with respect to the best-fit model. Some of our models are undernormalized, the best-fit τ is smaller than 0.01 which would lead to a reionization redshift of $z < 2$ and χ^2 for these models is larger than 3000 (the χ^2 for the best-fit WMAP-5 model is at roughly 2650). We fixed τ for those models at $\tau = 0.01$ and show them with dashed lines.

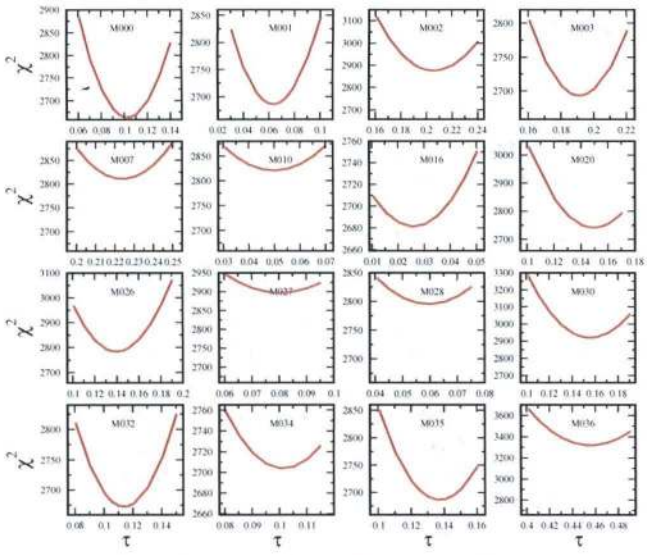

FIG. 5.— Sweep through τ for 16 models for which $\tau > 0.01$. The lowest value on the y-axis is set to the χ^2 value for the best-fit WMAP-5.

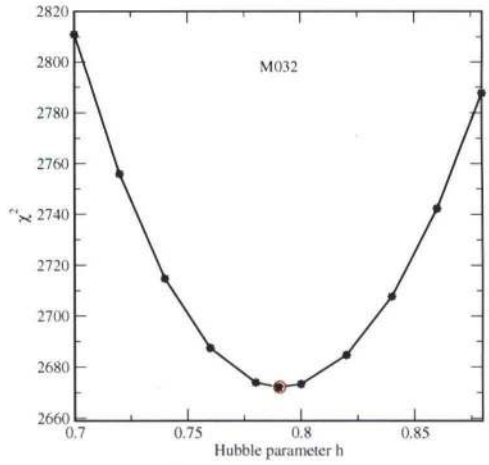

FIG. 6.— Sweep through h for model 32. The red circle marks the estimate for the Hubble parameter from assuming perfect knowledge of ℓ_A. It is in excellent agreement with the result we obtain from the WMAP-5 likelihood for the best-fit value of h for this model.

(1997). As CMB data improved the pivot point was shifted to smaller scales, closer to the middle of the range over which the spectrum is probed and where the normalization is best determined. In order to make closer connection with the initial fluctuations the amplitude not of the matter power spectrum but of the curvature or potential fluctuations has been adopted. These differ mostly by factors of growth and Ω_m. Anticipating future advances $k_p = 0.002\,\mathrm{Mpc}^{-1}$ was selected for the most recent CMB analysis by Komatsu et al. (2008) and the rms curvature fluctuation on this scale is now most commonly used as a normalization. With present CMB data the biggest uncertainties in the normalization are the near degeneracy with the optical depth, τ, and the uncertain growth of perturbations at low redshift due to the unknown equation of state of the dark energy, e.g. White (2006).

For our purposes however a normalization tied to the present day matter power spectrum and close to the non-linear scale has many advantages. Rather than introduce yet another convention, we therefore chose to use σ_8 as our normalization parameter. Of course, since all of the parameters of the models are specified one can compute any other parameter for our models. As an example, we have evaluated for each of the 38 models the best-fit value for τ using the likelihood code provided by the WMAP-5 team. The resulting TT power spectra are shown in Figure 4 as well as their ratios with respect to the best-fit WMAP-5 model. Some of our models are undernormalized and the resulting τ is smaller than 0.01 leading to reionization redshifts of $z < 2$. This undernormalization however is not a concern: we chose the 38 models to cover the parameter space overall well and not to provide fits close to the concordance cosmology. Figure 5 shows the best-fit value for τ for 16 models with $tau > 0.01$.

From the WMAP 5-year data, in combination with BAO, we have[4]

$$\begin{aligned} \omega_m &= 0.1347 \pm 0.0040 \quad (3\%) \\ \omega_b &= 0.0227 \pm 0.0006 \quad (3\%) \\ n_s &= 0.9610 \pm 0.0140 \quad (2\%) \end{aligned} \tag{7}$$

Current data restrict a constant equation of state for the dark energy to -1 with roughly 10% accuracy (for very recent results from supernovae see, e.g., Kowalski et al. 2008; for weak lensing see, e.g., Kilbinger et al. 2008; and for the latest constraints from clusters of galaxies, see Vikhlinin et al. 2008). Recent determinations put the normalization in the range $0.7-0.9$ with still rather large uncertainties (see, e.g. Vikhlinin et al. 2008for constraints from clusters, Voevodkin & Vikhlinin 2004 for a low estimate from clusters, Tegmark et al. 2007 for constraints from combined CMB and large scale structure data, and Evrard et al. 2008 for an extended discussion of recent results). Considering all these constraints and their uncertainties, we choose our grid to cover

$$\begin{aligned} 0.120 < \omega_m < 0.155 \\ 0.0215 < \omega_b < 0.0235 \\ 0.85 < n_s < 1.05 \\ -1.30 < w < -0.70 \\ 0.61 < \sigma_8 < 0.9 \end{aligned} \tag{8}$$

as shown in Table 1.

