THE ASTROPHYSICAL JOURNAL, 552:15–35, 2001 May 1

ON ESTIMATING THE QSO TRANSMISSION POWER SPECTRUM

LAM HUI,[1] SCOTT BURLES,[2] UROŠ SELJAK,[3] ROBERT E. RUTLEDGE,[4] EUGENE MAGNIER,[5] AND DAVID TYTLER[6]

Received 2000 May 3; accepted 2001 January 12

ABSTRACT

The Lyα forest has become an important tool for measuring the mass power spectrum at high redshifts ($z = 2$–4). A crucial intermediate step is the measurement of the transmission power spectrum. We present new methods to minimize the systematic and random errors for such a measurement and discuss their implications for observing strategies. Sources of systematic errors explored include metal line contamination and continuum fitting. We advocate the technique of trend removal in place of traditional continuum fitting: here a spectrum is normalized by its (smoothly varying) mean rather than its continuum; this method is easily automated and removes biases introduced by continuum fitting. Moreover, trend removal can be easily applied to spectra where continuum fitting is difficult, such as when the resolution or signal-to-noise ratio (S/N) is low, or for spectra at high redshifts. We further show that a measurement of the continuum power spectrum (plus a related quantity) using trend removal, from either low-redshift quasar spectra or the red side of Lyα, can be used to constrain the amount of spurious large-scale power introduced by the uncertain continuum and in principle allows the removal of such contamination and thereby expanding scales probed to larger ones. We also derive expressions for the shot noise bias and variance of the power spectrum estimate, taking into account the non-Poissonian nature of the shot noise and the non-Gaussianity of the cosmic fluctuations. An appropriate minimum variance weighting of the data is given. Finally, we give practical suggestions on observing strategy: the desired resolution and S/N for different purposes and instruments, as well as how to distribute one's finite observing time among quasar targets. Also discussed is the quasar spectroscopic study of the Sloan Digital Sky Survey (SDSS), which has the potential to measure the power spectrum accurate to better than 1% per mode ($\Delta k \sim 10^{-4}$ s km^{-1}). The techniques presented here will be useful for tackling the anticipated issues of shot noise and continuum contamination.

Subject headings: cosmology: observations — intergalactic medium — large-scale structure of universe — methods: data analysis — quasars: absorption lines

1. INTRODUCTION

Recent theoretical research on the low column density ($N_{\rm HI} \lesssim 10^{14.5}$ cm^{-2}) Lyα forest at redshifts $z \sim 2$–4 points toward a picture in which the forest consists largely of mildly nonlinear fluctuations of a smooth intergalactic medium (e.g., Bi, Boerner, & Chu 1992; Cen et al. 1994; Zhang, Anninos, & Norman 1995; Reisenegger & Miralda-Escudé 1995; Hernquist et al. 1996; Miralda-Escudé et al. 1996; Muecket et al. 1996; Bi & Davidsen 1997; Bond & Wadsley 1997; Croft et al. 1997; Hui, Gnedin, & Zhang 1997; Hui & Gnedin 1997; see Rauch 1998 for a review and further references). This provides the motivation to analyze the quasar (QSO) absorption spectrum as a continuous field with fluctuations, rather than as a collection of discrete absorption lines. The two-point correlation or its Fourier transform, the power spectrum, comes to mind as a useful and common statistic used in other areas such as the microwave background or galaxy large-scale structure. Indeed, its application to QSO spectra has been discussed by a number of authors (Zuo & Bond 1994; Miralda-Escudé et al. 1996; Bi & Davidsen 1997; Cen et al. 1998). Recently, Croft et al. (1998, 1999; see also Hui 1999; McDonald et al. 2000) have shown that the mass power spectrum can be recovered from the QSO transmission power spectrum, from which one could further deduce cosmological parameters such as Ω_m (Weinberg et al. 1999; Phillips et al. 2000). There exist at present a large number of high-quality QSO spectra (e.g., Hu et al. 1995; Lu et al. 1996; Kirkman & Tytler 1997; Cristiani et al. 1997; Kim et al. 1997; Rauch et al. 1997), which makes this an exciting field of research. Upcoming quasar surveys such as the Sloan Digital Sky Survey (SDSS) and the Anglo-Australian Telescope Two Degree Field (AAT2DF) will enlarge the database significantly.

Here we take the view that the QSO transmission power spectrum/correlation is interesting in its own right and focus on how to best measure it from the observed QSO spectra, independent of the underlying physical picture of the forest. The two major questions are (1) what the main sources of systematic errors are and what the best ways to bring them under control are, and (2) how to estimate the shot noise and to best combine data with different signal-to-noise ratios (S/Ns) (random errors).

Let us start by defining the transmission power spectrum and correlation function. Two possibilities arise. One of them we call the unnormalized power spectrum $P_{\rm un}$/two-

[1] Institute for Advanced Study, Olden Lane, Princeton, NJ 08540; and NASA/Fermilab Astrophysics Center, Fermi National Accelerator Laboratory, Batavia, IL 60510; lhui@ias.edu.
[2] Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637.
[3] Physics Department, Princeton University, P.O. Box 708, Princeton, NJ 08544.
[4] Department of Astronomy, California Institute of Technology, Pasedena, CA 91125.
[5] Canada-France-Hawaii Telescope, 65-1238, Mamalahoa Highway, Kamuela, HI 96743.
[6] Center for Astrophysics and Space Sciences, University of California, San Diego, MS 0424, La Jolla, CA 92093-0424.

point correlation ξ_{un} (D. H. Weinberg 1998, private communication; McDonald et al. 2000):

$$\xi_{un}(u) = \langle f(u')f(u'+u)\rangle\ , \quad P_{un}(k) = \int \xi_{un}(u)e^{-iku}\,du\ , \quad (1)$$

where f is the transmission defined by $f = e^{-\tau}$ with τ being the optical depth (the absorption is then $1-f$), u or u' is the observed velocity (or redshift or wavelength) along a line of sight, and k is its Fourier counterpart. The angular brackets denote ensemble averaging.

The other we call the normalized power spectrum P/two-point correlation ξ (e.g., Zuo & Bond 1994):

$$\xi(u) = \langle \delta_f(u')\delta_f(u'+u)\rangle\ , \quad P(k) = \int \xi(u)e^{-iku}\,du\ , \quad (2)$$

where $\delta_f = (f-\bar{f})/\bar{f}$, with $\bar{f}$ being the mean transmission $\langle f\rangle$.

We will almost exclusively focus on the latter but will discuss at some point the pros and cons of the two, especially with regard to systematic errors. Unless otherwise stated, hereafter "power spectrum/correlation" refers to the normalized version.

The layout of the paper is as follows. In § 3 we provide a brief overview of how the raw data output (a two-dimensional CCD image) is reduced to a one-dimensional QSO spectrum. Note that the quantity f above is never observed directly. It is important to have a description of how the whole data reduction procedure works, which is sometimes hard to find in the literature. We give some illustrations by showing simulated spectra with various realistic levels of noise.

In § 4 we discuss the estimation of the power spectrum and two-point correlation, beginning with the introduction of the quadratic estimator in § 4.1. An important point pertaining to the estimation of the two-point correlation is raised here: most estimators employed in the literature are not optimal; an alternative is given here that is an analog of the Landy-Szalay estimator (Landy & Szalay 1993) introduced originally for galaxy surveys. Aside from this point, we focus exclusively on the estimation of the power spectrum. In § 4.2 we discuss three sources of systematic errors: continuum fitting, gaps, and metal absorption lines. Particular attention is paid to issues related to continuum fitting. We advocate in § 4.2.1 trend removal to replace traditional continuum fitting, which avoids the latter's pitfalls. We further propose in § 4.2.2 that the power spectrum of the continuum can be estimated using trend removal as well, which offers us a way to measure accurately the transmission power spectrum on large scales where the continuum fluctuations might be important. In § 4.2.3 we discuss the effects of gaps and unremoved metal lines. We then turn our attention in § 4.3 to random errors. We emphasize here that the shot noise is not exactly Poisson distributed because of the particular way the data are reduced. We point out the importance of subtracting the shot noise bias correctly, describe a systematic way of assigning error bars to the power spectrum, and introduce minimum variance weighting techniques to combine data of different qualities. Some results here are stated without justification. The aim in this section is to summarize useful results for readers who might not be interested in details of the derivations, which are provided in the appendices. The techniques used in the appendices should be of broad interest, e.g., the issue of generally non-Poissonian shot noise might be relevant for galaxy power spectrum estimation.

Finally, we conclude in § 5. We summarize here our recipe for transmission power estimation; readers who would like a quick overview of our methods can skip directly to this section and only refer back to the relevant sections to fill in the details. We give general advice on observing strategies and discuss in particular analysis issues relevant for the SDSS.

Before we begin, let us first make a few clarifying remarks about some of our notation and terminology.

2. TERMINOLOGY AND NOTATION: AVERAGING AND AVERAGES

In this paper we refer to two different kinds of fluctuations that should be clearly distinguished. Take the observed photon (or electron) count from a quasar as an example, $\hat{N}_Q$. As one moves along a spectrum, the photon count fluctuates for two very different reasons. First, it fluctuates because the universe is intrinsically inhomogeneous, giving rise to nonuniform absorption; we will refer to these as cosmic fluctuations. Second, it fluctuates because the observed photon count is a discrete realization of the underlying cosmic signal (Poisson fluctuation is the canonical example but not the only possible one); we will refer to these as discrete fluctuations.

We define two different kinds of averages corresponding to these two different kinds of fluctuations. The discrete average of the observed photon count is denoted by $\langle \hat{N}_Q\rangle_D \equiv \tilde{N}_Q$. In other words, $\hat{N}_Q$ constitutes a discrete realization of the underlying cosmic signal $\tilde{N}_Q$. This signal $\tilde{N}_Q$ itself suffers from cosmic fluctuations, and we will denote its ensemble average by $\langle \tilde{N}_Q\rangle = \bar{N}_Q$. A fixed quasar continuum is assumed in this ensemble average, i.e., it is the fluctuation in the spectrum caused by intervening absorption that constitutes the cosmic signal we are after. Finally, we will sometimes use $\langle\rangle$ to implicitly stand for $\langle\langle\rangle_D\rangle$, e.g., $\hat{N}_Q$ actually means $\langle\langle \hat{N}_Q\rangle_D\rangle$, which is the same as $\langle \tilde{N}_Q\rangle = \bar{N}_Q$.[7]

To recap:

1. $\hat{N}_Q$ is the directly observed quasar photon count.
2. $\tilde{N}_Q \equiv \langle \hat{N}_Q\rangle_D$ refers to the idealized quasar photon count if one has infinite S/N.
3. $\bar{N}_Q \equiv \langle \tilde{N}_Q\rangle = \langle\langle \hat{N}_Q\rangle_D\rangle$ refers to the quasar photon count if one has infinite S/N and if one averages over all possible cosmic fluctuations keeping the continuum fixed, e.g., by taking the same quasar and putting it at all possible orientations in the sky. For instance, if $\tilde{N}_Q = N_C e^{-\tau}$, where N_C is the true continuum and τ is the optical depth, then $\bar{N}_Q = N_C\langle e^{-\tau}\rangle$, where $\langle e^{-\tau}\rangle$ is the mean transmission.

Note that, when we use the term "discrete," it is not implied that $\hat{N}_Q$ is necessarily an integer, although it is derived from some integral quantity such as the electron/photon count. We use the term "discrete fluctuations" instead of the usual Poisson fluctuations because, as we will see, $\hat{N}_Q$ is often not Poisson distributed, i.e., $\langle \hat{N}_Q^2\rangle_D - \langle \hat{N}_Q\rangle_D^2 \neq \tilde{N}_Q = \langle \hat{N}_Q\rangle_D$ (see § 3.1). The term "shot noise" is often used to describe Poissionian discrete fluctuations,

[7] Two exceptions in the use of $\langle\rangle$: in § 4.2.2 we use $\langle\rangle$ to include averaging over an ensemble of different continua, and in § 4.3 we use $\langle\rangle_{kk}$ to denote averaging over a shell of Fourier modes.

but in this paper we will use it more broadly to include non-Poissonian discrete fluctuations as well.

Finally, we note that we use the term "quasar counts" throughout this paper to refer to the photon counts from a quasar, rather than the number of quasars in a given patch of sky.

3. DATA REDUCTION: FROM THE CCD IMAGE TO THE QSO SPECTRUM

3.1. *A Brief Description*

We discuss briefly here aspects of the data processing necessary for understanding the noise properties of the reduced quasar spectrum. The reader is referred to Horne (1986), Zuo & Bond (1994), Barlow & Sargent (1997), Rauch et al. (1997), and Cen et al. (1998) for more discussions.

The raw data consist of a two-dimensional (spatial and spectral) array of counts (data numbers or photon counts converted from them) from a CCD image. The one-dimensional array of estimated quasar counts in the spectral direction (as a function of velocity, redshift, or wavelength) is obtained by collapsing the data in the spatial direction in the following fashion:

$$\hat{N}_Q^\alpha = \sum_{i,\beta} W^{\alpha\beta} W^{i\beta} (\hat{N}_{\rm RAW}^{i\beta} - \tilde{N}_B^{i\beta}) \,. \tag{3}$$

We have introduced and will stick with the following notations: the Latin index such as i and the Greek index such as α denote the spatial and spectral coordinates, respectively, of a CCD pixel (there are in fact a few exceptions, which should be clear from the context); $\hat{N}_Q^\alpha$ is our estimated quasar count, $\hat{N}_{\rm RAW}^{i\beta}$ is the raw count, and $\tilde{N}_B^{i\beta}$ is the mean background count that includes the sky and the readout offset; $W^{i\beta}$ is a weighting of the spatial pixels for each spectral coordinate β; and $W^{\alpha\beta}$ represents a rebinning of the spectral pixels that is sometimes done to achieve, for instance, a linear wavelength scale. Note that the α and i dimensions do not necessarily align with the two perpendicular axes of the CCD chip. The optical setup could be such that the spectrograph slit appears tilted at an angle to the CCD axes. We use the hat to emphasize the fact that the quantity of interest is a random variable with fluctuations. The overline tilda denotes a discrete average, e.g., $\tilde{N}_B^{i\beta} = \langle \hat{N}_B^{i\beta} \rangle_D$, where $\langle \rangle_D$ denotes discrete averaging.

Implicitly assumed in the above formulation is that the discrete average $\tilde{N}_B^{i\beta}$ is known, which is of course not strictly true, but since a typical slit covers a significant number of pixels that do not have any quasar photons, and since the background is often quite uniform, the discrete average can be estimated to high accuracy. Note also that the readout offset can be measured using short exposures with closed shutters or from the CCD overscan region.

The weighting $W^{i\beta}$ typically has nontrivial spectral dependence (β) to remove at least two artifacts: variations in detector efficiency across the chip and a nonflat blaze. The former is usually estimated in a procedure called flat-fielding by shining a lamp into the detection system. The latter arises because of the nontrivial shape of a diffraction order. This can be partially estimated in the flat-fielding procedure but is best done using a spectrophotometric standard star, usually a white dwarf. While the correction for the first artifact should be quite accurate, the blaze removal is often approximate. Any residual that is not correctly removed will show up in the form of a nontrivial effective continuum. We will see in § 4.2 perhaps some evidence of it. We assume in this paper that such artifacts show up as fluctuations on large scales (since the blaze itself is smoothly and slowly varying across a given order) but not on small scales (we will quantify the scales later on).

To make the above concrete, the raw count is given by

$$\hat{N}_{\rm RAW}^{i\beta} = \hat{N}_B^{i\beta} + \hat{N}_Q^{i\beta} \,, \tag{4}$$

where the quasar contribution has the following discrete average:

$$\tilde{N}_Q^{i\beta} \equiv \langle \hat{N}_Q^{i\beta} \rangle_D = g_{\rm ps}^{i\beta} g_b^\beta \tilde{N}_Q^\beta \,, \tag{5}$$

where $g_{\rm ps}^{i\beta}$ is the point-spread function, which describes how the light from the quasar gets spread out in the spatial direction i at a given spectral coordinate β, and g_b^β accounts for the variation of the blaze and quantum efficiency as a function of wavelength. The symbol $\tilde{N}_Q^\beta$ denotes the underlying quasar count (or cosmic signal, i.e., discrete averaged).

Many different rebinning kernels $W^{\alpha\beta}$ (eq. [3]) are possible. The simplest choice is of course no rebinning with $W_{\alpha\beta} = \delta_{\alpha\beta}$. There are several possible choices of the weighting $W^{i\beta}$ (eq. [3]), but any sensible choice has to satisfy the requirement that $\langle \hat{N}_Q^\alpha \rangle_D = \tilde{N}_Q^\alpha$, up to some constant normalization factor. This assumes that artificial fluctuations introduced by the blaze or detector efficiency are correctly taken out. If not, it shows up effectively as part of the continuum.

We give two examples of $W^{i\beta}$ here. The first is basically a uniform weighting over the spatial pixels that correspond to a given spectral coordinate:

$$W^{i\beta} = \frac{1}{g_b^\beta \sum_j g_{\rm ps}^{j\beta}} \,, \tag{6}$$

where the range of i, or j, is chosen to lie within, say, some fraction of the quasar seeing disk. There is sometimes an additional complication due to cosmic-ray hits, which will be discussed below.

The second is a minimum variance weighting (different from minimum variance weighting for measuring the power spectrum; § 4.3) over the spatial pixels, introduced by Horne (1986):

$$W^{i\beta} = \frac{(1/g_b^\beta)(g_{\rm ps}^{i\beta}/V_{\rm RAW}^{i\beta})}{\sum_j (g_{\rm ps}^{j\beta})^2 / V_{\rm RAW}^{j\beta}} \,, \tag{7}$$

where $V_{\rm RAW}^{j\beta}$ is the variance in the raw count,

$$V_{\rm RAW}^{j\beta} = \langle (\hat{N}_{\rm RAW}^{j\beta})^2 \rangle_D - \langle \hat{N}_{\rm RAW}^{j\beta} \rangle_D^2 = \tilde{N}_Q^{j\beta} + V_B^{j\beta} \,,$$
$$V_B^{j\beta} = \tilde{N}_S^{j\beta} + V_{\rm RO}^{j\beta} \,, \tag{8}$$

where $V_B^{j\beta}$, the background variance, has two contributions: the sky variance $\tilde{N}_S^{j\beta}$ and the readout variance $V_{\rm RO}^{j\beta}$. A word of caution is necessary here regarding the second weighting. The raw variance, $V_{\rm RAW}^{j\beta}$, depends on the underlying cosmic signal or quasar count ($\tilde{N}_Q^{j\beta}$, i.e., discrete averaged), which is not directly observable (the discrete-averaged sky count and the true readout variance are also strictly speaking not directly observable, but they can be estimated quite accurately because they are relatively uniform and can be observed over a larger number of pixels). Modeling $V_{\rm RAW}^{j\beta}$ using the measured raw count (i.e., using $\hat{N}_Q^{j\beta}$ instead of $\tilde{N}_Q^{j\beta}$ in eq. [8]) could lead to a biased estimation of $\tilde{N}_Q^\beta$. Horne (1986) suggested an iterative scheme to avoid this problem, but implementations of this weighting should be checked for a possible bias.

