THE Lyα FOREST POWER SPECTRUM FROM THE SLOAN DIGITAL SKY SURVEY

PATRICK MCDONALD,1,2 UROŠ SELIJAČIĆ,2 SCOTT BURLES,3 DAVID J. SCHLEGEL,4,5 DAVID H. WEINBERG,6 RENYUE CEN,6 DAVID SHIH,2 JOOP SCHAYE,7 DONALD P. SCHNEIDER,8 NETA A. BARCALL,4 JOHN W. BRIGGS,9 J. BRINKMANN,10 ROBERT J. BRUNNER,11 MASATAKA FUKUGITA,12 JAMES E. GUNN,7 ŽELJKO IVEZIĆ,4,13 STEPHEN KENT,14 ROBERT H. LUPTON,4 AND DANIEL E. VANDEN BERK8

Received 2004 May 3; accepted 2005 June 16

ABSTRACT

We measure the power spectrum, \( P_F(k, z) \), of the transmitted flux in the Ly\( \alpha \) forest using 3035 high-redshift quasar spectra from the Sloan Digital Sky Survey. This sample is almost 2 orders of magnitude larger than any previously available data set, yielding statistical errors of \(~0.6\%\) and \(~0.005\%\) on, respectively, the overall amplitude and logarithmic slope of \( P_F(k, z) \). This unprecedented statistical power requires a correspondingly careful analysis of the data and of possible systematic contaminations in it. For this purpose we reanalyze the raw spectra to make use of information not preserved by the standard pipeline. We investigate the details of the noise in the data, resolution of the spectrograph, sky subtraction, quasar continuum, and metal absorption. We find that background sources such as metals contribute significantly to the total power and have to be subtracted properly. We also find clear evidence for \( \text{Si} \equiv \text{correlations with the Ly}\( \alpha \) forest and suggest a simple model to account for this contribution to the power. Our tests indicate that any residual systematics in the analysis are unlikely to affect the inference of cosmological parameters from \( P_F(k, z) \). These results should provide an essential ingredient for all future attempts to constrain modeling of structure formation, cosmological parameters, and theories for the origin of primordial fluctuations.

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

Online material: color figures, machine-readable tables

1. INTRODUCTION

Although the Ly\( \alpha \) forest was discovered many decades ago (Lynds 1971), it has only recently emerged as one of the prime tracers of the large-scale structure in the universe. The high-resolution measurements using the Keck HIRES spectrograph (Vogt et al. 1994) have been largely reproduced using hydrodynamical simulations (Cen et al. 1994; Zhang et al. 1995; Hernquist et al. 1996; Theuns et al. 1998) and semianalytical models (Gnedin & Hui 1998). The picture that has emerged from these studies is one in which the neutral gas responsible for the absorption is in a relatively low-density, smooth environment, which implies a simple connection between the gas and the underlying dark matter. The neutral fraction of the gas is determined by the interplay between the recombination rate (which depends on the temperature of the gas) and ionization caused by ultraviolet photons. Photoionization heating and expansion cooling cause the gas density and temperature to be tightly related, except where mild shocks heat up the gas. This leads to a tight relation between the absorption and the gas density. Finally, the gas density is closely related to the dark matter density on large scales, while on small scales the effects of thermal broadening and Jeans smoothing must be included. In the simplest picture described here, all of the physics ingredients are known and can be modeled. The fact that one can trace the fluctuations over a range of redshifts (\( 2 \leq z \leq 6 \) using ground-based spectrographs) and over a range of scales, which are typically smaller than the scales of other tracers, is the main strength of this method. It becomes particularly powerful when combined with cosmic microwave background (CMB) anisotropies or other tracers that are sensitive to larger scales. Such a combination is sensitive to the shape of the primordial spectrum of fluctuations, which is one of the few observationally accessible probes of the early universe. These observations are therefore directly testing the models of the early universe such as inflation.

Ly\( \alpha \) forest observations and constraints on cosmology have been explored by several groups in the past. Most of the analyses focused on the power spectrum, \( P_F(k) \), of the fluctuations in the Ly\( \alpha \) forest flux,

\[
\delta_F(\lambda) = \exp[-\tau(\lambda)]/\langle\exp(-\tau)\rangle - 1,
\]

where \( \tau \) is the optical depth to Ly\( \alpha \) absorption. The first such work was by Croft et al. (1998), followed by McDonald et al.
of $k$-range is determined by the continuum fluctuations on the low end and spectral resolution at the high end. We note also that the useful range is limited not only by these uncertainties, which are related to the data analysis, but also by the uncertainties in the theoretical modeling and/or additional astrophysical effects. We will address these latter issues in more detail in a separate publication. However, we do not completely decouple the theory from the data analysis. For example, when discussing the importance of systematic errors it is useful to understand how they would affect cosmological results like the slope and amplitude of the matter power spectrum, so much of our discussion of systematics is devoted to this issue.

The common usage of the term Ly$\alpha$ forest is to describe the Ly$\alpha$ absorption by neutral hydrogen in the relatively low-density bulk of the intergalactic medium (IGM). In this paper we include damped-Ly$\alpha$ systems (DLAs) in the definition of the “forest,” so it includes all i-Ly$\alpha$ absorption. We could try to remove DLAs before measuring $P_F(k, z)$, because they are more difficult to simulate than the lower optical depth absorption; however, we believe the advantage of removal is illusory. If the DLAs were located randomly within the IGM (which they certainly are not completely), it would be simple to include them in the theory using their known column density distribution. If they are not located randomly, the regions obscured by DLAs in the spectra are special, so the effect of removing the DLAs still must be understood using simulations. We leave the handling of the effects of DLAs as a problem for the theory, which we will address elsewhere.

Absorption by metals is also difficult to simulate accurately, so we would like to remove this contribution to $P_F(k, z)$. This is relatively easy to do for transitions with wavelength $\lambda \gtrsim 1300$ Å, but it is basically impossible for transitions with $\lambda \lesssim \lambda_{\alpha}$, because the metal features always appear mixed with H i-Ly$\alpha$. We will subtract the power measured in the rest wavelength range $1268 \, \lambda_{\text{rest}} < 1380$ Å from our measurement of the power in the forest [$\lambda_{\text{rest}} \equiv \lambda_{\text{observed}}/(1 + z_q)$, where $z_q$ is the redshift of the quasar]. This subtraction removes the effect of transitions with longer wavelength, but we leave shorter wavelength transitions as part of the forest. The only significant shorter wavelength contaminant that we have identified is Si $\text{ii}$ absorption at 1206.50 Å, and we develop a simple and effective way to account for this in the theory. We refer to our final background-subtracted power spectrum as $P_F(k, z)$, and use $P_{\text{H}}(k, z)$ for the raw power measured in the interval $\lambda_1 < \lambda_{\text{rest}} < \lambda_2$. We are using the range $1041 \, \lambda < \lambda_{\text{rest}} < 1185$ Å for the Ly$\alpha$ forest.

The outline of this paper is as follows. In §2, we describe the selection of our data set and the preparation of spectra for the measurement of $P_F(k, z)$. In §3 we describe the method used to measure the power spectrum and estimate the error bars, test the procedure and give the basic results. We perform consistency checks on the results and discuss systematic errors in §4, which is followed by a brief recipe for using our results in §5, and conclusions in §6.

2. DATA SELECTION AND PREPARATION

We describe the sample of quasar spectra that we use in §2.1. In §2.2 we explain how we remove broad absorption line (BAL) quasars from the sample. In §2.3 we explain how we combine spectra from different exposures for the same quasar and use the differences between exposures to understand the noise in the data. We discuss the resolution of the spectra in §2.4. Finally, in §2.5 we discuss how we divide each spectrum by an estimate of the quasar continuum, the expected mean absorption level in the spectrum, and a spectral calibration vector (see below), to
produce the vectors of transmission-fluctuation estimates, $\mathbf{\delta}_T$, for each quasar, from which we will measure $P_F(k, z)$.

2.1. SDSS Observations and Sample Selection

The Sloan Digital Sky Survey (York et al. 2000) uses a drift-scanning imaging camera (Gunn et al. 1998) and a 640 fiber double spectrograph on a dedicated 2.5 m telescope. It is an ongoing survey to image 10,000 deg$^2$ of the sky in the SDSS $ugriz$ AB magnitude system (Fukugita et al. 1996; Stoughton et al. 2002) and to obtain spectra for $\sim 10^6$ galaxies and $\sim 10^5$ quasars. The astrometric calibration is good to better than 0.1" per coordinate (Pier et al. 2003), and the photometric calibration is accurate to 3% or better (Hogg et al. 2001; Smith et al. 2002). The data sample used in this paper was compiled in summer of 2002 and is a combination of data releases 1 (Abazajian et al. 2003) and 2 (Abazajian et al. 2004).

About 13% of the spectroscopic survey targets are quasar candidates selected based on their colors (Richards et al. 2002). The magnitude limit for UV-excess objects is $i = 19.1$, while additional high-redshift candidates ($z > 3$) are targeted to $i = 20.2$. Fibers are allocated according to a tiling algorithm (Blanton et al. 2003), with the galaxy sample and the quasar sample being the top priorities. The remaining 8% of fibers serve for calibration purposes.

SDSS spectra are obtained using plates holding 640 fibers, each of which subtends 3" on the sky; the spectra cover 3800–9200 Å. The pixel width is a slowly varying function of wavelength but is typically $\sim 70$ km s$^{-1}$. The resolution also varies but is typically also $\sim 70$ km s$^{-1}$ rms (i.e., the resolution is 1800 $< R < 2100$ and there are $\sim 2.4$ pixels per FWHM resolution element). All quasars have multiple spectra, usually taken one after the other (timescales of a fraction of an hour), so the quasar variability can be ignored (in the opposite case it would act as an additional source of noise). The co-added spectra in the official SDSS release use local spline interpolation onto a uniform grid of pixels of width $\Delta \log_{10}(\lambda) = 0.0001$ and do not guarantee the noise to be uncorrelated. We therefore redo this step starting from the individual exposures. This is discussed in more detail below. Spectral flux errors per pixel in most cases are about $1 \times 10^{-17}$ erg s$^{-1}$ cm$^{-2}$ Å$^{-1}$. Redshifts are automatically assigned by the SDSS spectral classification algorithm, which is based on $\chi^2$ fitting of templates to each spectrum (D. J. Schlegel et al. 2005, in preparation).

We limit ourselves to quasars with redshift $z_q > 2.3$ when measuring the power in the Ly$\alpha$ forest region of spectra, so that each spectrum contains a significant stretch of the Ly$\alpha$ forest above the detector cutoff at 3800 Å (which corresponds to Ly$\alpha$ absorption at $z = 2.12$). We use the sample compiled in summer of 2002, cut down to 3035 spectra by eliminating some plates of questionable quality, some spectra where two different redshift estimation codes disagree, and some BAL quasars (see below). Figure 1 shows the redshift distribution of the data.

The dashed, red histogram shows the distribution of quasar redshifts. The solid, black histogram shows the distribution of quasar redshifts. The solid, black histogram shows the distribution of quasar redshifts. The solid, black histogram shows the distribution of quasar redshifts. The solid, black histogram shows the distribution of quasar redshifts. The solid, black histogram shows the distribution of quasar redshifts. The solid, black histogram shows the distribution of quasar redshifts.

The regions we use to measure the Ly$\alpha$ forest power and background power are indicated by vertical dotted lines, along with a couple of alternate regions that we will discuss (note that the background and Ly$\alpha$ forest observed in the same quasar spectrum correspond to different redshifts). Figure 2 shows an example SDSS spectrum of a $z = 3.7$ quasar with unusually high S/N.

This spectrum is unusual in that most have lower S/N, and most quasars are at lower redshift.

We employ an additional sample of $\sim 8000$ spectra with $z_q < 2.3$, so that we can study the full observed wavelength range, 3800 Å $< \lambda < 9200$ Å, outside the confusion of the Ly$\alpha$ forest. As we discuss in §3.4, we compute a non-negligible background power term (probably mostly metal absorption), by measuring the power in the wavelength range 1268 Å $< \lambda_{\text{rest}} < 1380$ Å. Using only the primary sample, we would not be able to compute this term for observed wavelengths below $\sim 4400$ Å.
We remove several wavelength regions from our analysis because of calibration problems: \( \lambda < 3800 \text{ Å} \), \( 5575 \text{ Å} < \lambda < 5583 \text{ Å} \), \( 5888 \text{ Å} < \lambda < 5894 \text{ Å} \), \( 6296 \text{ Å} < \lambda < 6308 \text{ Å} \), and \( 6862 \text{ Å} < \lambda < 6871 \text{ Å} \) (the last two have no direct effect on the results we present). Most of these problems are due to strong sky lines.

2.2. BAL Removal

Our sample was initially examined by eye, and the most extreme broad absorption line (BAL) quasars were removed (see Hall et al. (2002) for a discussion of BALs). When we first measured the background power in the region \( 1409 \text{ Å} < \lambda_{\text{rest}} < 1523 \text{ Å} \), we found that the most extreme outliers in power were still obvious BALs (this was not true of the Ly\( \alpha \) forest region). To test the importance of these systems to our Ly\( \alpha \) forest power measurement, we removed a further 147 quasars using the following automated method: Each spectrum is smoothed by a Gaussian with rms width 280 km s\(^{-1}\). The continuum within the region \( 1420 \text{ Å} < \lambda_{\text{rest}} < 1535 \text{ Å} \) is redefined by dividing by the mean flux-to-continuum ratio in the region. A quasar is identified as a BAL quasar if the region \( 1420 \text{ Å} < \lambda_{\text{rest}} < 1535 \text{ Å} \) contains a 2000 km s\(^{-1}\) long continuous set of pixels that all fall more than 20% above or below our estimated continuum (we initially identified wide regions with flux above the continuum out of simple curiosity but found that these are in practice almost always obvious BAL quasars where the continuum has been biased low by the BAL feature). We iterate the continuum redefinition twice, computing the new mean after throwing out pixels more than 20% below the previous mean, but this makes almost no difference to the results. Note that the 280 km s\(^{-1}\) smoothing was applied to allow easier identification of BALs in noisy spectra. As we show below, this BAL cut makes essentially no difference to our \( P_F(k, z) \) result, although it does have a noticeable effect on the power measurement in the region \( 1409 \text{ Å} < \lambda_{\text{rest}} < 1523 \text{ Å} \).

2.3. Combining Exposures and Calibrating the Noise

SDSS obtains multiple (at least three) exposures for each quasar. We combine the individual exposure spectra to produce a single spectrum, using a nearest-grid-point method that produces uncorrelated noise and a reasonably well-defined sampling window. For each pixel we record estimates of wavelength, quasar flux, resolution, sky flux, read noise, and two different total noise estimates. The first noise estimate, which we will call \( \sigma_p \) (\( p \) for pipeline), is computed using the error array given for the exposure spectra by the spectral reduction pipeline. The second noise estimate, which we call \( \sigma_c \) (\( c \) for component), is computed by summing the read noise and the noise implied by estimates of the number of photons corresponding to the quasar flux and sky flux. The two noise estimates generally do not agree, but this is not a problem for us because we ultimately recalibrate the noise (next). Finally, we record \( \chi^2/v \) for each pixel, computed by treating the determination of the combined flux value for each pixel as a one-parameter fit to the measurements given by the different exposures. Examples of the more important of these quantities in Ly\( \alpha \) forest regions are shown in Figure 3.

For comparison to the sky and quasar flux levels, we have converted the Gaussian read noise into the flux of photons that would contribute the same noise variance. Several elements of Figure 3 (e.g., the estimation of the quasar continuum and \( \sigma_w \)) will be described later in this paper.

The noise estimate from the standard SDSS pipeline is only approximate. The accuracy of the noise estimate required for our purpose is much higher than anticipated when the pipeline was developed. For this reason we use the differences between single-exposure spectra for the same quasar to determine the noise properties of the data. We construct difference spectra by combining the flux-calibrated exposures with alternating sign for each exposure, i.e., we use exactly the same procedure that we normally use to produce combined spectra from the exposures, except half of the exposures are subtracted instead of added, so the mean result is zero (we drop the last exposure when there is an odd number; this is not the most efficient method possible, but we do not need it to be). The result is a direct measure of the exposure-to-exposure changes. We measure the power spectrum of these difference spectra using the method described in § 3.1, including noise subtraction based on the pipeline noise estimates for the pixels. The result is shown in Figure 4 (points with error bars).