In order to solve for the full set of cosmological parameters we impose a constraint that $\ell_A \equiv \pi d_{\rm ls}/r_s = 302.4$, where $d_{\rm ls}$ is the distance to the last scattering surface and r_s is the sound horizon. Observationally this is known to 0.3%, and models which significantly violate this equality are poor fits to the CMB data (see Figure 6). Unfortunately the sound horizon, like the epoch of last scattering, can be defined in a number of different ways which differ subtly. Specifically we use the fit to the redshift of last scattering of Eq. (23) of Hu & White (1997) and use Eq. (B6) of Hu & Sugiyama (1995) for the sound horizon. With these choices we find models with ω_m and ω_b in the range preferred by WMAP do indeed provide good fits to the WMAP data. This is demonstrated for model 32 in Figure 6.

The procedure is then as follows. For every specified ω_m and ω_b we compute r_s and hence the required $d_{\rm ls}$ to fit the ℓ_A constraint. We adjust h, at fixed spatial curvature and ω_m, until the model reproduces the required $d_{\rm ls}$. Knowing h and ω_m then gives us Ω_m and hence $\Omega_{\rm de}$, as shown in Table 2.

Finally, we generated a model '0' which has parameters close to the current best fit from CMB and large-scale structure (e.g. Komatsu et al. (2008)). This model has $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, $\omega_b = 0.0224$, $n = 0.97$ and $\sigma_8 = 0.8$ and can be used as an independent check of the interpolation in the range of most interest.

[4] See http://lambda.gsfc.nasa.gov/

TABLE 1
PARAMETERS FOR 38 MODELS

#	ω_m	ω_b	n_s	$-w$	σ_8	#	ω_m	ω_b	n_s	$-w$	σ_8
0	0.1296	0.0224	0.9700	1.000	0.8000	19	0.1279	0.0232	0.8629	1.184	0.6159
1	0.1539	0.0231	0.9468	0.816	0.8161	20	0.1290	0.0220	1.0242	0.797	0.7972
2	0.1460	0.0227	0.8952	0.758	0.8548	21	0.1335	0.0221	1.0371	1.165	0.6563
3	0.1324	0.0235	0.9984	0.874	0.8484	22	0.1505	0.0225	1.0500	1.107	0.7678
4	0.1381	0.0227	0.9339	1.087	0.7000	23	0.1211	0.0220	0.9016	1.261	0.6664
5	0.1358	0.0216	0.9726	1.242	0.8226	24	0.1302	0.0226	0.9532	1.300	0.6644
6	0.1516	0.0229	0.9145	1.223	0.6705	25	0.1494	0.0217	1.0113	0.719	0.7398
7	0.1268	0.0223	0.9210	0.700	0.7474	26	0.1347	0.0232	0.9081	0.952	0.7995
8	0.1448	0.0223	0.9855	1.203	0.8090	27	0.1369	0.0224	0.8500	0.836	0.7111
9	0.1392	0.0234	0.9790	0.739	0.6692	28	0.1527	0.0222	0.8694	0.932	0.8068
10	0.1403	0.0218	0.8565	0.990	0.7556	29	0.1256	0.0228	1.0435	0.913	0.7087
11	0.1437	0.0234	0.8823	1.126	0.7276	30	0.1234	0.0230	0.8758	0.777	0.6739
12	0.1223	0.0225	1.0048	0.971	0.6271	31	0.1550	0.0219	0.9919	1.068	0.7041
13	0.1482	0.0221	0.9597	0.855	0.6508	32	0.1200	0.0229	0.9661	1.048	0.7556
14	0.1471	0.0233	1.0306	1.010	0.7075	33	0.1399	0.0225	1.0407	1.147	0.8645
15	0.1415	0.0230	1.0177	1.281	0.7692	34	0.1497	0.0227	0.9239	1.000	0.8734
16	0.1245	0.0218	0.9403	1.145	0.7437	35	0.1485	0.0221	0.9604	0.853	0.8822
17	0.1426	0.0215	0.9274	0.893	0.6865	36	0.1216	0.0233	0.9387	0.706	0.8911
18	0.1313	0.0216	0.8887	1.029	0.6440	37	0.1495	0.0228	1.0233	1.294	0.9000

Note. — The five basic parameters for the 38 models in our grid. See text for definitions.

2.1.3. *The Resulting Design*

Based on the above considerations, we are now able to generate a space-filling design for the five parameters of interest. We restrict ourselves to 37+1 cosmologies and we will show in the remainder of the paper that this number is indeed sufficient to generate an accurate emulator. The resulting cosmological models are listed in Table 1 where we give the values for the basic parameters. In Table 2 we give in addition a few derived parameters: Ω_m, $\Omega_{\rm de}$ (recall, that we assume flatness), h as derived from our constraint on ℓ_A, and $d_{\rm ls}$.