Pixels affected by cosmic-ray hits, which are usually easy to identify because of their wild fluctuations and spiky nature, are dealt with in two different ways, depending on the severity. For a given spectral coordinate, if only a small fraction of the corresponding spatial pixels are affected, the weighting in equation (6) or equation (7) is simply modified by allowing i and j to range only over the unaffected spatial pixels. However, if all or most corresponding spatial pixels are affected, then all recorded counts at that spectral coordinate are discarded, leaving a gap in the reduced quasar spectrum. Gaps could result also because of metal line removal (an alternative would be to fit for the metal line and subtract, instead of simply discarding the pixels) or defects in the CCD.

Finally, the (random) error array output at the end of the data reduction corresponds to an estimate of

$$\sqrt{\langle(\hat{N}_Q^\alpha - \tilde{N}_Q^\alpha)^2\rangle_D} = \sqrt{\sum_{i,\beta} (W^{\alpha\beta}W^{i\beta})^2 V_{\rm RAW}^{i\beta}}\,, \qquad (9)$$

where we have assumed that the noise fluctuations are independent among the pixels. We emphasize that in practice the error array is only an estimate of the above quantity because the true $V_{\rm RAW}^{i\beta}$ is unknown but is estimated using the observed raw counts (using $\hat{N}_Q^{j\beta}$ instead of $\tilde{N}_Q^{j\beta}$ in eq. [8]).

It is clear from the above discussion that in general fluctuations in $\hat{N}_Q^\alpha$ are non-Poissonian, in the sense that $\langle\hat{N}_Q^2\rangle_D - \langle\hat{N}_Q\rangle_D^2 \neq \langle\hat{N}_Q\rangle_D = \tilde{N}_Q$. There are several reasons for this. First, $\hat{N}_Q$ suffers from additional discrete fluctuations from the background counts. Second, the weighting $W^{\alpha\beta}$ and $W^{i\beta}$ are in general nontrivial (i.e., not unity). A very simple example of the effect of nonunit weights is the following: suppose we multiply a Poisson variable by a factor of 2 and call the result $\hat{y}$; it is easy to see that $\langle\hat{y}^2\rangle_D - \langle\hat{y}\rangle_D^2 = 2\langle\hat{y}\rangle_D \neq \langle\hat{y}\rangle_D$.

For the rest of this paper we will pick for simplicity the weighting kernels $W_{\alpha\beta} = \delta_{\alpha\beta}$ and $W_{i\beta}$ as given by equation (6). In the appendices we will indicate where some of our expressions have to be modified to account for more general weightings.

3.2. *Simulated QSO Spectra*

For illustrations, and for later analyses, we have generated several different QSO spectra. The underlying noiseless (theoretical) transmission ($f = e^{-\tau}$) is shown in the bottom panel of Figure 1, and its associated power spectrum is shown in the top panel of the same figure. This is drawn from a standard cold dark matter (SCDM) simulation discussed in Gnedin (1998), which made use of the Hydro-PM algorithm developed by Gnedin & Hui (1998). The cell size (comoving size of 10 h^{-1} kpc) is small enough to resolve the effective Jeans scale and so should retain all small-scale structures. However, the box size is unrealistically small (comoving size of 2.56 h^{-1} Mpc), which means that a significant amount of large-scale power is missing. For most of our investigations here, it is not necessary that the simulations are highly realistic, but our simulated transmission power spectrum is in fact broadly consistent with an observed one. The long line of sight in Figure 1 is generated by shooting a ray at some oblique angle through the simulation box and allowing it to wrap around the box several times, but never repeating itself. The mean redshift here is $z = 2.85$. The ionizing background is chosen to give $\langle e^{-\tau}\rangle = 0.64$ (Press, Rybicki, & Schneider 1993).

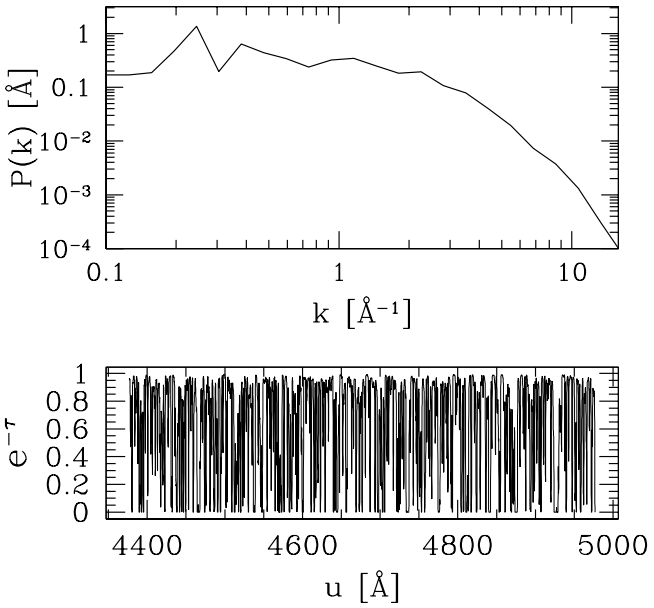

FIG. 1.—Lower panel shows the transmission $e^{-\tau}$ as a function of wavelength u taken from a SCDM simulation, with no noise added. The upper panel shows the corresponding (normalized) transmission power spectrum (eq. [2]). All subsequent simulated spectra in this paper are based on this one, with various levels of noise, contamination, etc., added.

An example of a somewhat realistic reduced QSO spectrum ($\hat{N}_Q^\alpha$ in eq. [3]) and its error array (eq. [9]) can be found in the bottom two panels of Figure 2 (ignore the other two panels for the moment). They are generated based on the prescriptions given in § 3.1, assuming that $W^{\alpha\beta} = \delta^{\alpha\beta}$ and $W^{i\beta}$ are given by equation (6). Briefly speaking, what we do is first to generate an array of g^α that represents $g_b^\alpha \sum_j g_{\rm ps}^{j\alpha}$ (i.e., we do not actually simulate the full two-dimensional CCD image; the spatial dimension is collapsed into g^α). Then we create a Poisson realization of the (intermediate) quasar count $g^\alpha N_C^\alpha e^{-\tau_\alpha}$, where N_C is the continuum and $e^{-\tau_\alpha}$ is predicted by our cosmological model. We similarly create a Poisson realization of the background count g^αconst, where the constant represents some fraction less than 1, and then subtract from it its Poisson mean. The end result is then added to the above quasar count. Lastly, we divide by g^α to obtain the reduced quasar count $\hat{N}_Q^\alpha$.

Note that the overall level of the reduced quasar count can be scaled up or down (because we are not interested in the absolute brightness of the quasar), provided that the error array is scaled accordingly to conserve S/N (count divided by square root of the variance). This example resembles a high-quality Keck HIRES spectrum, with S/N reaching up to 100 at certain pixels. It is composed of 12 echelle orders, 50 Å each (e.g., the instrument HIRES on the Keck telescope is an echelle spectrograph that consists of two diffraction gratings crossed at 90° to each other; see Vogt et al. 1994). The pixel size is 0.05 Å with a resolution FWHM of 0.125 Å. The example represents a case in which a relative calibration (but not necessarily absolute fluxing) between the orders has been attempted.[8] The dashed line in the bottom panel shows the input continuum.

The error array shows a lot of variations. A model of the random error as Gaussian distributed with uniform S/N that is sometimes used in the literature misses much struc-

[8] Some of the data we work with are in this form, but a relative calibration, whenever possible, would be desirable.

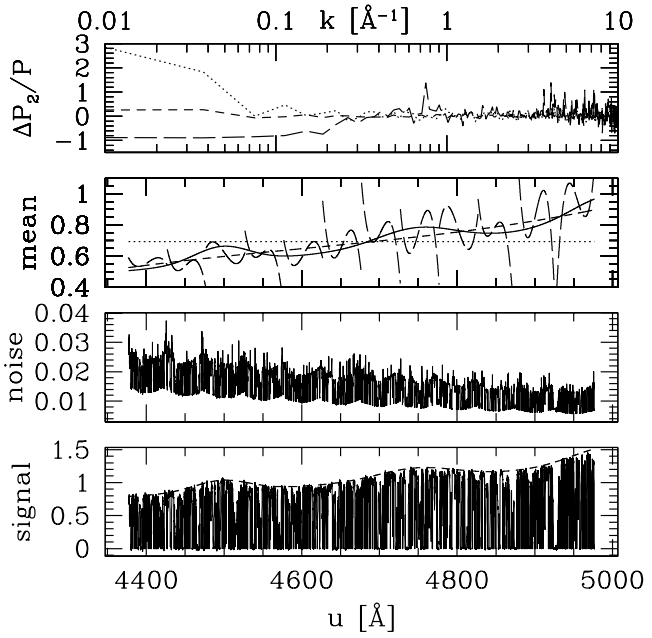

FIG. 2.—Bottom two panels show the reduced quasar spectrum (signal) with its noise array (eq. [9]), which resembles a high-resolution (8 km s^{-1} FWHM) echelle spectrum with 12 orders, each 50 Å long. It is assumed that a relative calibration between the different orders has been attempted. The spectrum is generated from the one in Fig. 1, with suitable noise added as described in § 3.2. Note that the overall normalization of the signal or noise is arbitrary but the ratio of the two is not. The dashed line of the bottom panel shows the input continuum level. The second panel from the top shows the recovered mean transmission for three different choices of the mean transmission basis ($\boldsymbol{p}$; eq. [16]): dotted line for a flat model continuum/mean ($p^0 = 1$ only); short-dashed line for a basis of polynomials up to the third order ($p^0 \ldots p^3$); and long-dashed line for a basis that is the same as that for the short-dashed line except that each echelle order is fitted separately. The solid line gives the true mean. The top panel shows the result of applying the power spectrum estimator $\hat{P}_2$ (eq. [12a]) with uniform weighting (eq. [14]). The symbol $\Delta P_2/P$ stands for $(\hat{P}_2 - P)/P$, where P is the true power spectrum. The different lines correspond to different choices of the mean basis, labeled as in the second panel.

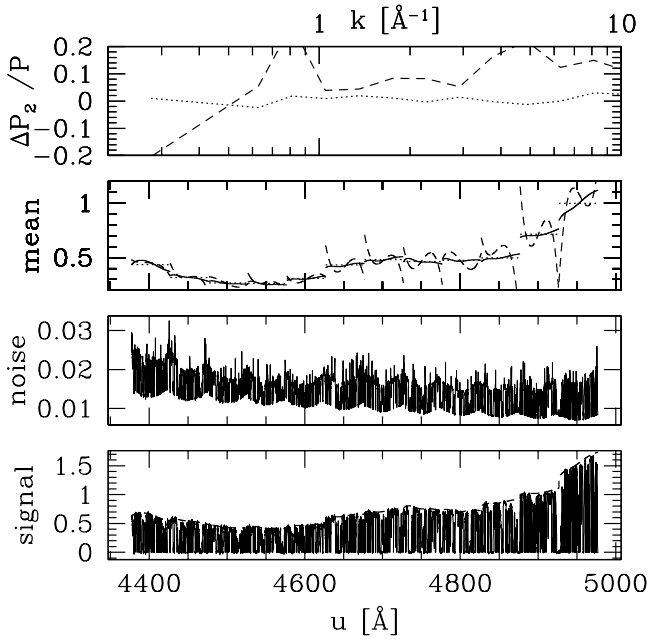

FIG. 3.—Similar to Fig. 2, except that a relative calibration between different echelle orders has not been done, and the input continuum is taken from continuum fits to an observed quasar spectrum. The dotted line in the top two panels corresponds to the case in which the continuum is modeled as flat for each order. The dashed line is where polynomials up to the third order are used to fit for the mean transmission in each order. The solid line in the second panel from the top shows the true mean transmission.

ture. About 3% of the spectrum consists of gaps that arise as a result of severe cosmic-ray hits. The spikes in the error array correspond to wavelengths at which a fraction, but not all, of the corresponding spatial pixels are affected by cosmic-ray hits. They also take up 3% of the spectrum. It is easy to see how these spikes arise from equations (6) and (9). At wavelengths where some of the spatial pixels are thrown out because of cosmic-ray hits, $W_{i\beta}$ is enhanced because the sum over j in its denominator is restricted to fewer pixels. Since it is the square of $W_{i\beta}$ that enters into the variance, a modest enhancement becomes a spike. Note how for each echelle order, the S/N drops toward the two ends. This is because of the blaze function that tends to suppress the flux at the ends. Note also that the S/N has a general decline toward the blue. This is due to a combination of a falling continuum and decreasing detector efficiency.

Sometimes, a relative calibration between echelle orders is either difficult or simply not attempted. An example is shown in Figure 3. Note how the continuum is broken into 12 discontinuous pieces.[9] These are taken from continuum fits to actual data.

[9] In some cases in which the different echelle orders overlap, there could be two jumps at each order junction.

An example of data with much poorer quality is shown later in Figure 7. The pixel size is 0.5 Å and the FWHM is 1.17 Å. The S/N is about 10 times worse than the two examples above. Such a spectrum could be the output of, say, a low-dispersion single-grating spectrograph, which does not have the characteristic division into short pieces as in the case of the echelle spectrograph.

All other simulated data in this paper are slight variations of the above, which will be described in turn at the appropriate places.

4. ESTIMATING THE POWER SPECTRUM/TWO-POINT CORRELATION

4.1. *The Quadratic Estimator*

Given the one-dimensional array of estimated quasar counts $\hat{N}_Q^\alpha$ (eq. [3]), how should one go about estimating the two-point correlation or the power spectrum?

A common practice is first to continuum fit, i.e., to estimate the continuum level $\hat{N}_C^\alpha$ and divide $\hat{N}_Q^\alpha$ by it to obtain an estimate of the transmission $\hat{f} = \hat{N}_Q^\alpha/\hat{N}_C^\alpha$. Then, the estimators for the unnormalized two-point correlation and power spectrum (eq. [1]) are

$$\hat{\xi}_{\rm un}(u) = \sum_{\alpha,\beta} w^{\alpha\beta}(u)\hat{f}^\alpha\hat{f}^\beta ,$$

$$\hat{P}_{\rm un}(k) = \sum_{\alpha,\beta} w^{\alpha\beta}(k)\hat{f}^\alpha\hat{f}^\beta - b(k) , \qquad (10)$$

where $b(k)$ subtracts out the shot noise (i.e., a bias) and $w^{\alpha\beta}(u)$ and $w^{\alpha\beta}(k)$ are weighting kernels for which we will give some examples shortly (to be distinguished from $W^{\alpha\beta}$ in eq. [3]). These are commonly called quadratic estimators for the simple fact that they make use of the data $\hat{f}^\alpha$ in quadratic combinations.

Defining the mean transmission to be $\bar{f}$, the obvious extensions of the above estimators, for the normalized two-point correlation and power spectrum (eq. [2]; unless otherwise stated, the two-point correlation or power spectrum with no qualifications refers to the normalized version), are

$$\hat{\xi}_1(u) = \sum_{\alpha,\beta} \frac{w^{\alpha\beta}(u)(\hat{f}^\alpha - \bar{f})(\hat{f}^\beta - \bar{f})}{\bar{f}^2} ,$$

$$\hat{P}_1(k) = \sum_{\alpha,\beta} \frac{w^{\alpha\beta}(k)(\hat{f}^\alpha - \bar{f})(\hat{f}^\beta - \bar{f})}{\bar{f}^2} - b(k) . \quad (11)$$

However, the form of the power spectrum or two-point correlation estimator given above suggests an interesting variation that allows us to avoid continuum fitting altogether: $(\hat{f}^\alpha - \bar{f})/\bar{f}$ can be estimated instead by $(\hat{N}^\alpha - \bar{N}^\alpha)/\bar{N}^\alpha$, where $\bar{N}^\alpha$ is the mean count defined by $\bar{N}^\alpha \equiv \langle \hat{N}^\alpha \rangle$. Here $\langle\rangle$ denotes the cosmic average, i.e., this corresponds to averaging out the cosmic fluctuations in f, for a fixed continuum ($\langle \hat{N}^\alpha \rangle = N_C^\alpha \bar{f}$, where N_C^α is the true continuum).[10] Note that the mean count is dependent upon α because of the slowly varying continuum. We will discuss how to estimate $\bar{N}^\alpha$ shortly. The key here is that the absolute level of the continuum gets divided out by definition. Hence, let us define the following alternative estimators of the two-point correlation and power spectrum:

$$\hat{\xi}_2(u) = \sum_{\alpha,\beta} w^{\alpha\beta}(u)\hat{\delta}_f^\alpha \hat{\delta}_f^\beta ,$$

$$\hat{P}_2(k) = \sum_{\alpha,\beta} w^{\alpha\beta}(k)\hat{\delta}_f^\alpha \hat{\delta}_f^\beta - b(k) , \quad (12a)$$

$$\hat{\delta}_f^\alpha \equiv (\hat{N}_Q^\alpha - \bar{N}_Q^\alpha)/\bar{N}_Q^\alpha . \quad (12b)$$

This alternative power spectrum estimator is what we will focus on, but we will also briefly investigate the behavior of the estimators in equations (10) and (11).[11]

It remains to specify what $w^{\alpha\beta}(u)$, $w^{\alpha\beta}(k)$, and $b(k)$ are. The simplest choice is to use uniform weighting, i.e., for the two-point correlation, it corresponds to

$$w^{\alpha\beta}(u) = \frac{\Theta^{\alpha\beta}(u)}{\sum_{\mu\nu} \Theta^{\mu\nu}(u)} , \quad (13)$$

where $\Theta^{\alpha\beta}(u)$ is equal to 1 if the two pixels α and β are separated by a distance u (or more generally, the distance falls into a bin that is centered around u with some finite width) and 0 otherwise. Using the above $w^{\alpha\beta}(u)$ corresponds to simply counting all pairs separated by a distance u, normalized by the total number of pairs.

With the above weighting, equation (12) is analogous to an estimator of the two-point correlation introduced by Landy & Szalay (1993) for galaxy surveys: $(DD - 2DR + RR)/RR$, if one identifies DD with $\sum_{\alpha\beta} w^{\alpha\beta}(u)\hat{N}^\alpha\hat{N}^\beta$, DR with $\sum_{\alpha\beta} w^{\alpha\beta}(u)\hat{N}^\alpha\bar{N}^\beta$, and RR with $\sum_{\alpha\beta} w^{\alpha\beta}(u)\bar{N}^\alpha\bar{N}^\beta$ and assumes that $\bar{N}^\alpha$ varies very slowly with α on the scale of interest u (the analogy becomes exact in the limit of a uniform $\bar{N}^\alpha$). As shown by Landy & Szalay (1993; see also Szapudi & Szalay 1998; Dodelson, Hui, & Jaffe 1997), a common alternative estimator $DD/RR - 1$ (equivalent to the estimator used by, e.g., Zuo & Bond 1994; Cen et al. 1998) is actually less desirable as it gives a larger variance compared to $(DD - 2DR + RR)/RR$.

