Mon. Not. R. Astron. Soc. **322**, 561–575 (2001)

Semi-analytic approach to understanding the distribution of neutral hydrogen in the Universe

T. Roy Choudhury,★ T. Padmanabhan★ and R. Srianand★
IUCAA, Post Bag 4, Ganeshkhind, Pune 411 007, India

Accepted 2000 October 11. Received 2000 September 15; in original form 2000 May 15

ABSTRACT

Analytic derivations of the correlation function and the column density distribution for neutral hydrogen in the intergalactic medium (IGM) are presented, assuming that the non-linear baryonic mass density distribution in the IGM is lognormal. This ansatz was used earlier by Bi & Davidsen to perform one-dimensional simulations of lines of sight and analyse the properties of absorption systems. We have taken a completely analytic approach, which allows us to explore a wide region of the parameter space for our model. The analytic results have been compared with observations to constrain various cosmological and IGM parameters, whenever possible. Two kinds of correlation functions are defined: (i) along the line of sight (LOS); and (ii) across the transverse direction. We find that the effects on the LOS correlation owing to changes in cosmology and the slope of the equation of state of the IGM, γ, are of the same order, which means that we cannot constrain both the parameters simultaneously. However, it is possible to constrain γ and its evolution using the observed LOS correlation function at different epochs provided that one knows the background cosmology. We suggest that the constraints on the evolution of γ obtained using the LOS correlation can be used as an independent tool to probe the reionization history of the Universe. From the transverse correlation function, we obtain the excess probability, over random, of finding two neutral hydrogen overdense regions separated by an angle θ. We find that this excess probability is always less than 1 per cent for redshifts greater than 2. Our models also reproduce the observed column density distribution for neutral hydrogen, and the shape of the distribution depends on γ. Our calculations suggest that one can rule out $\gamma > 1.6$ for $z \simeq 2.31$ using the column density distribution. However, one cannot rule out higher values of γ at higher redshifts.

Key words: intergalactic medium – quasars: absorption lines – large-scale structure of Universe.

1 INTRODUCTION

The nature and evolution of the initial power spectrum of density fluctuations could be obtained by studying the distribution of objects at different scales and different epochs. The formalism for studying the formation of dark matter (DM) structures is well established, as they are collisionless particles interacting only through gravity, and has been extensively studied using the large cosmological *N*-body simulations. However, in order to model the evolution of baryonic structures like galaxies, groups of galaxies, etc., one needs to incorporate all the hydrodynamical processes, heating, cooling, star formation, etc., in the *N*-body simulations. Because of such complications, our understanding of the formation of baryonic structures has been limited.

Among the various baryonic structures, the regions where one can neglect the star formation are comparatively easier to study. Two such areas are (i) low-amplitude fluctuations in the intergalactic medium (IGM), where the star formation rate is very low, and (ii) the intracluster medium, where one studies processes over large scales and thus the star formation details can be neglected. Hence considerable effort has been given in understanding these two types of structures.

The baryonic matter distribution at $z \leq 5$ is well probed through the absorption signatures that it produces on the spectra of the distant QSOs. It is widely believed that, while the metal line systems (detected through Mg II or C IV doublets) seen in the QSO spectra could be associated with the haloes of the intervening luminous galaxies (Bergeron & Boisse 1991; Steidel 1993), most of the low neutral hydrogen column density absorption lines (commonly called 'Lyα' clouds) are believed to be due to low-amplitude baryonic fluctuations in the IGM.

★ E-mail: tirth@iucaa.ernet.in (TRC); paddy@iucaa.ernet.in (TP); anand @iucaa.ernet.in (RS)

Semi-analytical as well as hydrodynamical simulations are consistent with the view that the Lyα clouds are small-scale density fluctuations (Bond, Szalay & Silk 1988; Cen et al. 1994; Zhang, Anninos & Norman 1995; Hernquist et al. 1996; Miralda-Escudé et al. 1996; Bi & Davidsen 1997; Riediger, Petitjean & Mücket 1998; Theuns, Leonard & Efstathiou 1998a; Theuns et al. 1998b; Davé et al. 1999) that are naturally expected in any standard structure formation models. This idea was supported by the detection and the evolution of the weak clustering among Lyα clouds in redshift space (Cristiani et al. 1995; Srianand 1997; Khare et al. 1997). Subsequently it was realized that the thermal history of the Lyα line-forming regions depends on (i) the epoch of reionization (equation of state), (ii) the rate of photoionization, and (iii) adiabatic cooling. One can in principle neglect shocks and other processes that are important only in the highly non-linear regime. However, a simple linear evolution of the densities will fail to produce the saturated Lyα systems, and one needs to incorporate non-linearities in the model.

As a first step, one can model the non-linear evolution of the baryonic fluctuations that produce Lyα clouds using one of the several approximations, such as (i) the Zeldovich approximation (Doroshkevich & Shandarin 1977; McGill 1990; Hui, Gnedin & Zhang 1997), (ii) the lognormal approximation (Bi 1993; Gnedin & Hui 1996; Bi & Davidsen 1997), or (iii) the power-law approximation (Bi, Ge & Fang 1995) (strictly speaking, the baryonic fluctuations are calculated here using the linear theory). In all these cases the baryon density is estimated from the DM density by some rule, and the neutral fraction is estimated by considering the equilibrium between the rate of photoionization owing to background radiation and the rate of recombination estimated from the temperature defined through the equation of state. All these models depend on various IGM parameters such as the intensity of the background radiation, equation of state and density-averaged temperature, as well as cosmological parameters like $\Omega_{\rm m}$, Ω_Λ, etc.

Observationally the statistical properties of the Lyα absorption lines are quantified through the column density distribution, correlation functions and their dependence on the mean redshift. The clustering properties of the Lyα absorption lines are studied through the two-point correlation function obtained either (i) in redshift space using the lines detected along a single line of sight (LOS), which we call the 'LOS correlation function', or (ii) among the absorption lines detected along the LOSs toward a few closely spaced QSOs, which we call the 'transverse correlation function'. In either case the observed spectra are decomposed into clouds using 'Voigt' profile fits. Although this process smoothes the density field over the width of the lines, the average effect owing to thermal broadening is taken care of by the Voigt profiles. One can also compute the two-point correlation function of the observed flux in different pixels. As this process does not decompose the actual density fields into cloudlets, in order to analyse the data the models should incorporate the thermal broadening and blending of contributions from different density fluctuations (Croft et al. 1999; McDonald et al. 2000). Most of the existing studies concentrate on obtaining constraints on the cosmological parameters using the observed statistical properties. Comparatively much less effort is directed to understanding how the observed quantities depend on the physical conditions in the IGM.

In this work we make a preliminary attempt to investigate the dependence of the observable quantities on various parameters of the models using a simple analytic approach. We derive analytic relations for the two-point correlation function among the Lyα clouds and the column density distribution using a lognormal approximation. These equations are used with the observed Voigt profile fitted data to get constraints on different IGM and cosmological parameters. In Section 2, we treat the non-linear evolution with a simple ansatz proposed by Bi & Davidsen (1997) for the baryonic density fluctuations, and derive analytic expressions for the correlation function along the LOS and in the transverse direction and the column density distribution. The model parameters used to obtain various results are discussed in Section 3. In Section 4, we study the correlation function at different redshifts for different structure formation models, and for different values of the IGM parameters such as the density-averaged temperature and the equation of state. We compare some of our results with the existing observational data. We also present the results for the column density distribution and study its dependence on various cosmological and IGM parameters. The results are summarized in Section 5.

2 ANALYTIC MODEL

The linear density contrast for DM in comoving k-space, for a particular redshift z is given by

$$\delta_{\rm DM}(\boldsymbol{k}, z) = D(z)\delta_{\rm DM}(\boldsymbol{k}, 0), \tag{1}$$

where $D(z)$ is the linear growth factor for the density contrast, normalized such that $D(0) = 1$. If we assume the linear density contrast to be a Gaussian random field, then the corresponding linearly extrapolated power spectrum $P_{\rm DM}(k)$ is defined by

$$\langle \delta_{\rm DM}(\boldsymbol{k}, 0)\delta_{\rm DM}(\boldsymbol{k}', 0)\rangle = (2\pi)^3 P_{\rm DM}(k)\delta_{\rm Dirac}(\boldsymbol{k} - \boldsymbol{k}'). \tag{2}$$

The power spectrum is a function only of the magnitude of $\boldsymbol{k}$, because of the isotropy of the background universe.

The linear density contrast for baryons in the IGM can be obtained from the DM density contrast by smoothing over scales below the Jeans length. We use the relation (Fang et al. 1993)

$$\delta_{\rm B}(\boldsymbol{k}, z) = \frac{\delta_{\rm DM}(\boldsymbol{k}, z)}{1 + x_{\rm b}^2(z)k^2}, \tag{3}$$

where

$$x_{\rm b}(z) = \frac{1}{H_0}\left[\frac{2\gamma k_{\rm B} T_{\rm m}(z)}{3\mu m_{\rm p}\Omega_{\rm m}(1+z)}\right]^{1/2} \tag{4}$$

is the Jeans length; $T_{\rm m}$ and μ are the density-averaged temperature and mean molecular weight of the IGM respectively; $\Omega_{\rm m}$ is the cosmological density parameter of total mass and γ is the ratio of specific heats. Strictly speaking, equation (3) is valid only for the case where $x_{\rm b}$ is independent of z, but it is shown by Bi, Börner & Chu (1992) that equation (3) is a good approximation for $\delta_{\rm B}(\boldsymbol{k}, z)$ even when $x_{\rm b}$ has a redshift dependence. The linear density contrast in real comoving space, $\delta(\boldsymbol{x}, z)$, is the Fourier transform of equation (3).