The two-dimensional projection of the design is shown in Figure 7. The upper part of the triangle shows the five input

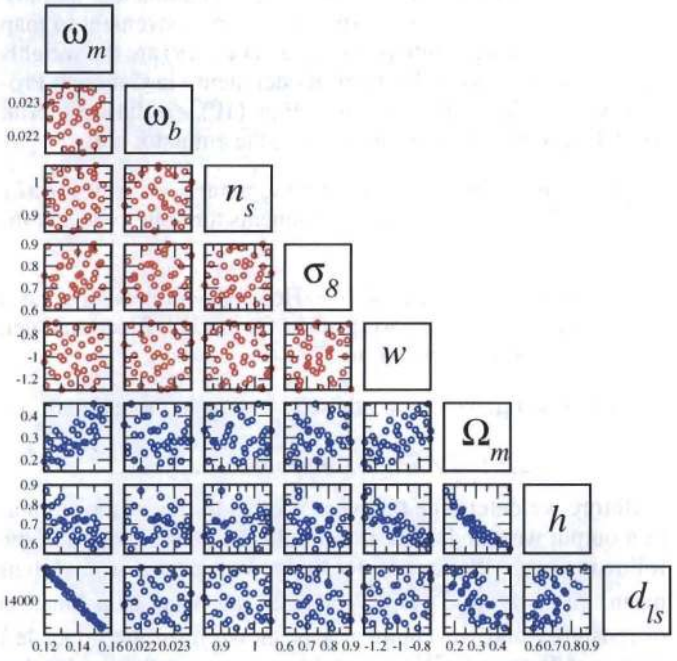

FIG. 7.— Resulting design for the five parameters under consideration in red, projected into two dimensions. The blue points show three derived parameters: Ω_m, h, and $d_{\rm ls}$.

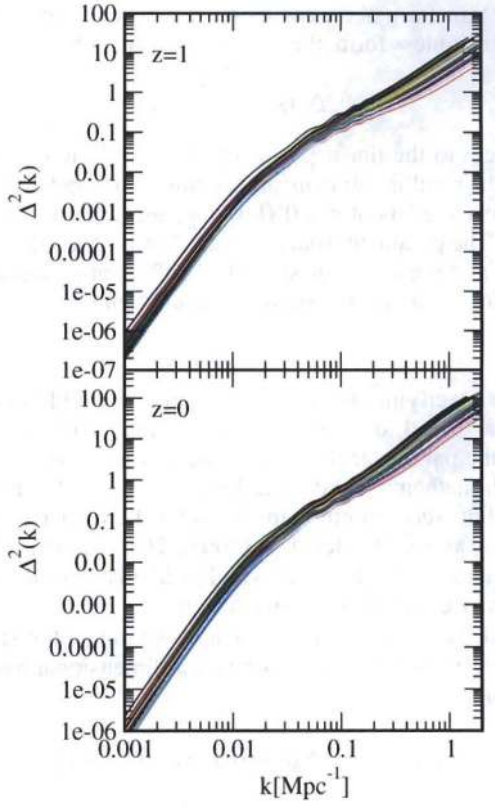

FIG. 8.— Dimensional power spectra for the 32 cosmologies specified in Table 1 at $z = 0$ (lower panel) and $z = 1$ (upper panel).

TABLE 2
DERIVED PARAMETERS FOR 38 MODELS

#	Ω_m	$\Omega_{\rm de}$	h	$d_{\rm ls}$	$\tau(\chi^2)$	#	Ω_m	$\Omega_{\rm de}$	h	$d_{\rm ls}$	$\tau(\chi^2)$
0	0.2500	0.7500	0.7200	10.25	0.1	19	0.1940	0.8060	0.8120	14.24	< 0.01 (7712)
1	0.4307	0.5693	0.5977	13.59	0.064	20	0.3109	0.6891	0.6442	14.27	0.15
2	0.4095	0.5905	0.5970	13.80	0.205	21	0.2312	0.7688	0.7601	14.14	< 0.01 (21579)
3	0.2895	0.7105	0.6763	14.10	0.19	22	0.3317	0.6683	0.6736	13.70	< 0.01 (11139)
4	0.2660	0.7340	0.7204	13.99	< 0.01 (5569)	23	0.1602	0.8398	0.8694	14.48	< 0.01 (4478)
5	0.2309	0.7691	0.7669	14.11	< 0.01 (2756)	24	0.1854	0.8146	0.8380	14.21	< 0.01 (13138)
6	0.3059	0.6941	0.7040	13.66	< 0.01 (19318)	25	0.4558	0.5442	0.5724	13.76	< 0.01 (3033)
7	0.3310	0.6690	0.6189	14.31	0.225	26	0.2804	0.7196	0.6931	14.05	0.14
8	0.2780	0.7220	0.7218	13.84	< 0.01 (4320)	27	0.3357	0.6643	0.6387	14.04	0.08
9	0.3707	0.6293	0.6127	13.93	< 0.015 (2845)	28	0.3988	0.6012	0.6189	13.66	0.06
10	0.3131	0.6869	0.6695	13.98	0.05	29	0.2516	0.7484	0.7067	14.32	< 0.01 (2809)
11	0.2790	0.7210	0.7177	13.82	< 0.01 (3928)	30	0.2810	0.7190	0.6626	14.37	0.155
12	0.2235	0.7765	0.7396	14.43	< 0.01 (5901)	31	0.3791	0.6209	0.6394	13.62	< 0.01 (17774)
13	0.3974	0.6026	0.6107	13.77	< 0.01 (11549)	32	0.1922	0.8078	0.7901	14.47	0.115
14	0.3289	0.6711	0.6688	13.74	< 0.01 (11803)	33	0.2634	0.7366	0.7286	13.96	<0.01 (2829)
15	0.2363	0.7637	0.7737	13.89	< 0.01 (9905)	34	0.3532	0.6468	0.6510	13.71	0.1
16	0.1981	0.8019	0.7929	14.40	0.025	35	0.3990	0.6010	0.6100	13.77	0.135
17	0.3586	0.6414	0.6305	13.94	< 0.01 (5012)	36	0.2949	0.7051	0.6421	14.41	0.455
18	0.2578	0.7422	0.7136	14.22	< 0.01 (5641)	37	0.2796	0.7204	0.7313	13.71	< 0.01 (2971)

Note. — The derived parameters, obtained from the basic parameters listed in Table 1 by applying the constraint on ℓ_A. The distance to last scattering is in Gpc, all other parameters are dimensionless. See text for details.

parameters in red, demonstrating a good sampling of the parameter space. The blue points show projections of three of the derived parameters, Ω_m, h, and $d_{\rm ls}$.