With this being said, we are going to focus our attention on the power spectrum from now on, although most of our treatments below can be applied to the two-point correlation as well. For the power spectrum, the simplest choice of uniform weighting corresponds to

$$w^{\alpha\beta}(k) = \frac{\mathscr{L}}{\mathscr{N}^2} R^{\alpha\beta}(k) , \quad R^{\alpha\beta}(k) \equiv \frac{1}{n_k} \sum_{k'} e^{-ik'(u^\alpha - u^\beta)} , \quad (14a)$$

$$b(k) = \frac{\mathscr{L}}{\mathscr{N}^2} \sum_\alpha \frac{q^\alpha \bar{N}_Q^\alpha + V_B^\alpha}{(\bar{N}_Q^\alpha)^2} ,$$

$$q^\alpha \equiv \sum_i (W^{i\alpha})^2 g_{\rm ps}^{i\alpha} g_b^\alpha ,$$

$$V_B^\alpha \equiv \sum_i (W^{i\alpha})^2 V_B^{i\alpha} , \quad (14b)$$

where $\mathscr{L}$ is the total length of the spectrum (in whatever units one prefers) and $\mathscr{N}$ is the total number of spectral pixels. $R^{\alpha\beta}(k)$ represents an average of the Fourier basis over some bin or band in k-space, i.e., we estimate the power spectrum at a bin centered at k by averaging over contributions from each k' that belongs to the bin (n_k is the total number of modes in it).[12] The symbol $b(k)$ represents the shot noise contribution to the power that has to be subtracted off, $g_{\rm ps}^{i\alpha}$ and g_b^α represent the point-spread function and the blaze as in equation (5), and V_B^α is the background contribution to the shot noise (including the sky and readout; see eq. [8]). Note how a part of the shot noise depends on the reciprocal of the mean quasar count, as expected for Poisson fluctuations, quite analogous to the shot noise of galaxy distributions.[13] However, the factor q^α, which arises from nontrivial weighting of CCD pixels ($W^{i\alpha}$; eq. [3]), signifies that the shot noise is not strictly Poisson distributed. Moreover, there are extra contributions to the shot noise from the background counts, which are generally absent in the case of galaxy surveys. Derivations of the above statements are given in Appendix A.

The corresponding power spectrum estimator obeys

$$\langle \hat{P}_2(k) \rangle = \int \frac{dk'}{2\pi} G(k, k')P(k') ,$$

$$G(k, k') \equiv \sum_{\alpha\beta} w^{\alpha\beta}(k) e^{ik'(u^\alpha - u^\beta)} , \quad (15)$$

where G is a window function that resembles, for $k \gg 1/\mathscr{L}$, a delta function centered at $k = k'$ with a width of the order of $1/\mathscr{L}$. The normalization of $w^{\alpha\beta}(k)$ in equation (14a) ensures that $\int dk'\, G(k, k')/2\pi = 1$. See Appendix A for a derivation.

It should be emphasized, however, that the above statements are strictly true only if one ignores uncertainties in the mean count $\bar{N}_Q^\alpha$, i.e., $\bar{N}_Q^\alpha$ is not known a priori but is instead estimated from the same data from which one is trying to measure correlations. We will discuss this further in § 4.2.2. It suffices to say here that our results in this section remain valid as long as one stays away from scales comparable to the entire length of the quasar spectrum.

[10] It is implicitly assumed that discrete averaging has been carried out before cosmic averaging, i.e., we use $\langle\rangle$ interchangeably with $\langle\langle\rangle_D\rangle$. See § 2.

[11] Note that in both estimators, the absolute brightness of the quasar gets divided out, which is as it should be, since we are interested in the fluctuations caused by the intervening intergalactic medium rather than the absolute brightness of the quasar itself.

[12] See Seljak (1998) and Bond et al. (1998) for discussions on precautions one should take on binning.

[13] See, e.g., Feldman et al. (1994).

4.2. *Systematic Errors*

4.2.1. *Continuum Fitting versus Trend Removal*

The power spectrum estimator $\hat{P}_2$ in equation (12), on which we are going to focus most of our attention, requires an estimate of mean count $\bar{N}_Q^\alpha$. The mean count is not strictly uniform because of a slowly fluctuating continuum, i.e., $\bar{N}_Q^\alpha = N_C^\alpha \bar{f}$, where N_C^α is the continuum and $\bar{f}$ is the (flat) mean transmission. We assume that $\bar{N}^\alpha$ has the following form:

$$\bar{N}_Q^\alpha = \sum_a C^a p^{a\alpha} , \tag{16}$$

where p^0 is a constant, p^1 is the first-order polynomial ($p^{1\alpha} = u^\alpha$), p^2 is the second-order polynomial [$p^{2\alpha} = (u^\alpha)^2$], and so on. The coefficients C^a need to be estimated from the quasar counts $\hat{N}_Q^\alpha$. Note that most of our following arguments would go through for a different set of basis functions. One key assumption we will exploit is that $\bar{N}^\alpha$ is slowly fluctuating, so that we can truncate the above series at relatively low orders. Continuum fitting in practice makes the same assumption.

To estimate C^a, we use a linear estimator:

$$C^a = \sum_\alpha M^{a\alpha} \hat{N}_Q^\alpha , \tag{17}$$

where $\boldsymbol{M}$ is a matrix to be specified. Comparing equations (16) and (17), it is not hard to see that $\boldsymbol{M}$ has to satisfy $\sum_\alpha M^{a\alpha} p^{b\alpha} = \delta^{ab}$. The simplest choice is to adopt, in vector notation, $\boldsymbol{M} = (\boldsymbol{p}\boldsymbol{p}^T)^{-1}\boldsymbol{p}$, where $\boldsymbol{p}\boldsymbol{p}^T$ in component form is $(pp^T)^{ab} = \sum_\alpha p^{a\alpha} p^{b\alpha}$. In summary, this means that our estimator for the mean quasar count is

$$\bar{N}_Q^\alpha = \sum_\beta L^{\alpha\beta} \hat{N}_Q^\beta , \quad \boldsymbol{L} \equiv \boldsymbol{p}^T (\boldsymbol{p}\boldsymbol{p}^T)^{-1} \boldsymbol{p} , \tag{18}$$

where $\boldsymbol{L}$ in component form reads $L^{\alpha\beta} \equiv \sum_{ab} p^{a\alpha} \tilde{p}^{ab} p^{b\beta}$ with $\tilde{p}^{ab}$ being the inverse of the matrix $(pp^T)^{ab}$. More sophisticated versions of the above can be found in Rybicki & Press (1992). Our experience is that the simple version given here suffices because the shape of the true quasar continuum is quite uncertain anyway.

Note the crucial differences between traditional continuum fitting and an estimation of the mean count as described above. The above method makes no reference to the absolute level of the continuum, i.e., the count level where there is supposedly no absorption. Continuum fitting in practice often involves human intervention (eyeballing) in the identification of such a level. In contrast, equation (18) is straightforward to implement and automate. The mean count is then used to normalize the quasar count as in equation (12b) before the power spectrum is estimated (eq. [12]). We call this procedure trend removal to distinguish it from traditional continuum fitting. Trend removal is widely used in other areas (e.g., Press, Rybicki, & Hewitt 1992a; Rybicki & Press 1992; Tegmark et al. 1998). It is akin to the estimation of, say, the long-term trend of the stock market in the midst of all its daily fluctuations.

Equation (18), together with equations (12) and (14), completely specifies the main power spectrum estimator we advocate. Several tests follow.

Test 1.—In Figure 2 we show the effect of different choices of the mean transmission basis $\boldsymbol{p}$ (eq. [16]). The simulated spectrum is of Keck-HIRES quality, with an S/N as high as 100 at certain pixels, and it assumes that one has a good relative calibration between the different echelle orders (12 in all), i.e., an almost ideal, state-of-the-art observed spectrum. The second panel from the top shows the recovery of the mean transmission. The solid line represents the true (input) mean. The rest shows the recovered mean for different bases: the dotted line for a basis consisting of p^0 only (a constant, i.e., the continuum or the mean is modeled as completely flat), the short-dashed line for a basis consisting of polynomials up to the third order, and the long-dashed line also for a basis of polynomials up to the third order but with coefficients fitted separately for each echelle order. The short-dashed line seems to give the best match to the true mean. However, none of them are perfect because the true mean does not, by choice, have a polynomial shape. This is what is likely to happen in practice: lacking a good understanding of the physics that determines the continuum shape of any given quasar, the best one can do is to pick a reasonable basis that contains enough freedom to describe the general features of the continuum, but not so much freedom that one overfits.

The important question is what impact the choice of basis has on power spectrum estimation. This is illustrated in the top panel of Figure 2, where the fractional error in the measured power spectrum is shown. The one that gives the best match to the true power spectrum is indeed the one where a simple basis of $p^0, \ldots, p^3$ is used for the whole length of the simulated spectrum. The biggest effect of underfitting (*dotted line*) or overfitting (*long-dashed line*) the mean transmission is on the power spectrum estimation on large scales. They cause, respectively, over- or underestimation of the large-scale power spectrum. An additional effect is that overfitting tends to introduce spurious power on small scales as well: witness the enhanced fluctuations in the error on small scales for the long-dashed line. We will see this more clearly in the next test.

Without any prior knowledge of the intrinsic continuum shape of an observed quasar, how does one decide if one is overfitting or underfitting? One way is to look at the region of the observed spectrum redward of the Lyα emission line, which is free of the forest, and the continuum is therefore relatively easy to reconstruct. Assuming that the general level of continuum fluctuation is the same both redward and blueward of Lyα, one can then gain an idea of what a good mean transmission basis might be. Low-redshift QSO spectra, where the continuum can be quite easily recovered even blueward of Lyα, can also be used to gauge the scales at which the continuum fluctuates. Naturally, one could also use such spectra to check the assumption that continuum fluctuations have similar characteristics redward and blueward of Lyα (more on this below).

Test 2.—In Figure 3 we took the continuum fits to an observed quasar spectrum and used them as the input continuum for our simulation. The simulated spectrum here represents a case in which no relative calibration between echelle orders has been attempted, which is often the case. This is why the continuum in the bottom panel is broken up into 12 pieces. The second panel from the top again illustrates the recovery of the mean transmission: the dotted line for a flat model continuum for each order, and the dashed line for a basis of polynomials up to the third order, also separately for each order. The solid line is the true mean transmission. The top panel shows the accuracy of the corresponding power spectrum estimations. The assumption of a simple flat continuum for each order gives a power spectrum that is accurate to $\sim 1\%$. On the other hand, over-

fitting with up to third-order polynomials not only causes an underestimation of power on large scales but also creates spurious power on small scales.

Combining Figures 2 and 3 (note that they show the power spectrum estimation on different scales), the lessons are (1) it is better to err on the side of underfitting the mean, which tends to overestimate the power on large scales, but leave the power on small scales relatively unaffected (this relies crucially on the fact that the continuum has fluctuations only on large scales); and (2) without sufficient prior knowledge of the true shape of the continuum, one can at least make conservative statements about the small-scale power, but the large-scale power is likely prone to systematics, unless some correction is made.

One additional comment: the input continuum in Figure 3, which is taken from fits to actual data, certainly seems to suggest that the observed continuum has fluctuations on scales of an echelle order (~ 50 Å). (We will quantify this better in § 4.2.2.) It is unclear whether this is truly due to the intrinsic continuum or whether it is an artifact of imperfect blaze removal or flat-fielding (see § 3.1). If it is the former, then ~ 50 Å represents a fundamental limit beyond which one cannot reliably measure the transmission power spectrum, at least not without some additional prior knowledge of the true continuum (which is what we will discuss in § 4.2.2). If it is the latter, then in principle one should be able to do better and extend the range of reachable scales to larger ones. Which is the case remains to be seen.

Test 3.—In Figure 4 we show the effect of traditional continuum fitting, which requires some degree of eyeballing. The same simulated spectrum as in Figure 3 is given to an observer (one of the authors) with no knowledge of the input continuum. Note that the second panel from the top now shows the actual continuum level rather than the mean transmission level. The estimated continuum actually matches the true one surprisingly well. However, one can still see that the continuum is generally underestimated. In the top panel we show the accuracy of two different power estimates. The long-dashed line corresponds to an estimate of the unnormalized power spectrum as defined in equation (1) (the estimator is eq. [10]). There is clearly an $\sim 5\%$ positive bias here because of the underestimation of the continuum. One way to correct for it is of course to use simulations—apply exactly the same procedure to the observed data and the simulated data and see how much bias results—but the size of the bias is likely to be model dependent. A simple alternative way to cure this problem is to measure the normalized power spectrum instead, using the continuum-fitted data, i.e., using the estimator in equation (11). This is shown with a short-dashed line. It has an accuracy of $\sim 1\%$, comparable to the dotted line in Figure 3. In view of this, it seems that bypassing continuum fitting altogether and proceeding simply with trend removal is desirable.

Test 4.—The failure of traditional continuum fitting is more dramatic in cases in which there is a lot of absorption, e.g., at high redshifts. In Figure 5 a simulated spectrum is shown with the ionizing background adjusted to give a mean transmission of 0.39, which is about the observed value at $z = 4$ (Press et al. 1993). The continuum is more seriously underestimated, leading to an overestimate of the unnormalized power spectrum by $\sim 20\%$ (*upper long-dashed line*). The normalized power spectrum, estimated either using the continuum-fitted data (eq. [11]) or using directly the trend-removed data (eq. [12]), is much more accurately measured.

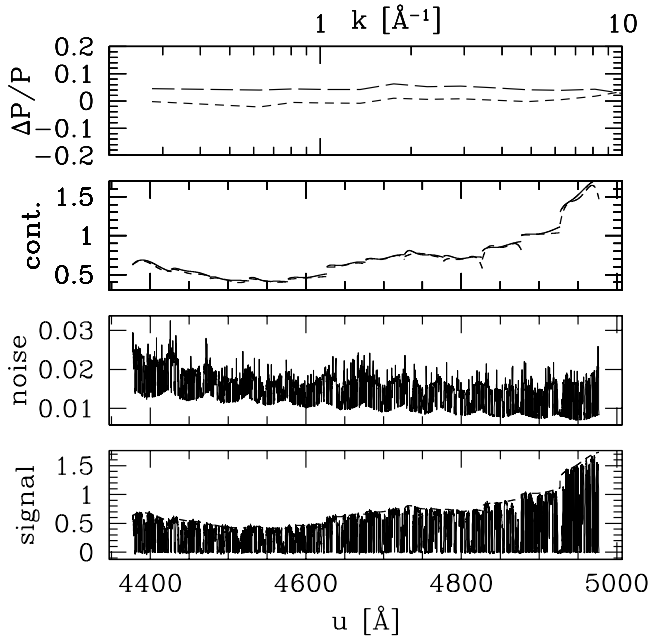

FIG. 4.—Bottom two panels are the same as in Fig. 3. The solid line in the second panel from the top shows the true continuum level (as opposed to the true transmission mean level as before). The dashed line is a continuum fit to the simulated spectrum. In the top panel, $\Delta P/P$ denotes $(\hat{P}_{\rm un} - P_{\rm un})/P_{\rm un}$ (the unnormalized power spectrum; eqs. [10] and [1]) for the upper long-dashed line, while it denotes $(\hat{P}_1 - P)/P$ (the normalized power spectrum, but estimated using the continuum-fitted data; eqs. [11] and [2]) for the lower short-dashed line.

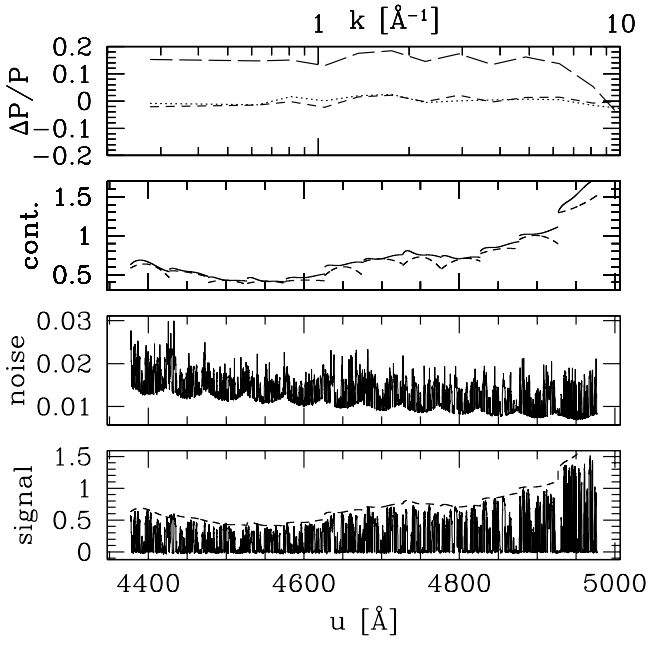

FIG. 5.—Similar to Fig. 4, except that the mean transmission is lower (0.39 instead of 0.64). The solid and dashed lines in the second panel from the top represent, respectively, the true continuum level and the continuum fits. In the top panel, the upper long-dashed line shows $(\hat{P}_{\rm un} - P_{\rm un})/P_{\rm un}$ (error for the unnormalized power spectrum estimated from eq. [10]), the lower short-dashed line shows $(\hat{P}_1 - P)/P$ (error for the normalized power spectrum estimated using continuum-fitted data; eq. [11]), and the dotted line shows $(\hat{P}_2 - P)/P$ (error for the normalized power estimated using trend removal with a flat trend for each echelle order; eq. [12]).

Test 5.—Another example in which traditional continuum fitting fails is shown in Figure 6. This is based on the same spectrum as in Figure 1, but convolved with a Gaussian of 1.17 Å FWHM and with much poorer S/N compared to the simulated spectra above. This is likely not the product of an echelle spectrograph, hence there is no division into 12 orders. We repeat the exercise of continuum fitting and then power spectrum measurement as before. Interestingly, the significant discrete fluctuations due to the low S/N here cause an overestimation (unlike in tests 3 and 4) of the continuum level and thus an underestimation of the unnormalized power spectrum. Once again, the normalized power spectrum does not suffer from the same problem. Note the somewhat large fluctuations of the estimated power; this is largely due to the high level of shot noise.

Figure 7 shows the measurement of power spectrum through trend removal instead. A third-order polynomial is used to estimate the mean transmission. The resulting (normalized) power spectrum estimate (eq. [12]) is of comparable accuracy to that using the continuum-fitted data. We also show in the top panel as a dotted line the power spectrum estimate without shot noise subtraction (eq. [14]). Clearly, shot noise introduces a bias of the order of 10% here. We will have some more to say about this in § 5.

Tests 4 and 5 above drive home the point that the bias of an estimate of the unnormalized power spectrum from continuum-fitted data is highly variable. It depends on the redshift, resolution, and S/N of the data.