In principle, to study the properties of the IGM one has to take into account the non-linearities in the density distribution and various physical processes such as shocks, radiation field, cooling, etc. However, detailed hydrodynamical modelling of the IGM has shown that most of the low column density Lyα absorption (i.e. $N_{\rm H\,I} \leq 10^{14}\,{\rm cm}^{-2}$) is produced by regions that are either in the linear or in the weakly non-linear regime (Cen et al. 1994; Zhang et al. 1995; Hernquist et al. 1996; Miralda-Escudé et al. 1996;

Theuns et al. 1998a,b; Davé et al. 1999). The lower envelope of the column density, $N_{\rm H\,I}$, versus the thermal velocity dispersion, b [given by $b = (2k_{\rm B}T/m_{\rm p})^{1/2}$], scatter plot (Schaye et al. 1999, 2000) suggests that there is a well-defined relationship between the density and the temperature of the IGM (Hui & Gnedin 1997). Thus it is possible to model low column density systems using a simple prescription for the non-linear density field and an equation of state.

In this work, we take into account the effect of non-linearity by assuming the number density distribution of the baryons, $n_{\rm B}(\boldsymbol{x},z)$, to be a lognormal random field

$$n_{\rm B}(\boldsymbol{x}, z) = A\,{\rm e}^{\delta_{\rm B}(\boldsymbol{x},z)}, \tag{5}$$

where A is a constant to be determined. The mean value of $n_{\rm B}(\boldsymbol{x}, z)$ is given by

$$\langle n_{\rm B}(\boldsymbol{x}, z)\rangle \equiv n_0(z) = A\langle {\rm e}^{\delta_{\rm B}(\boldsymbol{x},z)}\rangle. \tag{6}$$

Since $\delta_{\rm B}(\boldsymbol{x}, z)$ is a Gaussian random field, one can write

$$\langle {\rm e}^{\delta_{\rm B}(\boldsymbol{x},z)}\rangle = {\rm e}^{\Delta^2(z)/2}, \tag{7}$$

where

$$\Delta^2(z) = \langle \delta_{\rm B}^2(\boldsymbol{x}, z)\rangle = D^2(z)\int \frac{{\rm d}^3k}{(2\pi)^3}\frac{P_{\rm DM}(k)}{[1 + x_{\rm b}^2(z)k^2]^2}. \tag{8}$$

Hence

$$A = n_0(z)\,{\rm e}^{-\Delta^2(z)/2} \tag{9}$$

and

$$n_{\rm B}(\boldsymbol{x}, z) = n_0(z)\exp\left[\delta_{\rm B}(\boldsymbol{x}, z) - \frac{\Delta^2(z)}{2}\right]. \tag{10}$$

The lognormal distribution was introduced by Coles & Jones (1991) as a model for the non-linear matter distribution in the Universe. This ansatz has several interesting features.

(i) It can be seen that the matter density given by equation (10) is always positive, even when $\delta_{\rm B} \rightarrow -\infty$, unlike any polynomial function of $\delta_{\rm B}$. When the density contrast is small ($\delta_{\rm B} \ll 1$), equation (10) reduces to $n_{\rm B}/n_0 \simeq 1 + \delta_{\rm B}$, which is just what we expect from linear theory.

(ii) On small scales, equation (10) becomes the isothermal hydrostatic solution, which describes highly clumped structures like intracluster gas, $n_{\rm B} \propto \exp(-\mu m_{\rm p}\psi_{\rm DM}/\gamma k_{\rm B}T)$, where $\psi_{\rm DM}$ is the DM potential (Sarazin & Bahcall 1977). The lognormal function can be thought of as the simplest function that links these two extreme regions smoothly.

(iii) One can also think of the lognormal distribution as the kinematic model for the density field. If one assumes that the initial density and velocity fields are Gaussian, and extrapolates the continuity equation into non-linear regimes, treating the velocity field as linear, it turns out that the non-linear density field obtained in such a manner follows the lognormal distribution (Coles & Jones 1991).

(iv) Bi & Davidsen (1997) have tested the distribution against hydrodynamical simulations, and found a reasonable match between them. The lognormal assumption has also been used to model the IGM in numerical simulations (Bi 1993; Bi & Davidsen 1997), and is found to be working well in reproducing the observations. In particular, the simulation results match well with the observed column density distribution and number density of the Lyα absorption lines, the probability distribution of the b-parameter, etc. (see Bi & Davidsen 1997).

We shall also discuss briefly a more general argument as to why the lognormal distribution should be the natural choice in a large class of phenomena. There is a wide class of quantities, denoted by f, the time evolution of which can be characterized by the following property – the change in the value of f at some instant t_i is proportional to its value at that instant, with the proportionality factor being a random variable. In mathematical notation, this can be written as $f(t_{i+1}) = f(t_i) + \varepsilon_i f(t_i)$, where ε_i is the random variable. [Some examples of such phenomena in the sociological context are (i) the rich getting richer through fluctuations in the stock market, and (ii) more facilities being provided for people who already have them.] A similar situation can occur in structure formation scenarios also. The regions that have high density, because of stronger gravitational attraction, have a better chance of acquiring more mass. Let us denote the density field at some particular point at a given epoch t_i by $n(t_i)$ and postulate the evolution:

$$n(t_{i+1}) = n(t_i) + \varepsilon_i n(t_i) = (1 + \varepsilon_i)n(t_i). \tag{11}$$

We can now write $n(t_{i+1})$ in terms of some initial density field $n(t_0)$:

$$n(t_{i+1}) = (1 + \varepsilon_i)(1 + \varepsilon_{i-1})\ldots(1 + \varepsilon_0)n(t_0). \tag{12}$$

Taking the logarithm of both sides:

$$\ln[n(t_{i+1})] = \sum_{j=0}^{i}\ln(1 + \varepsilon_j) + \ln[n(t_0)]. \tag{13}$$

It is clear that when the time interval $(t_{i+1} - t_i)$ is small, the mass acquired within that interval will also be very small. Hence we expect that $\varepsilon_i \ll 1$. Then the above expression becomes

$$\ln\left[\frac{n(t_{i+1})}{n(t_0)}\right] = \sum_{j=0}^{i}\varepsilon_j. \tag{14}$$

This means that $\ln[n(t)/n(t_0)]$ is a sum of a large number of uncorrelated random variables. Using the central limit theorem, we can conclude that it follows a Gaussian distribution or, equivalently, $n(t)$ follows a lognormal distribution. This suggests that it may be reasonable to try an ansatz that the distribution of the non-linear baryonic density field is lognormal.

As an aside, we just mention that our analysis described here can easily be carried out for any other local ansatz for the non-linear baryonic density. [The results for a power-law assumption in which $n_{\rm B} \propto (1 + \delta)^p$ will be discussed in a later paper.]

Once we have obtained the total baryonic density, the fraction of hydrogen in the neutral form, f, in the IGM can be obtained by solving the ionization equilibrium equation

$$\alpha[z, T(z)]n_{\rm p}n_{\rm e} = J(z)n_{\rm H\,I}, \tag{15}$$

where $\alpha[z, T(z)]$ is the radiative recombination rate and $J(z)$ is the rate of photoionization for hydrogen at redshift z (Black 1981); $n_{\rm p}$, $n_{\rm e}$ and $n_{\rm H\,I}$ are the number densities of protons, electrons and neutral hydrogen, respectively. For simplicity, we assume that hydrogen is the only element present in the IGM and neglect the presence of helium and other heavier elements. In such a case, we have $n_{\rm e} = n_{\rm p}$. [This relation is not valid in the presence of helium or other heavier elements. If we had taken their presence into account, we would have got $n_{\rm e} = \kappa n_{\rm p}$, where κ is a constant.

Usually, $1 \leq \kappa \leq 1.2$, because the amount of helium and heavier elements in the IGM is small compared with hydrogen. Since we do not know $J(z)$ beyond an accuracy of 10–20 per cent, we can always absorb κ into $J(z)$.] Let us define the neutral fraction of hydrogen, f, by

$$f = \frac{n_{\rm H\,I}}{n_{\rm B}} = \frac{n_{\rm H\,I}}{n_{\rm H\,I} + n_{\rm p}}. \tag{16}$$

Hence we get from equation (15)

$$\frac{(1-f)^2}{f} = \frac{J(z)}{\alpha[z, T(z)] n_{\rm B}}. \tag{17}$$

In general, one can solve this equation and determine f as a function of $n_{\rm B}$. This expression simplifies for two extreme cases. For $f \ll 1$, we get

$$f = \frac{\alpha[z, T(z)] n_{\rm B}}{J(z)}, \tag{18}$$

and for $f \sim 1$ we get

$$f = 1 - \sqrt{\frac{J(z)}{\alpha[z, T(z)] n_{\rm B}}}. \tag{19}$$

Hence we have

$$n_{\rm H\,I}(\boldsymbol{x}, z) = \begin{cases} \dfrac{\alpha[z, T(z)]}{J(z)} n_{\rm B}^2(\boldsymbol{x}, z) & (\text{if } n_{\rm H\,I} \ll n_{\rm B}), \\ n_{\rm B}(\boldsymbol{x}, z) - \sqrt{\dfrac{J(z) n_{\rm B}(\boldsymbol{x}, z)}{\alpha[z, T(z)]}} & (\text{if } n_{\rm H\,I} \sim n_{\rm B}). \end{cases} \tag{20}$$

The ionization conditions in the Lyα absorbers are similar to those of H II regions with $f \simeq 10^{-4}$. Thus, from now on, we concentrate only on the case $n_{\rm H\,I} \ll n_{\rm B}$.