The key statistical observable discussed in this paper is the density fluctuation power spectrum $P(k)$, the (ensemble-averaged) Fourier transform of the two-point-density correlation function. In dimensionless form, the power spectrum can be written as

$$\Delta^2(k) \equiv \frac{k^3 P(k)}{2\pi^2}, \tag{9}$$

equivalent to the linear power spectrum in Equ. (5). Figure 8 shows the resulting dimensionless power spectra for the 38 cosmological models at $z = 0$ (lower panel) and at $z = 1$ (upper panel). The parameter space is overall well covered by these 38 models and the coverage should be sufficient to accommodate upcoming weak lensing survey measurements.

2.2. *Emulation*

After specifying the design, we can now build the emulator which will lead to predictions for the matter power spectrum within the parameter priors we specified in the design. For an in-depth mathematical description in the cosmological context of building such an emulator we refer the reader to Habib et al. (2007) and Schneider et al. (2008). Here we will explain the major ideas behind the process and explicitly show the emulator performance for our 37 model design.

In order to construct the emulator, we model the simulation output from the 37 models using a p_η-dimensional basis representation:

$$\eta(k;\theta) = \sum_{i=1}^{p_\eta} \phi_i(k) w_i(\theta) + \epsilon, \quad \theta \in [0,1]^{p_\theta}. \tag{10}$$

Here, $\eta(k,\theta)$ represents the power spectrum, which depends on the wavenmuber k and five cosmological parameters denoted by θ. It turns out to be more convenient to model the power spectrum as

$$\eta(k;\theta) = \frac{\Delta^2(k)}{2\pi k^{3/2}}, \tag{11}$$

which enhances the wiggles in the power spectrum due to the baryons. The dimensionality p_η refers the number of orthogonal basis vectors $\{\phi_1(k), ..., \phi_{p_\eta(k)}\}$. As we will show later, in our case $p_\eta = 6$. The parameter p_θ is the dimensionality of our parameter space – with 5 cosmological parameters we have $p_\theta = 5$. As we mentioned earlier it is more convenient to map the parameter ranges into [0,1] space. The $w_i(\theta)$ are the weights of the basis vectors and we will model them via Gaussian Process (GP) models. The last term in Eqn. (10), ϵ is the error term. Therefore, our main tasks in building the emulator are:

- Construct a suitable set of orthogonal basis vectors $\phi_i(k)$. In our case, principal components turn out to be an efficient representation.
- Model the weights $w_i(\theta)$. Here, GP models have been proven to be very well suited for smooth functions such as the power spectrum.

In the following, we discuss these two steps in more detail.

2.2.1. *Principal Component Analysis*

Before we determine the basis vectors to model the simulation output we standardize the simulation power spectra in the following way. We first center the power spectra around their mean, given by $\frac{1}{m}\sum_{j=1}^{37} \eta_j$. The resulting mean as a function of redshift z and wavenumber k is shown in in the upper left corner of Figure 9. We would like to remind the reader that we divide $\Delta^2(k)$ by $2\pi k^{3/2}$ which leads to the flattening of the power spectrum at high k. It also enhances the baryonic wiggles

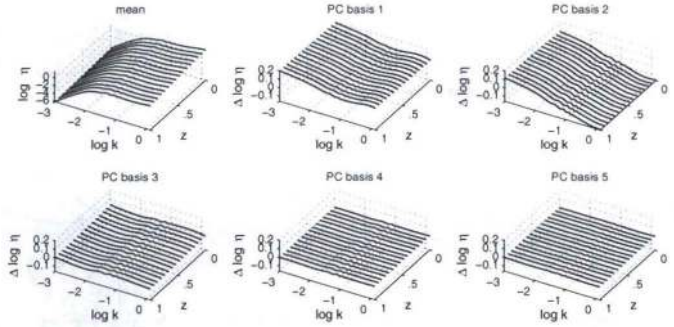

FIG. 9.— Mean (upper left corner) and five principal component (PC) bases derived from the output from the 38 HALOFIT power spectra. The third axis shows the time evolution of the mean and the five principal components between redshift $z = 1$ and $z = 0$. While the first two principal component bases show a lot of variation, the fourth and fifth are already almost flat, indicating that the inclusion of even higher principal component bases would not improve the quality of our emulator and the underlying GP model anymore.

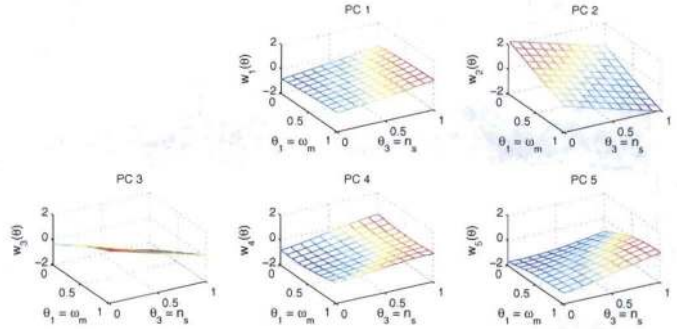

FIG. 10.— Principal component weights w_1 to w_5 as a function of two cosmological parameters, ω_m and n_s. The cosmological parameters are displayed in the normalized [0..1] space.

as can be seen in Figure 9. After having centered the simulations around the mean, we scale the output by a single value so that its variance is 1.