There have been in the literature discussions of an alternative method to normalize the quasar count: normalizing by the maximum value of the continuum-fitted count, instead of by the mean count (e.g., McDonald et al. 2000). Note that this procedure is also sensitive to the S/N and resolution of the data. For instance, max $(\hat{N}_Q^\alpha/\hat{N}_C^\alpha) = 1.4$ in Figure 6, while max $(\hat{N}_Q^\alpha/\hat{N}_C^\alpha) = 1.12$ in Figure 4, where $\hat{N}_C^\alpha$ is the estimated continuum. They share exactly the same underlying cosmic signal, but the former has a higher level of discrete fluctuations and poorer resolution. For reference, the true maximum transmission should be 0.99. This means that one should take care to simulate the noise properties correctly (e.g., Rauch et al. 1997).

Lastly, we should emphasize that while trend removal seems to be more desirable than traditional continuum fitting for the particular application here, continuum fitting is still very useful for other purposes, which we will discuss in §§ 4.2.2 and 5. However, a fully automated procedure for continuum fitting is clearly desirable.

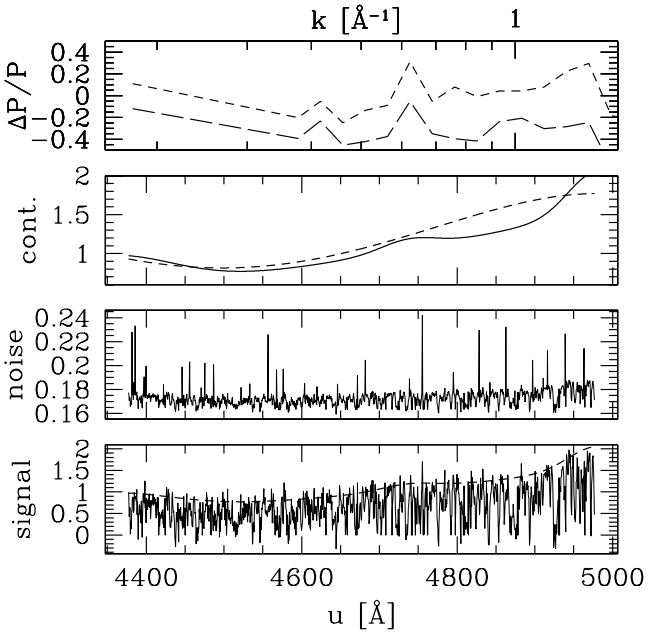

FIG. 6.—Bottom two panels show a low-resolution (FWHM = 1.17 Å), somewhat noisy simulated spectrum based on the one in Fig. 1. The solid line in the second panel from the top is the true continuum (not mean transmission), whereas the dashed line corresponds to a continuum fit to the data in the bottom panel. In the top panel, $\Delta P/P$ denotes $(\hat{P}_{un} - P_{un})/P_{un}$ (the unnormalized power spectrum; eqs. [10] and [1]) for the lower long-dashed line, while it denotes $(\hat{P}_1 - P)/P$ (the normalized power spectrum, but estimated using the continuum-fitted data; eqs. [11] and [2]) for the upper short-dashed line.

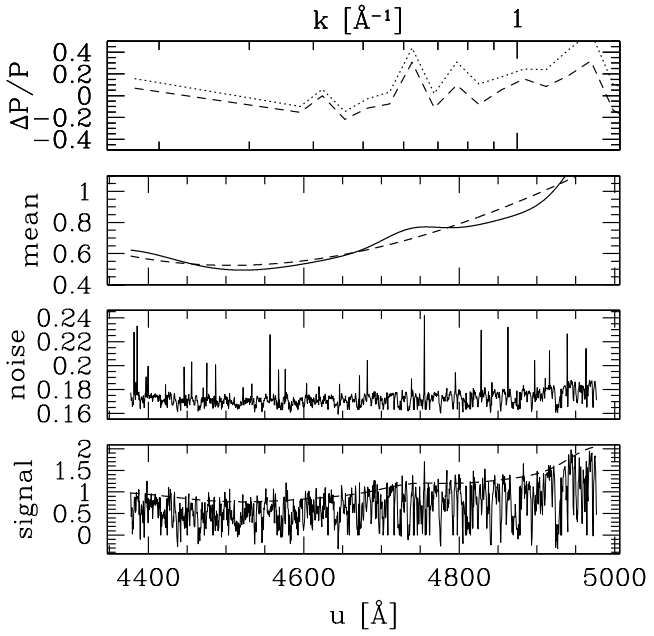

FIG. 7.—Bottom two panels are the same as in Fig. 6. The second panel from the top shows the recovery of the mean transmission assuming a model composing of polynomials up to the third order (*dashed line*). The solid line shows the true mean. The dashed line in the top panel represents $(\hat{P}_2 - P)/P$, the error for the normalized power spectrum estimated using eq. (12). The dotted line shows the power spectrum estimate if shot noise is not subtracted.

4.2.2. *A Bonus of Trend Removal: Power Correction on Large Scales*

As is clear from some of the previous tests in § 4.2.1, the power spectrum measured on large scales (i.e., scales comparable to the typical scales where the continuum has fluctuations) could contain spurious contributions from the continuum, the size of which depends somewhat on the continuum/mean shape model one assumes. The strategy adopted in § 4.2.1 is a conservative one: assume a model for the continuum that is as simple (or smooth) as possible, perform trend removal, and the resulting power spectrum would reflect the true transmission power spectrum at least on small scales, but not necessarily on large scales.

Can we do better? The answer is yes, under certain assumptions that we will make explicit shortly, and it illustrates an added benefit of trend removal as introduced in

§ 4.2.1. Readers not interested in the details can skip directly to the end of this section where two examples of how our technique of power correction works are given (Figs. 9 and 10).

Let us start by recalling the power spectrum estimator in equation (12), but focusing now on the fact that the true $\bar{N}^\alpha$ is unknown and has to be estimated using equation (18), which assumes implicitly that the true mean count obeys equation (16), which is of course only a reasonable guess. Let us use $\hat{\bar{N}}^\alpha_Q$ to denote the estimated mean count, which generally differs from the true mean count $\bar{N}^\alpha_Q$. We have used $\bar{N}^\alpha$ somewhat sloppily before when we really meant $\bar{N}^\alpha_Q$, e.g., equation (18). In other words, equation (18) should be more accurately written as

$$\hat{\bar{N}}^\alpha_Q = \sum_\beta L^{\alpha\beta}\hat{N}^\beta_Q\ , \quad \boldsymbol{L} \equiv \boldsymbol{p}^T(\boldsymbol{p}\boldsymbol{p}^T)^{-1}\boldsymbol{p}\ , \tag{19}$$

where $\boldsymbol{p}$ represents the basis functions. Similarly, the estimator for the power spectrum in equation (12) should be more accurately written as

$$\hat{\xi}_2(u) = \sum_{\alpha,\beta} w^{\alpha\beta}(u)\hat{d}^\alpha_f\,\hat{d}^\beta_f\ ,$$

$$\hat{P}_2(k) = \sum_{\alpha,\beta} w^{\alpha\beta}(k)\hat{d}^\alpha_f\,\hat{d}^\beta_f - b(k)\ , \tag{20a}$$

$$\hat{d}^\alpha_f \equiv \frac{\hat{N}^\alpha_Q - \hat{\bar{N}}^\alpha_Q}{\hat{\bar{N}}^\alpha_Q} = \frac{\sum_\gamma D^{\alpha\gamma}\hat{N}^\gamma_Q}{\sum_\eta L^{\alpha\eta}\hat{N}^\eta_Q}\ , \quad D^{\alpha\gamma} \equiv \delta^{\alpha\gamma} - L^{\alpha\gamma}\ . \tag{20b}$$

We now assume that the following quantities are small: $\hat{d}^\alpha_f$ and $(\sum L^{\alpha\gamma}\hat{N}^\gamma_Q - \bar{N}^\alpha_Q)/\bar{N}^\alpha_Q$. The second quantity tells us how far off our estimate of the mean is from the true mean, while the first contains contributions both from the fluctuation in transmission and from the second quantity. Therefore, putting equations (20b) and (20a) together, the lowest order contributions to the expectation value of the estimator $\hat{P}_2(k)$ are

$$\langle\hat{P}_2(k)\rangle = \sum_{\alpha\beta\gamma\eta} w^{\alpha\beta}(k)\left\langle D^{\alpha\gamma}D^{\beta\eta}\,\frac{\bar{N}^\gamma_Q\,\bar{N}^\eta_Q}{\bar{N}^\alpha_Q\,\bar{N}^\beta_Q}\right\rangle \times (1 + \langle\hat{\delta}^\gamma_f\,\hat{\delta}^\eta_f\rangle) \tag{21a}$$

$$= P_C(k) + \int \frac{dk'}{2\pi}\,P(k')G_n(k,\,k')\ ,$$

$$P_C(k) \equiv \sum_{\alpha\beta\gamma\eta} w^{\alpha\beta}(k)\left\langle D^{\alpha\gamma}D^{\beta\eta}\,\frac{\bar{N}^\gamma_Q\,\bar{N}^\eta_Q}{\bar{N}^\alpha_Q\,\bar{N}^\beta_Q}\right\rangle\ , \tag{21b}$$

$$G_n(k,\,k') \equiv \sum_{\alpha\beta\gamma\eta} w^{\alpha\beta}(k)e^{ik'(u^\gamma - u^\eta)}\left\langle D^{\alpha\gamma}D^{\beta\eta}\,\frac{\bar{N}^\gamma_Q\,\bar{N}^\eta_Q}{\bar{N}^\alpha_Q\,\bar{N}^\beta_Q}\right\rangle\ , \tag{21c}$$

where we have retained the old definition of $\hat{\delta}^\gamma_f$ as $(\hat{N}^\gamma_Q - \bar{N}^\gamma_Q)/\bar{N}^\gamma_Q$ (eq. [12]). The above gives an idea of how biased the estimator $\hat{P}_2(k)$ is. Note that we have used $\langle\rangle$ here to include, in addition to the ensemble averaging as explained in § 2, an averaging over the ensemble of possible continua (which changes $\bar{N}^\alpha_Q$ because it is directly proportional to the continuum count N^α_C). We have assumed that the fluctuations in the continuum are uncorrelated with fluctuations in the cosmic signal $\hat{\delta}^\gamma_f$. We have also ignored the shot noise contributions [e.g., $b(k)$] and will continue to do so for the rest of this section because the scales where the continuum contamination could be a problem are typically large enough that shot noise is subdominant.

The term $P_C(k)$ can be viewed as the power spectrum of the continuum fluctuation. This is fluctuation in the sense of $D^{\alpha\gamma}\bar{N}^\gamma_Q = \bar{N}^\alpha_Q - L^{\alpha\gamma}\bar{N}^\gamma_Q$. This fluctuation would vanish if our trend removal procedure were so accurate that the continuum shape is exactly reproduced. The term $G_n(k, k')$ is the effective window function, replacing the one in equation (15), which does not take into account the error involved in trend removal. The desirable normalization condition $\int G_n(k, k')dk'/(2\pi) = 1$ no longer holds with the choice of $w^{\alpha\beta}$ in equation (14a). We have instead

$$\int G_n(k,\,k')\,\frac{dk'}{2\pi} = \frac{1}{\mathcal{N}}\sum_{\alpha\beta\gamma} R^{\alpha\beta}(k)D^{\alpha\gamma}D^{\beta\gamma}\,\frac{(\bar{N}^\gamma_Q)^2}{\bar{N}^\alpha_Q\,\bar{N}^\beta_Q} \equiv 1 + \epsilon_G(k)\ , \tag{22}$$

where $R^{\alpha\beta}(k)$ is defined in equation (14a).

Assuming for now that $P_C(k)$ and $\epsilon_G(k)$ can be measured from a suitable ensemble of continua, we propose the following alternative estimator to $\hat{P}_2(k)$, which removes the bias due to continuum contamination:

$$\hat{P}_3(k) = \frac{\hat{P}_2(k) - P_C(k)}{1 + \epsilon_G(k)}\ . \tag{23}$$

The above gives us an unbiased estimate of the *windowed* power spectrum. The window is effectively $G_n(k, k')/[1 + \epsilon_G(k)]$, which has the desirable normalization. We will not attempt further improvements such as deconvolution in this paper.

A useful alternative estimator, in cases in which P_C dominates the bias in $\hat{P}_2(k)$, is

$$\hat{P}_4(k) = \hat{P}_2(k) - P_C(k)\ . \tag{24}$$

The above estimator gets rid of most of the bias in the estimator $\hat{P}_2(k)$ if $P_C(k)/P(k) \gg \epsilon_G(k)$. An interesting corollary is that, under such a condition, the bias in $\hat{P}_2(k)$ is positive since $P_C(k)$ is positive definite. Needless to say, this statement breaks down if $P_C(k)$ is not the dominant source of bias, or if the fractional error in the mean estimation is large (see, e.g., Fig. 2).

It is interesting to compare our derivation above with the well-known one for the integral constraint bias in galaxy surveys (e.g., Peebles 1980; Landy & Szalay 1993; Bernstein 1994; Tegmark et al. 1998). The integral constraint arises from the fact that the mean density of a galaxy survey has to be estimated from the same survey from which one is also measuring the power spectrum. The fact that the power spectrum estimator involves a nontrivial nonlinear combination of the data gives rise to a bias (see Hui & Gaztañaga 1999), quite analogous to our derivation here. However, in the standard derivations, it is assumed that the shape of the mean density is known (often taken to be uniform), and therefore $P_C(k)$ effectively vanishes, whereas $\epsilon_G(k)$ can be nonnegligible on scales comparable to the size of the survey but is otherwise small. The reader is referred to Bernstein (1994) and Hui & Gaztañaga (1999) for discussions on higher order contributions to the integral constraint.

How should one estimate $P_C(k)$ and $\epsilon_G(k)$? Given an ensemble of continua (with counts represented by N^α_C), our procedure is to replace $\bar{N}^\alpha_Q$, which appears in the definitions of $P_C(k)$ and $\epsilon_G(k)$ (eqs. [21b] and [22]), with N^α_C and compute the corresponding ensemble averages. Note that

$\bar{N}_Q^\alpha = N_C^\alpha \langle e^{-\tau} \rangle$, but $\langle e^{-\tau} \rangle$, which is taken to be constant over the finite redshift range of interest, is divided out in the relevant definitions of $P_C(k)$ and $\epsilon_G(k)$.

The hard question is of course how to obtain a suitable ensemble of continua. The first thing one might try is to measure the power spectrum of the continuum fits [i.e., $P_C(k)$, or more generally, both $P_C(k)$ and $\epsilon_G(k)$] from exactly the same regions from which one attempts to measure the transmission power spectrum. While this can give us a crude idea of how significant the continuum power spectrum is, it is not entirely satisfactory because part of what has been ascribed to the continuum might actually be large-scale fluctuations in the cosmic signal δ_f^α that we are after, or vice versa.

The second option that comes to mind is to measure the continuum power spectrum from regions where the continuum determination is relatively secure. Two possibilities are (1) low-redshift quasar spectra where the forest is much less dense and (2) regions of spectra that lie redward of Lyα emission. The working hypothesis is that the continuum power spectrum in these two regions is the same as, or at least similar to, the one in the region where we attempt to estimate the transmission power spectrum (the forest of interest). There is no guarantee that the hypothesis is valid. For instance, regarding possibility (1), the continuum power could systematically evolve with redshift. In fact, it probably does: assuming that the statistical properties of the quasar continuum *in rest frame* do not evolve with redshift, the observed continuum power would evolve as $P_C(k, z_1) = P_C[k(1 + z_0)/(1 + z_1), z_0]$. One could in principle constrain such redshift evolution with a sufficiently large sample of low-redshift quasar spectra. Regarding possibility (2), it is not unreasonable to expect that the continuum power is higher on the red side compared to the blue side because there are generally more broad emission lines on the red side (see, e.g., Peterson 1997; Blandford, Netzer, & Woltjer 1990; see below for caveats and a counterexample, however). An upper bound on the blue continuum power is by itself interesting because one can then obtain a conservative estimate of how much spurious power is introduced by the continuum into one's forest power measurements. Furthermore, systematic differences between the red and blue continuum power can be studied and quantified with a sufficiently large sample of low-redshift quasars.

In Figure 8 we show the continuum power spectrum measured from the continuum estimates on both sides of the Lyα emission of a quasar at $z = 3$ (QSO 1157+3143). The continuum estimates are shown in the bottom two panels. After fitting a flat mean to each echelle order, we compute the continuum power spectrum just as if this were the forest, and the results from the red side and blue side are shown as solid and dotted lines, respectively, in the top panel. The two power spectra look similar. However, we emphasize that because of the lack of small-scale power in the continuum, most of the power on small scales ($k \gtrsim 1$ Å^{-1}) that we see in Figure 8 is likely aliased from large scales. We will not attempt to perform a deconvolution to obtain the true small-scale power; it suffices to note here that the true small-scale power can only be smaller than what is shown in the figure. Also shown in the top panel is $\epsilon_G(k)$, on both sides of Lyα, which are basically indistinguishable from each other. Note that $\epsilon_G(k) \ll 1$. The second panel from the top shows the fractional difference between P_C from the red and blue sides, which is about 10%, with the

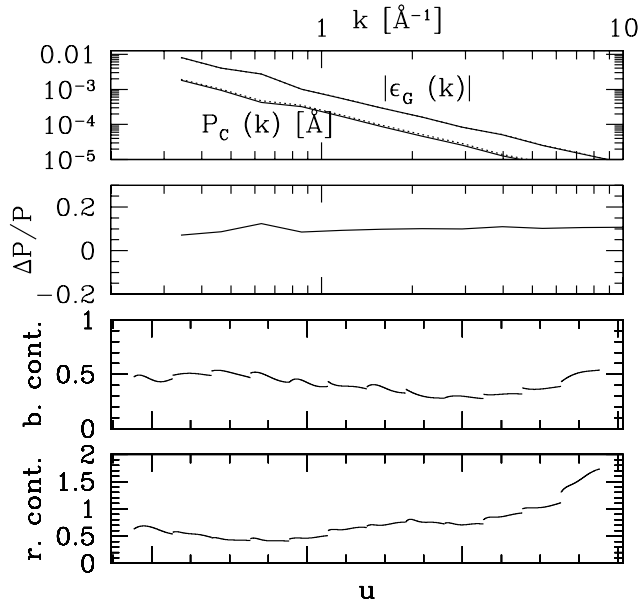

FIG. 8.—Bottom two panels show continuum fits taken from the red (*bottom panel*) and blue (*second panel from bottom*) sides of Lyα emission from the spectrum of a quasar at redshift $z = 3$. The top panel shows the power spectra of the red and blue continua (eq. [21b]) with solid and dotted lines, respectively. They are very close to each other; the panel below it shows their fractional difference $[P_C(\text{blue}) - P_C(\text{red})]/P_C(\text{red})$. Also shown in the top panel is the quantity $\epsilon_G(k)$ (eq. [22]), with solid and dotted lines denoting its values on the red and blue sides. They differ by only a few percent.

blue continuum power systematically higher than the red one. The results here, though drawn from admittedly a very small sample, are quite interesting for several reasons:

1. The excess of the blue continuum power spectrum over the red one is consistent with the hypothesis that some of the fluctuations in the forest have been wrongly assigned to the continuum during the continuum fitting process on the blue side. In other words, the true blue continuum power spectrum should be lower than the dotted line in the top panel of Figure 8. An upper bound on the true blue continuum power spectrum is already very useful. One can use it to quantify how much, and on what scales, one should worry about spurious continuum power introduced into estimation of the transmission power. One can compare Figure 8 with the theoretical expectation in Figure 1 and see that the spurious power must be negligible for $k \gtrsim 0.3$ Å^{-1}. This explains why the determination of the transmission power spectrum from both the continuum-fitted data and the trend-removed data is very accurate in examples like Figure 5, as long as one considers the *normalized* power. Unfortunately, the pieces of continuum we examine are not long enough to yield useful information on larger scales or smaller k values. If one takes a crude extrapolation, the continuum power spectrum (or more accurately its upper bound) might become nonnegligible compared to the transmission power spectrum at $k \sim 0.1$ Å^{-1}. However, one must keep in mind that the theoretical transmission power spectrum in Figure 1 is likely underestimated at small k values because the simulation lacks large-scale power. Nonetheless, there should be a genuine flattening of the transmission power spectrum at large scales. In any case, the first point to bear in mind is that an upper bound on the continuum power spectrum is useful as a conservative esti-

mate of the possible spurious power.[14]

2. Further, one can test the hypothesis that the excess in blue continuum power is due to contamination from the forest; if this is true, one expects the red and blue continuum power spectra to converge, as one goes to lower redshift quasars, because presumably the blue continuum power spectrum should be less affected by the forest at lower redshifts. Even if their difference does not converge to zero (as suggested by the larger number of broad emission lines on the red side) but to some small but finite value, this is still a useful exercise because it gives us an idea of how different the red and blue continuum power spectra can be. If we can determine the blue continuum power spectrum to an accuracy of, say, 10% and use this to correct for the transmission power spectrum on large scales, this is already a significant improvement over not correcting for the large-scale power or simply throwing away the information on large scales altogether. For instance, if the blue continuum power does become comparable with the transmission power at $k \sim 0.1$ Å^{-1}, not subtracting off the spurious power would result in a fractional error of 100%, while subtracting off an approximate blue continuum power accurate to 10% reduces the error by an order of magnitude.