We take the temperature dependence of the recombination coefficient α to be given by (Rauch et al. 1997)

$$\alpha[z, T(z)] = \alpha_0 \left[\frac{T(z)}{10^4\,{\rm K}}\right]^{-0.7}, \tag{21}$$

where $\alpha_0 = 4.2 \times 10^{-13}\,{\rm cm^3\,s^{-1}}$. This relation is a good approximation for α in the temperature range relevant for the Lyα forest. The temperature T is related to the baryonic density $n_{\rm B}$ through the equation of state. We assume a polytropic equation of state $p \propto \rho^\gamma \propto n_{\rm B}^\gamma$, or equivalently

$$T(z) = T_0(z)[n_{\rm B}(z)/n_0(z)]^{\gamma-1}, \tag{22}$$

where

$$n_0(z) = \frac{\Omega_{\rm baryon}\rho_{\rm c}}{\mu_{\rm b} m_{\rm p}}(1+z)^3 \tag{23}$$

is the mean baryonic number density at redshift z. $\rho_{\rm c}$ is the critical matter density at the present epoch, given by

$$\rho_{\rm c} = 1.8791 \times 10^{-29} h^2\,{\rm cm^{-3}}, \tag{24}$$

and $\mu_{\rm b} m_{\rm p}$ is the mass per baryonic particle. Then the H I density becomes

$$n_{\rm H\,I}(\boldsymbol{x}, z) = F(z)\left[\frac{n_{\rm B}(\boldsymbol{x}, z)}{n_0(z)}\right]^{\beta}, \tag{25}$$

where

$$F(z) = \alpha_0 n_0^2(z)\left[\frac{T_0(z)}{10^4\,{\rm K}}\right]^{-0.7} J^{-1}(z) \tag{26}$$

and

$$\beta = 2.7 - 0.7\gamma. \tag{27}$$

(We note, in passing, that β becomes negative if $\gamma > 3.86$.) We can write the H I density in terms of the linear baryonic density contrast:

$$n_{\rm H\,I}(\boldsymbol{x}, z) = F_1(z) \exp[\beta \delta_{\rm B}(\boldsymbol{x}, z)], \tag{28}$$

where

$$F_1(z) = F(z)\,{\rm e}^{-\beta\Delta^2(z)/2}. \tag{29}$$

It is clear from equation (28) that the H I distribution at a particular redshift is also described by a lognormal distribution. All the statistical quantities regarding H I can be derived from this in a straightforward manner.

2.1 Correlation function for neutral hydrogen

One of our main interests is the correlation function

$$\langle n_{\rm H\,I}(\boldsymbol{x}, z) n_{\rm H\,I}(\boldsymbol{x}', z')\rangle = F_1(z)F_1(z')\langle \exp\{\beta[\delta_{\rm B}(\boldsymbol{x}, z) + \delta_{\rm B}(\boldsymbol{x}', z')]\}\rangle, \tag{30}$$

from which several useful quantities can be obtained. Since $\delta_{\rm B}$ is a Gaussian random field, we can write, using equation (7),

$$\langle \exp\{\beta[\delta_{\rm B}(\boldsymbol{x}, z) + \delta_{\rm B}(\boldsymbol{x}', z')]\}\rangle = \exp\left\{\frac{\beta^2}{2}[\Delta^2(z) + \Delta^2(z') + 2Q(\boldsymbol{x}, \boldsymbol{x}'; z, z')]\right\}, \tag{31}$$

where

$$Q(\boldsymbol{x}, \boldsymbol{x}'; z, z') = \langle \delta_{\rm B}(\boldsymbol{x}, z)\delta_{\rm B}(\boldsymbol{x}', z')\rangle. \tag{32}$$

Simple algebra gives

$$Q(\boldsymbol{x}, \boldsymbol{x}'; z, z') \equiv Q(\boldsymbol{x} - \boldsymbol{x}'; z, z') = D(z)D(z')\int \frac{{\rm d}^3k}{(2\pi)^3} \frac{P_{\rm DM}(k)\,{\rm e}^{{\rm i}\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')}}{[1 + x_{\rm b}^2(z)k^2][1 + x_{\rm b}^2(z')k^2]}. \tag{33}$$

One also notes from equations (8) and (33) that $\Delta^2(z) = Q(0; z, z)$. We can now write equation (30) as

$$\langle n_{\rm H\,I}(\boldsymbol{x}, z) n_{\rm H\,I}(\boldsymbol{x}', z')\rangle = F_2(z)F_2(z')\,{\rm e}^{\beta^2 Q(\boldsymbol{x}-\boldsymbol{x}'; z, z')}, \tag{34}$$

where

$$F_2(z) = F_1(z)\,{\rm e}^{\beta^2\Delta^2(z)/2}. \tag{35}$$

One needs to normalize the quantity $\langle n_{\rm H\,I}(\boldsymbol{x}, z) n_{\rm H\,I}(\boldsymbol{x}', z')\rangle$, to obtain the correlation function $\xi_{\rm H\,I}(\boldsymbol{x} - \boldsymbol{x}'; z, z')$ for H I. A natural way of normalizing the correlation would be to use the definition

$$1 + \xi_{\rm H\,I}(\boldsymbol{x} - \boldsymbol{x}'; z, z') = \frac{\langle n_{\rm H\,I}(\boldsymbol{x}, z) n_{\rm H\,I}(\boldsymbol{x}', z')\rangle}{\langle n_{\rm H\,I}(\boldsymbol{x}, z)\rangle\langle n_{\rm H\,I}(\boldsymbol{x}', z')\rangle}. \tag{36}$$

Since $\langle n_{\rm H\,I}(\boldsymbol{x}, z)\rangle = F_2(z)$, we get

$$\xi_{\rm H\,I}(\boldsymbol{x} - \boldsymbol{x}'; z, z') = {\rm e}^{\,\beta^2 Q(\boldsymbol{x}-\boldsymbol{x}'; z, z')} - 1, \tag{37}$$

with Q given by equation (33).

All the analysis above is valid if one can probe any scale with arbitrary accuracy. It turns out, however, that one cannot obtain information about scales smaller than some particular value,

owing to various observational constraints. While observing along a LOS, it will be impossible to probe velocity scales less than the spectroscopic limit owing to thermal broadening and the blending of spectral lines. Similarly, while observing across the transverse direction, the peculiar velocities of individual points will constrain the velocity resolution which we have not taken into account in the above analysis. If Δv is the smallest scale that one can probe, then the corresponding limit in redshift space is

$$\Delta z = \frac{\Delta v}{c}(1+z). \tag{38}$$

This means that we will not be able to probe below a comoving length-scale given by

$$\Delta x(z) = d_H(z)\Delta z, \tag{39}$$

where

$$d_H(z) = c\left(\frac{\dot{a}}{a}\right)^{-1} = \frac{c}{H_0}[\Omega_\Lambda + \Omega_{\rm m}(1+z)^3 + \Omega_k(1+z)^2]^{-1/2}, \tag{40}$$

$$\Omega_k = 1 - \Omega_{\rm m} - \Omega_\Lambda. \tag{41}$$

This effect can be included in our calculation by smoothing over all length-scales smaller than $\Delta x(\bar{z})$, where

$$\bar{z} = \tfrac{1}{2}(z + z') \tag{42}$$

is the average redshift. We use a Gaussian window of width $\sigma_x(\bar{z}) = \Delta x(\bar{z})$, and get a smoothed version of Q in equation (33). In Fourier space, this smoothing will introduce an extra Gaussian term in the integrand, and our smoothed Q will be

$$Q_{\rm smooth}\ (\boldsymbol{x} - \boldsymbol{x}'; z, z') = D(z)D(z') \times \int \frac{{\rm d}^3k}{(2\pi)^3} \frac{P_{\rm DM}(k)\, {\rm e}^{-k^2\sigma_x^2(\bar{z})/2}\, {\rm e}^{{\rm i}\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')}}{[1 + x_{\rm b}^2(z)k^2][1 + x_{\rm b}^2(z')k^2]}. \tag{43}$$

The angular integrations can be carried out trivially, and we get

$$Q_{\rm smooth}\ (\boldsymbol{x} - \boldsymbol{x}'; z, z') = \frac{D(z)D(z')}{2\pi^2} \times \int_0^\infty {\rm d}k \frac{P_{\rm DM}(k)k^2\, {\rm e}^{-k^2\sigma_x^2(\bar{z})/2}}{[1 + x_{\rm b}^2(z)k^2][1 + x_{\rm b}^2(z')k^2]} \frac{\sin kX}{kX}, \tag{44}$$

where $X = |\boldsymbol{x} - \boldsymbol{x}'|$. The final integration can be done numerically, once the DM power spectrum is given.