We obtain a clear detection of power, where there should be none if the spectra differ only by the noise estimate from the pipeline that is being subtracted. If we assume that the noise has been underestimated by a constant factor, and fitted for that factor using the error covariance matrix estimated by bootstrap resampling, we find a decent fit: \( \chi^2 = 141.6 \) for 107 degrees of freedom (dof) (formally, this fit is not good because \( \chi^2 \) is unlikely to be this high by chance). This fit uses our usual points in 0.0013 s km\(^{-1}\) < \( k \) < 0.02 s km\(^{-1}\). The best-fit value of the excess noise contribution is 16.1 ± 0.4% of the original noise estimate, indicating that the rms noise was underestimated by 8%. The best fit and goodness of fit do not change if we add points on larger scales. The quality of the fit begins to degrade as we add points with larger \( k \), but the best-fit value changes by only 1% (in power) out to the Nyquist frequency of the data. Of course, we have no reason to expect a single redshift-independent factor to describe the relation between the true and pipeline noise, so the formally bad \( \chi^2 \) is not a fundamental problem. We check for systematic change with redshift by allowing a power-law dependence, \( P_{\text{residual noise}} \propto [3.75/(1 + z)]^d \) but find no significant detection (\( d = 0.07 \pm 0.20 \)). Our final method will effectively account for evolution anyway, as described below.

A \( k \) dependence different than expected for white noise could be a problem for us, so we check for this by fitting for a power-law dependence, \( P_{\text{residual noise}} \propto (k/b_0)^{q} \) (with \( b_0 = 0.0074 \text{ s km}^{-1} \)), finding \( b = -0.111 \pm 0.025 \), a significant detection (\( \chi^2 \) is now a reasonable 123.3 for 106 dof). Allowing a running of the power law, \( P_{\text{residual noise}} \propto (k/b_0)^{q} \exp^{q/2 \ln (k/b_0)} \), does not improve the fit (\( c = -0.046 \pm 0.066 \)). The slope we find corresponds to a ~20% change in 16% of the noise power at the extremes of our \( k \) range, i.e., only ~3% of the total noise power, which is a relatively small fraction of the Ly\( \alpha \) forest power except at the highest \( k \) (see Fig. 11 below). We henceforth assume that the extra noise is proportional to \( k^{-0.111} \) rather than white (this makes <1% difference in the final results except for the one highest \( k \), lowest \( z \) point where the difference is 2%).

How accurate is this noise estimate based on differences between exposures? Our difference spectra will contain a component of the Ly\( \alpha \) forest power if the calibration between exposures is not perfect. The power in this term would be suppressed relative to the Ly\( \alpha \) forest power by the fractional calibration error squared, so it would be very small unless the exposure-to-exposure calibration errors were quite large. The fact that a simple one-parameter extra-noise model fits reasonably well, in the face of variation in redshift, noise amplitude, and \( k \), argues against calibration errors being a big problem. More convincingly, we measure nearly the same excess noise contribution (14.2 ± 0.5%) and slope (\( b = -0.135 \pm 0.028 \)) in the region...
Fig. 3.—Examples of the chunks of spectra used to measure power, with (a, b) showing quasars at \( z_q = (3.24, 2.45) \) over the rest wavelength range \( 1113 \, \AA < \lambda_{\text{rest}} < 1185 \, \AA \), and (c) showing a quasar at \( z_q = 3.30 \) over the rest wavelength range \( 1041 \, \AA < \lambda_{\text{rest}} < 1113 \, \AA \). Top panel: quasar flux (solid line), sky flux (dotted line), our continuum estimate (short-dashed line), and the read noise as an equivalent photon flux (long-dashed line). Middle panel: S/N level shown as a ratio of our continuum to the different rms noise levels (see text), \( \sigma_c \) (solid line), \( \sigma_p \) (dotted line), and \( \sigma_r \) (dashed line). Bottom panel: Calibration correction vector, \( \bar{S} \) (dotted line), rms resolution in units of 100 km s\(^{-1}\) (dashed line), and evolution of the mean transmission fraction, \( \bar{F}(z) \) (solid line). The perfect degeneracy in our analysis between the overall normalization of the continuum and \( \bar{F}(z) \) has been broken arbitrarily, so only the evolution of \( \bar{F}(z) \) is meaningful (see text). [See the electronic edition of the Supplement for a color version of this figure.]

1268 Å < \( \lambda_{\text{rest}} \) < 1380 Å as we do in the Ly\( \alpha \) forest. This argues against any connection to leaking Ly\( \alpha \) forest power. Note that the effective absolute level of noise in the 1268 Å < \( \lambda_{\text{rest}} \) < 1380 Å region is about half that in the Ly\( \alpha \) forest region, so this test shows that the fraction of extra noise does not depend strongly on the noise level itself.

Pixels in different exposures are not perfectly aligned, and misalignment can allow Ly\( \alpha \) forest power to leak into our difference spectra. To test this alternative explanation for the apparent excess noise in the spectra, we split the spectra into two

groups with approximately equal weight, based on the rms misalignment in the forest region (the alignment is known from the wavelength calibration of the exposures, which is thought to be practically perfect). We find the same excess noise power in both the poorly aligned group (16.1 ± 0.6%, \( b = -0.086 \pm 0.036 \)) and the better aligned group (15.3 ± 0.6%, \( b = -0.123 \pm 0.036 \)), suggesting that the excess power is not due to misalignment. Furthermore, the presence of a similar level of excess noise power outside of the forest region again argues against leakage. We therefore believe that our noise estimate is
considerably more accurate than the noise estimate from the SDSS pipeline.

In our initial power spectrum analysis we multiplied the noise-power estimated from the pipeline errors by the factor 1.16 for all spectra; however, when we split the data based on the mean value of $\chi^2/\nu$ for the exposure combination (see § 4.4) we found subsamples disagreed significantly on the $P_R(k, z)$ results. We eliminated this problem by estimating the noise-correction factor individually for each spectrum, by fitting to the power in the difference spectrum for that quasar. The mean extra power from these fits is still close to 16%, but there is considerable scatter. When we use these individual estimates, the correlation between measured $P_R(k, z)$ and $\chi^2/\nu$ disappears, i.e., the mean value of $\chi^2/\nu$ for a spectrum’s exposure combination was a good indicator of the amount by which the noise in each spectrum was misestimated. Note that there are statistical errors in these noise estimates for each spectrum, of the same order as the error for which we are trying to correct; however, there is no systematic bias associated with these errors, and the random error they contribute is automatically included in our final bootstrap errors. In fact, including the spectrum-by-spectrum noise estimate reduces the bootstrap errors slightly on small scales, verifying that these estimates are on average more accurate than the original noise estimates. It is not known why the noise is misestimated by the standard pipeline. Tests at this level have not been done before.

Our final data product will be a measurement of $P_R(k, z)$ binned in $k$ and $z$, i.e., a matrix $P_{R,ij}$, where $i$ labels bins with $z_i$ and $j$ labels bins with $k_j$. We will also give the noise power that was subtracted, $P_{N,ij}$, in the same bins. We suggest allowing a 5% rms freedom in the noise amplitude in each $z$ bin when performing model fits, i.e., for each bin subtract $f_i P_{N,ij}$ from $P_{R,ij}$, and add $(f_i/0.05)^2$ to $\chi^2$. This is probably overly conservative for any one bin, but implies a combined freedom $\pm 0.05/3$ (for nine bins) on an overall noise misestimation. This seems prudent, even though it is not really required by any test we have performed.

2.4. Accuracy of the Resolution

The resolution of the SDSS spectra is estimated using lines from calibration lamps mounted on the telescope structure. Shifts of the detector pixel grid relative to a fixed observed wavelength frame during an exposure, which we will call flexure, are expected to be the dominant source of error in this spectral resolution estimate. We tried estimating the rate of shifting for each pixel by differencing the wavelength calibrations of adjacent exposures (this calibration is determined very precisely for each exposure using the positions of sky lines). The implied extra smoothing only changes the power by $\sim 2\%$ at our highest $k$ bin.

The strong sky line at 5577 Å provides a good opportunity to measure the resolution more directly (note that the spectral wavelengths are in vacuum, and heliocentric, so this and other sky lines generally appear shifted from their standard wavelength). We measure the power spectrum in $\sim 3000$ sky spectra in the range $5560 \lesssim \lambda < 5598 \text{ Å}$. If the sky line has negligible width and the smoothing has a Gaussian shape with rms width $R$, the power spectrum should be proportional to $W^2(k, R, l) = \exp \left[-(kR)^2/2\right] \left[\sin (k l/2)/(k l/2)^2\right]^2$, where $l$ is the pixel width (the pixelization effect is subdominant but not negligible). In Figure 5 we show the measured power averaged over all the sky spectra after dividing each individual measurement by $W^2(k, R, l)$, where $R$ and $l$ are the local values (they are a good approximation constant over the range we are looking at), and also dividing each measurement by the value at a low $k$ where the resolution should not have any effect.

The result is remarkably close to unity, indicating that the estimated resolution is an accurate representation of the true resolution. What are the small wiggles? Figure 6 shows an example of the region we Fourier transform to measure the power.

We believe the small features to the sides of the main line are OH lines at 5564 and 5589 Å (Slanger et al. 2003). We test this explanation for the wiggles by constructing mock sky spectra that simply have a delta function at 5579 Å and two more with 0.003 times the main line’s amplitude at 5566 and 5591 Å, convolved with the resolution and pixelization (0.003 was chosen to give the best fit to the wiggles). The red, dotted line in Figure 5 shows the same resolution test using the mocks. We see that the wiggles are essentially perfectly reproduced. In conclusion, the resolution profile appears to be perfectly Gaussian, with exactly the width expected from the given resolution. There is apparently no room for even a 2% level effect from flexure. We are prevented from performing the same kind of measurement using other sky lines by similar features that are always much larger relative to the central line.

We suggest that fits to $P_R(k, z)$ include a multiplicative uncertainty on the overall power, of the form $\exp(\alpha k^2)$, where $\alpha$ is a single parameter in the fit subject to the rms constraint $\sigma_v = (7 \text{ km s}^{-1})^2$. This allows for a $\sim 2\%$ change in the smoothing kernel at our highest $k$, similar to our estimate of the error from flexure. This error estimate is somewhat arbitrary, but the evidence we have presented suggests that it should be smaller, so our estimate is conservative.
Fig. 5.—Resolution test. The solid line with error bars shows the power measured in \( \sim3000 \) sky spectra in the range \( 5560 \lesssim \lambda < 5598 \) \( \text{Å} \) (dominated by the strong sky line at 5577 \( \text{Å} \)) divided by the asymptotic small \( k \) power and by the estimated resolution/pixelization kernel \( W^2(k) \) for each spectrum. If the resolution estimate was perfect, and the sky line was narrow and the only flux present, this division would give exactly 1. The long-dashed line shows the power not divided by error bars. We see that the results are different from what we expect, especially at small \( k \). The dotted line shows the result of our test for mock spectra constructed with a Gaussian at 5579 \( \text{Å} \) and two more at 5566 and 5591 \( \text{Å} \) with 0.03 times its amplitude, representing OH lines. The short-dashed line shows \( \exp\left( -0.7k \text{ km s}^{-1} y^2 \right) \). The vertical, dotted lines bound the \( k \) region in which we will present Ly\( \alpha \) forest results, while the horizontal, dotted line just guides the eye to 1. [See the electronic edition of the Supplement for a color version of this figure.]

Note that this resolution test, and the noise calibration, cannot be used directly with the standard pipeline spectra, where the exposures are combined in a different way. The reader may be confused at this point about how our spectra differ from the standard publicly available set, so we give the following summary:

1. Our nearest–grid-point combination of the exposures produces uncorrelated noise in pixels (to the extent that the noise in the exposures was uncorrelated, which is expected from the way they are extracted), while the standard pipeline uses a local splining procedure that does a good but not perfect job of preventing noise correlation.

2. When combining exposures we record the effect of different pixel sizes, misalignment of the pixels, and flexure of the detector during exposures, which can influence the effective resolution.

3. We record the contribution of quasar flux, sky flux, and read noise to the total noise in each pixel. Knowing the contribution from quasar flux is important if pixel-by-pixel noise weighting is to be used, because the correlation between flux level and noise amplitude can lead to biases (see the end of § 2.5).

4. We correct for the bias in the exposure combination associated with cross-correlation between the noise variance level in exposure pixels and the quasar flux in the pixel (a different incarnation of the problem alluded to in the previous point).

5. The noise is recalibrated for each spectrum by differencing the exposures. The noise variance in the standard pipeline exposure spectra is underestimated by on average 16%, on top of any error related to the exposure combination, and the power measured in the difference spectra is slightly tilted relative to white noise.

The last point is the most important.

2.5. Determination of the Continuum and Mean Absorption Level

Our goal is to measure the power spectrum of the fluctuations in the transmitted flux fraction through the IGM, \( \delta_\ell(\lambda) = F(\lambda)/\bar{F} - 1 \), where \( F(\lambda) = \exp(-\tau(\lambda)) \) and \( \tau(\lambda) \) is the Ly\( \alpha \) forest optical depth (as defined in § 1). However, the spectrum of each quasar is the product of \( F \) and the quasar continuum (note that we use ‘‘continuum’’ to refer to all the flux emitted by the quasar, including emission lines), further complicated by errors in the detector calibration and absorption by longer wavelength transitions. The details of the procedure we use to separate these contributions will be presented elsewhere; here we give the basic idea and key results that are relevant for the flux power spectrum determination.

We use an iterative procedure to determine the components of the data model

\[
\hat{f}^i = A_\delta \bar{C}^i(\lambda_{\text{rest}})(1 + \delta_C^i)\bar{S}^i(\lambda^i)(1 + \delta_\ell^i)\bar{F}(\lambda^i)(1 + \delta_F^i) + n^i, \tag{2}
\]

where \( \hat{f}^i \) is the raw flux in pixel \( i \), \( n^i \) is the noise, \( A_\delta \) is the overall normalization of the quasar spectrum, \( \bar{C}^i(\lambda_{\text{rest}}) \) is the mean quasar continuum shape, \( \delta_C \) is the fluctuation around the mean continuum, \( \bar{S}^i(\lambda) \) is a mean generalized calibration vector (this includes wavelength-dependent calibration errors in the SDSS spectra, but also the mean absorption by metal lines with resonance wavelength \( \gtrsim 1300 \) \( \text{Å} \)), \( \delta_\ell \) is the fluctuation around \( \bar{S} \), such as individual metal lines or variable calibration errors, \( \bar{F}(\lambda) \) is the mean Ly\( \alpha \) forest absorption at a given redshift, and \( \delta_F \) is the fluctuation in Ly\( \alpha \) forest absorption. Note that here, as most places in this paper, \( z^i = \lambda^i/\lambda_{\alpha} - 1 \) is the redshift of gas that
would produce Lyα absorption in the pixel, not the redshift of the quasar. Briefly, we determine \( A_q \) for each spectrum, the global functions \( C(\lambda_{\text{rest}}), S(\lambda), \) and \( F(z) \), and a set of principal component analysis (PCA) eigenvectors that describe \( \delta_C \) by, in turn, computing each component of the model from all the spectra while holding all the others fixed. For example, \( F(z) \) is estimated in bins of \( z \) by averaging \( f_i A_q C(\lambda_{\text{rest}}(1 + \delta_C)) S(\lambda) F(z) \) over all the pixels in the Lyα forest that fall in each bin. We separate \( S(\lambda) \) from \( F(z) \) by measuring \( S(\lambda) \) in the rest wavelength range 1268 Å < \( \lambda_{\text{rest}} < 1380 \) Å, i.e., outside the Lyα forest. A few iterations suffice to determine all the components of the model independently. Three examples of the results are shown in Figure 3. The full details of this procedure will be presented in a separate paper focused on a precise determination of \( F(z) \).