The next step is the principal component analysis (PCA) which will lead to the orthogonal basis vectors $\phi_i(k)$ for modeling the simulation output following Eqn. (10). For this step, we write the standardized power spectra in an $n_\eta \times m$ matrix, where n_η denotes the number of wave vectors for each power spectrum and $m = 37$ is the number of simulations. The matrix then reads:

$$y_{\rm sims} = [\eta_1; ...; \eta_{37}]. \quad (12)$$

Following Habib et al. (2007), we now apply a singular value decomposition to the simulation output matrix $y_{\rm sims}$ giving

$$y_{\rm sims} = UDV^T, \quad (13)$$

where U is an $n\eta \times m$ orthogonal matrix, D is a diagonal $m \times m$ matrix holding the singular values, and V is an $m \times m$ orthonormal matrix. The PC basis matrix Φ_η is now defined to be the first p_η columns of $[UD/\sqrt{m}]$ and the principal component weights are given by $[\sqrt{m}V]$.

In order to model the nonlinear matter power spectrum, we find that five principal components are sufficient to capture all information. Therefore we have $p_\eta = 5$ and $\Phi_eta = [\phi_1; \phi_2; \phi_3; \phi_4; \phi_5]$. The resulting five PC bases are shown in Figure 9 as a function of z and k. The fourth and the fifth component are already very flat which leads us to the above conclusion that 5 PCs are sufficient for our analysis.

2.2.2. *Gaussian Process Modeling*

In the next step we have to model the PC weights $w_i(\theta)$ in Eqn. (10). Here we will use a Gaussian Process (GP) model approach. The GP model is a nonparametric regression model which is particular well suited for interpolation of smooth functions of parameter space. A major feature of the GP is that it relies on local interpolation, hence the smoothness is important and an good sampling scheme which evenly covers the parameter space under investigation. The GP (also called Gaussian random functions) is simply a generalization of the Gaussian probability distribution. While the Gaussian distribution describes random variables which are scalars or vectors, the Gaussian process is the generalization to functions. (For an excellent introduction to Gaussian processes, see, Rasmussen & Williams 2006.) While the Gaussian distribution is specified simply by a scalar mean μ or a mean vector and a covariance matrix, the GP is specified by a mean function and a covariance function. In our case, each PC weight $w_i(\theta), i = 1, ..., 5$ is modeled as a mean-zero GP

$$w_i(\theta) \sim GP(0, \lambda_{wi}^{-1} R(\theta, \theta'; \rho_{wi})). \quad (14)$$

(The symbol $\sim$ here means distributed according to) Here λ_{wi} is the marginal precision of the process and the covariance function is given by:

$$R(\theta, \theta'; \rho_{wi}) = \prod_{l=1}^{p_\theta} \rho_{wil}^{4(\theta_l - \theta'_l)^2}. \quad (15)$$

The parameter ρ_{wil} controls the spatial range for the lth input dimension of the process $w_i(\theta)$. Under this parametrization, ρ_{wil} gives the correlation between $w_i(\theta)$ and $w_i(\theta')$ when the input conditions θ and θ' are identical, except for a difference in 0.5 in the lth component.

2.2.3. *Emulator Performance*

In order to evaluate the accuracy of the emulator we generate a second set of power spectra with HALOFIT within the prior parameter ranges. For this set we choose the input cosmologies randomly, still insuring that the constraint on the Hubble parameter is satisfied. We then predict the results for those cosmologies with the emulator scheme and compare them to the HALOFIT output, the "truth". The results for the residuals are

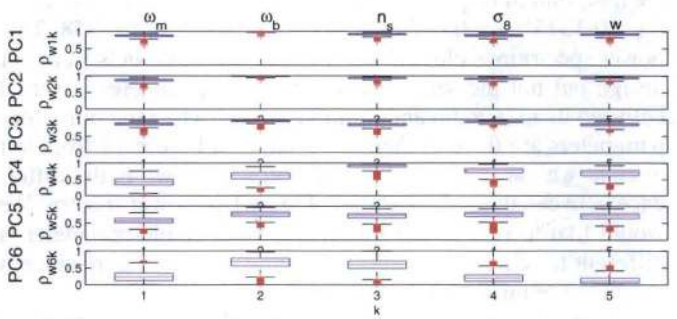

FIG. 11.— Boxplots of posterior samples for each ρ_{wil} for the nonlinear matter power spectrum. Boxplots offer a convenient way of showing the distribution of data by just five numbers. The blue box itself contains 50% of the data, the lower edge indicates the 25th percentile and the upper edge the 75th percentile of the data set. The red (center) line denotes the median. If the red line is not at the center of the box, the data is skewed. The black lines (or whiskers) extend out to the full range of the data. With our parametrization, a box value close to 1 indicates that the parameter is inactive, i.e., the PC is not changing much under the variation of that parameter.