At this stage, the relations derived above can be used for any $\boldsymbol{x}$, $\boldsymbol{x}'$, z, z'. As we mentioned earlier, if one observes the H I along a particular LOS, then one is probing different regions of the IGM at different redshifts. The position x will be related to the redshift z by the relation

$$x(z) = \int_0^z d_H(z')\, {\rm d}z', \tag{45}$$

where $d_H(z)$ is given by equation (40). Then the LOS correlation function is given by

$$\xi_{\rm H\,I}^{\rm LOS}(z, z') = {\rm e}^{\beta^2 Q_{\rm LOS}[l(z,z');z,z']} - 1, \tag{46}$$

where

$$Q_{\rm LOS}[l(z,z');z,z'] = \frac{D(z)D(z')}{2\pi^2} \int_0^\infty {\rm d}k \frac{P_{\rm DM}(k)k^2\, {\rm e}^{-k^2\sigma_x^2(\bar{z})/2}}{[1 + x_{\rm b}^2(z)k^2][1 + x_{\rm b}^2(z')k^2]} \frac{\sin kl}{kl} \tag{47}$$

and

$$l(z, z') = x(z) - x(z'). \tag{48}$$

It should be stressed that $\xi_{\rm H\,I}^{\rm LOS}(z, z') \neq \xi_{\rm H\,I}^{\rm LOS}(z - z')$. This means that one cannot rigorously define a power spectrum from the LOS correlation function because the correlation is a function of *two* variables z and z'. *In other words, the LOS power spectrum does not exist in the strict sense.* However, one can get an approximate LOS power spectrum for a small redshift range around any mean redshift. We have already defined the average redshift in equation (42). We define a redshift difference

$$\Delta z = z - z' \tag{49}$$

and evaluate the correlation function for a particular value of $\bar{z}$ as a function of Δz, i.e.

$$\xi_{\rm H\,I}^{\rm LOS}(\bar{z}, \Delta z) = {\rm e}^{\beta^2 Q_{\rm LOS}(\bar{z},\Delta z)} - 1, \tag{50}$$

where

$$Q_{\rm LOS}(\bar{z}, \Delta z) = \frac{D(\bar{z} + \Delta z/2)D(\bar{z} - \Delta z/2)}{2\pi^2} \int_0^\infty {\rm d}k \left\{ \frac{\sin[kd_H(\bar{z})\Delta z]}{kd_H(\bar{z})\Delta z} \times \frac{P_{\rm DM}(k)k^2\, {\rm e}^{-k^2\sigma_x^2(\bar{z})/2}}{[1 + x_{\rm b}^2(\bar{z} + \Delta z/2)k^2][1 + x_{\rm b}^2(\bar{z} - \Delta z/2)k^2]} \right\}. \tag{51}$$

For small Δz, one can use equation (39) to write the correlation as a function of Δx, Fourier transform the correlation, and get the power spectrum. Such a power spectrum will depend on the value of $\bar{z}$. We stress again that this power spectrum is approximate in the sense that it exists only for $\Delta z \ll \bar{z}$.

The transverse correlation is observed at some particular redshift ($z = z'$), along the transverse direction. Then

$$\xi_{\rm H\,I}^{\rm trans}(l_\perp; z) = {\rm e}^{\beta^2 Q_{\rm trans}(l_\perp;z)} - 1, \tag{52}$$

$$Q_{\rm trans}(l_\perp; z) = \frac{D^2(z)}{2\pi^2} \int_0^\infty {\rm d}k \frac{P_{\rm DM}(k)k^2\, {\rm e}^{-k^2\sigma_x^2(z)/2}}{[1 + x_{\rm b}^2(z)k^2]^2} \frac{\sin kl_\perp}{kl_\perp}, \tag{53}$$

where $l_\perp$ is the comoving distance along the transverse direction. For a given redshift, the transverse correlation is only a function of $l_\perp$. Hence one can obtain the power spectrum from $\xi_{\rm H\,I}^{\rm trans}$ following the usual methods.

2.2 Column density distribution

One of the other statistics that observers use to quantify the distribution of Lyα absorption lines is the column density distribution. Indeed, one can get the analytic expression for this using the formalism developed so far in this work. Note that Voigt profile fitting to the absorption lines is used to get the observed column density distribution. Here we use a method called 'density-peak ansatz' (DPA), discussed by Gnedin & Hui (1996) and Hui et al. (1997), to derive an analytic expression for the column density distribution.

Suppose we are looking at the IGM along any one direction, at some redshift z. Then the linear density field $\delta_{\rm B}^{(1{\rm D})}(x, z)$ along that LOS will be described by a one-dimensional Gaussian random field. DPA assumes that each density peak in the comoving space is associated with an absorption line, and one can assign a definite column density to each of them. In the articles referred to above, each density peak is fitted with a Gaussian, and the column density is calculated using

$$N_{\rm H\,I} \propto \int_{\rm peak} n_{\rm H\,I}(x)\, {\rm d}x. \tag{54}$$

In such a case, there is a definite correlation between the value of

the density field at the peak, and the effective width of the absorber (which is determined by the correlation between the density field and its second derivative at the peak, and is fixed once the fitting function for the density peak is given). We, however, take a simpler approach (described below) in assigning the column density to a density peak.

The coherence scale of the distribution is defined as (Bardeen et al. 1986)

$$R^* \equiv \frac{\sigma_1}{\sigma_2}, \tag{55}$$

where σ_1 and σ_2 are defined in equation (A3) (see Appendix A). This length is a measure of the distance between two successive zeros for the one-dimensional Gaussian random field. Since this is the relevant scale for the distribution of zero-crossing, we expect the effective length-scale of a peak to be a fraction of R^*. Then the column density corresponding to a particular peak will be

$$N_{\rm H\,I} \propto n_{\rm H\,I}({\rm peak})R^* = n_{\rm H\,I}({\rm peak})R^*\epsilon, \tag{56}$$

where $n_{\rm H\,I}$(peak) is the H I number density at the peak and ϵ is the proportionality constant, which can be used as a free parameter in comparing with observations. We have assumed ϵ to be independent of $N_{\rm H\,I}$, which means that the column density is directly proportional to the peak density. Using this prescription for obtaining the column density from the H I density, we can easily obtain the relation between $N_{\rm H\,I}$ and the total baryonic overdensity $n_{\rm B}/n_0$ using equation (20). For the case $n_{\rm H\,I} \ll n_{\rm B}$, the relation is given by

$$\frac{n_{\rm B}({\rm peak})}{n_0} = \left[\frac{N_{\rm H\,I}}{R^*\epsilon F(z)}\right]^{1/\beta}. \tag{57}$$

The relation between $N_{\rm H\,I}$ and $\delta_{\rm B}^{(1{\rm D})}$ is then given by

$$\delta_{\rm B}^{(1{\rm D})}({\rm peak}) = \frac{1}{\beta}\ln\left[\frac{N_{\rm H\,I}}{R^*\epsilon F(z)}\right] + \frac{\Delta^2}{2}. \tag{58}$$

Given this relation, it is straightforward to obtain the quantity $dN_{\rm pk}/(dz dN_{\rm H\,I})$, defined as the number of clouds (peaks) per column density interval per redshift interval. For completeness, we give the relevant calculations in the Appendix.

3 MODEL PARAMETERS

In this section we discuss the various model parameters used in obtaining the results. The parameters defining the model can be divided into two categories: (i) cosmological parameters; and (ii) parameters related to the IGM.

The first set of parameters comprises those that determine the background cosmology. We assume that the background universe is described by the Friedmann–Robertson–Walker (FRW) metric. We have considered four different cosmological models with the parameters listed below.

[SCDM]: $\Omega_{\rm m} = 1$, $\Omega_\Lambda = 0$, $h = 0.5$;
[OCDM]: $\Omega_{\rm m} = 0.35$, $\Omega_\Lambda = 0$, $h = 0.5$;
[LCDM1]: $\Omega_{\rm m} = 0.35$, $\Omega_\Lambda = 0.65$, $h = 0.5$;
[LCDM2]: $\Omega_{\rm m} = 0.35$, $\Omega_\Lambda = 0.65$, $h = 0.65$.

The next cosmological input that is required is the form of the DM power spectrum. We take the following form for $P_{\rm DM}(k)$ (Efstathiou, Bond & White 1992):

$$P_{\rm DM}(k) = \frac{Ak}{\{1 + [ak + (bk)^{1.5} + (ck)^2]^\nu\}^{2/\nu}}, \tag{59}$$

where $\nu = 1.13$, $a = (6.4/\Gamma)\,h^{-1}$ Mpc, $b = (3.0/\Gamma)\,h^{-1}$ Mpc, $c = (1.7/\Gamma)\,h^{-1}$ Mpc and $\Gamma = \Omega_{\rm m}h$. The normalization parameter A is fixed through the value of σ_8 (the rms density fluctuation in spheres of radius $8\,h^{-1}$ Mpc). We take the values of σ_8 to be given by (Eke, Cole & Frenk 1996)

$$\sigma_8 = \begin{cases} (0.52 \pm 0.04)\Omega_{\rm m}^{-0.46+0.10\Omega_{\rm m}} & ({\rm if}\ \ \Omega_\Lambda = 0), \\ (0.52 \pm 0.04)\Omega_{\rm m}^{-0.52+0.13\Omega_{\rm m}} & ({\rm if}\ \ \Omega_\Lambda = 1 - \Omega_{\rm m}). \end{cases} \tag{60}$$

The next set of parameters is related to the physical conditions in the IGM. The parameters that we need to describe the IGM are γ, $T_{\rm m}$, $\Omega_{\rm baryon}$, $J(z)$ and $T_0(z)$. All these quantities were defined in Section 2. Besides these, we have also introduced a parameter ϵ while modelling the column density distribution of the IGM. This is taken as a free parameter, to be fixed through observations.