In preparation for measuring the power spectrum, we divide each quasar spectrum by our estimate of \( A_q C(\lambda_{\text{rest}})(1 + \delta_C) S(\lambda) F(z) \). The power in the \( S(\lambda) \) and \( F(z) \) terms is completely negligible (always <0.5% of \( P_F(k, z) \) and usually much less). Values of \( \delta_C \) are represented for each quasar by \( N \) PCA eigenvectors. We have tried several different values for \( N \) ranging from 0 to 13, and find that the power spectrum results depend slightly (but not critically) on the value, as we discuss below. For our final results we use \( N = 0 \), i.e., we only divide by a mean continuum, although we will also show how using \( N = 13 \) affects the cosmological fit results. We do not use the PCA continua because we are not satisfied with their robustness, and division by them frequently actually increases the resulting power slightly. This may indicate that the error introduced by allowing additional freedom in the continuum is larger than the continuum fluctuations that we are trying to remove. Our study of PCA continua was primarily aimed at determining \( F(z) \) rather than the power spectrum, so we cannot be certain that the PCA method could not be used productively in a power spectrum measurement if it were more carefully optimized for that application. Because we know that our continuum estimate (which involves an extrapolation from outside the Lyα forest region) is not perfect within the Lyα forest, we further divide each chunk of spectrum that will be used to make a power spectrum estimate by its mean (optimally computed considering both observational noise and absorption variance). We call our resulting observed data vector \( \delta_f = \delta_F + \delta_S + \delta_n \), where \( \delta_F \) is the normalized noise fluctuation and we are ignoring the cross terms between \( \delta_F, \delta_S, \) and \( \delta_n \). As we describe in detail below, \( \delta_S \) is treated as a random noise background and its statistical properties are determined by measuring the power spectrum in the rest wavelength range 1268 Å < \( \lambda_{\text{rest}} < 1380 \) Å, where \( \delta_F \equiv 0 \) (and \( F \equiv 1 \)).

A small but non-negligible detail of our procedure is hidden within our description of the normalization of the spectra. When we estimate the mean to divide by, we weight the computation optimally using the covariance matrix, \( C_{ij} \), of the pixels (see further in our explanation of the power spectrum measurement below). The matrix \( C \) includes Lyα forest fluctuations and measurement noise. We do not use our best estimate of the measurement noise directly for the weighting, because the noise variance estimate is correlated with the measured flux in the pixel, which leads to a bias: the mean is underestimated because lower flux pixels have lower noise. The original noise estimate is \( \sigma_p^2 = \gamma (f_{\text{quasar}} + f_{\text{sky}}) + \sigma_{\text{real noise}}^2 \), where \( f_{\text{quasar}} \) is the flux from the quasar, \( f_{\text{sky}} \) is the flux from the sky, and \( \gamma \) accounts for the conversion between the units of flux and photons (this description is slightly idealized since the reduction of two-dimensional CCD data to a spectrum introduces some complications). To remove the correlation between flux and noise, we subtract \( \gamma f_{\text{quasar}} \) from \( \sigma_p^2 \) and add \( \gamma (f_{\text{quasar}}) \), where \( f_{\text{quasar}} = A_q C(\lambda_{\text{rest}})(1 + \delta_C) S(\lambda) F(z) \). We call the final result \( \sigma_x \) (with \( x \) for weight; see Fig. 3 for some comparisons of the noise estimates). The estimate of \( \gamma \) we have from the spectral reductio pipeline is not perfect, so our replacement of the correlated part of the noise amplitude is imperfect. We make a final, very small, correction to the mean estimation based on a direct computation of the cross-correlation between the flux and noise amplitude. We use the same decorrelated noise amplitudes for weighting the power spectrum extraction (discussed below); however, the bias is completely insignificant in that case [i.e., \( P_F(k, z) \) computed using the original noise estimates for weighting is practically identical].

3. POWER SPECTRUM DETERMINATION

The high precision of the \( P_F(k, z) \) measurement obtainable using the SDSS data sample requires unprecedented (in this field) care in the design and testing of the procedure used to produce it. We describe the basic method that we use to extract a power spectrum and estimate the errors in §3.1. In §3.2 we present the test of the full method as implemented in our code, using mock data sets. In §3.3 we give the raw result for the measurement of power in the Lyα forest region.

We aim to measure the power spectrum of \( \delta_F \), representing the correlation of fluctuations in the Lyα forest absorption only; however, the covariance matrix of the data vector \( \delta_f \) is

\[
\langle \delta_1 \delta_1 \rangle = \langle \delta_F \delta_F \rangle + \langle \delta_S \delta_S \rangle + \langle \delta_n \delta_n \rangle
\]

(the three components of \( \delta_F \) as we have defined it should be uncorrelated). The noise term in equation (3) is relatively easy to compute and subtract. We estimate and subtract most of \( \langle \delta_S \delta_S \rangle \) by defining

\[
P_F(k, z) = P_{1041, 1185}(k, z) - P_{1268, 1380}(k, z),
\]

where \( z \) is always defined by \( z = \lambda / \lambda_{\alpha} - 1 \), so that we are subtracting power measured in the same observed wavelength ranges, not the same quasar spectrum (we remind the reader that we have defined \( P_{\lambda_1, \lambda_2} \) to mean the power measured in the range \( \lambda_1 < \lambda_{\text{rest}} < \lambda_2 \)). As it appears in \( P_{1268, 1380}(k, z) \), \( z \) is the redshift of gas that would produce Lyα absorption in this part of the quasar spectrum, if it were not at a higher redshift than the quasar, i.e., \( z \) is really just an indicator of observed wavelength. The subtraction in equation (4) will completely remove the power due to transitions with \( \lambda > 1380 \) Å, including Si iv (a doublet absorbing at rest wavelengths 1393.75 and 1402.77 Å) and C iv (another doublet at 1548.20 and 1550.78 Å). Note that this subtraction of metal power is exact, not an approximation [except for the approximation that \((1 + \delta_F)(1 + \delta_S) \equiv 1 + \delta_F + \delta_S \)], because we are determining the metal power in exactly the same observed wavelength range as the Lyα forest power from which it is being subtracted, i.e., the same gas, at the same redshift, is doing the absorbing both inside and outside the forest, so the absorption will have identical statistical properties. This background subtraction will also remove any strictly observed-wavelength–dependent power introduced by the detector, such as spectrum-to-spectrum variations in the calibration of the detector. We implement it in §3.4.

3.1. Core Method

In this subsection we describe our method for extracting the power spectrum, \( P_F(k, z) \), from any selected rest wavelength
range $X$ (§ 3.1.1), and estimating its statistical uncertainty (§ 3.1.2).

### 3.1.1. Band-Power Estimation

We estimate $P_X(k)$ using the quadratic estimation method, which is essentially a fast iterative implementation of the maximum likelihood estimator (we follow the expressions as given in Seljak 1998). This method is optimal for a Gaussian probability distribution. While the power spectrum estimates are not Gaussian distributed, the deviations are small, as shown below. We measure the power in flat bands with edges given by $\log_{10}(k) = -(4.2 + 0.1)i$, where $i$ ranges from 0 to 30 (to produce 29 bands), although we will not give results for some of the large- and small-scale bands when we think they are unreliable. Defining $\delta_S = \delta_{S_X} + \delta_{S_n}$, where $\delta_{S_X}$ are the fluctuations we are measuring (e.g., $\delta_{S_S} = \delta_{S_E} + \delta_{S_5}$ within the forest) and $\delta_{S_n}$ are the normalized noise fluctuations, a band-power estimate, $\hat{P}_k$, for each chunk of spectrum is given by

$$
\hat{P}_k = \frac{1}{2} \sum_l F_{kk}^{-1}(\delta_S \delta_f^{-1} Q_k - b_k), \quad (5)
$$

where $C = \langle \delta_f \delta_f^T \rangle = S + N$, $S = \langle \delta_{S_X} \delta_{S_X} \rangle$, $N = \langle \delta_{S_n} \delta_{S_n} \rangle$, $Q_k = \partial S / \partial P_k$, $F_{kk} = 1/2 \text{tr}(C^{-1} Q_k C^{-1} Q_k)$, and $b_k$ is the noise bias

$$
b_k = \text{tr}(C^{-1} Q_k C^{-1} N). \quad (7)
$$

Note that we could include the background power explicitly in these equations as a noise source when measuring the power in the Ly$\alpha$ forest region, but we ignore this because its contribution is too small to change the weighting significantly. We will subtract it from the estimates later. The noise subtraction term, $b_k$, is computed using the pipeline noise estimates, $\sigma_{S_n}$ (not $\sigma_{S_X}$), with the amplitude corrected as discussed above based on the differences between exposures. In principle, $S$ in these equations should be the true covariance matrix; however, as we discuss below, we use the measured covariance from a previous iteration of the power spectrum determination instead.

Except in a few cases that we will identify as they arise, when we set out to measure the power in a defined rest-wavelength region (e.g., $1041 \, \text{Å} < \lambda_{\text{rest}} < 1185 \, \text{Å}$ for the Ly$\alpha$ forest region) we first use equation (5) to estimate the power separately in halves of the region in each spectrum (e.g., $1041 \, \text{Å} < \lambda_{\text{rest}} < 1113 \, \text{Å}$ and $1113 \, \text{Å} < \lambda_{\text{rest}} < 1185 \, \text{Å}$). Our choice of half-spectra is a compromise between competing desires for resolution in redshift and wavenumber. The full length of the forest in a spectrum corresponds to a redshift interval, $\Delta z \simeq 0.4$, that is unnecessarily large. While the precision of the measured power spectrum would support smaller than half-spectrum chunks to give finer redshift resolution than $\Delta z \simeq 0.2$, the shorter chunks would limit the $k$-space resolution. Note that we could have used full chunks and still achieved the same $z$-resolution by more carefully applying the estimator equation, as we discuss below, but this would increase the computational time without much improvement in the final errors on the scales of relevance.

After computing estimates $\hat{P}_k$ for each half-spectrum, we perform a weighted average to determine $P_X(k, z)$ in redshift bins centered at $z_i = 2 + 0.2i$, where $i = 1.13$ (in this paper we generally only discuss results up to $i = 9$, where the statistical errors are smallest, although we include up to $i = 11$ in Table 3). Each bin is the average of the power in all the half-spectra for which the redshift of the central pixel falls within $\pm 0.1$ of the bin center (we discuss below how we correct for an asymmetric distribution of data within a bin). We combine sets of estimates using the Fisher matrices (eq. [6]) for the weighting. In principle this means that we sum the quantity in parentheses in equation (5) over all estimates and multiply the result by the inverse of the sum of the Fisher matrices for each individual estimate. Our procedure would be optimal for Gaussian data, which the Ly$\alpha$ forest is not; however, when we use the Gaussian approximation to compute the errors on the measured power, the results are not much different from the more accurate bootstrap errors (see § 3.1.2), so we conclude that our method is not far from optimal.

Whenever we have a finite length of spectrum, there will be mixing between the power in different bins. Variable noise or gaps in the data will produce more mixing. This mixing is described in terms of a window matrix, which is given by the Fisher matrix in equation (6). In our standard procedure, the power spectrum estimates in equation (5) are multiplied by the inverse of the Fisher matrix and are thus deconvolved with the window, which removes the mixing of other modes into the bin one is estimating (however, the bins are still correlated). This method thus produces a diagonal window matrix, so each combined estimate of $P_X(k, z)$ represents exactly the range of $k$ corresponding to its bin. Our tests below show that there is no practical problem with instability in the Fisher matrix inversion (the window matrices are close to diagonal to begin with). A diagonal window matrix is desirable from a theoretical standpoint because our ability to compute the power spectrum from simulations is limited at both low $k$ (by limited box size) and high $k$ (by simulation resolution and complexity of physics). In the few cases where we use the power without deconvolution, we are using the estimator $N_k(FP_X)_k$, where $N_k \equiv (\sum_k F_{kk}^{-1})^{-1}$ (Seljak 1998).

To compute the weight matrix $C$, we need an estimate of $S$, i.e., the power spectrum we are trying to measure. We solve this problem by computing $P_X(k, z)$ iteratively. The first estimate is made assuming $S = 0$. In subsequent iterations we compute $S$ from the previous estimate of $P_X(k, z)$. This procedure converges quickly (the difference between $S = 0$ and a reasonable estimate of $S$ is significant, but once $S$ is in the right ballpark it does not matter what it is exactly). We add a large constant (10.0) to all elements of the weight matrix to remove all direct sensitivity of our power measurement to the mean of the chunk. This makes very little difference to the results on the scales we present. We are, however, still sensitive to the mean estimate from when we divided the spectrum by it. Even if the mean estimates are correct on average, the statistical error on the mean for each spectrum can still lead to a bias. If the errors on the mean estimate were small and uncorrelated with the fluctuations in the flux field, the bias would be $1 + 3 \sigma_m^2$, where $\sigma_m$ is the error on the mean [to lowest order in $\sigma_m$, i.e., the bias is $1/(1 + \delta_m^2) \simeq 1 + 3 (\delta_m^2)$, where $\delta_m$ is the fractional error in the mean, and $\langle \delta_m^2 \rangle = \sigma_m^2$]. We divide each estimate by this factor as part of our standard procedure; however, as we discuss below when we test our code on mock spectra, this approximation is not sufficient and we will need to include another small, $k$-dependent, factor determined numerically using the mock spectra (this is the only use of the mock spectra other than for testing).

The reader may at this point be wondering what redshift the resulting $P_X(k, z)$ should be taken to represent, i.e., $z_i$ is not necessarily the center of weight of the data, and neither is the mean redshift of the pixels in the bin, considering the rather
complicated weighting in equation (5). In fact, the effective redshift is not even the same for each $k$-bin in the same $z$-bin. We resolve this question—$P_X(k, z)$ represents the power spectrum at precisely $z_i$ (to first order)—in our construction of $S^{ab} = \langle \delta_i \delta_j \rangle$ and $Q^{ab} = \partial S^{ab} / \partial P_k$, where $a$ and $b$ label pixels at redshifts $z_a$ and $z_b$, and $i$ labels the redshift bin in which this chunk of spectrum falls. To account for the evolution from $z_a$ and $z_b$ to $z_i$, we define a power spectrum growth factor, $D_k(z) = [(1 + z)/(1 + z_i)]^{\alpha_k}$, where

$$\alpha_k = \frac{d \ln [P_X(k, z)]}{d \ln (1 + z)} \bigg|_{z_i} \approx \frac{\ln (P_{k,j+1}/P_{k,j-1})}{\ln ((1 + z_{j+1}),(1 + z_{j-1}))}$$

(we use a one-sided derivative estimate instead of eq. [8] for the first and last redshift bins). Now $Q^{ab} = D_k(z_{ab})Q^{ab}_{k,z_i}$, where $z_{ab} = (z_a + z_b)/2$ and $Q^{ab}_{k,z_i}$ is computed as if the pixels were located at the center of the bin. Finally, $S^{ab} = \sum_k Q^{ab}_{k,z_i} P_k(z_i)$. This correction may be difficult to understand intuitively at first, but it is really quite simple. The modification of $Q$ just corrects the power spectrum estimate for the excess (dearth) of power that we expect for pixels in the high-redshift (low-redshift) ends of the bin. The correction to $S$ affects the weighting, simply producing a more accurate $S$ at the redshifts of the pixels in question.

Note that an alternative method would be to treat the points $P_X(k, z_i)$ as simply parameters of a continuous power spectrum defined by some form of interpolation. This would mean $S^{ab}$ would have nonzero derivative with respect to more than one of the power spectrum bins (e.g., usually two for linear interpolation). This method would be elegant and probably produce narrower effective window functions in the $z$ direction; however, it will increase the correlation in the $z$ direction between measurement errors, because the same pixels would contribute to more than one power spectrum point. Since this more sophisticated method would allow long chunks of spectra to be used without degrading our $z$ resolution, it would be most useful if we were trying to measure the power on even larger scales.

How does our method compare to the straightforward Fourier transform (FT) method? The basic FT method is to project the data vector, $\delta_i$, onto a set of modes of the form $d_{\alpha,\beta} = \exp (i k_n \Delta y_i)$ and to simply compute the variance of the amplitudes of all the modes with $k$ in some bin, i.e.,

$$\hat{P}_{FT} \propto \sum_{\alpha} \sum_{\beta} |d_{\alpha,\beta}|^2 \delta_{\alpha,\beta},$$

where $k_{\text{min}}$ and $k_{\text{max}}$ define the bin, and the discrete spacing of $k_n$ is somewhat arbitrary (the natural spacing is $\Delta k = 2\pi/L$, where $L$ is the length of the spectrum, but nothing prevents one from choosing more finely spaced $k$-values). Our estimator, equation (5), can be cast in a similar form, i.e., as a projection of the data vector onto a set of modes, and a sum of the squares of the mode amplitudes. We require that the mode amplitudes are statistically independent, which makes their computation equivalent to a computation of Karhunen-Loève eigenmodes (see, e.g., Tegmark et al. 1997). Figure 7 shows the two most important modes for our bin with $0.00126 < k < 0.00158$ s km$^{-1}$, for the chunk of spectrum shown in Figure 3a.