Obviously, more testing using observed data is warranted, particularly on the estimation of red and blue continuum power as a function of redshift. This will be carried out in a separate paper. One natural question that might occur to the reader is whether a universal continuum power spectrum actually exists, given the large observed variations in the continuum from one quasar to another. It suffices to note that given an ensemble, the averaged power spectrum is always a well-defined quantity. The tricky part is to make sure the ensemble from which one estimates the continuum power spectrum has the same averaged continuum power spectrum as the ensemble of continua in the forest regions of interest. As a simple example, one might want to make sure that the same proportions of radio-loud quasars are included in both ensembles. This is probably desirable if one uses the working hypothesis that low-redshift blue continuum power is similar to high-redshift blue continuum power, as suggested above. Alternatively, if the hypothesis that blue and red continuum power spectra resemble each other irrespective of redshift turns out to be a reasonable one, the simplest way to make sure one has the right ensemble is to use both sides of Lyα for any give quasar: use the blue side for its forest and the red side for its continuum.

With all of the above caveats in mind, let us illustrate the technique of power correction with two idealized examples, where it is assumed that the right ensemble of continua is in hand.

In Figure 9 we show in the bottom panel a simulated spectrum with a somewhat unusual continuum (*middle panel*) with a fair amount of fluctuations. We generate a set of 10 different continua and impose each on our underlying cosmic absorption to obtain a set of 10 different simulated spectra (only one of which is shown in the figure). We compute the power spectrum using $P_2(k)$ as in equation (20). The resulting fractional error from the true transmission power spectrum is shown as a solid line in the top panel. There is clearly a lot of spurious power on large scales as a result of the imperfectly estimated mean count, which reflects the wild fluctuations in the continuum. We then apply the power spectrum corrections: the dotted line shows $\hat{P}_3(k)$ from equation (23), while the dashed line shows $\hat{P}_4(k)$ from equation (24). One can see that subtracting the continuum power spectrum $P_C(k)$ alone removes most of the spurious power.

To make the example realistic, we have multiplied the continua in the forest region by a power law that goes like $(u^\alpha)^{0.96}$ (i.e., the "blue" continuum) and similarly multiplied the continua from which we actually estimate the continuum power by a power law of $(u^\alpha)^{-0.01}$ (i.e., the hypothetical "red" continuum). This is meant to mimic a possible turnover of the quasar continuum around Lyα (see, e.g., Zheng et al. 1998 for evidence of a turnover around Lyβ). We have in mind a situation in which the continuum power spectrum P_C is estimated from the red side of Lyα. Clearly, the fact that the mean trends on the blue and red are different does not present an obstacle.

In Figure 10 we show a similar version of the above, but with much noisier data and poorer resolution, and a mean

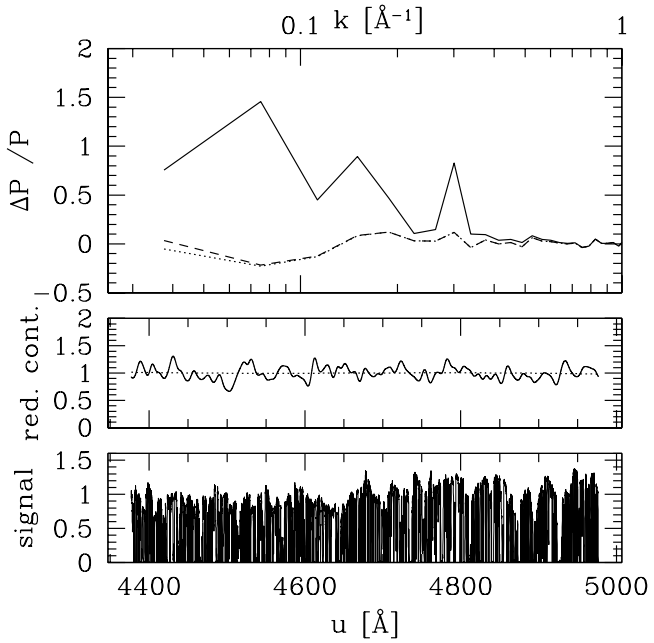

FIG. 9.—Demonstration of power spectrum corrections on large scales. The theoretical spectrum from Fig. 1 is multiplied by a set of 10 different continua (one of which is shown in the middle panel), convolved with a Gaussian of 0.125 Å FWHM and with a small amount of noise added (similar to Fig. 2). The resulting 10 simulated spectra (one of which is shown in the bottom panel) are analyzed, and the resulting power spectrum fractional error is shown in the top panel. The solid line shows error in the power spectrum estimate with no corrections applied (eq. [12] or eq. [20]), the dotted line shows the error using the estimator $\hat{P}_4$ in eq. (24), and the dashed line shows the error using $\hat{P}_3$ from eq. (23).

[14] One should be aware of a possible pitfall of the above argument, however. It is not impossible that the opposite can happen, that one underfits the blue continuum and ends up underestimating the blue continuum power. This is probably not the case here, where the data from which the blue continuum is estimated have high S/N and resolution (similar to the simulated spectrum in Fig. 4, where it can be seen that the continuum fit tends to have features that follow the forest). Underfitting the blue continuum is more likely for low-resolution data, although even there the situation is not clear: underfitting would result in underestimation of the continuum power on small scales, but not necessarily on large scales. Obviously, more tests are needed.

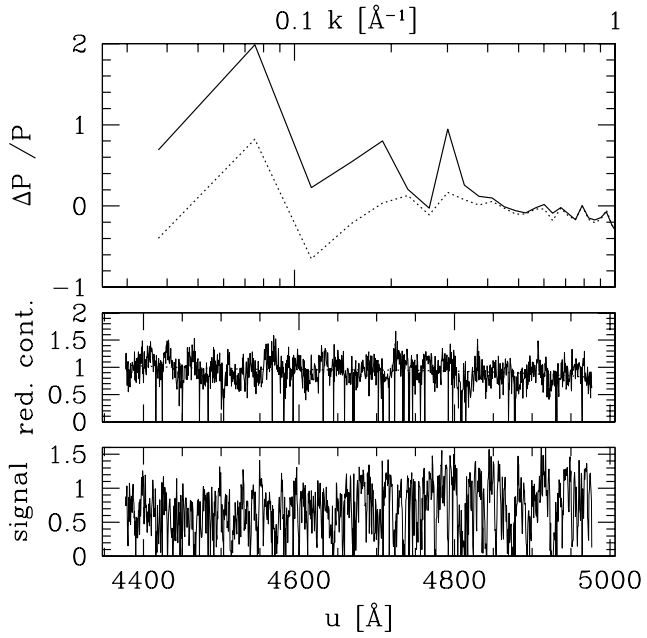

FIG. 10.—Similar to Fig. 9, except that the noise level and resolution resemble instead those of Fig. 6. Note how the noise added makes the continuum (*middle panel*), from which we estimate the continuum power spectrum, quite noisy as well. In the top panel, the solid line shows the fractional error in the power spectrum estimate with no corrections applied (eq. [12] or eq. [20]), and the dotted line shows the error using the estimator $\hat{P}_4$ from eq. (24).

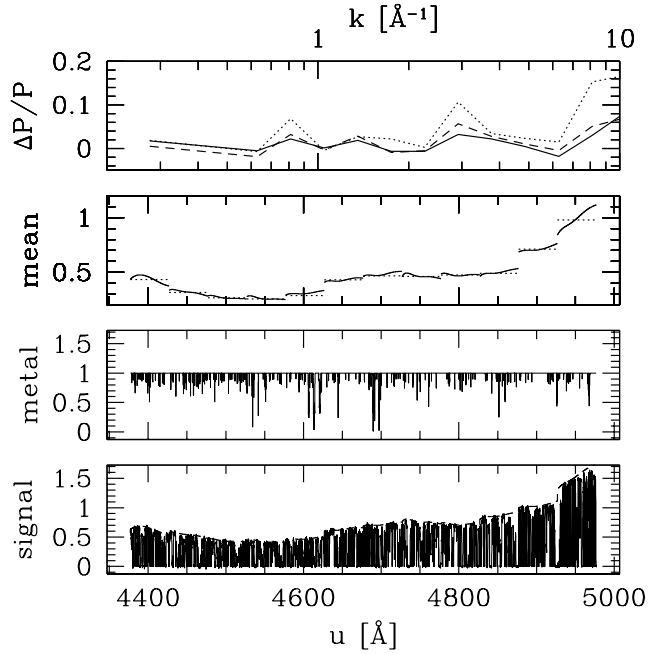

FIG. 11.—Bottom panel shows a simulated spectrum with resolution and S/N similar to Fig. 3, except that metal absorption lines shown in the second panel have been added. The y-axis of the second panel is $e^{-\tau}$, where τ is the optical depth due to metal absorption. The second panel from the top shows the true mean transmission and the recovered mean transmission assuming a flat trend for each order (*dotted line*). The dotted line in the top panel is $(\hat{P}_2 - P)/P$ for the case in which no attempt is made to cut out the metal lines; the dashed (solid) line is the same fractional error for the normalized power spectrum for the case in which all metal lines with $\tau > 1$ ($\tau > 0.4$) are discarded and the corresponding gaps are filled in via interpolation.

power law of $(u^z)^{1.5}$ and a mean of $(u^z)^{-0.9}$ have been imposed on the continua on the blue and red sides, respectively. The same technique works here as well.

One last point we should make: when the quantities $P_C(k)$ and $\epsilon_G(k)$ are estimated from some ensemble of continua, they in general receive shot noise contributions. We have ignored shot noise here, assuming that the scales where power correction is most interesting are sufficiently large that shot noise is unimportant. This should be checked on a case-by-case basis.

4.2.3. *Gaps and Metal Absorption Lines*

There are at least two other possible sources of systematic errors in addition to that due to continuum fitting. Gaps are quite common in observed spectra as a result of defects in the CCD, incomplete spectral coverage, or cosmic-ray hits. Fortunately, since they are at known locations, we can either consider only those parts of the spectrum that are between the gaps (for instance, when the gaps are large) or interpolate to fill in the gaps (for instance, when the gaps are small). The latter is what we have implicitly done in all of the tests mentioned in § 4.2.1, where 3% of the pixels are assumed to be discarded because of cosmic-ray hits. The hits are typically 1 to a few pixels wide, and we simply fill them in by linearly interpolating the counts from neighboring pixels. Clearly, we can recover the power spectrum quite well in spite of the need to interpolate.

A more challenging problem is possible systematics due to the presence of metal absorption lines. Shown in the bottom panel of Figure 11 is a simulated spectrum with resolution and S/N very similar to that of Figure 3, except that metal absorption lines as shown in the panel above have been added on top of the Lyα forest. This list of lines is taken from a quasar spectrum redward of Lyα (HS 1103+6416, $z = 2.19$). The mean transmission is estimated by assuming a flat trend for each echelle order as before. What is interesting is the dotted line in the top panel, demonstrating the creation of spurious power by the metal lines. The dashed (solid) line shows fractional error in the power spectrum estimate if all metal lines with $\tau > 0.4$ ($\tau > 1$) are assumed to be "detected" and therefore cut out and treated as gaps as before (i.e., interpolated across). Such a procedure seems to eliminate much of the spurious power. In practice, sufficiently strong metal lines should be identifiable by their narrow widths.

Figure 12 shows that metal absorption lines in data with lower resolution and poorer S/N have a relatively small effect on the accuracy of the power spectrum estimation. It would obviously be desirable to repeat the above exercises with a larger sample of metal lines to obtain a better estimate of the average level of metal contamination.

4.3. *Random Errors: Shot Noise Bias, Variance, and Minimum Variance Weighting*

Random errors arise first of all from (cosmic) sampling fluctuations and secondly from electron/photon counting, which can be traced to fluctuations in the intrinsic quasar counts, the sky counts, and the readout (see § 3.1). We will summarily refer to the latter as shot noise. Shot noise affects two aspects of power spectrum estimation.

First, shot noise introduces a bias that has to be subtracted off. This is the term $b(k)$ in equation (14b). We will give here a more general expression for $b(k)$ suitable for different weightings ($w^{\alpha\beta}$). As we have demonstrated in Figure 7, shot noise bias subtraction can be important for low-S/N data. We will return to this point in § 5.

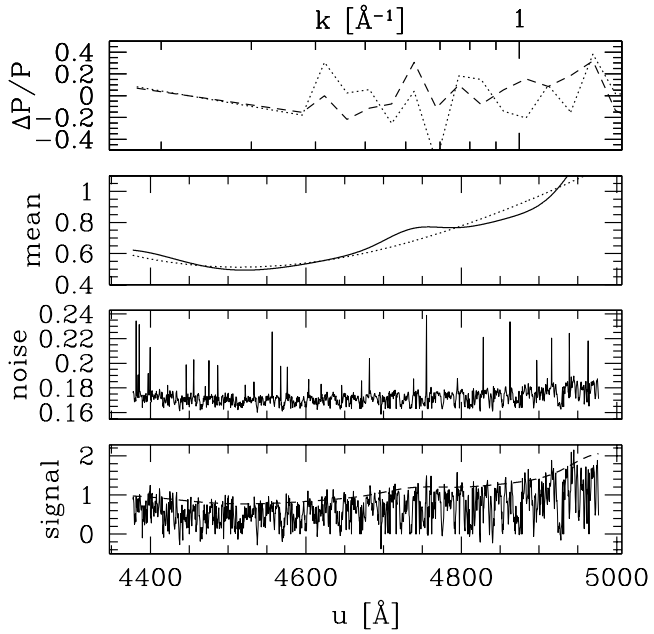

FIG. 12.—Power spectrum measurements from low-resolution and noisy quasar spectra (similar to Fig. 6), with (*dotted line*) and without (*dashed line*) metal contamination.

Second, shot noise, together with cosmic fluctuations, determines the variance of the power spectrum estimate. We will give the expression for the variance in this section and then address the question of how to best combine data with different levels of S/N to minimize the variance.

The power spectrum estimator we will focus on is given in equation (12). It is assumed that trend removal as explained in § 4.2.1 has been performed. We ignore uncertainties due to the unknown continuum in this section. Here we do not limit ourselves to the choice of uniform weighting (eq. [14]) as we have done so far. Let us rewrite $w^{\alpha\beta}(k)$ in equation (12) as

$$w^{\alpha\beta}(k) = \bar{w}^{\alpha\beta}(k) R^{\alpha\beta}(k) , \tag{25}$$

where $R^{\alpha\beta}(k)$ is given in equation (14) and is an average of the Fourier basis over some bin centered at k, with width Δk.

It can be shown that the variance of such a bin-averaged power estimate is given by (Appendix B)

$$V(k) \equiv \langle [\hat{P}_2(k) - P(k)]^2 \rangle = \frac{\mathcal{N}^3}{\mathcal{L}^2} \sum_\alpha [w^{\alpha\alpha}(k)]^2 E^\alpha(k) , \tag{26a}$$

$$E^\alpha(k) \equiv \frac{2}{n_{\bar{k}}} \left[P(k) + \frac{\mathcal{L}}{\mathcal{N}} \frac{q^\alpha \bar{N}_Q^\alpha + V_B^\alpha}{(\bar{N}_Q^\alpha)^2} \right]^2 + \frac{1}{\mathcal{L}} \langle T \rangle_{kk} + 4 \langle B \rangle_{kk} \frac{q^\alpha}{\bar{N}_Q^\alpha} \frac{1}{\mathcal{N}} + 2 \langle P \rangle_{kk} \left(\frac{q^\alpha}{\bar{N}_Q^\alpha} \right)^2 \frac{\mathcal{L}}{\mathcal{N}^2} + 4P(k) \frac{q'^\alpha}{(\bar{N}_Q^\alpha)^2} \frac{\mathcal{L}}{\mathcal{N}} + \frac{\mathcal{L}^2}{\mathcal{N}^3} \frac{q''^\alpha}{(\bar{N}_Q^\alpha)^3} , \tag{26b}$$

assuming that the k of interest satisfies $k \gg 1/\mathcal{L}$ and that the width of the bin Δk also satisfies $\Delta k \gg 1/\mathcal{L}$, where $\mathcal{L}$ is the length of the spectrum. This is sometimes referred to as the classical limit in the case of galaxy power spectrum measurement (Feldman, Kaiser, & Peacock 1994; Hamilton 1997a). We will not consider larger scales here because measurements on such scales are likely dominated by systematic rather than random errors.