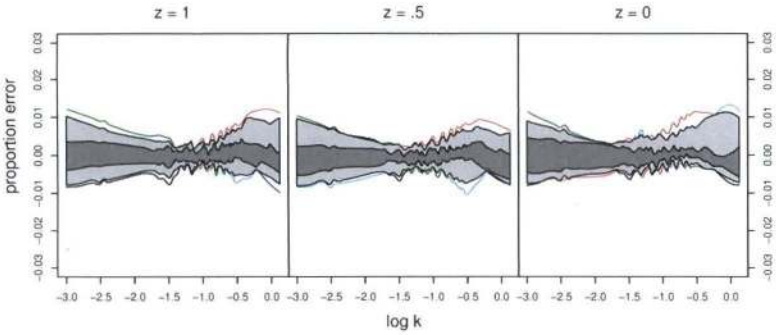

FIG. 12.— Emulator performance at three redshifts, $z = 1$, 0.5 and 0 (left to right). The emulator is tested on 10 additional HALOFIT runs within the parameter priors. Shown is the error of the emulator compared to the HALOFIT results. The central gray region contains the middle 50% of the residuals, the wider light gray region, the middle 90%. The errors are sub-percent level.

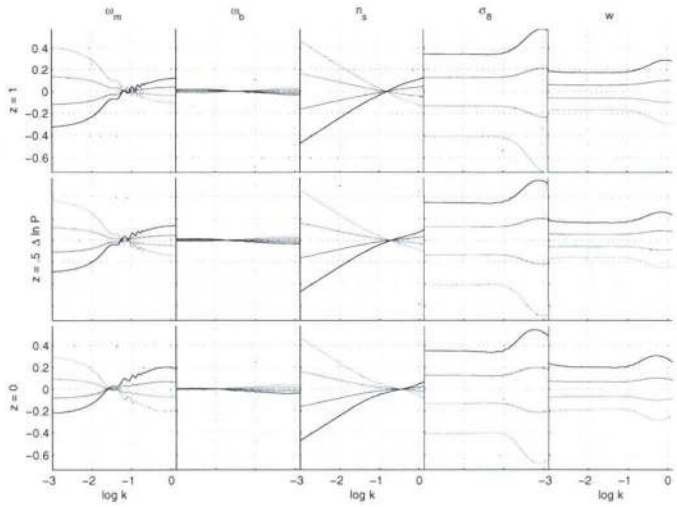

FIG. 13.— Sensitivity analysis for each of the five parameters at redshift $z = 0$, 0.5, and 1 (bottom to top). The y-axis shows the deviation of the log of the power spectrum from the nominal spectrum where each parameter is set at its midpoint. The light to dark lines correspond to the smallest parameter setting to the biggest. This plots shows that, e.g. ω_b is not influencing the power spectrum much and therefore will not be very well constraint from matter power spectrum observations alone.

shown in Figure 12 for three redshifts, $z = 0$, 0.5, and 1. The middle 50% of the residuals (dark gray band) are accurate to 0.5% or better over the whole k-range and for all three redshifts. All predictions have errors less than 1%. This result shows that a simulation set with an as small number as 37 cosmologies is sufficient to produce a power spectrum emulator accurate at 1%.

In a previous paper (Habib et al. 2007) several more convergence tests were shown. These tests demonstrated that the emulator performance improves considerably (by an order of magnitude) if either the number of simulation training runs is increased or the parameter space under consideration is decreased. In the present paper we follow the second strategy and restrict the priors as much as possible to obtain accurate results.

2.3. *Sensitivity Analysis*

After the emulator has been built it can be used to explore the behavior of the power spectrum within the parameter priors in more detail, the so-called sensitivity analysis can be carried out. Sensitivity is used here with respect to the changes of the power spectrum under the variation of the cosmological input parameters.

An example of this is shown in Figure 13. We show the variations of the power spectrum at three redshifts $z = 0$, $z = 0.5$, and $z = 1$ (bottom to top). The reference power spectrum is the power spectrum which we obtain if we fix every parameter at the midpoint of its prior range, so in this case for the cosmology $\omega_m = 0.1375$, $\omega_b = 0.02215$, $n_s = 0.95$, $w = -1$, $\sigma_8 = 0.758$. This power spectrum is close to the mean of the 37 models from our design but not the same. Next, only one parameter is varied between its maximum and minimum value while the other four parameters are fixed at their midpoints. In Figure 13 from left to right we vary: ω_m, ω_b, n_s, σ_8, and w. Shown is the difference between natural logarithm of these two power spectra. We would like to remind the reader that the Hubble parameter is different for each power spectrum shown in this figure since it is newly optimized for each cosmology.

The results contain information about the scales at which the power spectrum is most sensitive to each parameter and about parameter degeneracies. For example, it is clear that ω_b is not changing the power spectrum much at any scale or any redshift and therefore will not be well constrained from power spectrum measurements alone. We will show this point in more detail in Section 2.4. In the quasi-linear to non-linear regime at $k \sim 0.1 - 1\ h\mathrm{Mpc}^{-1}$, the power spectrum holds a lot of information about σ_8 and w, but degeneracies become an issue. Very large scales are particularly sensitive to the spectral index and ω_m. In the statistics literature, parameters that change the quantity of interest a lot are often called active parameters.

The sensitivity analysis also allows the targeted augmentation of simulation designs. If the accuracy of the emulator is not sufficient for the problem of interest, one would like to improve it by adding additional simulations. These simulations would then involve variations of the most active parameters and keep the other parameters more or less fixed. The augmentation of existing designs is an active field of research in statistics.

2.4. *Calibration*

As the last step we perform a complete analysis of a synthetic nonlinear power spectrum and determine the underlying cosmology. Since the emulator is a stand-alone routine, it is easily coupled to any standard Markov Chain Monte Carlo (MCMC) code such as CosmoMC (Lewis & Bridle 2002). Another route, which we choose here, is to extend our framework to incorporate not only the simulation outputs but also the real data itself.