In this case two modes differing primarily by a phase shift, analogous to $\sin (k x)$ and $\cos (k x)$, contain most of the information, because our bin width is approximately $2\pi/L$. We see that the difference between our modes and a simple sine wave is not dramatic; there is a little bit of edge tapering (downweighting the edges to make the effective window on the data more compact in Fourier space) and some straightforward down weighting of the most noisy pixels. Curiously, there seems to be an additional effect where pixels adjacent to an edge are given extra weight, possibly as a way of compensating for missing data (this is seen more clearly in spectra where a narrow gap is present in the middle of the data). The picture is similar for bins with larger $k$, except of course that there are increasingly many important modes as the width of our bins increases (the bins have a fixed width $\Delta \log (k)$, but the relevant mode width is $\Delta k$). For more discussion of the quadratic estimator see, e.g., Tegmark (1997).

The method we adopt is optimal for Gaussian fields and therefore guarantees that no other method can surpass it. An additional advantage is that within this formalism window and covariance matrices are automatically computed. For continuous spectra with few gaps and near-uniform noise one does not necessarily expect an FT method to be significantly worse. In practice the noise level is slowly varying across the spectrum, so averaging all the pixels uniformly is not optimal and degrades performance. Another advantage is that with our method each pixel pair has its own effective redshift and the correlations for a given pair are then interpolated to the redshift of interest using the appropriate evolution. In the FT method the whole spectrum is Fourier transformed first, so the redshift information is preserved only in an averaged sense, but a priori it is not clear how this average is defined.

3.1.2. Bootstrap Error Estimation

While the Fisher matrix obtained during the estimation process would give the error matrix for $P_X(k, z)$ if the data were Gaussian, we cannot reliably assume this. Our solution is to compute a bootstrap error matrix by the standard procedure (Press et al. 1992). From our data set of $N$ spectra, we form a bootstrap data set by selecting $N$ spectra at random, with replacement. The covariance matrix of $P_X(k, z)$ is taken to be $M^B = \langle \Delta \hat{P}^B_X \Delta \hat{P}^B_X \rangle$, where...
where \( \Delta \bar{P}_k = \bar{P}_k - \langle \bar{P}_k \rangle \), \( \bar{P}_k \) is an estimate of the power in the \( k \)th bin from a bootstrap data set and angle brackets denote the average over bootstrap realizations. We generally use 4000 realizations, after checking that this produces convergence in the result. We assume that the error correlations extend only one bin off diagonal in the \( z \) direction, because the spectrum of a single quasar practically never contributes to nonadjacent bins.

We have no compelling reason to believe that this method of computing the error bars will give rigorously correct results. Considering the large number of off-diagonal elements that must be estimated, one worry is that a particular linear combination of the bins may accidentally vary very little in our data set, so it will appear to have an unrealistically small error. Our tests on mock spectra (§3.2.2) show no sign of this problem. Still, to be conservative we apply one tweak to \( M \) after it is computed, in an attempt to inoculate it against the possibility. We perform a singular-value decomposition on \( M \), which produces a set of independent vectors and their variances. We then compute the variances of the same vectors under the Gaussian approximation, using the Fisher matrix. If the bootstrap variance is smaller than the Gaussian variance, we replace it with the Gaussian variance. Finally, we transform back to \( M \). The tests on mock samples described below give us confidence that our procedure is reliable.

3.2. Tests on Mock Data Sets

We validate our procedure as implemented in code by applying it to mock data sets. Many iterations of these tests were required to produce results that show no serious problems in the error estimation or the power spectrum estimation itself. Testing the results on realistically created mock data is absolutely essential for measurements of such high precision. In §3.2.1 we describe our procedure for generating the mock spectra. We test our bootstrap error estimates in §3.2.2. Finally, we test the power spectrum estimation procedure for systematic errors in §3.2.3.

3.2.1. Generating Mock Spectra

We generate mock spectra by combining the auxiliary information we have for each observed spectrum (e.g., our continuum estimate, noise estimate, sky estimate, etc.) with a simplified version of the Bi et al. (1992) model for the \( \exp (-\tau) \) field, which results in realistic looking spectra.

For each observed quasar we start with the term we divide by before computing the power spectrum, \( A_0 C(\lambda_{\text{rest}})(1 + \delta_c)S(\lambda) \) (see eq. [2]). We multiply this by \( \exp (-\tau) \) (generated as described below), smooth the result using the resolution from the observed spectrum, and sample the result onto the observed grid of pixels. This produces a noise-free version of the flux we would observe coming from this quasar. We add flux from the sky as estimated for the observed spectrum and transform the total flux to the number of photons that would be expected in each pixel. We generate a Poisson deviate with this mean, add the appropriate Gaussian read noise for each pixel, transform back to the original flux units, and subtract the sky flux estimate to obtain an observed (noisy) quasar spectrum. The results of this procedure for each observed quasar are written into files in the same format as the observed spectra, so exactly the same code can be used to measure the power in the mock spectra.

To generate the \( \exp (-\tau) \) fields we use a simple model that is arranged to give roughly the correct power spectrum as a function of \( k \) and \( z \), and the correct mean absorption as a function of redshift. For each observed spectrum, we start by generating a Gaussian random field, \( \bar{b}_{i,0} \), on a very long, relatively finely spaced grid (65,536 cells with width 7 km s\(^{-1} \), to be precise), with power spectrum

\[
P_{\delta}(k) = \frac{1 + [0.01 \text{ s km}^{-1}/k_0]^{\nu w}}{1 + (k/k_0)^{\nu w}} \exp \left[-(kR_0)^2\right],
\]

where \( k_0 = 0.001 \text{ s km}^{-1} \), \( \nu = 0.7 \), and \( R_0 = 5 \text{ km s}^{-1} \) [this \( P_{\delta} \) was chosen after some experimentation because it produces a final flux power spectrum with approximately the same \( k \) dependence as the observed \( P_{\delta}(k, z) \)]. An arbitrary cell in this grid is chosen to correspond to the redshift of the quasar, and the evolution of the amplitude of the power spectrum with redshift is imposed by the transformation \( \tilde{a}_i = a(z_i)\delta_{i,0} \) with \( a(z_i) = 58.6/(1 + z_i)^{4/\nu w} \), where the form of \( a(z) \) was chosen so that the final flux power spectrum would evolve like the observed one. Next we make the squared lognormal transformation \( n_i = [(\exp (\bar{b}_i - \sigma_i^2/2))^2] \), where \( \sigma_i^2 \) is computed from the input power spectrum, including the amplitude factor (the factor \( \sigma_i^2/2 \) in the exponential just fixes the mean of the lognormal field to 1). We smooth the \( n \) field with a Gaussian filter with rms width \( R_s = 20 \text{ km s}^{-1} \) and multiply it by a factor 0.374 \((1 + z_i)^{4/\nu w}\) to produce a field \( \tau \) (this redshift evolution factor produces roughly the observed redshift evolution of \( F \)). The mock transmitted flux in each grid cell is then \( F_i = \exp (-\tau_i) \), which is sampled as described above.

The procedure described above leads to realistic looking spectra of the Ly\( \alpha \) forest. We have verified that it generates a bispectrum that is within a factor of 2 of the one measured in \( N \)-body simulations. The main advantages of this procedure over the \( N \)-body simulation approach when generating the mock spectra are that it is faster, so one can make an arbitrary number of independent realizations, and that the simulated spectra can be of arbitrary length, important to eliminate any periodicity effects (this would be impossible with simulations, where a typical box size is much shorter than the total length of a single spectrum). Both of these advantages are critical for a high-precision test. We determine the true \( P_{\delta}(k, z) \) by a simple FFT of extremely long \( \exp (-\tau) \) fields (without redshift evolution).

3.2.2. Tests of the Error Estimates

Without accurate statistical errors it is difficult to identify systematic problems, so we first test our bootstrap procedure for estimating the errors. Note that there is no reason to expect bootstrap errors to be perfect (there is even some ambiguity in how exactly the bootstrapping should be done when the data do not consist of statistically identical objects). Regardless of systematic errors in the method, the only difference between the power spectra measured from two mock data sets that differ only in the random seed that was used to create them should be the statistical errors that we estimate. We test our error bars by generating 10 different sets of mock data and computing \( \chi^2 \) for the differences between each of them and their error-weighted mean, using the bootstrap error bars and the 108 points in 0.0013 s km\(^{-1} \) < \( k < 0.02 \text{ s km}^{-1} \), and \( 2.1 < z < 3.9 \). This is effectively a fit of 108 parameters to 1080 data points, with 972 dof. The total \( \chi^2 \) is 939, perfectly consistent with a random fluctuation around the mean, and strongly disfavoring an underestimation of the errors by more than a couple percent.

3.2.3. Tests of the Power Estimates

We can now search for systematic errors. To enhance the statistical significance of any errors, we average our 10 sets of mock spectra to form a single, more precise measurement. The result is shown in Figure 8.
The results look reasonably good; however, we find an unacceptably bad $\chi^2 = 346$ for the comparison between our measured power spectrum and the true power spectrum (there are 108 dof). To quantify the systematic problem, we first assume the bias has the form $B(k) = B_0 [k/0.0067 \text{ s km}^{-1}]^\nu \equiv P_{\text{measured}}/P_{\text{input}}$ and fit for the values of $B_0$ and $\nu$ that minimize $\chi^2$ in the comparison. We find $B_0 = 1.0036 \pm 0.0014$ and $\nu = -0.0173 \pm 0.0013$ with $\chi^2 = 173$ for 106 dof (the pivot point $k_0 = 0.0067 \text{ s km}^{-1}$ was chosen to make the errors independent; the amplitude coefficient would be larger if we were not already dividing by $k$). The combination of slope and amplitude errors corresponds to a 3.1% excess of power at our largest scale, $k = 0.0014 \text{ s km}^{-1}$, and a 1.3% underestimate at $k = 0.018 \text{ s km}^{-1}$. We find some less significant dependencies by generalizing the fitting formula even more to

$$B(k, z) = B_0 a^\nu(z) \left( \frac{k}{k_0} \right)^{\nu+1/2} \ln \left( k/k_0 \right) + \zeta \ln \phi(z),$$

where $a(z) = (1 + z_0)/(1 + z)$, with $z_0 = 2.85$. The parameters are $B_0 = 1.0073 \pm 0.0016$, $\mu = 0.049 \pm 0.012$, $\nu = -0.0195 \pm 0.0015$, $\eta = -0.0157 \pm 0.0038$, and $\zeta = -0.026 \pm 0.012$, with $\chi^2 = 135$. Where does this bias come from? We expect some bias related to the division of each chunk of spectrum by its overall mean (not because of an integral constraint suppression of large-scale power—our estimator should take care of that—but because of statistical error in the estimate of the mean that we divide by). When we measure the power without this division by the mean, which we can only do using mock spectra, we find significantly smaller corrections—small enough to ignore when model fitting.

We expect that this bias should be present when we use real observed spectra, so we will correct for it by dividing the measured power by $B(k, z)$. We describe its effect on the amplitude and slope of the power spectrum below (Table 1).

### 3.3. Raw Power

Figure 9 shows the raw power measured in our standard Lyα forest rest wavelength range, 1041 Å < $\lambda_{\text{rest}}$ < 1185 Å.

All the figures in this subsection show $P_{0.0015, 1185}$, not the background subtracted power $P_F$. Our normalization convention is

$$\langle \delta^2 \rangle = \int_\infty^- \frac{dk}{2\pi} P(k).$$

We usually plot the dimensionless quantity $\Delta^2(k) \equiv \pi^{-1} k P(k)$, the contribution to the variance per unit $\ln k$.

Figure 10 shows the fractional errors on all of the measured points. The errors are less than 5% for most of the points, and frequently as small as 3%. If we were only estimating a single amplitude parameter by combining all these points then its error would be 0.6%. An overall logarithmic slope would have an error ±0.005. The errors on the largest scales are increased somewhat by the diagonalization of the window matrix.

Figure 11 shows the ratio of subtracted noise power to measured signal power ($P_{0.0015, 1185}$) for each point. The noise power is significant (20%–30%) on all scales but diverges at high $k$ where the resolution suppresses the absorption power. The lowest-redshift bin has the most noise, due to the lower Lyα forest power combined with extra noise at the short wavelength end of the spectra.

Figure 12 shows our window matrix (at $z = 2.6$), which we proceed to diagonalize. The matrix is reasonably close to diagonal already, with large contributions only from adjacent bins. It is useful to diagonalize the matrix at this stage, rather than waiting until the model-fitting stage, because this allows us to compute bootstrap errors directly for the final bins (the bootstrap error calculation and window matrix diagonalization do not perfectly commute).

Figure 13 shows the ratio of the bootstrap errors to the errors estimated assuming the data are Gaussian. We did not apply the Gaussian floor to the bootstrap errors when making this figure. Typically, the bootstrap errors are 0%–20% larger than the Gaussian errors. Figure 14 shows examples of the estimated correlation between the errors, at $z = 2.6$. We note that diagonalizing the window matrix noticeably reduces the error correlations. The bootstrap errors are, in contrast to the Gaussian errors, noticeably correlated $\left( \langle \Delta P_i \Delta P_j \rangle / \sigma_{P_i} \sigma_{P_j} \right) \sim 0.0 – 0.2$ when $|i – j| > 1$, where $i$ and $j$ label the bins) across the full $k$ range. These differences between bootstrap and Gaussian errors are not necessarily an indication of intrinsic non-Gaussianity in the absorption fluctuations. Possible alternative explanations for the differences include the uncertainty in the mean flux value that each chunk of spectrum is divided by and the uncertainty in the noise-subtraction term for each chunk, neither of which are included in the Gaussian estimate and both of which would increase the error in a way that is correlated across $k$ bins.

### 3.4. Background Subtraction

Our background subtraction is the power in the wavelength range 1268 Å < $\lambda_{\text{rest}}$ < 1380 Å. Figure 15 shows $P_{1268, 1380}$ and $P_{0.0015, 1185}$ for comparison.