The symbol n_k denotes the number of modes within the k bin of interest; $\mathcal{N}$ is the number of pixels in the length $\mathcal{L}$; and q^α, $\bar{N}_Q^\alpha$, and V_B^α are as defined in equation (14b). The quantities q'^α and q''^α are analogous to q^α:

$$q'^\alpha \equiv \sum_i (W^{i\alpha})^3 g_{\rm ps}^{i\alpha} g_b^\alpha , \quad q''^\alpha \equiv \sum_i (W^{i\alpha})^4 g_{\rm ps}^{i\alpha} g_b^\alpha . \tag{27}$$

The symbols $\langle T \rangle_{kk}$, $\langle B \rangle_{kk}$, and $\langle P \rangle_{kk}$ represent the shell-averaged trispectrum, bispectrum, and power spectrum, respectively ($\langle\rangle$ here is to be distinguished from ensemble average discussed in § 2):

$$\langle T \rangle_{k_1 k_2} \equiv \frac{1}{n_{k_1} n_{k_2}} \sum_{k_1'} \sum_{k_2'} T(-k_1', k_1', -k_2', k_2') , \tag{28a}$$

$$\langle B \rangle_{k_1 k_2} \equiv \frac{1}{n_{k_1} n_{k_2}} \sum_{k_1'} \sum_{k_2'} B(-k_1' - k_2', k_1', k_2') , \tag{28b}$$

$$\langle P \rangle_{k_1 k_2} \equiv \frac{1}{n_{k_1} n_{k_2}} \sum_{k_1'} \sum_{k_2'} P(k_1' + k_2') , \tag{28c}$$

where the sum over k_1' extends over modes within the bin centered at k_1, and similarly for k_2'. The trispectrum T and bispectrum B are Fourier transforms of the four- and three-point correlation functions, defined in an analogous manner to equation (2).

The variance as given in equation (26) contains contributions from both cosmic fluctuations and discrete fluctuations (see § 2). The terms such as $P(k)^2$ and $\langle T \rangle_{kk}$ arise because of intrinsic fluctuations of the cosmic signal from one part of the universe to another; these terms are present even if one has data with arbitrarily high S/N. The terms containing $\bar{N}_Q^\alpha$ arise because of discrete fluctuations; these we will loosely refer to as shot noise.

As we have emphasized in §§ 3.1 and 4.1, the shot noise contributions to the random error are not exactly Poisson distributed. The shot noise contributions (ignoring cosmic sample fluctuations) in equation (26b) would all be simply $1/\bar{N}_Q^\alpha$ if $\hat{N}_Q^\alpha$ were strictly a Poisson variable. We have additional fluctuations in $\hat{N}_Q$ due to the background (sky and readout) and also due to nonunity weights used in reducing the data (eq. [3]; see also end of § 3.1).

Given equation (26), it is simple to derive a weighting $\bar{w}^{\alpha\beta}(k)$ that minimizes the variance $V(k)$, subject to the constraint that the effective window (G as defined in eq. [15]) is properly normalized. This is most simply derived by minimizing the following Lagrangian:

$$L(k) = V(k) - \lambda \left[\int G\left(k, k'\right) \frac{dk'}{2\pi} - 1 \right] , \tag{29}$$

where λ is a Lagrange multiplier. Differentiating the above with respect to $\bar{w}^{\alpha\beta}(k)$ and setting the result to zero, we obtain

$$\bar{w}^{\alpha\beta}(k) = \frac{[E^\alpha(k) E^\beta(k)]^{-1/2}}{M(k)} , \quad M(k) \equiv \sum_\mu [E^\mu(k)]^{-1} \frac{\mathcal{N}}{\mathcal{L}} . \tag{30}$$

The corresponding shot noise subtraction, instead of equation (14b), would then be

$$b(k) = \sum_\alpha \bar{w}^{\alpha\alpha}(k) \frac{q^\alpha \bar{N}_Q^\alpha + V_B^\alpha}{(\bar{N}_Q^\alpha)^2} , \tag{31}$$

where q^α, N_Q^α, and V_B^α are as defined in equation (14b).

In summary, the minimum variance estimator of the power spectrum is

$$\hat{P}_5(k) = \frac{\sum_{\alpha,\beta} R^{\alpha\beta}(k)[E^{\alpha}(k)^{1/2}\hat{\delta}_f^{\alpha}][E^{\beta}(k)^{1/2}\hat{\delta}_f^{\beta}]}{M(k)} - b(k)\,, \quad (32)$$

where $b(k)$ is given by equation (31), $E^{\alpha}(k)$ and $M(k)$ are given in equations (26) and (30), and $R^{\alpha\beta}(k)$ is as in equation (14a).

The minimum variance estimator can be understood simply as follows: $\hat{\delta}_f^{\alpha}$ at each pixel is weighed by $1/[E^{\alpha}(k)]^{1/2}$ before the array is Fourier transformed, squared, and grouped to form band power estimates. Note that the above estimator reduces to the one with uniform weighting (eq. [14]) if $E^{\alpha}(k)$ is independent of α, e.g., when sample/cosmic variance is significantly larger than shot noise $[P(k) \gg (\mathscr{L}/\mathscr{N})(q^{\alpha}\bar{N}_Q^{\alpha} + V_B^{\alpha})/(\bar{N}_Q^{\alpha})^2]$. It is important to note that the weighting as a function of α is determined by $\bar{N}_Q^{\alpha}$ rather than, say, $\hat{N}_Q^{\alpha}$. Downweighing pixels with a lot of absorption (hence relatively low $\hat{N}_Q^{\alpha}$) would be the wrong thing to do, since the fluctuation in absorption is the signal that we are after. The proper procedure is to downweigh pixels with an overall lower mean count $\bar{N}_Q^{\alpha}$.

Unfortunately, the minimum variance weighting given above is difficult to implement because one needs to specify simultaneously P, B, and T, in addition to the level of shot noise. A common simplification is to use the Gaussian approximation in which $E^{\alpha}(k)$ is approximated as

$$E^{\alpha}(k) \sim \frac{2}{n_{\bar{k}}}\left[P(k) + \frac{\mathscr{L}}{\mathscr{N}}\frac{q^{\alpha}\bar{N}_Q^{\alpha} + V_B^{\alpha}}{(\bar{N}_Q^{\alpha})^2}\right]^2 \quad (33)$$

(see, e.g., Hamilton 1997a). Note that in addition to ignoring the bispectrum and trispectrum terms, the above also ignores certain power spectrum terms that are mixed with shot noise—the last three terms in equation (26)—which is equivalent to assuming that either the shot noise or the correlation is sufficiently weak. With the above approximation, one can start with some initial P, use the minimum variance weighting scheme to get a first measurement of P, and iterate subsequently (Bond, Jaffe, & Knox 1998). Analogous (Gaussian) power spectrum estimators for galaxy surveys and microwave background experiments have been widely discussed in the literature (e.g., Feldman et al. 1994; Vogeley & Szalay 1996; Tegmark, Taylor, & Heavens 1997; Hamilton 1997a; Tegmark et al. 1998; Bond et al. 1998; Seljak 1998).

We will not attempt to address here the important question of how significant the non-Gaussian contributions are. A proper treatment will involve the analysis of a large number of simulations or a large data set, which we hope to present in a future paper. It suffices to say that the very nonlinear mapping from the density field to $e^{-\tau}$ will likely introduce some degree of non-Gaussianity, even if the initial density field is Gaussian.

The use of observed data to study this issue is particularly interesting and deserves some more comments. In principle, since different QSO sight lines typically sample independent regions of the universe, one can estimate the variance of the transmission power spectrum and hence infer the importance of the non-Gaussian contributions, using directly the fluctuations in power spectrum estimates from one sight line to another. However, one should keep in mind that shot noise also contributes to the variance. Since different lines of sight generally have different S/Ns, the sight line–to–sight line fluctuations in power spectrum estimates should be interpreted with care. In a data set of several quasars, it is possible that the quasar-to-quasar fluctuations are dominated by a few with low S/N, and their mean square would give an overestimate of the true power spectrum variance.

We show in Figure 13 an example in which the data consist of six high-quality spectra (similar to Fig. 3) and six others with S/Ns about 20 times smaller. The bottom panel shows the power spectrum estimated with uniform weighting (eq. [14]), while the top panel represents the power spectrum estimated with minimum variance weighting using the Gaussian approximation. The (1 σ) error bars are theoretical and are estimated using equations (26a) and (33). This illustrates how our weighting scheme can reduce the error bars at high values of k where shot noise is important. Lacking information on the non-Gaussian nature of the power spectrum variance, we advocate the Gaussian weighting scheme (eq. [33]) as a rational way to combine data with different levels of S/N to reduce the variance, even though it does not necessarily achieve minimum variance.

In combining the different spectra with different S/Ns, we have weighed the power spectrum estimate of each line of sight by its inverse variance, which is an obvious generalization of the minimum variance weighting introduced above. For instance, supposing we have two separate lines of sight A and B, we could combine the two power spectrum estimates $\hat{P}_2^A$ and $\hat{P}_2^B$ in the following way, assuming that the two lines of sight are independent:

$$\hat{P}_2(k) = \frac{[\hat{P}_2^A(k)/V_A(k)] + [\hat{P}_2^B(k)/V_B(k)]}{[1/V_A(k)] + [1/V_B(k)]}\,, \quad (34)$$

where $V_A(k)$ and $V_B(k)$ are estimated with the same P but could have different levels of shot noise. The noisier quasar spectrum is naturally downweighed.

Lastly, we should emphasize that the above discussion does not address the issue of cross variance between power

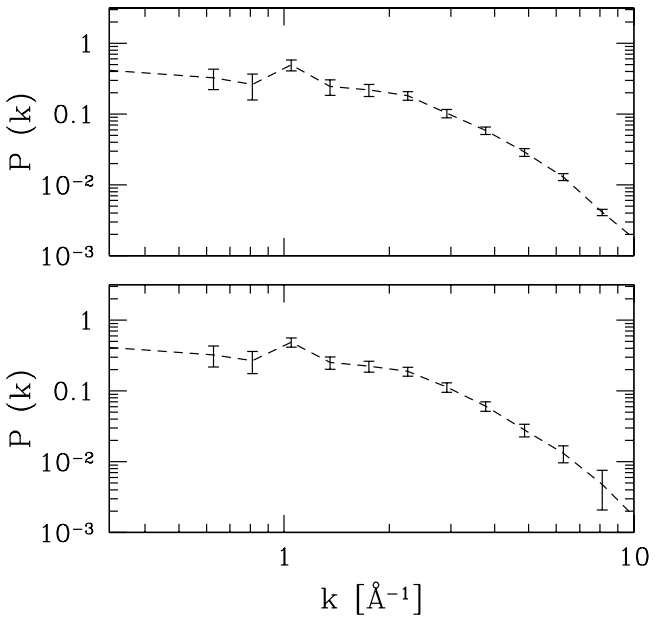

FIG. 13.—Power spectrum estimation using uniform weighting (*bottom panel*; eq. [14]) vs. minimum variance weighting (*top panel*; eq. [30]). The simulated QSO spectrum consists of 12 segments, half of which have comparable S/N to Fig. 3 and half of which have ~20 times lower S/N.

spectrum estimates at two different wave bands, which is introduced by the non-Gaussian terms (Meiksin & White 1999; Scoccimarro, Zaldarriaga, & Hui 1999). Hamilton (2000) introduced a scheme that simultaneously diagonalizes and minimizes the covariance. However, it makes specific assumptions about the form of the trispectrum and bispectrum, the validity of which for the forest remains to be checked.

5. DISCUSSION

Our recipe for measuring the transmission power spectrum is summarized here.

1. Given an array of reduced quasar counts $\hat{N}_Q^\alpha$, identified metal lines should be removed, especially the strong ones ($\tau > 1$). Small gaps in the spectrum (e.g., due to cosmic-ray hits removal) can be (linearly) interpolated over, while large gaps should be avoided (§ 4.2.3).

2. The mean quasar counts ($\bar{N}_Q^\alpha$) are estimated using equation (18). The mean basis (the functional form of the mean trend) should be chosen to be as smooth as possible: underfitting is better than overfitting (see tests 1 and 2 of § 4.2.1). In practice, it appears that a flat mean suffices for short spectra (~ 50 Å), while polynomials up to the third order can be used for longer spectra (~ 500 Å). One can gain an idea of what a reasonable basis is using the red side of Lyα or low-redshift QSO spectra where the continuum can be seen more clearly.

3. The trend-removed and normalized fluctuation $\hat{\delta}_f^\alpha$ is defined according to equation (12b), and the power spectrum is estimated using the quadratic estimator in equation (12a). Different weightings are possible; the simplest that we recommend is given in equation (14). A more sophisticated weighting scheme that can reduce the random error is given by equations (30), (31), (32), and (33). If one is interested in the real-space correlation function instead, the recommended weighting is equation (13); this gives a smaller variance compared to other estimators commonly used in the literature. We emphasize that the shot noise bias [$b(k)$ in eq. (14b) or eq. (31)] should be subtracted correctly, especially for noisy data.

4. If better control over systematic errors on large scales ($\gtrsim 30$ Å) introduced by the unknown continuum is desired, the techniques outlined in § 4.2.2 can be used. The corresponding estimator is given in equation (23), which requires an estimate of the continuum power spectrum P_C (eq. [21b]) and an additional correction factor ϵ_G (eq. [22]). This procedure requires the identification of an appropriate set of continua (see discussions in § 4.2.2). Even if one is not interested in the power spectrum on large scales, we recommend this procedure as a consistency check that the spurious power introduced by the continuum is negligible on the scales of interest.

What implications does the above have for one's observing strategy? To discuss this question, we need to take a closer look at the issue of shot noise. The shot noise enters in two different places in the above discussion. First, it contributes to the variance (random error) of the power spectrum estimate (eq. [26]). Second, it appears as a bias in the power spectrum estimate that we have to subtract off [e.g., $b(k)$ in eqs. (12a) and (14)].

In the literature on power spectrum measurement, shot noise subtraction has been largely ignored (e.g., Croft et al. 1998; see McDonald et al. 2000 for an alternative approach where shot noise is simulated rather than subtracted). Let us estimate how important it is. The expression in the simplest case of uniform weighting is given in equation (14b) (see eq. [31] for more complicated weightings), which can be rewritten as

$$b(k) = \frac{\Delta u}{\mathcal{N}} \sum_\alpha \frac{\sum_i (W^{i\alpha})^2 (g_{\rm ps}^{i\alpha} g_b^\alpha \bar{N}_Q^\alpha + V_B^{i\alpha})}{(\bar{N}_Q^\alpha)^2} , \tag{35}$$

where Δu is the size of a pixel, $\mathcal{N}$ is the total number of pixels, and the rest of the symbols are as defined in § 3.1: i is the pixel label in the spatial direction and α in the spectral direction, $\bar{N}_Q^\alpha$ is the mean reduced quasar count, $V_B^{i\alpha}$ is the background variance, $W^{i\alpha}$ is a weighting, and $g_{\rm ps}^{i\alpha}$ and g_b^α are the point-spread function and blaze function, respectively (eqs. [6] and [5]). An important observation is that the numerator within the summation is closely related to the variance array, which is often given along with a spectrum (eq. [9]):[15]

$$\mathrm{var}(\alpha) = \sum_i (W^{i\alpha})^2 (\hat{N}_Q^{i\alpha} + V_B^{i\alpha}) . \tag{36}$$

The quantity $\hat{N}_Q^{i\alpha}$ is of course different from $g_{\rm ps}^{i\alpha} g_b^\alpha \bar{N}_Q^\alpha$, which we need to estimate the shot noise, but since we are in practice interested in an average over all pixels, it turns out that the following estimate of the shot noise is accurate to within a percent for all cases we have tested:

$$b(k) \sim \frac{\Delta u}{\mathcal{N}} \sum_\alpha \frac{\mathrm{var}(\alpha)}{(\bar{N}_Q^\alpha)^2} . \tag{37}$$

Without the above approximation, an exact estimate of the shot noise would require the knowledge of $\hat{N}_Q^\alpha$, $\sum_i (W^{i\alpha})^2 g_{\rm ps}^{i\alpha} g_b^\alpha$, and $\sum_i (W^{i\alpha})^2 V_B^{i\alpha}$.

Equation (37) provides a useful means of estimating the shot noise (see Appendix A on shot noise estimation under more complicated circumstances, i.e., with nontrivial rebinning or weighting). One can simplify further by making a crude approximation in relating $b(k)$ to the typical S/N of the data (which is often quoted at the continuum) through

$$b(k) \sim \left(\frac{\Delta u}{\bar{f}}\right) (\mathrm{N/S})^2 , \tag{38}$$

which can be justified if one ignores the part of the variance due to the sky and readout. We find that this simple rule of thumb generally provides an underestimate of the shot noise (particularly at low S/N where the background counts become important) but is accurate to within about a factor of 2.

Figure 14 summarizes some useful information for devising an observing strategy, based on our estimate of the shot noise in equation (38) above. The solid line shows the mean observed transmission power spectrum at $z = 3$ taken from McDonald et al. (2000).[16] The two horizontal dotted lines show the level of shot noise expected for the two extremes of the kinds of observations we are likely to encounter: the bottom corresponds to very high S/N observations with HIRES quality resolution (e.g., Hu et al. 1995; Kirkman & Tytler 1997; Rauch et al. 1997), while the dotted line on the top corresponds to low-S/N observations

[15] See paragraph after eq. (9) and § 2 on the distinction between $\tilde{N}_Q^{i\alpha}$ and $\hat{N}_Q^{i\alpha}$.

[16] We divide the unnormalized power spectrum (eq. [1]) of McDonald et al. (2000) by the square of their measured mean transmission to obtain the normalized power spectrum given in Fig. 14. See tests 4 and 5 in § 4.2.1 on the bias of the unnormalized power spectrum.

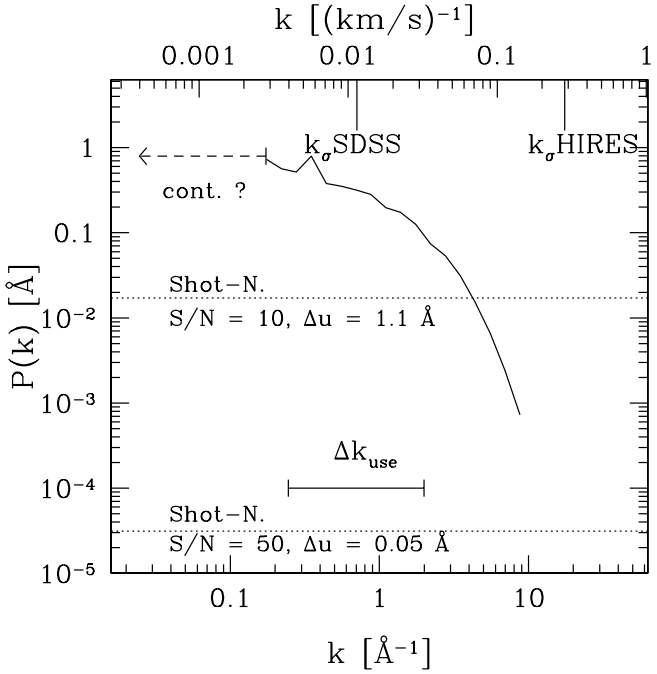

FIG. 14.—Solid line is the mean observed transmission power spectrum at $z = 3$ from McDonald et al. (2000). Its continuation as a dashed line at small k values indicates the scales at which continuum might introduce significant uncertainties, and so the power is not shown. The two dotted lines at top and bottom show the level of shot noise at two extremes. The top one resembles the quality expected for SDSS spectra, while the bottom resembles HIRES Keck spectra of bright quasars. The interval of $\Delta k_{\rm use}$ at the bottom indicates the range of scales that are currently used to recover the linear mass power spectrum. The long tick marks at the top indicate the resolution of SDSS and typical Keck HIRES spectra. Different units of k are related by $[k/(\mathrm{km\ s^{-1}})^{-1}] = [(1+z)/4](k/\mathrm{\AA}^{-1})/61.67 = \{(1+z)/[\Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda]^{1/2}\}(k/h\ \mathrm{Mpc}^{-1})/100$.

expected for a large number of quasars in the SDSS. We emphasize that the shot noise level does not depend on the resolution per se but on the pixel size for a given S/N. SDSS is expected to produce $\sim$1000 QSO spectra at S/N = 20 pixel^{-1} (QSOs at $z > 2.7$, where the redshift limit is determined by the blue limit of the spectrograph, 3800 Å), $\sim$10,000 at S/N = 15, and $\sim$30,000 at S/N = 7, corresponding to i' magnitude cuts at 18, 19, and 20, respectively (Fukugita et al. 1996; Fan 1999). The pixel size of SDSS is quite uniform in velocity 70 km s^{-1}, which is equivalent to 1.13 Å at 4864 Å (Lyα at $z = 3$).