For our example, we assume the ability to measure the power spectrum from observations at 1% accuracy in the quasi-linear and nonlinear regime. On large scales, the statistical error increases up to 10% due to sample variance. These error magnitudes are only rough estimates but they are sufficient to demonstrate the accuracy of our framework. The synthetic power spectrum is generated from a HALOFIT output at $z = 0$ (The same synthetic data set was used in Heitmann et al. (2008)). We pick 34 points from the power spectrum, spaced according to what future surveys will provide. We then move each of these points off the base power spectrum according to a Gaussian distribution with variance specified by the error estimate outlined above. The resulting mock data points and the underlying power spectrum are shown in Figure 14.

3. CONCLUSIONS

Over the last three decades we have entered a new era in cosmology. From order of magnitude estimates, we have now mea-

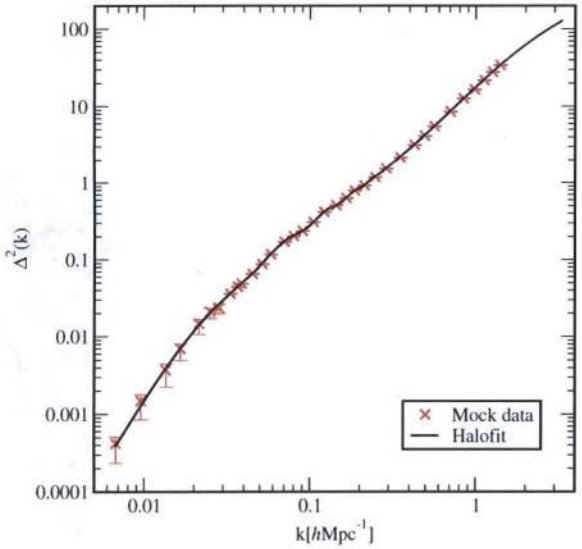

FIG. 14.— Synthetic power spectrum from a HALOFIT run. We choose 34 data points and assign error bars to them. For the smallest scales we assume 1% accuracy while on larger scales we increased the error bars to take cosmic variance into account.

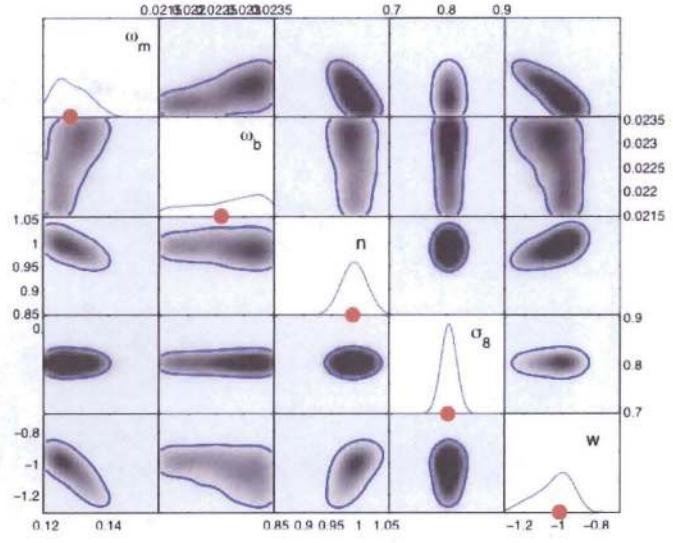

FIG. 15.— Likelihoods for the five input parameters. The red dots mark the true values. The constraints on ω_b are rather weak, but that is expected since the nonlinear power spectrum has not much constraining power on the baryon fraction. All other parameters are constraint rather well. This example demonstrates that with only 37 power spectra a reliable MCMC framework can be build.

surements at 10% accuracy or better, which have revealed one of the biggest mysteries in physics today: a dark energy leading to the acceleration of the expansion of the Universe. In order to understand the origin, nature, and dynamics of this dark energy – or to prove that the acceleration is due to a modification of gravity on the largest scales – we have to go the next step and obtain measurements at the 1% accuracy level. At the same time, our theoretical predictions have to be at least as accurate. Three major probes of dark energy – baryon acoustic oscillations, weak lensing, and clusters – are based on measurements of the large scale structure in the Universe. In order to obtain precise predictions for these probes, expensive, nonlinear simulations have to be carried out and we have to find ways to extract the needed information from a limited number of such simulations.

In this paper, we demonstrated, that if we have very accurate simulations, we in fact can produce such predictions schemes from just tenth of high-accuracy simulations. We have concentrated in this paper on the nonlinear matter power spectrum, but our scheme will work as well for, e.g., the cluster mass function. We have introduced a set of 38 cosmologies, the Coyote Universe named after the computer cluster it was carried out on, and showed that from these simulations we can generate an emulator for the nonlinear power spectrum which is accurate at the 1% level or better. In order to get such high accuracy from a small number of simulation inputs we have made use of (i) an interpolation method based on a sophisticated simulation design method and GP modeling which has been developed and refined in the statistics community over the last decade to address problems of the nature described here; (ii) the excellent parameter constraints we have from CMB measurements, which allow us to base our emulator on relatively narrow parameter priors and therefore ease the interpolation task.

This paper is the second in a series of three papers with the final goal to provide a high-precision emulator for the nonlinear power spectrum out to $k \sim 1\ h\mathrm{Mpc}^{-1}$. In the first paper Heitmann et al. (2008) we have demonstrated that we can achieve the required accuracy to predict the power spectrum from N-body simulations. The current paper introduces the cosmologies underlying the Coyote Universe and demonstrates that our prediction scheme can achieve 1% accuracy from only 37 cosmological models. The third paper and final paper will present results from the simulation suite discussed in this paper and will include a power spectrum emulator that will be publicly released.