The bump at $k \sim 0.013 \text{ s km}^{-1}$ in $P_{1268, 1380}$ is probably due to the $C iv$ doublet at separation 499 km s$^{-1}$. The bump at
## Table 1

| Variant*                          | $\delta \ln \Delta^2$ | $\sigma_{\ln \Delta^2}$ | $\delta n_{\text{eff}}$ | $\sigma_{n_{\text{eff}}}$ | $\Delta \chi^2$ $^b$ | $\chi^2$ $^c$ |
|-----------------------------------|------------------------|--------------------------|--------------------------|--------------------------|-----------------------|----------------|
| Standard fit.                     | 0.00                   | 0.10                     | 0.024                    | 0                         | 129.7                 | 129.7         |
| No window diagonalization.        | -0.06                  | 0.10                     | -0.024                   | 0.021                     | 1.4                   | 129.5         |
| No Si iii correction.             | 0.04                   | 0.10                     | 0.017                    | 0.021                     | 0.7                   | 193.7         |
| $z$-independent Si iii.           | 0.00                   | 0.10                     | 0.00                     | 0.024                     | 0.0                   | 130.9         |
| Variable width Si iii.            | -0.02                  | 0.11                     | -0.006                   | 0.025                     | 0.1                   | 129.0         |
| Include Si iii–Si iii term.       | -0.02                  | 0.10                     | -0.003                   | 0.023                     | 0.0                   | 132.1         |
| $\sigma_{\text{noise power}}$ = 0.5% | 0.00                   | 0.10                     | -0.000                   | 0.024                     | 0.0                   | 130.3         |
| $\sigma_{\text{noise power}}$ = 50% | -0.08                  | 0.12                     | 0.008                    | 0.025                     | 1.1                   | 123.8         |
| No individual noise reestimation. | -0.02                  | 0.10                     | 0.002                    | 0.023                     | 0.1                   | 128.1         |
| Believe pipeline noise.           | -0.20                  | 0.11                     | 0.019                    | 0.021                     | 9.6                   | 129.9         |
| No Gaussian floor on errors.      | 0.01                   | 0.10                     | 0.002                    | 0.024                     | 0.0                   | 133.1         |
| Gaussian errors.                  | 0.02                   | 0.10                     | 0.001                    | 0.023                     | 0.1                   | 151.7         |
| Ignore error correlations.        | 0.04                   | 0.11                     | 0.002                    | 0.027                     | 0.2                   | 117.2         |
| No background subtraction.        | -0.20                  | 0.10                     | -0.022                   | 0.019                     | 3.6                   | 169.6         |
| Background 1409–1523 Å.           | 0.05                   | 0.10                     | 0.028                    | 0.025                     | 1.5                   | 133.2         |
| No background noise matching.     | -0.07                  | 0.10                     | -0.004                   | 0.021                     | 0.7                   | 142.3         |
| Previous, but use 1409–1523 Å.    | -0.06                  | 0.10                     | 0.008                    | 0.022                     | 1.7                   | 143.1         |
| No automated BAL cut.             | 0.01                   | 0.10                     | 0.003                    | 0.023                     | 0.0                   | 127.0         |
| Previous, but use 1409–1523 Å.    | 0.14                   | 0.10                     | 0.094                    | 0.025                     | 16.0                  | 156.6         |
| (70 km s$^{-1}$)$^2$ resolution error | -0.11                  | 0.14                     | 0.015                    | 0.024                     | 1.4                   | 126.4         |
| (0.7 km s$^{-1}$)$^2$ resolution error | 0.02                  | 0.10                     | -0.003                   | 0.024                     | 0.3                   | 130.4         |
| No continuum division.            | 0.02                   | 0.10                     | 0.002                    | 0.024                     | 0.1                   | 132.2         |
| PCA continuum division.           | 0.02                   | 0.10                     | -0.010                   | 0.024                     | 0.8                   | 139.1         |
| No reduced sensitivity to mean.   | -0.00                  | 0.10                     | -0.002                   | 0.024                     | 0.0                   | 130.5         |
| No bin-redshift correction.       | -0.02                  | 0.10                     | -0.006                   | 0.023                     | 0.1                   | 137.2         |
| Ignore $F_{-\sigma_p}$ correlation.| 0.00                   | 0.10                     | 0.002                    | 0.024                     | 0.0                   | 128.7         |
| Rescale $\sigma_p$                | 0.00                   | 0.10                     | 0.002                    | 0.024                     | 0.0                   | 129.0         |
| No code bias correction.          | 0.06                   | 0.10                     | -0.013                   | 0.024                     | 3.0                   | 132.3         |
| 8000 bootstrap sets.              | 0.00                   | 0.10                     | 0.000                    | 0.024                     | 0.0                   | 128.5         |
| Fixed nuisance parameters.        | 0.036                  | 0.021                    | -0.025                   | 0.012                     | ...                  | ...           |
| Optimistic $\tilde{F}$.           | 0.068                  | 0.062                    | 0.009                    | 0.019                     | ...                  | ...           |
| Optimistic $T_1$s, $\gamma$.      | 0.002                  | 0.082                    | -0.014                   | 0.021                     | ...                  | ...           |
| Optimistic $\tilde{D}$, $T_1$s, $\gamma$. | 0.002   | 0.051                    | -0.016                   | 0.018                     | ...                  | ...           |
| $P_F(k, z)$ errors divided by $\sqrt{3}$. | -0.051               | 0.081                    | -0.000                   | 0.015                     | ...                  | ...           |

* $z_p = 2.6$, $k_p = 0.009$ km s$^{-1}$.

* The meaning of each variant is explained in § 4.3.

* $\Delta \chi^2$ of the fitted parameters relative to the standard parameters, using the errors from the variant fit.

* $\chi^2$ for the fit (essentially unrelated to $\Delta \chi^2$).
$k \sim 0.003 \text{ s km}^{-1}$ may be due to the Si iv doublet at separation 1933 km s$^{-1}$. Figure 16 shows $P_{1268, 1380}/P_{1041, 1185}$.

We see that, even though the background power is a small fraction of the Ly$\alpha$ forest power, it is quite significant when compared to the small size of the errors on the Ly$\alpha$ forest power.

It is important to remember that even a small overall systematic error can be very significant if it covers many data points (e.g., a 1/2 $\sigma$ error over 100 points shifts the mean by 5 $\sigma$).

We are going to subtract the power in the range 1268–1380 Å from the Ly$\alpha$ forest power, but it is informative to measure the power at other places in the quasar rest frame for comparison. The range 1409–1523 Å includes C iv absorption (at 1548.2 and 1550.78 Å) but excludes Si iv (at 1393.75 and 1402.77 Å)
and shorter wavelength transitions. Figure 17 shows $P_{1409, 1523}/P_{1041, 1185}$.

If all of the power was coming from metal line absorption, the power in the range $1409 < \lambda_{\text{rest}} < 1523$ A should always be less than the power in the range $1268 < \lambda_{\text{rest}} < 1380$ A. As we see in Figure 18, which shows the difference in the background fractions, $(P_{1268, 1380} - P_{1409, 1523})/P_{1041, 1185}$, the power in $P_{1268, 1380}$ is greater than $P_{1409, 1523}$ except on large scales.

The difference on large scales suggests that there is tiny amount of power left in the quasar continua (in spite of our division by the mean continuum), which is larger in the range $1409–1523$ A than in the range $1268–1380$ A. Finally, Figure 19 shows $P_{1558, 1774}/P_{1041, 1185}$, past the wavelength of C IV absorption.

The reduction of power relative to shorter wavelengths is dramatic, but not surprising since C IV is the most common metal absorber. It does suggest, however, that most of the power is due to metals and not continuum fluctuations, unless the continuum in the range $1558–1774$ A has significantly less power relative to...
other intervals studied here (which is admittedly not inconceivable). It seems likely, although we are not certain, that the $z = 2.2$ background power has a noticeable contribution from measurement-related problems, because the alternative is a very sudden increase in metal absorption power.

What is the upshot from these studies? The metal absorption appears to contribute a small but significant amount of power, which should also appear in the Ly$\alpha$ forest region. We subtract this power from the power measured in the forest. There is some indication of measurement-related problems in our lowest-redshift bin. The power contributed by deviations of the quasar continua from their mean appears to be small.

While the idea that $P_{1268, 1380}$ contains almost exclusively power due to simple metal absorption seems plausible at this point, when we perform consistency checks in § 4.4 we find evidence that this is not the case. Splitting the data set used to measure $P_{1268, 1380}$ in half based on the noise level in each
from $P_{1041,1185}$ to obtain $P_F$, we use $\sigma_w$ computed in the 1041–1185 Å wavelength range to compute the appropriate subtraction term. Figure 20 shows the extra power subtracted through equation (13), beyond what we would subtract if we simply used $P_{1268,1380}(k, z)$ from Figure 16. It is typically less than 4% of the Ly$\alpha$ forest power, but rises to 10% at the highest $k$ for the lowest redshift.

The reader who is paying attention may complain that we have no compelling reason to believe that this source of noise-dependent power that we do not understand depends on noise in the same way inside and outside the Ly$\alpha$ forest region. This would be true, except that when we follow this prescription for background subtraction the differences in $P_{1041,1185}$ between subsamples are removed (see § 4.4). This would be a remarkable coincidence if our model for the subtraction was not substantially correct.

Note that the background power has much smaller (absolute) statistical errors than $P_{1041,1185}$, mostly because there is simply less power, but also because there are more quasars probing a given redshift interval.

4. CONSISTENCY CHECKS

In § 4.1 we describe how we use fits of theoretical models to the $P_F(k, z)$ results to help understand the importance of any systematic errors. We plot the correlation function of the Ly$\alpha$ forest in § 4.2 and use it to identify a significant contribution to $P_F(k, z)$ from Si in absorption. In § 4.3 we examine the effects of modifications of our procedure. In § 4.4 we break the data set up in many different ways to look for dependencies of $P_F(k, z)$ that should not exist. In § 4.5 we discuss the possibility that continuum fluctuations contribute significant power. Finally, we compare our results to past measurements in § 4.6.

4.1. Rudimentary Fitting of Theoretical Models

The ultimate purpose of measuring the Ly$\alpha$ forest power spectrum is to determine cosmological parameters by comparing the observed $P_F(k, z)$ to the predictions for different cosmological models. For the ΛCDM models supported by present observations, the universe is nearly Einstein–de Sitter at $z > 2$, so cosmology influences the Ly$\alpha$ forest almost entirely through the linear theory power spectrum of the mass fluctuations, $P_L(k, z)$, at $z \sim 3$ and $k \sim 0.01$ s km$^{-1}$ (roughly 1 comoving $h$ Mpc$^{-1}$, depending somewhat on the model). We usually parameterize $P_L(k, z)$ by its amplitude, $\Delta^2(k_p, z_p) \equiv k_p^3 P_L(k_p, z_p)/2\pi^2$, slope $n_{\text{eff}}(k_p, z_p) \equiv d \ln P_L/d \ln k|_{k_p, z_p}$, and curvature $\alpha_{\text{eff}}(k_p, z_p) \equiv d^2 \ln P_L/d \ln k|_{k_p, z_p}$, where we use $k_p = 0.009$ s km$^{-1}$ and $z_p = 2.6$ as the pivot points.

A full discussion of the details of theoretical modeling of $P_F(k, z)$ using numerical simulations is beyond the scope of this paper. Furthermore, the theory of the Ly$\alpha$ forest is perhaps less certain than the observations, so we want to present the observational results untarnished by theoretical interpretation. However, it is very useful to interpret the possible systematic errors in the appropriate context of cosmological model fitting. In other words: without model fitting, it is difficult to know how important a given change in $P_F(k, z)$ is.

In this paper we take a cautious approach to the theoretical model fitting—we perform fits to different estimates of $P_F(k, z)$ computed using modifications of the extraction procedure or different subsamples of the data; however, we do not give the central result, only the deviations in the results from the value obtained from our preferred $P_F(k, z)$. These deviations in fitting results should give the reader a useful indication of the importance of
systematic effects in the data, regardless of the reader’s opinion of the theory.

The simulations and fitting procedure that we use are described in McDonald et al. (2004), where we present the final result. We use an $\Lambda CDM$ transfer function and perform the fit with $\Delta^2$ and $H_{\text{eff}}$ as free parameters (because $\alpha_{\text{eff}} = dH_{\text{eff}}/d \ln k$ is not tightly constrained by the present Ly$\alpha$ forest data alone, we fix the primordial running $\alpha = dH/d \ln k$, not to be confused with $\alpha_{\text{eff}} \sim -0.2$, to zero). Unless otherwise specified, we perform the fits using the 108 $P_T(k, z)$ points in the ranges $0.0013 \text{ s km}^{-1} < k < 0.02 \text{ s km}^{-1}$ and $2.1 < z < 3.9$. We allow for some error in our noise estimate by permitting the noise subtraction terms to vary independently in each redshift bin by 5% (nine extra free parameters to fit for, constrained by Gaussian likelihood function with this rms width). We also allow a single overall parameter describing the squared resolution error to vary with rms constraint $(7 \text{ km s}^{-1})^2$.

The Ly$\alpha$ forest model in the simulations is controlled by the externally constrained functions $F(z)$, the mean absorption, $T_{1,4}(z)$, the temperature at overdensity $1.4$, $(\gamma - 1)(z)$, the logarithmic slope of the temperature-density relation, and a reionization parameter that we will call $x_{\text{rei}}$. The function $F(z)$ is described in our fits by the 10 parameter formula

$$F_i = \mathcal{F}_i,$$

where $i$ labels our nine redshift bins, $\mathcal{F}_i$ gives the arbitrarily normalized $z$ dependence and $\mathcal{F}$ is an overall normalization. We have performed a preliminary analysis using the formalism in §2.5 to measure $\bar{F}(z)$ from SDSS data and we use this to constrain the parameters $\mathcal{F}_i$ (the error on each redshift bin is $\sim 0.005$). Because the SDSS analysis does not constrain the overall normalization, we leave $\mathcal{F}$ free except for the additional constraint that we require $\mathcal{F}$ interpolated to $z = (3.9, 3.0, 2.4)$ to match the HIRES constraints $\bar{F} = (0.458 \pm 0.034, 0.676 \pm 0.032, 0.816 \pm 0.023)$ [see McDonald et al. 2000; we have modified the numbers slightly and increased the errors to allow for systematic uncertainties, as discussed in Seljak et al. 2003]. We parameterize $T_{1,4}(z)$ and $(\gamma - 1)(z)$ by quadratic functions of $z$ (3 parameters each) with the external constraints $T_{1,4} = (20, 100 \pm 3400, 20, 300 \pm 2400, 20, 700 \pm 2800)\text{K}$ and $\gamma - 1 = (0.43 \pm 0.45, 0.29 \pm 0.3, 0.52 \pm 0.14)$ at $z = (3.9, 3.0, 2.4)$ (see McDonald et al. 2001); we added 2000 K in quadrature to the temperature errors to allow for systematic errors). Note that there are other, sometimes more precise, measurements of $\bar{F}$ (Schaye et al. 2003; Bernardi et al. 2003) and the temperature-density relation (Schaye et al. 2000; Ricotti et al. 2000) in the literature; our choice of McDonald et al. (2000, 2001) for this example is simply for convenience. In our analysis paper (McDonald et al. 2004), we show that the result is insensitive this external constraint on $\bar{F}$ and somewhat sensitive to this constraint on the temperature-density relation, and we argue that we would rather not impose tighter constraints and risk introducing systematic error. The redshift of reionization and postionization temperature of the gas influence Ly$\alpha$ forest predictions because the smoothing of the gas on small scales depends on its pressure history. At the level of precision we care about, this dependence can be captured by a single parameter. In our modeling, we use $x_{\text{rei}}$ to interpolate between two reasonable boundaries, reionization heating of the gas to 25,000 K at $z = 7$ or to 50,000 K at $z = 17$, both of which are consistent with our temperature constraints $T_{1,4}(z)$. However, in this paper we fix $x_{\text{rei}}$, because it is weakly constrained by the data and the hard lower limit we have to impose on the redshift of reionization leads to non-Gaussian errors on the power spectrum parameters we are interested in (this is a problem of presentation, not of principle).

Figure 21 shows our first fit to the standard $P_T(k, z)$ results. The value of $\chi^2 = 193.7$ is much too high for approximately 106 dof (we are marginalizing over a large number of nuisance parameters, but these generally are externally constrained so they do not reduce the number of degrees of freedom). Including $\alpha_{\text{eff}}$ as a free parameter does not improve the fit significantly. It appears that much of the disagreement comes from bumps in the power spectrum, e.g., at $k \sim 0.003 \text{ s km}^{-1}$. This motivates us to look at the correlation function of the flux.

### 4.2. The Correlation Function and the Si iii Cross-Correlation

Sometimes features in the power spectrum are easier to understand by looking at the correlation function, $\xi(v) = \langle \delta(x) \delta(x+v) \rangle$ ($v$ is as usual a stand-in for wavelength differences, as is $x$ in this case). We show the normalized correlation function, $\xi(v)/\xi(0)$ for the first six redshift bins in Figure 22.

The correlation function shows the expected behavior—positive for small $v$, negative for large $v$—except for an obvious bump at $v \sim 2200 \text{ km s}^{-1}$. We focus on this bump in the inset panel of Figure 22. The most likely explanation seems to be cross-correlation between Ly$\alpha$ and Si iii absorption. Si iii absorbs at $\lambda = 1206.50 \text{ Å}$, so the Si iii absorption by gas at some point along the line of sight will appear in the spectrum separated by 2271 km s$^{-1}$ from the Ly$\alpha$ absorption by the same gas. We see that the bump in $\xi(v)$ appears at this separation, and note that the features that ruined our power spectrum fit appear at the expected multiples of $k = 2\pi/2271 = 0.0028 \text{ s km}^{-1}$. Note that this is the only correlation seen; another metal correlation one might expect to see is N v ($\lambda = 1238.8, 1242.8 \text{ Å}$), but there is no apparent feature at the corresponding velocity differences ($\Delta v \sim 5600, 6700 \text{ km s}^{-1}$), as seen in Figure 22.