Clearly, the importance of shot noise depends on the scales at which one is interested in measuring the power spectrum. A few interesting scales are shown in Figure 14. First, instrumental resolution imposes a high k limit beyond which one cannot reliably measure the transmission power spectrum. The resolution window is often characterized by a Gaussian with a given FWHM. The effect of such a resolution window on the power spectrum can be represented by $P(k) \rightarrow P(k)e^{-k^2/k_\sigma^2}$, where $k_\sigma = (8\ln 2)^{1/2}/\mathrm{FWHM} \sim 2.355/\mathrm{FWHM}$. Two representative k_σ values are shown as long tick marks at the top. Note that even at $k = k_\sigma$, the resolution window reduces the power by 63% and so has a nonnegligible effect. The Sloan FWHM is about 2.1 pixels, i.e., 147 km s^{-1}, or 2.4 Å at 4864 Å.

On the other hand, the range of scales that is currently being used to infer the *mass* power spectrum is indicated by the interval near the bottom $\Delta k_{\rm use}$. The high k limit is set by the scales at which the shape of the power spectrum is preserved in the transformation from mass to transmission (i.e., linear biasing; e.g., Croft et al. 1998). We can see that for high-quality Keck spectra, information from a whole decade of measurable scales is unused for the recovery of the mass power spectrum; it would be very useful to push the current analysis techniques to these scales, since power on these scales is of particular interest in constraining, e.g., neutrino properties (Hui et al. 1997; Croft, Hu, & Davé 1999). Such an effort would require disentangling the effects of peculiar velocities and thermal broadening, however. At the other end, the low-k limit of currently usable scales is set by the scales at which the continuum fluctuates. This is indicated by the dashed line at the top, where the transmission power spectrum is unknown.

From the above discussion, we can distill a few tips for observing/analysis.

1. To ensure that shot noise is subdominant, one might want to achieve S/N $\gtrsim [10\Delta u/\bar{f}P(k_{\rm int})]^{1/2}$, where $k_{\rm int}$ is the scale of interest and $\bar{f}$ is the mean transmission. The factor of 10 is somewhat arbitrary; this will ensure that the shot noise contribution to the power spectrum variance is no more than about 20% (under the Gaussian approximation; see eq. [33]), or the 1 σ error bar on the power spectrum would only be increased by 10% as a result of the contribution from shot noise. An important question is what $k_{\rm int}$ should be, which depends on at what scales one can usefully extract cosmologically interesting information. Current literature mainly focused on $k_{\rm int} \lesssim 2$ Å^{-1}, where $P \sim 0.06$ Å, therefore S/N $\sim 20(\Delta u/1\ \mathrm{\AA})^{1/2}$ would be sufficient. Since P rises with scale, shot noise would be even less important at smaller k values. Note that with very small Δu such as $\sim$0.05 Å, an S/N as low as 4–5 is acceptable. To give some examples, an S/N of 8 per 0.05 Å can be achieved with 1 hr of exposure using Keck/HIRES for a $V = 19$ quasar; on the other hand, an S/N of about 15 per 1.1 Å is expected with just slightly under an hour of exposure using the SDSS spectrograph for an $i' = 19$ quasar.

2. A corollary of focusing on only $k \lesssim 2$ Å^{-1} is that observations with $k_\sigma \gtrsim 3 \times 2$ Å^{-1} or a resolution FWHM of 0.3 Å or $R \sim 16{,}000$ at $z = 3$ are acceptable. The factor of 3 above (i.e., in 3×2 Å^{-1}) is somewhat arbitrary: it ensures that at $k = 2$ Å^{-1}, the resolution window does not reduce the power by more than 10%. If the resolution window is known accurately, or if one is willing to sacrifice information on the small scales close to $k \sim 2$ Å^{-1}, one could in principle consider lower resolutions. We would like to emphasize, however, that in principle the modes at $k > 2$ Å^{-1} could still contain very interesting cosmological information, even though the current attempts at recovering the mass power spectrum ignored them.

3. If shot noise is subdominant compared to the power spectrum, the only other limiting factor to the size of the random error is the total size of one's sample or the number of sight lines in it. Assuming that all sight lines have similar coverage with length $\mathscr{L}$, then the fractional error of a single k-mode (i.e., in a k bin of $2\pi/\mathscr{L}$) is given by $\delta P/P = C/(N_{\rm sight})^{1/2}$, where $N_{\rm sight}$ is the number of sight lines, assuming that they are independent, and $C = 1$ under the Gaussian approximation (eq. [33]) and a little larger than unity under more general circumstance (see, e.g., Meiksin & White 1999; Scoccimarro et al. 1999).

4. How should one distribute one's observing time among quasar targets to minimize the random error on the

transmission power spectrum? There are many possible versions of this problem. We will discuss two, giving an explicit solution for the first and only general expressions for the second. In the simplest case in which all the candidate quasar targets have similar magnitudes, given a finite amount of observing time, one can deduce the optimal total number of quasars one should target by minimizing

$$N_{\rm tot}^{-1}\left[P(k_{\rm int})+\frac{\Delta u}{\bar{f}}\frac{A}{t}\right]^2, \tag{39}$$

subject to $N_{\rm tot}\,t=T_{\rm tot}$, where $N_{\rm tot}$ is the total number of quasars targeted, $k_{\rm int}$ is the scale of interest, $T_{\rm tot}$ is the total amount of observing time one has, t is the amount of time one spends on each quasar, and $1/A$ is equal to $({\rm S/N})^2$ reached per unit exposure time. The above assumes equation (33) and that the sight lines are independent. The solution is easy to deduce: $N_{\rm tot}=[P(k_{\rm int})\bar{f}/\Delta u](T_{\rm tot}/A)$, or $t=A\Delta u/\bar{f}/P(k_{\rm int})$. A typical value for $1/A$ is $1/A\sim(1200\ {\rm hr\ \AA^{-1}})\times 10^{(19-{\rm mag})/2.5}({\rm aperture}/100\ {\rm m}^2)f_{\rm throughput}$, where $f_{\rm throughput}$ is about unity for Keck/HIRES and ~ 2.5 for the SDSS. Using again $k_{\rm int}\sim 2\ {\rm \AA^{-1}}$, for a 19th magnitude quasar, with an aperture of 6.25 $\rm m^2$ and assuming $\Delta u=1$ Å and $f_{\rm throughput}=3$, the exposure time is $t=14$ minutes only! The above prescription, however, only allows for just enough exposure time to reduce the shot noise to a level comparable to the cosmic/sampling variance [i.e., $P(k_{\rm int})\sim(\Delta u/\bar{f})({\rm N/S})^2$]; the sole aim is to maximize the number of quasars observed within a given length of time to beat down the sampling variance. The prescription would certainly be different if one has, for instance, a finite number of quasar targets, or if one has other purposes in mind, such as measuring the mean decrement, etc. (see earlier prescription for making shot noise subdominant, equivalent to multiplying t by about a factor of 10). A more general version of the above problem deals with a case in which the quasars span a range of magnitudes, i.e., A is no longer the same number for each quasar. A simple *Ansatz* is to assume $t=\alpha A\Delta u/\bar{f}/P(k_{\rm int})$, in other words, spending more time for fainter quasars because it takes longer to beat down the shot noise, except that we have a normalizing factor α that enforces the constraint of total observing time: $\alpha=T_{\rm tot}[\bar{f}P(k_{\rm int})/\Delta u]/\int_{A_{\rm min}}^{A}A'n(A')dA'$, where $n(A)dA$ is the number of quasars with A falling in the given range and $A_{\rm min}$ corresponds to the brightest quasar in one's sample. Then, we can determine how many quasars one should include, starting from the brightest one, or how faint one should go by minimizing $[\int_{A_{\rm min}}^{A}n(A')dA']^{-1}P(k_{\rm int})^2(1+1/\alpha)^2$ with respect to A.

The following is particularly relevant for SDSS or comparable observations:

1. In addition to contributing to the power spectrum variance, shot noise also contributes a bias that has to be subtracted off (see, e.g., Fig. 7). This is quite important for SDSS because, with greater than 10^4 sight lines, the survey has the capability of reducing the fractional error of the power spectrum to less than 1% per mode. Therefore, a bias of $\sim$3%–100%, depending on the scale of interest (as indicated by the top dotted line in Fig. 14), is not acceptable and should be subtracted off. We note that analyses so far in the literature (e.g., Croft et al. 1999; McDonald et al. 2000) focused on higher quality data where S/N $\sim$ 30, with Δu ranging from about 0.04 to 1 Å, and so according to equation (38) and Figure 14, the shot noise bias was about 1% of the power or smaller and therefore could be ignored, although a more careful check should be performed for some data sets with lower S/N.

2. The low resolution of SDSS spectra implies that it would be difficult to obtain useful information on scales $k\gg k_\sigma\sim 0.01\ ({\rm km\ s^{-1}})^{-1}$ or $0.7\ {\rm \AA^{-1}}$. On larger scales or smaller k values, two problems have to be reckoned with. For $k\sim 0.3k_\sigma-k_\sigma$, the resolution window suppresses the power by 10% or more; therefore, one needs to have an accurate measure of the resolution window to recover the true transmission power spectrum.[17] This can be achieved by using narrow metal lines or arc lines. There are relatively few sky lines in the relevant part of the spectrum.

3. For scales $k<0.004\ ({\rm km\ s^{-1}})^{-1}$ or $0.2\ {\rm \AA^{-1}}$, the effect of the continuum has to be properly taken into account, and the method of § 4.2.2 can be useful here. From Figure 14, it is clear that the range of scales accessible to SDSS would be quite limited unless a correction for continuum contamination is applied.

It is worth pointing out that while the above quoted numbers are all based on $z=3$, we do not expect them to change significantly for $z=2$ or $z=4$. This is in part because of the slow evolution of the transmission power spectrum: the growth of the mass spectrum with time is partially compensated by the lowering of the mean decrement (McDonald et al. 2000).

At least three issues remain to be explored in future work. As we have emphasized, the concept of correcting for continuum contamination in the transmission power on large scales as laid out in § 4.2.2 is an interesting one but requires more testing. An important check is the measurement of continuum power spectrum as a function of redshift on both sides of Lyα emission. Second, counts-in-cells analysis (i.e., measuring moments of the one-point probability distribution function), just like power spectrum analysis, requires shot noise subtraction, and since typically one considers cells at the limit of resolution, shot noise is likely nonnegligible except for high-S/N spectra. Counts-in-cells analysis provides a very interesting way to test the gravitational instability paradigm (Gaztañaga & Croft 1998; Hui 1998; Nusser & Haehnelt 2000) and should be done with care. Useful expressions will be presented elsewhere (L. Hui 2001, in preparation). Lastly, as is hopefully clear by now, the spirit of the methods presented in this paper is to avoid continuum fitting and replace it with trend removal. We have demonstrated that this is possible for measuring the transmission power spectrum. However, other quantities of interest such as the mean decrement require an estimate of the continuum, by definition. Furthermore, to interpret theoretically the transmission power spectrum in terms of the mass fluctuation, current methods require the measurement of the mean decrement to fix a free parameter in one's cosmological model, which is a combination of the ionizing background, the mean baryon density, and the mean temperature. Therefore, in a sense, the technique of trend removal only goes half-way in solving the problem of continuum fitting. Although we still recommend our method over continuum fitting because the transmission power spectrum is an unambiguous quantity that can and should

[17] We thank Rupert Croft for useful discussions on this point.

be determined as accurately as possible (not to mention the fact that continuum fitting is difficult with low S/N or low resolution, or at high z), there is clearly a need for an alternative method to bridge the gap between the measured transmission power and the theoretically interesting mass power. This will be explored in future publications (L. Hui & S. Burles 2001, in preparation; Zaldarriaga, Seljak, & Hui 2000).

We thank for useful discussions Len Cowie, David Kirkman, Patrick Petitjean, Michael Rauch, Wal Sargent, Don Schneider, David Weinberg, Don York, and the participants of the 1999 Haifa workshop on the intergalactic medium and large-scale structure. We also thank Nick Gnedin for supplying a simulation. Special thanks are due to Matias Zaldarriaga for pointing out the importance of aliasing and for many interesting discussions. This work was supported in part by the DOE and NASA grant NAG 5-7092 at Fermilab, as well as by NSF grant PHY-9513835. L. H. thanks the IAS for the Taplin Fellowship.

APPENDIX A

Our main aim here is to derive equation (15) for the estimator $\hat{P}_2(k)$, which is given by equations (12) and (14), with an eye toward generalization to $W^{i\beta}$ different from equation (6) and $W^{\alpha\beta} \neq \delta^{\alpha\beta}$. Derivations of results in § 4.3 on estimator variance and the minimum variance power spectrum estimator can be found in Appendix B. We will ignore the integral constraint and assume that $\bar{N}^\alpha$ is known to high accuracy.

We need first of all the correlation matrix $\langle \hat{\delta}_f^\alpha \hat{\delta}_f^\beta \rangle$. We will do it in two steps. First, let us derive $\langle \hat{\delta}_f^\alpha \hat{\delta}_f^\beta \rangle_D$. Rewriting $\hat{\delta}_f$ (eq. [12b]) as $(\hat{N}^\alpha - \tilde{N}^\alpha)/\bar{N}^\alpha + (\tilde{N}^\alpha - \bar{N}^\alpha)/\bar{N}^\alpha$, where $\tilde{N}^\alpha = \langle \hat{N}^\alpha \rangle_D$, it can be shown that

$$\langle \hat{\delta}_f^\alpha \hat{\delta}_f^\beta \rangle_D = \delta_f^\alpha \delta_f^\beta + \frac{1}{\bar{N}_Q^\alpha \bar{N}_Q^\beta} \langle (\hat{N}^\alpha - \tilde{N}^\alpha)(\hat{N}^\beta - \tilde{N}^\beta) \rangle_D$$

$$= \delta_f^\alpha \delta_f^\beta + \frac{1}{\bar{N}_Q^\alpha \bar{N}_Q^\beta} \sum_{\gamma i} W^{\alpha\gamma} W^{\beta\gamma} (W^{i\gamma})^2 (\langle \hat{N}_Q^{i\gamma} \rangle_D + V_B^{i\gamma}) , \tag{A1}$$

where δ_f^α is to be distinguished from $\hat{\delta}_f^\alpha$ in that it has only cosmic or sample fluctuations (eq. [2]), $\hat{N}_Q^{i\gamma}$ is a strictly Poisson variable with an average given by equation (5), and $V_B^{i\gamma}$ is the variance contributions from the sky and readout (eq. [8]).

Taking the cosmic mean of the above, we obtain the correlation matrix

$$\langle \hat{\delta}_f^\alpha \hat{\delta}_f^\beta \rangle = \xi(u^\alpha - u^\beta) + \frac{1}{\bar{N}_Q^\alpha \bar{N}_Q^\beta} \sum_\gamma W^{\alpha\gamma} W^{\beta\gamma} \sum_i [\langle (W^{i\gamma})^2 g_{\rm ps}^{i\gamma} g_b^\gamma \tilde{N}_Q^\gamma \rangle + \langle (W^{i\gamma})^2 \rangle V_B^{i\gamma}] . \tag{A2}$$

The second term on the right-hand side is the shot noise contribution that has to be subtracted off. Using the estimator in equation (12), with $w^{\alpha\beta}(k)$ given in equation (14), the correct shot noise subtraction is

$$b(k) = \frac{\mathscr{L}}{\mathscr{N}^2} \sum_{\alpha\beta\gamma} W^{\alpha\gamma} W^{\beta\gamma} \sum_i \frac{\langle (W^{i\gamma})^2 g_{\rm ps}^{i\gamma} g_b^\gamma \tilde{N}_Q^\gamma \rangle + \langle (W^{i\gamma})^2 \rangle V_B^{i\gamma}}{\bar{N}_Q^\alpha \bar{N}_Q^\beta} , \tag{A3}$$

where we have made use of the assumption that $W^{\alpha\gamma}$ is nonzero only for α and γ on very small separations and that the k of interest satisfies $k(u^\alpha - u^\gamma) \ll 1$ on such separations.

Note that for weightings such as the one given in equation (7), $W^{i\gamma}$ depends on $\tilde{N}_Q^\gamma$, which makes an estimation of the shot noise nontrivial. However, simplification results in two extreme limits: in the signal-dominated regime where $\tilde{N}_Q^{j\beta} \gg V_B^{j\beta}$, $W^{i\gamma}$ reduces to uniform weighting as in equation (6); and in the background-dominated regime where $\tilde{N}_Q^{j\gamma} \ll V_B^{j\gamma}$, $W^{i\gamma}$ reduces to $(1/g_b^\gamma)(g_{\rm ps}^{i\gamma}/V_B^{i\gamma})/[\sum_j (g_{\rm ps}^{i\gamma})^2/V_B^{i\gamma}]$, which is independent of $\tilde{N}_Q^\gamma$. In such cases, and for $W^{\alpha\gamma} = \delta^{\alpha\gamma}$, the above $b(k)$ reduces to the one given in equation (14b).