It is known that the value of γ, at any given epoch, depends on the reionization history of the Universe (Hui & Gnedin 1997). The value of γ and its evolution are still quite uncertain. Using Voigt profile fits to the observed Lyα absorption lines, one can in principle obtain the value of γ. Available observations are consistent with γ in the range $1.2 < \gamma < 1.7$ (Schaye et al. 2000) for $2 \le z \le 4.5$. As far as the evolution of γ is concerned, we shall treat it as independent of z. The density-averaged temperature is defined as

$$T_{\rm m} = \frac{\int \rho(T) T\, {\rm d}T}{\int \rho(T)\, {\rm d}T}. \tag{61}$$

Using the equation of state $T \sim \rho^{\gamma-1}$, we get

$$T_{\rm m} = \frac{2\gamma - 1}{\gamma}(T_{\rm max} - T_{\rm min}). \tag{62}$$

We take $T_{\rm m}$ to be in the range $10\,000 < T_{\rm m} < 60\,000$ K. The minimum value of $T_{\rm m}$ corresponds to the minimum temperature of the IGM, which is determined by the photoionization equilibrium. The maximum value of $T_{\rm m}$ corresponds to a b-parameter $\sim$31.7 km s^{-1} and is consistent with the minimum value of b observed at higher H I column densities (i.e. $10^{14.5}$ cm^{-2}). The evolution of $T_{\rm m}$ depends on how $T_{\rm max}$ and $T_{\rm min}$ evolve with redshift. One possibility is to take the adiabaticity relation $T_{\rm m} \sim (1+z)^{3\gamma-3}$. However, it has been argued that, since there is no conclusive evidence of the evolution of the temperature in IGM, one should treat the mean temperature of the IGM as constant (Bi et al. 1995). Hence, in this paper, we also consider the second possibility where $T_{\rm m}$ is independent of z. For the cases where we consider a small redshift range Δz around a mean redshift $\bar{z}$, the effect of the evolution of $T_{\rm m}$ is not very significant. Whenever we study the evolution of a quantity over a large redshift range, however, we have to take into account the various redshift dependences of $T_{\rm m}$.

We note that when we normalize the correlation function for neutral hydrogen, the parameters $\Omega_{\rm baryon}$, $J(z)$ and $T_0(z)$ cancel out. Hence knowledge of these parameters is not necessary for modelling the correlation function. However, in the case of the column density distribution, they appear as a combination $(\Omega_{\rm baryon}h^2)^2\,J^{-1}(z)T_0^{-0.7}(z)$ through $F(z)$ [equations (26) and (23)]. The values of these quantities are not known accurately, nor do we know how $J(z)$ and $T_0(z)$ evolve. In this work, we take $J(z)$ to be independent of z. We fix the value of the combination $(\Omega_{\rm baryon}h^2)^2\,J^{-1}$ to be $(0.026)^2 \times 10^{12}$ s, which is consistent with the values given by McDonald et al. (2000). One should note that any change in the values of the above parameters can be compensated (to some extent) by changing the value of ϵ,

which is a free parameter in our model for the column density distribution. For T_0 evolution, we shall consider two separate cases like T_m, i.e. (i) $T_0 \sim (1+z)^{3\gamma-3}$, and (ii) T_0 independent of z.

As we have discussed earlier, we need to smooth the power spectrum below some velocity because the blending of spectral lines makes it impossible to resolve the lines below a particular velocity. Typically this velocity is of the order of a few tens of $\mathrm{km\,s^{-1}}$. For definiteness, we take the smoothing velocity to be $\Delta v = 30\,\mathrm{km\,s^{-1}}$.

4 RESULTS

4.1 LOS correlation

We shall now compute the results for the H I correlation function along the LOS as a function of the velocity separation v, where v is related to Δz by $v = c\Delta z/(1+\bar{z})$.

The results for the LOS correlation function for different cosmological models are shown in Fig. 1. We have chosen typical values of $T_m = 40\,000\,\mathrm{K}$ and $\gamma = 1.5$, at a redshift of $\bar{z} = 2.5$. It can be seen that the correlation curves tend to flatten at low velocities, and go to zero at high velocities.

Miralda-Escudé et al. (1996) give the correlation function of the transmitted flux along a LOS (see the solid curve in their fig. 12a). We note that the correlation curves for the transmitted flux and the neutral hydrogen density need not be exactly the same. However, we expect that the broad features and the general trends should be alike. It turns out that the correlation curve for transmitted flux obtained from the simulations does have the same trend as our results.

We can investigate how the shape of $\xi_{\mathrm{LOS}}(v)$ depends on various parameters. We find that the $\xi_{\mathrm{LOS}}(v)$ curve falls approximately like a power law, $\xi_{\mathrm{LOS}}(v) \propto v^{-p}$, at velocities in the range 100–1000 $\mathrm{km\,s^{-1}}$. At lower velocities the curve is practically independent of v. We have given the value of p and the rms error on p for different cosmological models and for different IGM parameters in Table 1.

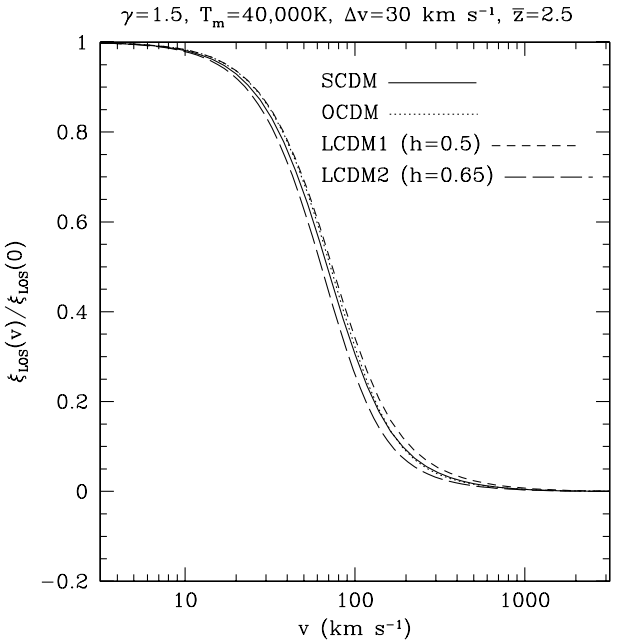

Figure 1. LOS correlation function as a function of velocity separation. Results for four different cosmological models at a mean redshift of $\bar{z} = 2.5$ are presented. The correlation function is normalized to unity at zero velocity separation.

The dependence of p on the IGM parameters can be understood easily. A higher value of T_m implies a larger x_b which, in turn, implies more smoothing of the power spectrum at low scales. However, the larger scales are more or less unaffected by x_b. Consequently, the correlation curve becomes flatter as we increase T_m. Also, for a fixed value of γ, the effect of the cosmological models on the shape of ξ_{LOS} is large for low T_m.

The effect of γ on ξ_{LOS} is twofold – increasing γ introduces more smoothing at low scales just like T_m, and there is also a reduction in the neutral hydrogen density fluctuations for given baryonic fluctuations (see equation 28). Both of these effects make the correlation curve flatter, which is what we see from Table 1. Furthermore, because of this twofold effect, γ affects p much more than T_m does. This point can be seen clearly in Fig. 2. In the left-hand panel, we have kept x_b constant, and shown the effect of changing only the neutral hydrogen density fluctuations. The middle panel shows the effect of changing only T_m or, equivalently, the Jeans scale x_b, without changing anything else. In the right-hand panel, the full effect of γ can be seen, as it changes both the Jeans scale and the neutral hydrogen density fluctuations. Since T_m does not have much effect on p, we can determine γ from observations of the LOS correlation even with ill-constrained values of T_m, provided that the cosmology is known from other studies.

Just like in the case of T_m, the effect of the cosmological models on p is large for low γ. The first reason for this is same as in the case of T_m – low γ implies less smoothing and hence DM fluctuations are more effective. The second reason is that, for low γ, the neutral hydrogen density fluctuations are larger for given linear baryonic density fluctuations. Thus a slight change in the DM fluctuations causes a significant change in the neutral hydrogen fluctuations (for a fixed x_b) which, in turn, makes the correlation function sensitive to the DM power spectrum.

As the universe evolves after reionization, it is possible that the value of γ increases (Hui & Gnedin 1997) and the LOS correlation becomes less and less sensitive to the DM power spectrum. Hence, if the reionization has occurred very early, it is extremely unlikely that one can fix the cosmological parameters from observations of ξ_{LOS}. We need to know not only the values of γ and T_m accurately, but also the value of p to an accuracy better than 10 per cent.

Finally, we comment on the effect of h on the shape of ξ_{LOS}. As one can clearly see from Table 1, the effect is quite significant ($\sim$12 per cent). The parameter h affects the LOS correlation function (equation 51) in three ways: (i) it changes the shape of $P_{\mathrm{DM}}(k)$ (see equation 59); (ii) it affects the value of x_b (equation 4);

Table 1. Power-law index, p, for LOS correlation function, defined by $\xi_{\mathrm{LOS}}(v) \propto v^{-p}$ in the velocity range 100–1000 $\mathrm{km\,s^{-1}}$ for $z = 2.5$. We also give the 1σ error in p.