A special acknowledgment is due to supercomputing time awarded to us under the LANL Institutional Computing Initiative. Part of this research was supported by the DOE under contract W-7405-ENG-36. SH, KH, DH, and CW acknowledge support from the LDRD program at Los Alamos National Laboratory. MJW was supported in part by NASA and the DOE. We would like to thank Dragan Huterer, Nikhil Padmanabhan, and Michael Schneider for useful discussions.

REFERENCES

Auld, T., Bridges, M., Hobson, M.P., & Gull, S.F. 2007, MNRAS, 376, L11
Auld, T., Bridges, M., & Hobson, M.P. 2008, MNRAS, 387, 1575
Brandbyge, J., Hannestad, S., Haugboelle, T. & Thomsen, B. 2008, arxiv:0802.3700
Bunn, E. & White, M. 1997, ApJ, 480, 6
Carlson J.W., et al., 2009, preprint
Crocce, M. & Scoccimarro, R. 2006, Phys. Rev. D, 73, 063520
Currin C., Mitchell T., Morris, M., & D. Ylvisaker 1991, J. Am. Stat. Assn. 86, 953
Evrard, A.E. et al. 2008, ApJ, 672, 122
Fendt, W. & Wandelt, B. 2007, ApJ, 654, 2
Fendt, W. & Wandelt, B. 2007, arXiv:0712.0194
Francis, M., Lewis, G., & Linder, E. 2007, MNRAS, 380, 1079
Habib, S., Heitmann, K., Higdon, D., Nakhleh, C., & Williams, B. 2007, Phys. Rev. D, 76, 083503
Hamilton, A.J.S., Kumar, P., Lu, E., Matthews, A. 1991, ApJ, 374, L1
Heitmann, K., Higdon, D., Nakhleh, C., & Habib, S. 2006, ApJ 646, L1
Heitmann, K., White, M., Wagner, C., Habib, S., & Higdon, D., arXiv:0812.1052
Hu, W. & White, M. 1997, ApJ, 479, 568
Hu, W. & Sugiyama, N. 1995, ApJ, 444, 489
Huterer, D. & Takada, M. 2003, Astrop. Physics, 23, 369
Jimenez, R., Verde ,L. Peiris, H., & Kosowsky, A. 2004, Phys. Rev. D, 70, 23005

Jing, Y.P., Zhang, P., Lin W.P., Gao L., & Springel, V. 2006, ApJ 640, L119
Kaplinghat, M., Knox, L., & Skordis, C. 2002, 578, 665
Kilbinger, M. et al. 2008, arXiv:0810.5129
Komatsu, E. et al. 2008, arXiv:0803.0547
Kowalski, M. et al, 2008, ApJ, 686, 749
Leary, S., Bhaskar, A., & Keane, A. 2003, Journal of Appl. Stat., 5, 585
Lewis, A. & Bridle, S. 2002, Phys. Rev. D, 66, 103511
Li, W. & Ye, K.Q. 2000, Journal of Statistical Planning and Inference, 90, 145
Linder, E. & White M. 2005, Phys. Rev. D72, 061304
Matsubara, T. 2008, Phys. Rev. D, 77, 063530
McKay, M., Beckman, R., & Conover, W. 1979, Technometrics, 21, 239
Morris, M.D. & Mitchell, T.J. 1995, Journal of Stat. Planning and Interference, 43, 381
Peacock, J.A. & Dodds, S.J. 1994, MNRAS, 267, 1020
Peacock, J.A. & Dodds, S.J. 1996, MNRAS, 280, L19
Perlmutter, P. et al. 1999, ApJ, 517, 565
Pietroni, M. 2008, arXiv:0806.0971
Rasumussen, C.E. & Williams, K.I. 2006, Gaussian Processes for Machine Learning, MIT Press, www.GaussianProcess.org/gpml
Riess A.G et al. 1998, AJ, 116, 1009
Rudd, D., Zentner, A., & Kravtsov, A. 2008, ApJ, 672, 19
Sacks, J., Welch, W.J., & Mitchell, T.J. 1989, Stat. Sci. 4, 409
Sacks, J., Schiller, S.B., & Welch, W.J. 1989, Technometrics, 34, 15
Santner, T.J., Williams, B.J., & Notz, W.I. 2003, The Design and Analysis of Computer Experiments, Springer, New York
Shewry, M. & Wynn, H. 1987, J. of Appl. Stat., 14, 2, 165
Schneider, M., Knox, L., Habib, S., Heitmann, K., Higdon, D., & Nakhleh, C., Phys. Rev. D, 78, 063529 (2008)
Smith, R.E., Peacock, J.A., Jenkins, A., White, S.D.M., Frenk, C.S., Pearce, F.R., Thomas, P.A., Efstathiou, G., & Couchman, H.M.P. 2003, MNRAS, 342, 1311
Tang, B. 1993, Am. Stat. Assoc., 88, 1392
Tegmark, M. & Zaldarriaga M. 2000, ApJ 544, 30
Tegmark, M. et al. 2006, Phys. Rev. D, 74, 123507
Vikhlinin, A. et al. 2008, "Chandra Cluster Cosmology Project II: Samples and X-Ray Data Reduction", ApJ in press.
Voevoedkin, A. & Vikhlinin, A. 2004, ApJ, 601, 610
White M. & Vale C. 2004, Astroparticle Physics, 22, 19
White, M. 2004, Astropart. Phys., 22, 211
White M. 2006, New Astronomy Reviews, 50
Ye, K.Q 1998, Journal of the Am. Stat. Assoc., 93, 1430
Zhan, H. & Knox, L. 2004, Astrophys. J., 616, L75