What should we do about this Si iii-Ly$\alpha$ cross-correlation, since the poor $\chi^2$ suggests that it is too significant to ignore? Our
first guess might be that the Si iii-Lyα correlation is a simple offset version of the Lyα-Lyα correlation, i.e., something like \( \xi_{\text{Si,}iii} - \xi_{\text{Lyα-Lyα}} \propto \xi_{\text{Lyα-Lyα}}(v) - 2271 \text{ km s}^{-1} \). The simplest way to model this is to assume that the Si iii structure is equal to that of the Lyα forest up to an overall normalization, \( \delta_P = \delta(v) + a \delta(v + v_3) \), where \( \delta(v) \) is for Lyα only and \( v_3 = 2271 \text{ km s}^{-1} \). The corresponding correlation function is

\[
\xi_P(v) = (1 + a^2)\xi(v) + a\xi(v + v_3) + a\xi(v - v_3),
\]

with corresponding power spectrum

\[
P_P(k) = (1 + a^2)P(k) + 2a \cos(v_3k)P(k),
\]

where unsubscripted \( \xi \) and \( P \) are understood to mean Lyα-Lyα. For our first fit to \( P_P(k, \ z) \) accounting for Si iii using equation (15), we assume \( a = \bar{F}/[1 - \bar{F}(z)] \), with \( \bar{F} \) as a single extra free parameter of the fit. We find a remarkable improvement in \( \chi^2 \), from 193.7 to 130.9. We find \( f \sim 0.011 \) (\( a \sim 0.04 \), depending on the redshift). The small value suggests that the relative contribution of Si iii to the autocorrelation is \( a^2 < 0.004 \), which will not affect our fit results significantly (see § 4.3). We thus only need to estimate the cross-correlation term. We also tried allowing a power-law \( 1 + z \) dependence for \( f \), but the improvement in fit, \( \Delta \chi^2 = 1.1 \), was not significant.

In the inset panel of Figure 22 we plot scaled \( \xi(v - 2271) \) to show how the shape of the bump compares to the zero-lag correlation. It is difficult to compare the shapes by eye, because of the slope of the underlying correlation, but it appears that this model explains the cross-correlation reasonably well. We can allow for a change in scale using the slightly more general form

\[
\xi_P(v) = \xi(v) + a\xi(v + v_3) + a\xi(v - v_3).
\]

Allowing \( s \) to vary freely only improves \( \chi^2 \) by 0.7 (note that the logarithmic \( k \)-binning that we use may reduce our ability to constrain these parameters). The error bars on other parameters may increase when we include \( z \) dependence of \( f \) and allow \( s \) to be free, so to be conservative one probably wants to leave them free even though they are not needed. In our standard fit in this paper, we allow \( f \) to have \( z \) dependence but fix \( s = 1 \). We show the improved fit to \( P_P(k, \ z) \) in Figure 23.

4.3. Modifications of the Procedure

The pipeline developed for this analysis includes many improvements and corrections that were added throughout the development. It is worth taking a step back to ask how important the various corrections are for the final result. Table 1 lists various modifications of our procedure (described individually below) and quantifies their effects on the fit results. In each case we show the change in the best-fit \( \Delta^2 \) and \( n_{\text{eff}} \) relative to our standard fit, and their error bars for comparison to the standard fit. We give \( \chi^2 \) to indicate the goodness of fit of the theory to the modified measured \( P_P(k, \ z) \). We reiterate that we are not asserting the correctness of the theory that we use in the fitting; we give these \( \chi^2 \)-values and other fitting results only to show trends. We list \( \Delta \chi^2 \) between the standard procedure best fit and the variant best fit to give an indication of how significant the deviation is (this is necessary because the errors are correlated so simply knowing \( \Delta^2 \) and \( n_{\text{eff}} \) and their errors does not give the full picture—see Fig. 26 for an example of the full error contours). Because the statistical fluctuations between these different fits should be small, a 1 or difference (or even less) should be interpreted as “significant,” however, since we believe that our
standard fit is better or more conservative than all of the vari-

ants, our systematic error should generally be smaller than the
deviation shown. Note that where applicable the changes in
procedure are only applied to the $P_{1041,1185}$ calculation, not the
$P_{1268,1380}$ result that is used in the background subtraction
(small changes in $P_{1268,1380}$ have no effect on the final results).

Our first variant is to not diagonalize the window matrix.
Figure 24 shows the measured power spectrum before and after
diagonalization. Figure 25 shows the ratio of the diagonal errors
after diagonalization of the window matrix to before diagonal-
ization. Not diagonalizing the window matrix does lead to a
significant change in the fitted parameter values, and the error on
$n_{\text{eff}}$ decreases by about 12%. We are, in effect, using information
from a wider range of scales, but this forces us to use theory
results outside their range of validity (e.g., at low $k$ we need to
extrapolate beyond the size of our simulation boxes). Note that
the change in error on $n_{\text{eff}}$, from 0.024 to 0.021, implies that we
should expect a random difference between the two results with
typical size $(0.024^2 - 0.021^2)^{1/2} = 0.012$, i.e., what might seem
like a surprisingly large part of the difference between the results
could be random.

As discussed above, the correction for Si $\text{iii}$-Ly$\alpha$ correlation is
very important to the goodness of our fit. It is less important for
the best-fit values, changing them only by 0.8 $\sigma$ for $n_{\text{eff}}$ and 0.4 $\sigma$
for $\Delta^2$. Normally, we allow a power-law dependence on redshift
for the amplitude of the Si $\text{iii}$-Ly$\alpha$ correlation, but removing this
freedom makes almost no difference. Allowing the correlation scale for the Si $\text{iii}$-Ly$\alpha$ correlation to be different than for Ly$\alpha$-
Ly$\alpha$ (freeing $s$ in eq. [16]; we usually fix this in this paper for
technical reasons) has only a very small effect. Including the Si $\text{iii}$-
Si $\text{iii}$ autocorrelation term (the $a^2$ part of eq. [15]) in the fit has
essentially no effect.

For our standard fit, we allow variation in the noise amplitude
at each redshift, represented by a multiplicative parameter sub-
ject to a 5% rms Gaussian constraint. Our fitting procedure then
marginalizes over this component. Reducing this constraint to
0.5% (i.e., fixing the noise) produces no change in our fit result and
does not even reduce the error bars noticeably. Leaving the noise
effectively free makes a noticeable difference in the fit results,
decreasing the amplitude by about 2/3 $\sigma$, increasing its
error by 20%, and decreasing $\chi^2$ to 123.8. Changes at this level
are expected when we remove the constraints on some parameters
and do not imply that the constraints were too small (i.e., we
are effectively removing 9 data points from the fit so we gen-
erally expect a decrease in $\chi^2$, increase in the error bars, and
some corresponding drift in the parameter values). Removing
our spectrum-by-spectrum noise estimation makes very little
difference. Finally, we note that if we did not correct the noise as
discussed in § 2.3, the results would change significantly. As
an example, we show the fit results in the $\delta \ln \Delta^2 - \delta n_{\text{eff}}$ plane
in Figure 26, for our standard case and these noise-related vari-
ants. We show the ratio of the power without individual noise
estimates for each quasar to our standard case in Figure 27.

Our requirement that the principal components of the error
matrix have at least the Gaussian level of variance makes no
difference to the fit results, although it improves $\chi^2$ for the fit
a little bit. Simply using the Gaussian error matrix instead of
bootstrap errors makes no difference to the fit results but in-
creases $\chi^2$ significantly. Ignoring the bootstrap error correla-
tions increases the error on $n_{\text{eff}}$ by about 12% and significantly
reduces $\chi^2$.

We normally use the range 1268 $\AA < \lambda_{\text{rest}} < 1380$ $\AA$ for our
background subtraction (i.e., subtract $P_{1268,1380}$). Removing the
background subtraction entirely reduces the inferred amplitude by
2 $\sigma$, and the slope by 1 $\sigma$, and results in a very large $\chi^2$ (the
error on $n_{\text{eff}}$ also decreases significantly, but this is mostly be-
bcause of the change in the best-fit values, not because of un-
certainty in the background subtraction). [Note that removing
the background subtraction, which increases $P_F(k, z)$, decreases
the inferred amplitude because the fitted $P_F$ decreases more than
enough to offset the increase in $P_F(k, z)$.] Clearly, the background
cannot be ignored. Using $P_{1409,1523}$ instead produces a somewhat
disturbingly large 0.028 (1.1σ) increase in \( n_{\text{eff}} \). We expect the longer wavelength range to give a less accurate estimate of the background power, because some metals are missing, but further investigation shows that most of this difference is probably caused by C IV BALs adding power to the 1409 Å < \( \lambda_{\text{rest}} < 1523 \) Å region. As we see in Table 1, removing the adjustment for noise dependence of the background (see eq. [13]) brings the two background regions closer together (this is reflected in Fig. 18). Adding the 147 BAL quasars identified by our automated algorithm leads to a huge discrepancy (0.094 in \( n_{\text{eff}} \)) when we use the \( P_{1409, 1523} \) background, but only when we adjust for noise level (without this the discrepancy for \( n_{\text{eff}} \), not shown in the table, is only 0.029). Note that the BAL cut makes essentially no difference to our standard fit using the \( P_{1268, 1380} \) background. All of these differences are easy to understand: First, BALs are known to be strongest in C IV absorption (Hall et al. 2002), so it is not surprising that we see the effects of BALs primarily in this wavelength region. Second, both our original by-eye and subsequent automated removal of BALs inevitably identify the features more easily in less noisy data, so the power from BALs naturally shows up when we intentionally use the noisier spectra for the background power. The fact that removing the 147 most obvious BALs has essentially no effect on our basic result gives us confidence that any remaining BAL features in the Ly\( \alpha \) forest and 1268 Å < \( \lambda_{\text{rest}} < 1380 \) Å regions are not significant.

To investigate the effect of a systematic uncertainty in the spectral resolution, we include in our fits an overall factor of the form \( \exp(\alpha k^2) \) multiplying the power spectrum, where \( \alpha \) is a free parameter. In our standard fit we impose an external constraint on \( \alpha, \pm(7 \text{ km s}^{-1})^2 \) rms. Essentially removing this freedom has no effect on the fit, while leaving \( \alpha \) essentially free increases the error on the amplitude by 40% and increases \( n_{\text{eff}} \) by 2/3 (the change in fitted amplitude is certainly consistent with drift from the increased error). As we show in Figure 5, our standard fit should be conservative.

Simply dividing each chunk of spectrum by its mean instead of also dividing by the mean continuum before estimating the power from the chunk makes little difference to the fit results. Division by the mean continuum actually increases the measured flux power by ~0%–2%, as we show in Figure 28.

We have performed a preliminary PCA analysis to try to model fluctuations around the mean continuum. When we use continua for each quasar composed of 13 PCA eigenvectors, our

![Fig. 26.—Fit results for variant noise treatments. The error bars show the 1σ error on each parameter. The ovals show \( \Delta \chi^2 = 2.3 \). Standard case (with 5% noise amplitude freedom in each redshift bin): solid lines. No individual noise estimate for each quasar: dotted lines. Noise amplitude freedom 0.5% (50%): short-dashed lines (long-dashed lines). The dot-dashed line shows the result using the pipeline noise estimates. [See the electronic edition of the Supplement for a color version of this figure.]](image1)

![Fig. 27.—Ratio of \( P_{1041, 1185} \) computed without quasar-by-quasar noise re-estimates (a constant 16% extra noise power was assumed instead) to the standard case. [See the electronic edition of the Supplement for a color version of this figure.]](image2)

![Fig. 28.—Ratio of \( P_{1041, 1185} \) computed without dividing by the continuum estimate to the standard case. [See the electronic edition of the Supplement for a color version of this figure.]](image3)
results change only a little ($n_{\text{eff}}$ is reduced by 0.4 $\sigma$), and $\chi^2$ increases, probably an indication of the unsatisfactory level of noise that we know remains in our estimates. As we see in Figure 29, the modification of adding a large constant to the weight matrix to make our measurement less sensitive to the mean of each chunk has little effect (the effect is larger on larger scales).

The line “no bin-redshift correction” in Table 1 refers to removing the correction for evolution in the power across the width of the redshift bins (see eq. [8]). We see (Fig. 30) that this correction mostly affects the lowest-redshift bin (where the low-$z$ edge of the bin is empty of data) and has little effect on the fit (not surprisingly, leaving out this correction increases $\chi^2$).

The line “ignore $F - \sigma_n$ correlation” in Table 1 shows the change in the fitted parameters if we naively use the given noise estimates for weighting without accounting for the fact that there is a correlation between the flux estimate and the noise amplitude estimate for each pixel. Figure 31 shows that the bias is a fairly constant 3%--5% increase in the flux power.

The difference is not actually caused by the change in weighting used in the power spectrum estimation; instead, the power is biased high because the estimation of the mean that each chunk of spectrum is divided by is biased low because low-flux pixels have smaller noise estimates. Ignoring this effect does not change our fit results. Normally, we base our estimation of the amount of the noise that is due to quasar flux on the separate estimates we have from the spectral reduction pipeline for the flux, sky, and read noise contributions; however, these estimates do not add up to the total noise reported by the pipeline. If we rescale the individual numbers to make them consistent with the total (not necessarily the correct thing to do) we see that the fit results are not changed significantly (the line “rescale $\sigma_n$”; we use $\sigma_r$ to refer to noise computed using the separate flux, sky, and read noise estimates), although the power does change by as much as 3% (Fig. 32; this difference would be a bit larger if we did not directly measure and correct for the cross-correlation between the noise amplitude and flux).

Finally, our power spectrum extraction code has a bias (partially related to division by the mean of each chunk of spectrum), that we correct for by dividing the result by equation (11). Removing this correction decreases the estimated $n_{\text{eff}}$ by about 1/2 $\sigma$ and increases the amplitude by a similar amount. The combined change is actually quite significant because it is transverse to the degeneracy direction for these parameters.

To summarize, most of the effects described above are small relative to the statistical errors on the final estimated parameters. We understand the cases where the difference is significant and...
fraction of the noise that is due to photon counting noise associated with flux from the quasar (see text) to the standard case. [See the electronic edition of the Supplement for a color version of this figure.]

expect that our standard method will be much more accurate than the difference between it and the variants (we show these variants to help the reader better understand our measurement). These tests give us confidence that the final results are very robust to small changes in the analysis pipeline.

How sensitive are these conclusions to our assumptions about the nuisance parameters, $F$, $T_1$, and $\gamma$, i.e., if these constraints improve in the future, will we need to worry more about systematic errors in $P_T (k, z)$? We investigate this by first fixing all the parameters in the fit (including removing the noise amplitude uncertainty, resolution uncertainty, and freedom in the Si III correction), so the only uncertainty is on $P_T (k, z)$. Table 1 (the “fixed nuisance parameters” line) shows that the error on the amplitude improves dramatically, by a factor of 5. The error on $n_{\text{eff}}$ improves by a factor of 2. So in principle the amplitude error can be improved a lot relative to potential systematic errors, and $n_{\text{eff}}$ improved as well (see also Mandelbaum et al. 2003). The next line (“optimistic $F$”), where we assume the HIRES constraint on $F$ is improved by a factor of 5, and the SDSS constraint by a factor of 2, shows that $\Delta^2$ is substantially degenerate with $F$ (as expected), but $n_{\text{eff}}$ is less degenerate. Improving the constraints on $T_1$ and $\gamma$ by factors of 5, in addition to the improved constraints on $F$, leads to little further improvement. Finally, for comparison, we tried simply reducing the errors on $P_T (k, z)$ by a factor of $\sqrt{3}$ and found that the error on $n_{\text{eff}}$ decreases by almost the same factor (1.6), but the error on $\Delta^2$ decreases less (a factor of 1.2). SDSS will collect a factor of $\sim$ 3 more data than we have in the present sample. We conclude that the error on $n_{\text{eff}}$ can easily be reduced by simply gathering more data, while improvements on $\Delta^2$ can be made by improving the errors on $F$.