		$\gamma = 1.2$	$\gamma = 1.7$
$T_m = 10\,000\,\mathrm{K}$	SCDM	2.03±0.015	1.78±0.004
	OCDM	2.29 ± 0.036	1.81 ± 0.012
	LCDM1	1.93 ± 0.017	1.63 ± 0.003
	LCDM2	2.19 ± 0.029	1.83 ± 0.008
$T_m = 60\,000\,\mathrm{K}$	SCDM	1.94 ± 0.003	1.68 ± 0.015
	OCDM	2.10 ± 0.014	1.62 ± 0.011
	LCDM1	1.83 ± 0.005	1.53 ± 0.012
	LCDM2	2.08 ± 0.013	1.71 ± 0.009

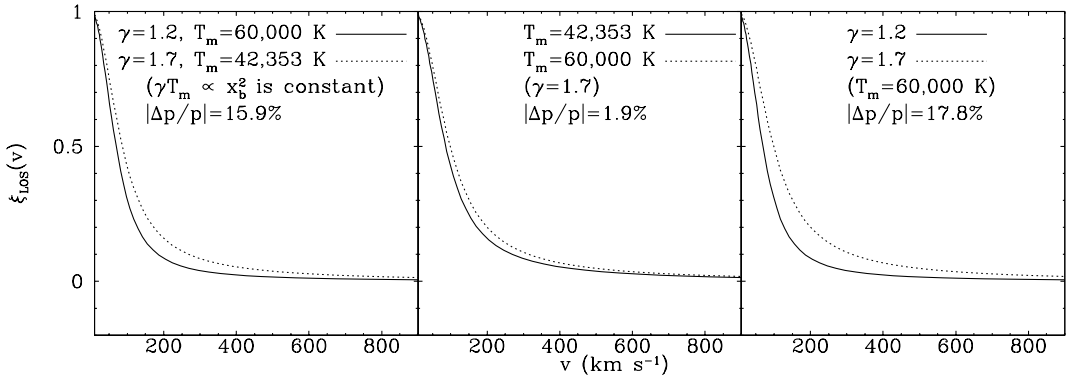

Figure 2. Comparison of the effects of γ and $T_{\rm m}$ on the power-law index of $\xi_{\rm LOS}(v)$. The plots are for the LCDM1 model. We give the relative change in the power-law index $|\Delta p/p|$ for the two curves in each panel. In the left-hand panel, $x_{\rm b}$ is kept constant. It shows the effect of changing the neutral hydrogen density fluctuations. The middle panel shows the effect of changing $x_{\rm b}$ without changing the neutral hydrogen density fluctuations (i.e. γ is constant). The right-hand panel shows the effect of γ, which affects both $x_{\rm b}$ and the neutral hydrogen density fluctuations.

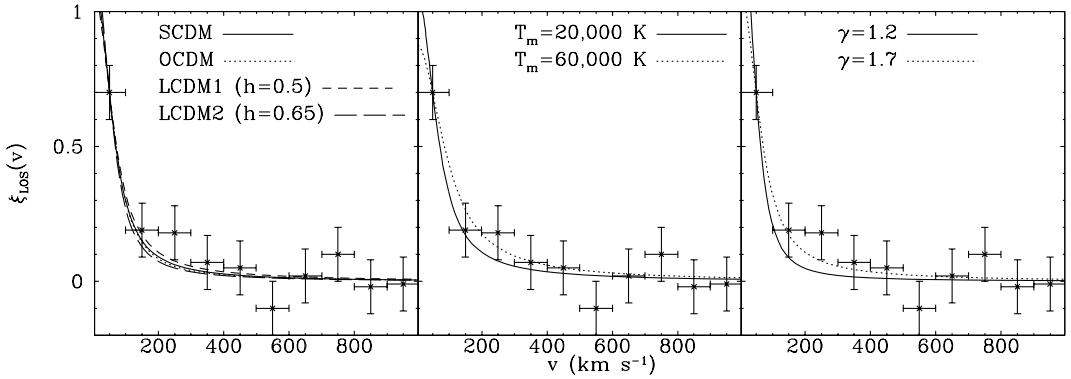

Figure 3. Comparison of the theoretical $\xi_{\rm LOS}$ with observational data (Cristiani et al. 1997). The theoretical curves have been normalized in such a way that they match with the observed data point at the lowest velocity bin. In the left-hand panel, the IGM parameters are fixed to be $T_{\rm m} = 20\,000$ K and $\gamma = 1.7$. In the middle panel, the cosmological model is LCDM1, and $\gamma = 1.7$. In the right-hand panel, again the cosmology is LCDM1, and $T_{\rm m} = 20\,000$ K.

and (iii) it affects the relation between distance and redshift. As a result, the values of $\sigma_x(\bar{z})$ and $d_H(\bar{z})$ in the integrand of equation (51) are modified. However, it turns out that the last two effects can be scaled out. To understand this more clearly, we rewrite equation (51) in a slightly modified form:

$$Q_{\rm LOS}(\bar{z},\Delta z) = \frac{D(\bar{z}+\Delta z/2)D(\bar{z}-\Delta z/2)}{2\pi^2} \times \int_0^\infty {\rm d}K \left\{ \frac{\sin[KD_H(\bar{z})\Delta z]}{KD_H(\bar{z})\Delta z} \times \frac{h^3 P_{\rm DM}(Kh) K^2\, {\rm e}^{-K^2\Sigma_x^2(\bar{z})/2}}{[1+X_{\rm b}^2(\bar{z}+\Delta z/2)K^2][1+X_{\rm b}^2(\bar{z}-\Delta z/2)K^2]} \right\}, \tag{63}$$

where $K \equiv k/h$ and

$$D_H(\bar{z}) \equiv d_H(\bar{z})h, \quad \Sigma_x(\bar{z}) \equiv \sigma_x(\bar{z})h, \quad X_{\rm b} \equiv x_{\rm b}h. \tag{64}$$

One can easily verify that all the three quantities defined above $(D_H, \Sigma_x, X_{\rm b})$ are independent of h. Hence $Q_{\rm LOS}(\bar{z},\Delta z)$ depends on h only through the combination $h^3 P_{\rm DM}(Kh)$. It is very difficult to study this function analytically. We have studied it using numerical methods and found that, for the power spectra used in this paper, its effect is to make the LOS correlation steeper when h is larger.

To compare our results with observational data, we take the data from Cristiani et al. (1997). The data consist of several QSO spectra at various redshifts, ranging from 1.7 to 3.7. This range is pretty large, and evolutionary effects will be significant in the data. We compare the observed LOS correlation (points with error bars) with the theoretical curve for the four cosmological models, and for various ranges of values of $T_{\rm m}$ and γ, in Fig. 3 for $\bar{z} = 2.5$. It should be noted that the observational data points were obtained using Lyα clouds with $\log(N_{\rm H\,I}/{\rm cm}^{-2}) > 14$. However, we have not used any such constraint while obtaining the analytical curves. As a preliminary check, we can see that the analytical curves have the broad features that are expected from the observational data. We hope to carry out a more detailed comparison with observations in a future publication.

We next check the redshift evolution of the LOS correlation function. For definiteness, we consider $\xi_{\rm LOS}$ at a particular velocity, $v = 100\,{\rm km\,s^{-1}}$, and study it as a function of $\bar{z}$. We have assumed that γ does not evolve with redshift. Since we are studying the evolution over a large redshift range, we have to consider both the possibilities for the evolution of $T_{\rm m}$ discussed in Section 3. The evolution curve closely follows a power-law dependence, i.e. $\xi_{\rm LOS}(\bar{z}) \propto (1+\bar{z})^{-q}$ in the range $1.5 < \bar{z} < 4.5$. We give the values of q and the rms error for different cases in Table 2.

It is clear that the cosmology has maximum influence when γ and $T_{\rm m}$ are small. The reason for this is same as that discussed in the case of Table 1.

Table 2. Power-law index, q, for the evolution of the LOS correlation function at velocity separation $100\,\mathrm{km\,s^{-1}}$, defined by $\xi_{\rm LOS}(\bar{z}) \propto (1+\bar{z})^{-q}$ in the redshift range 1.5–4.5. 1σ errors are given.

		$T_{\rm m} \sim (1+z)^{3\gamma-3}$	
		$\gamma = 1.2$	$\gamma = 1.7$
$T_{\rm m}(\bar{z}=2.5)$ =10 000 K	SCDM	4.00 ± 0.045	3.12 ± 0.025
	OCDM	5.69 ± 0.066	4.14 ± 0.039
	LCDM1	4.69 ± 0.062	3.56 ± 0.035
	LCDM2	5.28 ± 0.073	3.91 ± 0.042
$T_{\rm m}(\bar{z}=2.5)$ =60 000 K	SCDM	3.83 ± 0.042	3.38 ± 0.024
	OCDM	5.11 ± 0.057	4.18 ± 0.039
	LCDM1	4.44 ± 0.056	3.73 ± 0.034
	LCDM2	4.96 ± 0.067	4.08 ± 0.041
		$T_{\rm m}$ = constant	
		$\gamma = 1.2$	$\gamma = 1.7$
$T_{\rm m} = 10\,000$ K	SCDM	3.97 ± 0.045	2.98 ± 0.026
	OCDM	5.63 ± 0.064	3.89 ± 0.037
	LCDM1	4.66 ± 0.061	3.42 ± 0.035
	LCDM2	5.25 ± 0.072	3.75 ± 0.042
$T_{\rm m} = 60\,000$ K	SCDM	3.65 ± 0.039	2.72 ± 0.021
	OCDM	4.80 ± 0.049	3.24 ± 0.025
	LCDM1	4.25 ± 0.052	3.09 ± 0.028
	LCDM2	4.73 ± 0.061	3.33 ± 0.033

For a given cosmology and a given value of γ, the effect of $T_{\rm m}$ (or, equivalently, $x_{\rm b}$) on q is insignificant with the relative change in q being about 6–10 per cent. The reason for this is as follows: the effect of $x_{\rm b}$ is significant only for scales $<x_{\rm b}$. For the parameters that we are considering, the velocity scale corresponding to $x_{\rm b}$ is $\sim$10–30 $\mathrm{km\,s^{-1}}$. Hence, for the case where $v = 100\,\mathrm{km\,s^{-1}}$, the evolution will not be affected significantly by the Jeans length. We studied the evolution at a higher velocity scale ($v = 250\,\mathrm{km\,s^{-1}}$), and found that the effect of $x_{\rm b}$ was even less (the relative change in q was about 3 per cent at $v = 250\,\mathrm{km\,s^{-1}}$).