For $W^{\alpha\gamma} \neq \delta^{\alpha\gamma}$ (i.e., rebinning has been done; we continue to assume that $W^{i\gamma}$ is roughly independent of the signal), we can write $b(k)$ as

$$b(k) \sim \frac{\mathscr{L}}{\mathscr{N}^2} \sum_{\alpha\beta\gamma} \frac{W^{\alpha\gamma} W^{\beta\gamma}}{\bar{N}_Q^\alpha \bar{N}_Q^\beta} \sum_i (W^{i\gamma})^2 (g_{\rm ps}^{i\gamma} g_b^\gamma \langle \tilde{N}_Q^\gamma \rangle + V_B^{i\gamma}) , \tag{A4}$$

where the term under summation of i is simply the variance array of the pre-rebinned data (eq. [36]), and one can replace $\bar{N}_Q^\alpha \bar{N}_Q^\beta$ by $(\bar{N}_Q^\alpha)^2$, since $W^{\alpha\gamma} W^{\beta\gamma}$ is nonzero only if α and β are close together. Equation (37) is therefore replaced by the following if the data have been rebinned:

$$b(k) \sim \frac{\Delta u}{\mathscr{N}} \sum_{\alpha\beta} \frac{\mathrm{var}\ (\alpha, \beta)}{\bar{N}_Q^\alpha \bar{N}_Q^\beta} , \tag{A5}$$

where var(α, β) is the variance matrix of the rebinned data.

To complete our derivation, we need to show that the choice of $w^{\alpha\beta}(k)$ given in equation (14) has the correct normalization such that the window function satisfies $\int dk'\, G(k, k')/2\pi = 1$ (eq. [15]). Putting equation (14) into equation (12) and using the

correlation matrix given in equation (A2) together with the relation between the two-point function and the power spectrum in equation (2), it is not hard to see that $\langle \hat{P}_2(k) \rangle$ satisfies equation (15) with $G(k, k')$ given by $\sum_{\alpha\beta} w^{\alpha\beta}(k) e^{ik'(u_\alpha - u_\beta)}$. Using $\int dk'\, e^{ik'(u_\alpha - u_\beta)} = 2\pi\delta^{\alpha\beta}\mathcal{N}/\mathcal{L}$ then completes the derivation. One might want to explore more complicated data windowing (e.g., Press et al. 1992b; Hamilton 1997b), but since in practice uncertainties in the large-scale power estimate, where the survey window matters most, are likely dominated by the continuum, the simple choice we have adopted is probably adequate.

APPENDIX B

We derive here the band power variance given in equation (26). The power spectrum estimator is given in equation (12) with the matrix $w^{\alpha\beta}(k)$ given by equation (25). We ignore here the uncertainty in the estimation of the mean count $\bar{N}^\alpha_Q$.

The band power covariance can be written compactly as

$$C(k_1, k_2) \equiv \langle \Delta\hat{P}^\alpha(k_1) \Delta\hat{P}^\beta(k_2) \rangle = \sum_{\alpha\beta\gamma\eta} w^{\alpha\beta}(k_1) w^{\gamma\eta}(k_2) (\langle \hat{\delta}^\alpha_f \hat{\delta}^\beta_f \hat{\delta}^\gamma_f \hat{\delta}^\eta_f \rangle - \langle \hat{\delta}^\alpha_f \hat{\delta}^\beta_f \rangle \langle \hat{\delta}^\gamma_f \hat{\delta}^\eta_f \rangle) \,. \tag{B1}$$

The band power variance is simply the diagonal piece of this matrix: $C(k, k)$.

We can work out $\langle \hat{\delta}^\alpha_f \hat{\delta}^\beta_f \hat{\delta}^\gamma_f \hat{\delta}^\eta_f \rangle - \langle \hat{\delta}^\alpha_f \hat{\delta}^\beta_f \rangle \langle \hat{\delta}^\gamma_f \hat{\delta}^\eta_f \rangle$ using the same methodology as used in Appendix A for $\langle \hat{\delta}^\alpha_f \hat{\delta}^\beta_f \rangle$: rewrite $\hat{\delta}^\alpha_f$ as $(\hat{N}^\alpha - \tilde{N}^\alpha)/\bar{N}^\alpha + (\tilde{N}^\alpha - \bar{N}^\alpha)/\bar{N}^\alpha$, where $\tilde{N}^\alpha = \langle \hat{N}^\alpha \rangle_D$, and as before, take the discrete ensemble average $\langle \rangle_D$ before taking the cosmic average $\langle \rangle$. The result is

$$\begin{aligned}
C(k_1, k_2) = \sum_{\alpha\beta\gamma\eta} w^{\eta\gamma}(k_1) w^{\alpha\beta}(k_2) \Big\{ & \langle \delta^\alpha_f \delta^\beta_f \rangle \langle \delta^\gamma_f \delta^\eta_f \rangle + \langle \delta^\alpha_f \delta^\eta_f \rangle \langle \delta^\beta_f \delta^\gamma_f \rangle + \langle \delta^\alpha_f \delta^\gamma_f \rangle \langle \delta^\beta_f \delta^\eta_f \rangle + \langle \delta^\alpha_f \delta^\beta_f \delta^\gamma_f \delta^\eta_f \rangle_c \\
& + \frac{1}{\bar{N}^\eta_Q \bar{N}^\gamma_Q} \sum_{\sigma i} W^{\eta\sigma} W^{\gamma\sigma} \langle (W^{i\sigma})^2 [g^{i\sigma}_{\rm ps} g^\sigma_b \bar{N}^\sigma_Q (1 + \delta^\sigma_f) + V^{i\sigma}] \delta^\alpha_f \delta^\beta_f \rangle + [\text{1 other perm.: } (\alpha \leftrightarrow \gamma,\ \beta \leftrightarrow \eta)] \\
& + \frac{1}{\bar{N}^\beta_Q \bar{N}^\gamma_Q} \sum_{\sigma i} W^{\beta\sigma} W^{\gamma\sigma} \langle (W^{i\sigma})^2 [g^{i\sigma}_{\rm ps} g^\sigma_b \bar{N}^\sigma_Q (1 + \delta^\sigma_f) + V^{i\sigma}_B] \delta^\alpha_f \delta^\eta_f \rangle + [\text{3 other perm.: } (\beta \leftrightarrow \alpha,\ \gamma \leftrightarrow \eta), (\gamma \leftrightarrow \eta), (\beta \leftrightarrow \alpha)] \\
& + \frac{1}{\bar{N}^\beta_Q \bar{N}^\gamma_Q \bar{N}^\eta_Q} \sum_{\sigma i} W^{\beta\sigma} W^{\eta\sigma} W^{\gamma\sigma} \langle (W^{i\sigma})^3 [g^{i\sigma}_{\rm ps} g^\sigma_b \bar{N}^\sigma_Q (1 + \delta^\sigma_f) + \tilde{N}^{i\sigma}_S] \delta^\alpha_f \rangle + [\text{3 other perm.: } (\alpha \leftrightarrow \beta), (\alpha \leftarrow \gamma), (\alpha \leftrightarrow \eta)] \\
& + \frac{1}{\bar{N}^\alpha_Q \bar{N}^\beta_Q \bar{N}^\gamma_Q \bar{N}^\eta_Q} \sum_{\sigma i \chi j} W^{\alpha\sigma} W^{\beta\sigma} W^{\eta\chi} W^{\gamma\chi} \langle (W^{i\sigma})^2 (W^{j\chi})^2 [g^{i\sigma}_{\rm ps} g^\sigma_b \bar{N}^\sigma_Q (1 + \delta^\sigma_f) + V^{i\sigma}_B] [g^{j\chi}_{\rm ps} g^\chi_b \bar{N}^\chi_Q (1 + \delta^\chi_f) + V^{j\chi}_B] \rangle \\
& + \frac{1}{\bar{N}^\alpha_Q \bar{N}^\beta_Q \bar{N}^\gamma_Q \bar{N}^\eta_Q} \sum_{\sigma i \chi j} W^{\alpha\sigma} W^{\eta\sigma} W^{\beta\chi} W^{\gamma\chi} \langle (W^{i\sigma})^2 (W^{j\chi})^2 [g^{i\sigma}_{\rm ps} g^\sigma_b \bar{N}^\sigma_Q (1 + \delta^\sigma_f) + V^{i\sigma}_B] [g^{j\chi}_{\rm ps} g^\chi_b \bar{N}^\chi_Q (1 + \delta^\chi_f) + V^{j\chi}_B] \rangle \\
& + [\text{1 other perm.: } (\alpha \leftrightarrow \beta)] + \frac{1}{\bar{N}^\alpha_Q \bar{N}^\beta_Q \bar{N}^\gamma_Q \bar{N}^\eta_Q} \sum_{\sigma i} W^{\alpha\sigma} W^{\beta\sigma} W^{\eta\sigma} W^{\gamma\sigma} \langle (W^{i\sigma})^4 g^{i\sigma}_{\rm ps} g^\sigma_b \bar{N}^\sigma_Q (1 + \delta^\sigma_f) \rangle - \langle \hat{\delta}^\alpha_f \hat{\delta}^\beta_f \rangle \langle \hat{\delta}^\gamma_f \hat{\delta}^\eta_f \rangle \Big\} \,.
\end{aligned} \tag{B2}$$

The first set of terms (first line) arises from the shot noise–free part of $\hat{\delta}^\alpha_f$, namely, $(\tilde{N}^\alpha - \bar{N}^\alpha)/\bar{N}^\alpha$. The next two sets of terms (second and third lines) come from combinations of $\hat{\delta}^\alpha_f$ involving products of two shot noise terms with two shot noise–free terms. The next set of terms (fourth line) arises from products of three shot noise terms and a shot noise–free one. The next set of terms (fifth + sixth + seventh lines, except the very last term) comes from products of four shot noise terms. The last term corresponds to what has to be subtracted off to compute the covariance.

To make further progress, we assume that $W^{i\alpha}$ is independent of δ^α_f, which is strictly correct for $W^{i\alpha}$ given by equation (6) but only roughly so for equation (7) (see discussion in Appendix A). Then, taking the small wavelength limit in the sense that $k, \Delta k > 2\pi/\mathcal{L}$ (Δk is the size of a k bin), and making use of the fact that $W^{\alpha\sigma} W^{\beta\sigma}$ is only nonzero at separations $u^\alpha - u^\beta$ much less than $1/k$, where k is the wavenumber of interest, we obtain

$$\begin{aligned}
C(k_1, k_2) = \sum_\sigma (\bar{w}^{\sigma\sigma})^2 \frac{\mathcal{N}^3}{\mathcal{L}^2} \Big\{ & \frac{2}{n_{k_1}} \left[P(k_1) + \frac{\mathcal{L}}{\mathcal{N}} \sum_i \frac{1}{(\bar{N}^\sigma_Q)^2} \langle (W^{i\sigma})^2 (g^{i\sigma}_{\rm ps} g^\sigma_b \bar{N}^\sigma_Q + V^{i\sigma}_B) \rangle \sum_{\beta\gamma} W^{\beta\sigma} W^{\gamma\sigma} \right]^2 \delta^{k_1 k_2} \\
& + \frac{\langle T \rangle_{k_1 k_2}}{\mathcal{L}} + 4 \frac{1}{\mathcal{N}} \langle B \rangle_{k_1 k_2} \sum_i \frac{1}{\bar{N}^\sigma_Q} \langle (W^{i\sigma})^2 g^{i\sigma}_{\rm ps} g^\sigma_b \rangle \sum_{\beta\gamma} W^{\beta\sigma} W^{\gamma\sigma} \\
& + 2[P(k_1) + P(k_2)] \frac{\mathcal{L}}{\mathcal{N}^2} \sum_i \frac{1}{(\bar{N}^\sigma_Q)^2} \langle (W^{i\sigma})^3 g^{i\sigma}_{\rm ps} g^\sigma_b \rangle \sum_{\beta\eta\gamma} W^{\beta\sigma} W^{\eta\sigma} W^{\gamma\sigma} \\
& + 2 \langle P \rangle_{k_1 k_2} \frac{\mathcal{L}}{\mathcal{N}^2} \left[\frac{1}{\bar{N}^\sigma_Q} (W^{i\sigma})^2 g^{i\sigma}_{\rm ps} g^\sigma_b \sum_{\alpha\eta} W^{\alpha\sigma} W^{\eta\sigma} \right]^2 \\
& + \frac{\mathcal{L}^2}{\mathcal{N}^3} \frac{1}{(\bar{N}^\sigma_Q)^4} [(W^{i\sigma})^4 (g^{i\sigma}_{\rm ps} g^\sigma_b \bar{N}^\sigma_Q + \tilde{N}^{i\sigma}_S)] \sum_{\alpha\beta\gamma\eta} W^{\alpha\sigma} W^{\beta\sigma} W^{\eta\sigma} W^{\gamma\sigma} \Big\} \,.
\end{aligned} \tag{B3}$$

A few comments are in order. The terms in the second + fifth + sixth lines of equation (B2) are canceled by the last term of

equation (B2). The terms in the second line of equation (B2) contain contributions proportional to $B(k, -k, 0)$ that vanish. There is also a term from the fifth line of equation (B2) that is proportional to $P(0)$ that vanishes also. The shot noise terms in the first line of equation (B3) come, respectively, from the third + fifth + sixth lines of equation (B2). The rest of the terms in equation (B3) basically follow the order in which they are presented in equation (B2). Lastly, setting $W^{\alpha\beta} = \delta^{\alpha\beta}$ and $k_1 = k_2$ then recovers equation (26).

REFERENCES

Barlow, T. A., & Sargent, W. L. W. 1997, AJ, 113, 136
Bernstein, G. M. 1994, ApJ, 424, 569
Bi, H. G., Boerner, G., & Chu, Y. 1992, A&A, 266, 1
Bi, H., & Davidsen, A. F. 1997, ApJ, 479, 523
Blandford, R. D., Netzer, H., & Woltjer, L. 1990, Active Galactic Nuclei (Berlin: Springer)
Bond, J. R., Jaffe, A. H., & Knox, L. 1998, Phys. Rev. D, 57, 2117
Bond, J. R., & Wadsley, J. W. 1997, Proc. 12th Kingston Conf. (Halifax), in press (astro-ph 9703125)
Cen, R., Miralda-Escudé, J., Ostriker, J. P., & Rauch, M. 1994, ApJ, 437, L9
Cen, R., Phelps, S., Miralda-Escudé, J., & Ostriker, J. P. 1998, ApJ, 496, 577
Cristiani, S., D'Odorico, S., D'Odorico, V., Fontana, A., Giallongo, E., & Savaglio, S. 1997, MNRAS, 285, 209
Croft, R. A. C., Hu, W., & Davé, R. 1999, Phys. Rev. Lett., 83, 1092
Croft, R. A. C., Weinberg, D. H., Katz, N., & Hernquist, L. 1997, ApJ, 488, 532
———. 1998, ApJ, 495, 44
Croft, R. A. C., Weinberg, D. H., Pettini, M., Katz, N., & Hernquist, L. 1999, ApJ, 520, 1
Dodelson, S., Hui, L., & Jaffe, A. 1997, preprint (astro-ph 9712074)
Fan, X. 1999, AJ, 117, 2528
Feldman, H. A., Kaiser, N., & Peacock, J. A. 1994, ApJ, 426, 23
Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, AJ, 111, 1748
Gaztañaga, E., & Croft, R. A. C. 1998, MNRAS, 309, 885
Gnedin, N. Y. 1998, MNRAS, 299, 392
Gnedin, N. Y., & Hui, L. 1998, MNRAS, 296, 44
Hamilton, A. J. S. 1997a, MNRAS, 289, 285
———. 1997b, MNRAS, 289, 295
———. 2000, MNRAS, 312, 257
Hernquist, L., Katz, N., Weinberg, D. H., & Miralda-Escudé, J. 1996, ApJ, 457, L51
Horne, K. 1986, PASP, 98, 609
Hu, E. M., Kim, T.-S., Cowie, L. L., Songaila, A., & Rauch, M. 1995, AJ, 110, 1526
Hui, L. 1998, Proc. MPA/ESO Conf., Evolution of Large-Scale Structure, ed. A. Banday & R. Sheth (Garching: ESO), in press (astro-ph 9812293)
———. 1999, ApJ, 516, 519
Hui, L., & Gaztañaga, E. 1999, ApJ, 519, 622
Hui, L., & Gnedin, N. Y. 1997, MNRAS, 292, 27
Hui, L., Gnedin, N. Y., & Zhang, Y. 1997, ApJ, 486, 599
Kim, T.-S., Hu, E. M., Cowie, L. L., & Songaila, A. 1997, AJ, 114, 1
Kirkman, D., & Tytler, D. 1997, ApJ, 484, 672
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Lu, L., Sargent, W. L. W., Womble, D. S., & Takada-Hidai, M. 1996, ApJ, 472, 509
McDonald, P., Miralda-Escudé, J., Rauch, M., Sargent, W. L. W., Barlow, T. A., Cen, R., & Ostriker, J. P. 2000, ApJ, 543, 1
Meiksin, A., & White, M. 1999, MNRAS, 308, 1179
Miralda-Escudé, J., Cen, R., Ostriker, J. P., & Rauch, M. 1996, ApJ, 471, 582
Muecket, J. P., Petitjean, P., Kates, R. E., & Riediger, R. 1996, A&A, 308, 17
Nusser, A., & Haehnelt, M. 2000, MNRAS, 313, 364
Peebles, P. J. E. 1980, The Large-Scale Structure of the Universe (Princeton: Princeton Univ. Press)
Peterson, B. M. 1997, An Introduction to Active Galactic Nuclei (Cambridge: Cambridge Univ. Press)
Phillips, J., Weinberg, D. H., Croft, R. A. C., Katz, N., Hernquist, L., & Pettini, M. 2000, preprint (astro-ph 0001089)
Press, W. H., Rybicki, G. B., & Hewitt, J. N. 1992a, ApJ, 385, 404
Press, W. H., Rybicki, G. B., & Schneider, D. P. 1993, ApJ, 414, 64
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992b, Numerical Recipes (Cambridge: Cambridge Univ. Press)
Rauch, M. 1998, ARA&A, 36, 267
Rauch, M., et al. 1997, ApJ, 489, 7
Reisenegger, A., & Miralda-Escudé, J. 1995, ApJ, 449, 476
Rybicki, G. B., & Press, W. H. 1992, ApJ, 398, 169
Scoccimarro, R., Zaldarriaga, M., & Hui, L. 1999, ApJ, 527, 1
Seljak, U. 1998, ApJ, 503, 492
Szapudi, S., & Szalay, A. S. 1998, ApJ, 494, L41
Tegmark, M., Hamilton, A. J. S., Strauss, M. A., Vogeley, M. S., & Szalay, A. S. 1998, ApJ, 499, 555
Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, ApJ, 480, 22
Vogeley, M. S., & Szalay, A. S. 1996, ApJ, 465, 34
Vogt, S. S., et al. 1994, Proc. SPIE, 2198, 362
Weinberg, D. H., Croft, R. A. C., , Hernquist, L., Katz, N., & Pettini, M. 1999, ApJ, 522, 563
Zaldarriaga, M., Seljak, U., & Hui, L. 2000, ApJ, in press (astro-ph/0007101)
Zhang, Y., Anninos, P., & Norman, M. L. 1995, ApJ, 453, L57
Zheng, W., Kriss, G. A., Telfer, R. C., Grimes, J. P., & Davidsen, A. F. 1998, ApJ, 492, 855
Zuo, L., & Bond, J. R. 1994, ApJ, 423, 73