4.4. Subsamples of the Data

Another way to test for systematic errors is to search for internal discrepancies between the different subsamples of the same data. Of course, there are only a finite number of possible subsamples we can try, so this test cannot be fully exhaustive. In addition, with many such tests performed one must worry that some will give an apparently statistically significant deviation just by random chance. Table 2 shows results of splitting the data into roughly equal weight subsamples, defined by various properties of the spectra that, at least at first glance, should not be correlated with the measured power. In practice, we rank the spectra by the property of interest and split the sample into halves by requiring that the Gaussian errors on the $k = 0.007 \text{ s km}^{-1}$ point are equal for the two halves (the bootstrap errors will not be precisely equal). We list the probability of obtaining $\chi^2$ greater than the value computed by differencing the power spectra (these differences include the different background subtraction computed using eq. [13] for different noise levels). We also list the fitting parameter results for each subsample and give the probability for obtaining the observed level of difference between the fits. Because these subsamples are basically independent,

**TABLE 2**

| Split                      | $P_{\chi^2}$ (points) (%) | $\delta \ln \Delta^2_{e, h}$ | $\delta n_{\text{eff}, <}$ | $\delta \ln \Delta^2_{e, <}$ | $\delta n_{\text{eff}, >}$ | $P_{\chi^2}$ (fit) (%) |
|---------------------------|---------------------------|-------------------------------|-----------------------------|-------------------------------|-----------------------------|-------------------------|
| $\lambda_{\text{rest}}$   | 40                        | $-0.03 \pm 0.12$              | $-0.027 \pm 0.031$          | $0.04 \pm 0.11$               | $0.018 \pm 0.029$           | 51                      |
| Noise                     | 10                        | $-0.01 \pm 0.11$              | $0.020 \pm 0.028$           | $-0.01 \pm 0.12$              | $-0.001 \pm 0.030$          | 76                      |
| Noise (raw)               | 0.0006                    | $-0.05 \pm 0.11$              | $0.027 \pm 0.029$           | $-0.19 \pm 0.13$              | $-0.010 \pm 0.026$          | 61                      |
| Sky                       | 5.9                       | $-0.02 \pm 0.11$              | $-0.001 \pm 0.028$          | $0.02 \pm 0.12$               | $0.014 \pm 0.030$           | 93                      |
| $\sigma_w - \sigma_{\lambda}$ | 34                       | $0.02 \pm 0.11$              | $0.029 \pm 0.030$           | $-0.01 \pm 0.12$              | $-0.011 \pm 0.030$          | 49                      |
| Read noise                | 94                        | $0.08 \pm 0.11$              | $0.020 \pm 0.030$           | $-0.05 \pm 0.12$              | $-0.015 \pm 0.028$          | 68                      |
| Cont. $\chi^2_{/\nu}$     | 33                        | $-0.04 \pm 0.11$             | $0.017 \pm 0.028$           | $0.06 \pm 0.12$               | $0.001 \pm 0.030$           | 40                      |
| Resolution                | 73                        | $0.08 \pm 0.11$              | $0.036 \pm 0.031$           | $-0.08 \pm 0.12$              | $-0.025 \pm 0.028$          | 32                      |
| Flexure                   | 14                        | $0.10 \pm 0.11$              | $0.040 \pm 0.030$           | $-0.11 \pm 0.11$              | $-0.033 \pm 0.026$          | 19                      |
| Alignment                 | 29                        | $0.09 \pm 0.11$              | $0.031 \pm 0.030$           | $-0.09 \pm 0.11$              | $-0.031 \pm 0.027$          | 29                      |
| Exp. $\chi^2_{/\nu}$      | 65                        | $0.01 \pm 0.11$              | $-0.015 \pm 0.029$          | $0.00 \pm 0.12$               | $0.010 \pm 0.030$           | 63                      |
| Error on mean             | 56                        | $-0.07 \pm 0.10$             | $-0.014 \pm 0.030$          | $0.07 \pm 0.11$               | $0.018 \pm 0.031$           | 67                      |
| Error on $A_q$            | 40                        | $0.06 \pm 0.09$              | $0.020 \pm 0.027$           | $-0.01 \pm 0.12$              | $-0.014 \pm 0.030$          | 68                      |

* Probabilities may not be fully reliable because we have not demonstrated that $\chi^2$ is properly distributed.

b The subsample fit results cannot be combined to produce the result of the fit to the full data set because the underlying nuisance parameters were not required to be the same.

c The “noise (raw)” line shows the comparison without accounting for the noise dependence of the background.
deviations within the error bars are expected and are not an indication of systematic errors. We describe these subsample splits below.

The power we measure should be independent of the rest frame region of the quasar continuum in which it is measured. Figure 33 shows $P_{1041,1113}(k,z)$ and $P_{1113,1185}(k,z)$ to test this expectation.

The results look pretty similar, but to compare them quantitatively, we compute

$$
\chi^2 = (P_\Delta - P_c)^T(C_c + C_\Delta)^{-1}(P_\Delta - P_c),
$$

finding $\chi^2 = 111.0$ for 108 points. The agreement appears perfect. To compare the two in a different way, we perform separate fits of the linear mass power spectrum parameters $\Delta^2$ and $n_{\text{eff}}$ to $P_c(k,z)$ computed from $P_{1041,1113}(k,z)$ and $P_{1113,1185}(k,z)$. The results, given in the first line of Table 2, are consistent within the expected errors. This test provides some evidence that power from continuum fluctuations is not an important contribution to the total, beyond what we would expect from looking at the red side of the Ly$\alpha$ emission line. It is possible that the two halves of the forest could have significant extra continuum power, but if they do it has to be the same in each half.

We compute the weighted mean of the rms noise for each chunk of spectrum as we use it to estimate the power spectrum. A split based simply on this noise level, illustrated in Figure 34, produces a small but unambiguously significant discrepancy in the raw measurement of $P_{1041,1185}$. $\chi^2 = 185$, though the fit parameters agree within their errors (Table 2, row 3).

This discrepancy in power is the motivation for, and is largely removed by, our noise-dependent background subtraction procedure defined by equation (13). Figure 35 shows the power $P_{1268,1380}$ that is used for background subtraction, again sampled by noise level.

There is a clear difference in power, which is not isolated to a few wavenumbers or redshifts. Once the $P_{1268,1380}$ power is subtracted according to equation (13), we obtain $P_F$ estimates and corresponding fit parameters from the high- and low-noise subsamples that agree within the errors (Table 2, row 2). Since the fit parameters agree even without noise-dependent background subtraction, it appears that the discrepancy in raw $P_{1041,1185}$ power does not mimic a change in cosmological parameters, and our ultimate conclusions would therefore not change even if we did not implement it. Nonetheless, the origin
of the differences remains somewhat mysterious, since we went to
great effort to estimate the noise correctly.

Two more splits that yield discrepant $P_{1041,1185}$ but show no
ing of trouble after the noise-dependent background subtraction
are based on the ratio of the mean sky flux to the mean quasar
and $\sigma_w - \sigma_c$, the difference between the pipeline estimate
of the noise and the sum of our estimates of the quasar flux,
sky flux, and read noise components of the noise. We are not sure
what $\sigma_w - \sigma_c$ means, since we do not understand the source of
noise misestimation in the standard pipeline. Even without noise-
derpendent background subtraction, the fit results did not differ
significantly in either of these cases. They are almost surely
symptoms of the same noise-related problem discussed above.

The split based on read noise in the spectra shows good agree-
ment between the $P_{1113}$ measurements, even without noise-
derpendent background subtraction as does a split based on how
well the mean continuum matches the quasar spectrum outside
the Lyα forest, quantified by computing $\chi^2/\nu$ for the difference
between the continuum and spectrum (\textit{``cont.} $\chi^2/\nu$\textit{''} in Table 2).

Several other splits that show little or no sign of trouble are based
on the mean value of $\chi^2/\nu$ computed for each pixel when com-
bining exposures (this was the comparison that motivated our
spectrum-by-spectrum noise reestimation), the mean resolution,
the movement of the spectrum relative to the detector pixel grid
during the observation (\textit{``flexure''}), the alignment of the pixels in
the different exposures for the same spectrum (closely related to
flexure), the error on the overall normalization of the spectrum,
$A_g$ (see eq. [2]); this error is set by a combination of the noise level
outside the forest and the length of spectrum observed outside
the forest), and the error on the means computed for the forest
chunks (differences at fixed $z$ are related to the length of the
chunk and the noise in the forest).

Overall, the agreement between our subsamples is excellent,
both for the $P_{1041,1185}$ results and the fit results. In some cases this
agreement relies on the noise-dependent background subtraction,
which we would like to understand better (in no case does
the fit agreement rely on this).

4.5. Continuum Power

The power in the mean continuum, for the four different rest
frame regions identified in Figure 2, is shown in Figures 36a–
36d, relative to the Lyα forest power (the mean continuum power
was measured by replacing the quasar flux in each pixel by the
mean continuum level at that pixel).

The mean continuum is very well behaved over the $k$ range
that we use (0.0013–0.02 s km$^{-1}$), but its fluctuations quickly
become significant at $k \lesssim 0.001$ s km$^{-1}$. What little power the
mean continuum shows in our chosen $k$ range should be removed
when we divide the spectra by the continuum; it is only fluctua-
tions around the mean that matter.

We summarize our strong, but maybe not airtight, argument
for believing that continuum fluctuations are not corrupting our
measurement as follows:

1. The power in the mean continuum is small.
2. The results for the $P_{1041,1185}$ measurement in two halves of
   the forest region, $P_{1041,1113}$ and $P_{1113,1185}$, agree.
3. The power in the background regions, $P_{1268,1380}$
   and $P_{1409,1523}$, agree at the level we care about in the low-noise
data, as long as BALs (which mostly affect the latter region) are
removed.
4. Division by preliminary 13 eigenvector PCA estimates of
   the continua (i.e., including fluctuations around the mean) does
not change the results.

To be quantitatively important despite these arguments, the power
in quasar-to-quasar continuum fluctuations in the forest must be
substantially larger than the power in the mean continuum itself,
the continuum fluctuations in the forest must be substantially
different from those in the background regions (despite those
regions being similar to each other and the two halves of the for-
est being similar to each other), and our PCA analysis must be
substantially flawed. Further study is warranted, but a big effect
seems unlikely.

4.6. Comparison with Past Measurements

There are three $P_{k, z}$ measurements already in the litera-
ture, McDonald et al. (2000), Croft et al. (2002), and Kim et al.
(2004), all using at least some high-resolution data. Each uses its
own set of redshift bins, so to compare we need a way to inter-
polate our results to these redshifts. We do this by performing our
standard cosmological fit to all of the data (at first—later we will
remove some of the past results). This gives us a set of best-fit
model parameters that can be used to compute the power at any $k$
and $z$. Within the range of $k$ where we have SDSS measurements,
the fit is always dominated by the SDSS points. The fitted curves
always match the SDSS results to much better than the size of
the errors on the past results, meaning that, for the purpose of
comparison to the past results, the curves are simply a faithful
interpolation between the SDSS points. At $k > 0.02$ s km$^{-1}$, the fit
is effectively a weighted average of the past results, although
the constraint that it must match SDSS at lower $k$ has some
influence [our simulation predictions do not allow for sharp fea-
tures in $P_{k, z}$].

We first perform a fit to all the data with $k < 0.05$ s km$^{-1}$ and
$z > 2.1$, finding an atrociously bad $\chi^2 = 392$ for $\sim 238$ dof.
Removing McDonald et al. (2000), reduces $\chi^2$ by 53.4 (for
39 data points), removing Croft et al. (2002) reduces $\chi^2$ by 85.2
(for 65 points), and removing Kim et al. (2004) reduces $\chi^2$ by
123.3 (28 points). Clearly, there is gross disagreement between
Kim et al. (2004) and the other results. Figure 37 shows the Kim
et al. (2004) points at $z = 2.18$ and $z = 2.58$ (from their Table 5)
along with the fit prediction for them. Note that we include Si ii
contamination in the model as described in § 4.2, so the model
curves are not perfectly smooth.

We see large discrepancies, as we expect from the bad $\chi^2$. The
point at $k = 0.0012$ s km$^{-1}$, $z = 2.58$ is 5.9 $\sigma$ below the pre-
prediction (as well as any reasonable extrapolation of the other
Kim et al. 2004) points), and the points at increasingly high $k$ are
generally too high (at the highest $k$, this is a reflection of dis-
agreement with the other high-resolution data, but actually the
agreement is not much better if we only include SDSS in the fit,
because no model can fit the highest $k$ SDSS points and then
climb to match the higher $k$ Kim et al. (2004) points). To re-
assure the reader that we are not playing games with the fitted
curves, we also plot the SDSS points at $z = 2.2$ and 2.6.

Since the Kim et al. (2004) results clearly have some problem,
unless the other three measurements are all wrong (we will see
that, with one exception, the other three agree with each other),
we eliminate them from the rest of the comparison. A fit to SDSS,
McDonald et al. (2000), and Croft et al. (2002) gives $\chi^2 =
269$ for $\sim 210$ dof (still a bad fit). Removing McDonald et al.
(2000), reduces $\chi^2$ by 44.3 (39 points), while removing Croft
et al. (2002) reduces $\chi^2$ by 97.2 (for 65 points this reduction
would occur by chance only 0.6% of the time). Figure 38 shows the
Croft et al. (2002) points, along with the fit prediction for them.

The agreement is actually very good for four of the five redshift
bins, while the $z = 2.47$ points are obviously out of place (these 13
Fig. 36.—Power in the mean continuum relative to the Ly$\alpha$ forest power, for various rest wavelength intervals: (a) $1041$ Å $< \lambda_{\text{rest}} < 1185$ Å, (b) $1268$ Å $< \lambda_{\text{rest}} < 1380$ Å, (c) $1409$ Å $< \lambda_{\text{rest}} < 1523$ Å, (d) $1558$ Å $< \lambda_{\text{rest}} < 1774$ Å. The error bars show the fractional error on $P_{0441,1185}$ (without diagonalizing the window matrix because the diagonalization works poorly at the lowest $k$-values that we show). [See the electronic edition of the Supplement for a color version of this figure.]
points increase \( \chi^2 \) by 54). Figure 39 shows the McDonald et al. (2000) points, along with the fit prediction for them (for this fit we removed the \( z = 2.47 \) Croft et al. (2002) points).

The agreement is good, with the agreement at \( z = 2.41 \) disfavoring the anomalous Croft et al. (2002) \( z = 2.47 \) points; further investigation by R. Croft (2004 private communication) does not reveal any obvious error in this redshift bin that would explain the anomaly. Note that the agreement of McDonald et al. (2000) and Croft et al. (2002) at high \( k \) adds weight to the idea that something is seriously wrong with Kim et al. (2004). Kim et al. (2004) show some comparisons with past results, and claim they agree, but these comparisons used custom redshift bins (i.e., not the bins in their table) and were not high precision (e.g., they compare the Croft et al. 2002 points at \( z = 2.13 \) to a bin with \( z = 2.04 \), so evolution cancels some of the amplitude offset, and they call the apparent \( \sim 50\% \) difference at \( k \sim 0.04 \) s km\(^{-1} \) a “slight” excess).

After reading a preprint of this paper, Kim et al. (2004) reviewed their measurement and published an erratum explaining that their \( k \) bins were mislabeled. As we discuss in McDonald et al. (2004), this change improves the agreement, but there is still significant residual disagreement. We argue in McDonald et al. (2004) that unsubtracted background power in the Kim et al. (2004) measurement can account for the disagreement.