The effect of γ is threefold here – it affects the value of $x_{\rm b}$, the evolution of $x_{\rm b}$ (more precisely, this also depends on whether we evolve $T_{\rm m}$ or not) and the neutral hydrogen density fluctuations (through the value of β). We have already seen that the effect of changing the value of $x_{\rm b}$ is not very important. Increasing the value of γ will make the evolution of $x_{\rm b}$ more rapid. Since $x_{\rm b}$ appears in the denominator of the integrand in equation (51), the higher the value of γ, the more rapid is the decrease of $\xi_{\rm LOS}$ with increasing redshift. This feature alone will increase the value of q with increasing γ. Actually, however, q decreases when we increase γ because of the third effect of γ – increasing γ reduces the value of β (equation 27), and β appears in the exponential in equation (37). Hence the $\xi_{\rm LOS}(\bar{z})$ curves will decrease less rapidly when β is small, i.e. γ is large.

Finally, we note that the effect of whether $T_{\rm m}$ evolves or not becomes appreciable for large values of γ. This is obvious because the larger the value of γ, the more rapid is the evolution of $T_{\rm m}$. Furthermore, we have already argued that if the evolution of $T_{\rm m}$ is more rapid, then the value of q should increase (provided, of course, that the value of β remains unchanged). Hence the values of q are smaller when $T_{\rm m}$ is kept constant than when $T_{\rm m}$ has a redshift dependence. This can be verified from Table 2.

These effects are shown in Fig. 4. The curves are normalized in such a way that $\xi_{\rm LOS}(v = 100\,\mathrm{km\,s^{-1}}) = 0.21$ at $\bar{z} = 3.85$, which is taken from Cristiani et al. (1997). In the top row of Fig. 4, the value of $T_{\rm m}$ is fixed at a particular redshift (in this case, at $\bar{z} = 2.5$) and the values of $T_{\rm m}$ at other redshifts are calculated using the relation $T_{\rm m} \sim (1+z)^{3\gamma-3}$. In the bottom row, $T_{\rm m}$ is kept constant. We do not plot the effect of $T_{\rm m}$ because we find it to be very weak.

The point to be noted here is that *the effects arising from changes in cosmology and γ are of the same order, which means that we cannot constrain both of the parameters simultaneously.* Thus the above analysis clearly suggests that it will be very difficult to recover the power spectrum of density fluctuations uniquely from the Lyα absorption lines without knowing the IGM parameters. However, the cosmological parameters determined through studies such as the cosmic microwave background radiation (and other data) can be used to constrain the equation of state (using the plot on the right in Fig. 4), provided that we have some idea about the evolution of the Jeans length or, equivalently, $T_{\rm m}$.

Throughout this paper we have treated γ as being independent of z. However, there are indications that γ could change with z. Schaye et al. (2000) notice that the temperature of the IGM has a peak at $z \simeq 3$ and it decreases with decreasing redshift afterwards. They also notice that the slope of the equation of state become close to 1 at $z \simeq 3$, then increases with decreasing redshift. Theoretical calculations suggest that γ increases with time and the rate of evolution depends on the reionization epoch (Hui & Gnedin 1997). From Table 2 and Fig. 4 we can infer that when γ becomes larger the rate of growth of $\xi_{\rm LOS}$ at a given velocity decreases. *Thus our study clearly suggests that the evolution of $\xi_{\rm LOS}$ at a given velocity can be used as probe of the reionization and thermal history of the IGM once the cosmological model and the evolution of $x_{\rm b}$ are fixed.* We hope to study this in detail in a future publication.

4.2 Transverse correlation

In this section we present the results for the transverse correlation. As before, we consider the same cold dark matter (CDM) power spectrum, and essentially the same range of the IGM parameters. The smoothing velocity is taken to be $30\,\mathrm{km\,s^{-1}}$, which is the typical peculiar velocity of a blob in the IGM.

Given z, we calculate $\xi_{\rm trans}$ as a function of the transverse comoving distance $l_\perp$. One can then convert this length-scale to an angular scale θ through the following relations:

$$\theta = \frac{l_\perp}{d_{\rm a}^{\rm com}(z)}, \tag{65}$$

$$d_{\rm a}^{\rm com}(z) = \frac{c}{H_0\sqrt{|\Omega_k|}}\, S_k\left[x(z)\frac{H_0}{c}\sqrt{|\Omega_k|}\right], \tag{66}$$

where Ω_k is given by equation (41), $x(z)$ is given by equation (45) and

$$S_k(r) = \begin{cases} \sin r & (\text{if } \Omega_k < 0), \\ r & (\text{if } \Omega_k = 0), \\ \sinh r & (\text{if } \Omega_k > 0). \end{cases} \tag{67}$$

Instead of plotting the correlation function directly, we plot the quantity $\mathcal{P}_\theta(\theta)$, which is defined as follows. The excess probability, over random background, of finding two neutral hydrogen overdense regions separated by a comoving transverse

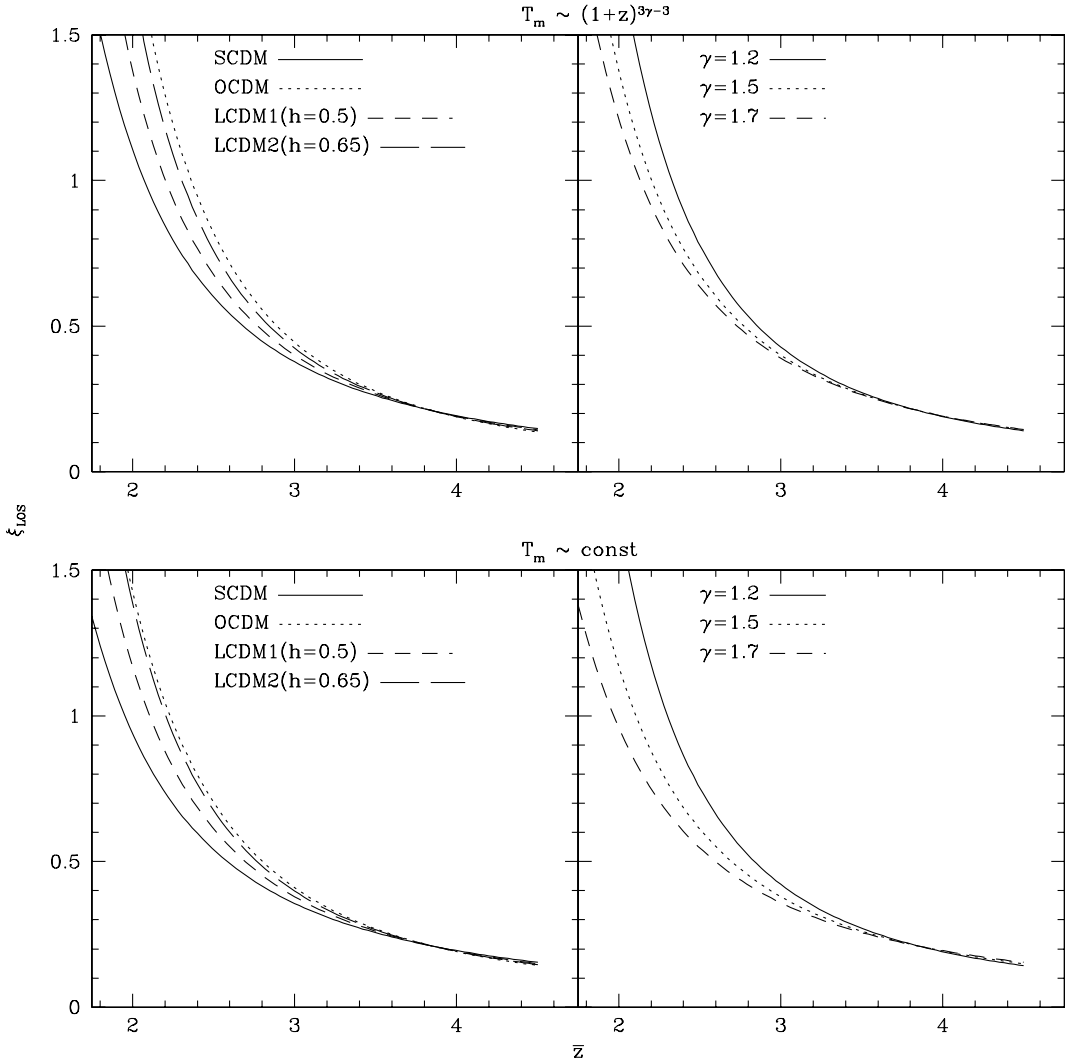

Figure 4. $\xi_{\rm LOS}(v = 100\,{\rm km\,s^{-1}})$ as a function of $\bar{z}$. The curves are normalized in such a way that $\xi_{\rm LOS}(v = 100\,{\rm km\,s^{-1}}) = 0.21$ at $\bar{z} = 3.85$, which is taken from Cristiani et al. (1997). In the left-hand plots, the IGM parameters are fixed at $T_{\rm m} = 40\,000\,{\rm K}$, $\gamma = 1.5$. In the right-hand plots, we have fixed the cosmological model to be LCDM1, and $T_{\rm m} = 40\,000\,{\rm K}$.