5. FINAL RESULTS TABLE AND DIRECTIONS FOR USE

Table 3 gives the primary power spectrum results. The columns are as follows: \( z \), the redshift of the bin; \( k \), the wavenumber of the bin; \( P_F(k, z) \), our final \( \text{Ly}_\alpha \) forest power spectrum result (along with the square roots of the diagonal elements of the error covariance matrix); \( P_{\text{noise}} \), the noise power that was subtracted from each bin; and \( P_{\text{background}} \), the background that was subtracted from each bin (\( P_{1268,1380} \) adjusted according to the amount of noise in the forest; eqs. [3], [4], and [13]). The quantity \( P_{\text{noise}} \) is just the noise subtracted from \( P_{1380} \) (a roughly comparable amount of noise was subtracted from the background, so to some degree these cancel in the final result). Table 4 gives the covariance matrix of the errors, \( \langle \Delta P_F^2 \Delta P_F^2 \rangle \). These results tables are also available online. The covariance matrix must be used in any serious quantitative fitting. When

---

15 See http://www.cita.utoronto.edu/~pmcdonal/LyaF/sdss.html.
[Table 3 continued]

| $z$  | $k$     | $P_r(k, z)$ | $P_{\text{noise}}$ | $P_{\text{background}}$ |
|------|---------|-------------|---------------------|--------------------------|
| 3.2  | 0.00891 | 27.03 ± 0.95 | 9.82 ± 1.49        |
| 3.2  | 0.01122 | 22.39 ± 0.84 | 10.58 ± 1.81       |
| 3.2  | 0.01413 | 18.64 ± 0.72 | 11.91 ± 1.40       |
| 3.4  | 0.01778 | 15.46 ± 0.67 | 14.37 ± 1.19       |
| 3.6  | 0.01413 | 56.04 ± 5.81 | 11.04 ± 4.06       |
| 3.4  | 0.01778 | 75.13 ± 5.30 | 11.23 ± 2.39       |
| 3.4  | 0.00224 | 56.28 ± 3.91 | 10.88 ± 2.07       |
| 3.4  | 0.00282 | 58.77 ± 3.37 | 10.94 ± 3.14       |
| 3.4  | 0.00355 | 52.14 ± 2.83 | 10.87 ± 2.61       |
| 3.4  | 0.00447 | 43.17 ± 2.20 | 10.70 ± 1.99       |
| 3.4  | 0.00562 | 41.53 ± 1.73 | 10.95 ± 2.04       |
| 3.4  | 0.00708 | 37.37 ± 1.43 | 11.28 ± 1.55       |
| 3.4  | 0.00891 | 32.72 ± 1.20 | 11.92 ± 1.77       |
| 3.4  | 0.01122 | 28.80 ± 1.17 | 13.18 ± 2.01       |
| 3.4  | 0.01413 | 22.64 ± 0.89 | 14.86 ± 1.64       |
| 3.4  | 0.01778 | 18.43 ± 0.80 | 18.20 ± 1.24       |
| 3.6  | 0.00141 | 79.22 ± 8.31 | 15.07 ± 2.25       |
| 3.6  | 0.00178 | 84.73 ± 8.25 | 14.83 ± 2.86       |
| 3.6  | 0.00224 | 74.61 ± 5.85 | 14.79 ± 3.26       |
| 3.6  | 0.00282 | 65.77 ± 4.97 | 14.43 ± 3.29       |
| 3.6  | 0.00355 | 65.99 ± 4.15 | 14.52 ± 2.25       |
| 3.6  | 0.00447 | 55.47 ± 3.38 | 14.18 ± 1.33       |
| 3.6  | 0.00562 | 49.46 ± 2.75 | 14.15 ± 1.28       |
| 3.6  | 0.00708 | 43.87 ± 2.17 | 14.44 ± 1.63       |
| 3.6  | 0.00891 | 40.46 ± 1.95 | 15.19 ± 1.99       |
| 3.6  | 0.01122 | 32.41 ± 1.65 | 15.90 ± 1.62       |
| 3.6  | 0.01413 | 26.26 ± 1.33 | 17.55 ± 1.27       |
| 3.6  | 0.01778 | 22.02 ± 1.26 | 21.61 ± 1.52       |
| 3.8  | 0.00141 | 118.73 ± 15.49 | 22.60 ± 6.89   |
| 3.8  | 0.00178 | 61.45 ± 9.80 | 20.90 ± 6.40       |
| 3.8  | 0.00224 | 77.11 ± 7.10 | 20.86 ± 5.07       |
| 3.8  | 0.00282 | 71.59 ± 7.41 | 20.71 ± 2.19       |
| 3.8  | 0.00355 | 77.33 ± 7.53 | 20.95 ± 2.51       |
| 3.8  | 0.00447 | 59.07 ± 4.58 | 20.36 ± 2.37       |
| 3.8  | 0.00562 | 57.60 ± 3.91 | 20.80 ± 2.61       |
| 3.8  | 0.00708 | 56.50 ± 3.70 | 21.72 ± 2.79       |
| 3.8  | 0.00891 | 42.82 ± 2.38 | 21.87 ± 2.58       |
| 3.8  | 0.01122 | 37.42 ± 2.32 | 23.73 ± 2.70       |
| 3.8  | 0.01413 | 30.08 ± 2.32 | 26.62 ± 2.48       |
| 3.8  | 0.01778 | 28.45 ± 1.92 | 34.34 ± 2.16       |
| 3.8  | 0.00141 | 123.91 ± 21.09 | 23.83 ± 8.10   |
| 3.8  | 0.00178 | 123.49 ± 18.39 | 22.86 ± 6.27   |
| 3.8  | 0.00224 | 96.54 ± 12.12 | 22.31 ± 4.53       |
| 3.8  | 0.00282 | 83.87 ± 9.90 | 21.41 ± 4.84       |
| 3.8  | 0.00355 | 71.95 ± 8.35 | 21.13 ± 3.71       |
| 4.0  | 0.00447 | 66.40 ± 5.86 | 21.16 ± 3.25       |
| 4.0  | 0.00562 | 68.63 ± 5.78 | 22.14 ± 3.63       |
| 4.0  | 0.00708 | 60.24 ± 4.91 | 22.78 ± 2.60       |
| 4.0  | 0.00891 | 52.32 ± 3.98 | 24.08 ± 3.17       |
| 4.0  | 0.01122 | 46.19 ± 3.66 | 26.76 ± 3.35       |
| 4.0  | 0.01413 | 31.83 ± 4.36 | 29.99 ± 2.25       |
| 4.2  | 0.00141 | 167.33 ± 42.60 | 20.94 ± 5.77    |
| 4.2  | 0.00178 | 131.55 ± 21.88 | 20.49 ± 4.33   |
| 4.2  | 0.00224 | 122.69 ± 18.37 | 20.51 ± 2.78   |
| 4.2  | 0.00282 | 134.66 ± 17.84 | 20.81 ± 3.69   |
| 4.2  | 0.00355 | 113.73 ± 13.62 | 20.57 ± 2.76   |
| 4.2  | 0.00447 | 84.39 ± 9.77 | 20.30 ± 2.23       |
| 4.2  | 0.00562 | 68.02 ± 7.22 | 20.44 ± 2.66       |
| 4.2  | 0.00708 | 77.30 ± 8.08 | 21.88 ± 1.77       |
| 4.2  | 0.00891 | 62.49 ± 6.13 | 23.86 ± 2.14       |
| 4.2  | 0.01122 | 52.11 ± 4.78 | 27.09 ± 2.34       |
| 4.2  | 0.01413 | 50.56 ± 4.83 | 34.33 ± 2.84       |
| 4.2  | 0.01778 | 37.86 ± 4.45 | 48.04 ± 2.45       |

Notes.—Units of $k$ are km$^{-1}$, power spectra have units km$^3$ s$^{-1}$. The error covariance matrix must be used for any quantitative fitting. Table 3 is also available in machine-readable form in the electronic edition of the *Astrophysical Journal Supplement*. 
using this table to constrain models, the following allowances should be made for residual systematic uncertainties:

1. Allow ±5% rms error on the noise power amplitude at each redshift. We do not have any reason to think that the error is really this large, but, considering the complications related to the noise, we think it is prudent to include it. Operationally, we suggest subtracting $f_i P_{\text{noise}}(k, z_i)$ from $P_F(k, z_i)$, where $f_i$ are free parameters in the fit (one for each redshift bin), and adding $\sum_i (f_i/0.05)^2$ to $\chi^2$.

2. Allow ±7 km s$^{-1}$ rms overall error on the resolution variance (i.e., the square of the rms width of the Gaussian resolution kernel). This is the expected size of the uncertainty due to flexure in the detector, although Figure 5 suggests that it may actually be smaller. Specifically, multiply $P_F(k, z)$ by $\exp(\alpha k^2)$, with $\alpha$ a free parameter in the fit, and add $[\alpha/(7 \text{ km s}^{-1})^2]^2$ to $\chi^2$.

3. Si $\equiv$ Lyα cross-correlation must be accounted for. We have suggested a simple method—assume the cross-correlation has the same form as the Lyα-Lyα autocorrelation up to an amplitude that is a free parameter, and possibly include freedom in the correlation width and/or redshift evolution of the amplitude—but others could be devised (e.g., including Si $\equiv$ in the simulated spectra through a parameterized semianalytic model).

6. CONCLUSIONS

We have analyzed a sample of 3035 quasar spectra measured by SDSS and covering the redshift range $2 < z < 4$. This data set is almost 2 orders of magnitude larger than previously available data sets. We have focused on the flux power spectrum in the redshift range $2.1 < z < 3.9$ and for modes $0.0013 \text{ s km}^{-1} < k < 0.02 \text{ s km}^{-1}$. The extraordinary size of the data sample leads to an order-of-magnitude reduction in errors compared to previous analyses. Consequently, to do justice to this data set requires a much more careful analysis than was needed in the past. To this end we have developed a new analysis pipeline using quadratic power spectrum estimation with near optimal performance. We applied this analysis to realistic mock spectra and demonstrated (after several tweaks) that the method performs as expected. We emphasize that realistic mock spectra are essential if one is to trust the results at the level of precision allowed by this data set. Our error estimation is based on bootstrap resampling, which works very well here because the individual quasars are independent of each other. The errors were tested against mock spectra and found to be accurate. We also compared the bootstrap errors to Gaussian errors, finding them to be in general less than 20% higher than Gaussian.

Given the small errors on the recovered flux power spectrum the required control of systematic effects must be improved correspondingly as well. A significant part of this paper is devoted to this issue. We find several sources of contamination present in the data and develop methods to remove them. Metal absorption for metals with rest wavelength transitions above $\sim 1300 \AA$, uncertainties in sky subtraction, and calibration errors can be subtracted essentially exactly by measuring the power on the red side of Lyα using the same observed wavelength range. We search for a contribution from metals with transitions close to Lyα using a correlation function analysis, assuming that they are correlated with hydrogen. We find clear evidence of Si $\equiv$ contamination and develop a simple and effective scheme to remove it. This procedure improves the $\chi^2$ of the fit from 194 to 129 for $\sim 104$ dof and is thus necessary for a satisfactory fit. We find no evidence of any other metal line contribution to the background subtracted flux power spectrum.

We reduce any contribution of the continuum to the flux power spectrum by dividing each spectrum by the mean quasar continuum. If contributions from quasar-to-quasar continuum differences are similar in different regions of the spectrum, then our subtraction of power from the red side of Lyα, as described above, should remove them. Several tests suggest that any residual contributions from continuum fluctuations are negligible. First, we measure the power in the mean continuum in several rest frame regions, finding it to be always small relative to our error bars, so power in quasar-to-quasar fluctuations has to be larger than power in the mean continuum itself to be significant. Second, measurements of the background in several rest frame regions place upper limits on the fluctuations in those regions. Third, a split of the Lyα forest region into two halves reveals no evidence that residual continuum fluctuations differ from one half to the other. Finally, estimating the continuum quasar-by-quasar using a principal component analysis does not change the power spectrum results significantly.

In § 4 we perform a series of tests to verify the robustness of the analysis against several modifications of the standard procedure and splits of the data. This reveals a correlation between the power in the red side and the average noise (and some other properties of the spectra that correlate with noise, like the amplitude of the sky flux relative to the quasar flux). While we do not have a detailed explanation for this effect, we are able to remove it by modifying the standard procedure to include this correlation. From our full battery of tests, we conclude that systematic effects in the power spectrum measurement are not likely to significantly affect the results of cosmological fitting [i.e., it is likely that some effects remain formally significant relative to the errors on $P_F(k, z)$, but the shape of these systematic errors does not seem to correspond to a change in the cosmological parameters]. This conclusion is further confirmed by the analysis of different subsets, which do not reveal any systematic deviations from those expected statistically.

In this paper we limit ourselves to the analysis of the flux power spectrum, without attempting to compare it to cosmological models. The results of this paper should thus be fairly noncontroversial and can be used by others who wish to perform their own cosmological analysis. Our own analysis will be presented in a separate publication, as will the cosmological implications that follow from it. We note that the expected error on the linear rms amplitude of fluctuations is $\sim 5\%$, and the error on the slope is $\sim 0.024$, both at the pivot point $k = 0.009 \text{ s km}^{-1}$. This should be compared to 10% error on the amplitude and 0.04 on the slope from the WMAP data at $k = 0.05 \text{ Mpc}^{-1}$ (Spergel et al. 2003). This data set provides very tight constraints on the amplitude and slope of the matter power spectrum. Many additional analyses can be performed using this data set, among them the mean flux evolution, cross-correlations between close pairs, and a bispectrum.
measurements. These will provide a wealth of additional information both on cosmology and on the state of the intergalactic medium at $2 < z < 4$, and they will allow us to test the basic picture of the Lyα forest that has emerged over the last decade. We believe that the unprecedented size of this data set will revolutionize our understanding of the high-redshift universe; this work is merely a first step in this endeavor.

Funding for the creation and distribution of the SDSS Archive has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the US Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is http://www.sdss.org/

The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, Korean Scientist Group, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, Princeton University, the United States Naval Observatory, and the University of Washington.

Some of the computations used facilities at Princeton provided in part by NSF grant AST-0216105, and some computations were performed at NCSA. Transition wavelengths used in this paper are from the Atomic Line List, http://www.pa.uky.edu/~peter/atomic.

J. S. is grateful for support from the W. M. Keck Foundation and NSF grant PHY-0070928. D. P. S. is supported by NSF grant AST03-07582. R. C. acknowledges grants AST-0206299 and NAG5-13381. S. B. is supported by NSF AST-0307705.

REFERENCES

Abazajian, K., et al. 2004, AJ, 128, 502
———. 2003, AJ, 126, 2081
Bernardi, M., et al. 2003, AJ, 125, 32
Bi, H. G., Boerner, G., & Chu, Y. 1992, A&A, 266, 1
Blanton, M. R., Lin, H., Lupton, R. H., Maley, F. M., Young, N., Zehavi, I., & Loveday, J. 2003, AJ, 125, 2276
Cen, R., Miralda-Escude, J., Ostriker, J. P., & Rauch, M. 1994, ApJ, 437, L9
Croft, R. A. C., Weinberg, D. H., Bohte, M., Burles, S., Hernquist, L., Katz, N., Kirkman, D., & Tytler, D. 2002, ApJ, 581, 20
Croft, R. A. C., Weinberg, D. H., Katz, N., & Hernquist, L. 1998, ApJ, 495, 44
Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, AJ, 111, 1748
Gnedin, N. Y., & Hamilton, A. J. S. 2002, MNRAS, 334, 107
Gnedin, N. Y., & Hui, L. 1998, MNRAS, 296, 44
Gunn, J. E., et al. 1998, AJ, 116, 3040
Hall, P. B., et al. 2002, ApJS, 141, 267
Hernquist, L., Katz, N., Weinberg, D. H., & Jordi, M. 1996, ApJ, 457, L51
Hogg, D. W., Finkbeiner, D. P., Schlegel, D. J., & Gunn, J. E. 2001, AJ, 122, 2129
Kim, T., Viel, M., Haehnelt, M. G., Carswell, R. F., & Cristiani, S. 2004, MNRAS, 347, 355 (erratum 351, 1471)
Lynd, R. 1971, ApJ, 164, L73
Mandelbaum, R., McDonald, P., Seljak, U., & Cen, R. 2003, MNRAS, 344, 776
McDonald, P., Miralda-Escudé, J., Rauch, M., Sargent, W. L. W., Barlow, T. A., & Cen, R. 2001, ApJ, 562, 52
McDonald, P., Miralda-Escudé, J., Rauch, M., Sargent, W. L. W., Barlow, T. A., & Cen, R., & Ostriker, J. P. 2000, ApJ, 543, 1
McDonald, P., et al. 2004, preprint (astro-ph/0407377)
Pier, J. R., Munn, J. A., Hindsley, R. B., Hennessy, G. S., Kent, S. M., Lupton, R. H., & Ivezić, Ž. 2003, AJ, 125, 1559
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes in C: The Art of Scientific Computing (2nd ed.; Cambridge: Cambridge Univ. Press)
Richards, G. T., et al. 2002, AJ, 123, 2945
Ricotti, M., Gnedin, N. Y., & Shull, J. M. 2000, ApJ, 534, 41
Schaye, J., Aguirre, A., Kim, T., Theuns, T., Rauch, M., & Sargent, W. L. W. 2003, ApJ, 596, 768
Schaye, J., Theuns, T., Rauch, M., Efstathiou, G., & Sargent, W. L. W. 2000, MNRAS, 318, 817
Seljak, U. 1998, ApJ, 506, 64
Seljak, U., McDonald, P., & Makarov, A. 2003, MNRAS, 342, L79
Slanger, T. G., Cosby, P. C., Osterbrock, D. E., Stone, R. P. S., & Misch, A. A. 2003, PASP, 115, 869
Smith, J. A., et al. 2002, AJ, 123, 2121
Spergel, D. N., et al. 2003, ApJS, 148, 175
Stoughton, C., et al. 2002, AJ, 123, 485
Tegmark, M. 1997, Phys. Rev. D, 55, 5895
Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, ApJ, 480, 22
Theuns, T., Leonard, A., Efstathiou, G., Pearce, F. R., & Thomas, P. A. 1998, MNRAS, 301, 478
Vogt, S. S., et al. 1994, Proc. SPIE, 2198, 362
York, D. G., et al. 2000, AJ, 120, 1579
Zaldarriaga, M., Scoccimarro, R., & Hui, L. 2003, ApJ, 590, 1
Zhang, Y., Anninos, P., & Norman, M. L. 1995, ApJ, 453, L57