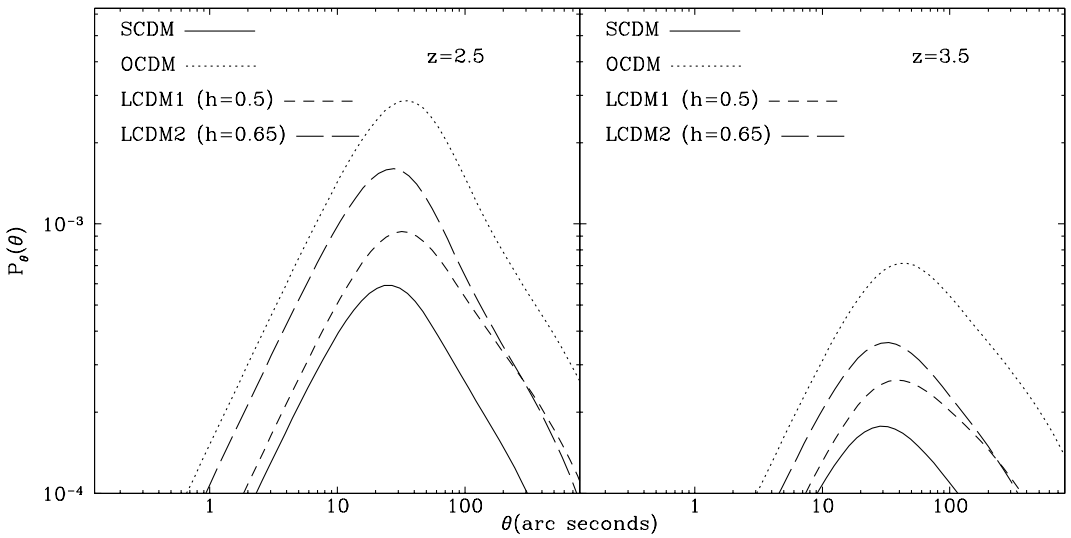

Figure 5. Plot of $\mathcal{P}_\theta(\theta)$. Results for four different cosmological models and for two different redshifts are presented. The IGM parameters are $\gamma = 1.5$, $T_{\rm m}(z = 2.5) = 40\,000\,{\rm K}$. $T_{\rm m}$ at redshift $z = 3.5$ is calculated using the relation $T_{\rm m} \sim (1 + z)^{3\gamma-3}$.

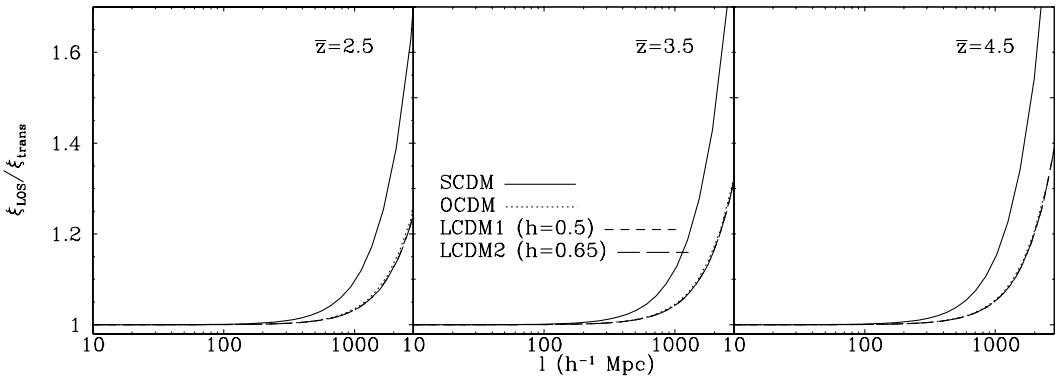

Figure 6. The ratio of $\xi_{\rm LOS}$ and $\xi_{\rm trans}$ as a function of comoving scale l. The results for four cosmological models and for three different mean redshifts are plotted. The curves for the LCDM1, LCDM2 and OCDM models nearly overlap. The model parameters relating to the IGM are $\gamma = 1.5$, $T_{\rm m}(\bar{z} = 2.5) = 40\,000$ K. $T_{\rm m}$ at the other two redshifts is calculated using the relation $T_{\rm m} \sim (1+z)^{3\gamma-3}$.

distance $l_\perp$ is

$$\mathcal{P}_{l_\perp}(l_\perp)\,{\rm d}l_\perp = \frac{\xi_{\rm trans}(l_\perp; z) 2\pi l_\perp\,{\rm d}l_\perp}{4\pi[d_{\rm a}^{\rm com}(z)]^2}. \tag{68}$$

Using equation (65) we get the excess probability over random background of finding two neutral hydrogen overdense regions separated by an angle θ as

$$\begin{aligned}\mathcal{P}_\theta(\theta)\,{\rm d}\theta &= \mathcal{P}_{l_\perp}(l_\perp)\frac{{\rm d}l_\perp}{{\rm d}\theta}{\rm d}\theta \\ &= \tfrac{1}{2}\,\xi_{\rm trans}(\theta)\theta\,{\rm d}\theta.\end{aligned} \tag{69}$$

From Fig. 5 it is clear that even the maximum excess probability of finding two H I overdense regions over an angular scale greater than few arcseconds is less than 1 per cent. Observationally, the distribution of H I along the transverse direction is probed by studying the common absorbers along the LOS towards closely spaced QSOs. The angular scales probed vary between a few arcseconds and a few arcminutes (Shaver, Boksenberg & Robertson 1982; Shaver & Robertson 1983; Smette et al. 1992, 1995; Dinshaw et al. 1994; Bechtold et al. 1994; Crotts et al. 1994; Bechtold & Yee 1994; D'Odorico et al. 1997; Petitjean et al. 1998). Based on our analysis, it is most likely that the common absorbers seen in the spectra of closely spaced QSOs are most likely to probe the transverse extent of the same overdense region rather than the clustering length-scale of separate regions.

4.3 Difference between $\xi_{\rm trans}$ and $\xi_{\rm LOS}$

It should be noted that, for a given mean redshift, the values of the LOS and the transverse correlation functions need not be the same. This is because, when we observe along one LOS, we actually sample different points at different redshifts. In contrast to this, the transverse correlation is calculated at the same redshift. The effect of evolution in the LOS correlation makes it different from the transverse correlation.

To illustrate this point more clearly, let us first assume that $x_{\rm b}$ does not evolve with z. Then, from equations (47) and (53), we see that, for a given length-scale l, the integrands in the two equations are identical. Hence we get

$$\frac{Q_{\rm LOS}}{Q_{\rm trans}} = \frac{D(\bar{z} + \Delta z/2)D(\bar{z} - \Delta z/2)}{D^2(\bar{z})}, \tag{70}$$

where $\bar{z}$ and Δz are defined in equations (42) and (49) respectively.

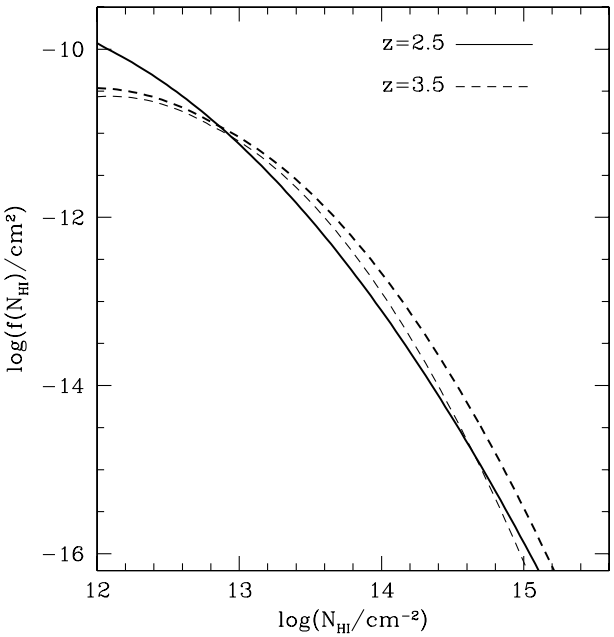

Figure 7. Column density distribution of neutral hydrogen plotted for different kinds of evolution of $T_{\rm m}$ and T_0 for two redshifts. The thicker lines represent the case where both of them are constant. For definiteness we have taken $T_{\rm m} = 40\,000$ K and $T_0 = 20\,000$ K. The thinner lines are for the case where both the temperatures vary as $(1+z)^{3\gamma-3}$. The values are $T_{\rm m}(z = 2.5) = 40\,000$ K and $T_0(z = 2.5) = 20\,000$ K. Obviously, the curves for the two cases overlap at $z = 2.5$. The cosmology is taken to be LCDM1, $\gamma = 1.7$ and $\epsilon = 0.3$.

The difference in the two correlation functions is now entirely due to the evolution of the power spectrum. Thus the two correlation functions will be nearly equal for small Δz but will start to differ from each other for large Δz. In the general case when $x_{\rm b}$ evolves with z, the difference will be much more prominent.This is indeed true, as one can see from Fig. 6. For scales below $200\,h^{-1}$ Mpc, the two correlation functions are nearly the same. Above such scales, however, the two functions start to differ appreciably. For the observations made at scales of 10–100 h^{-1} Mpc, our analytical calculation shows that one should not see any appreciable difference between the LOS and transverse correlations. This can be used as an important tool in determining the power spectrum (provided, of course, that we know the IGM parameters and the correlation function completely). As we have argued

earlier, one cannot get the power spectrum from the LOS correlation, but the power spectrum can be obtained from the transverse correlation in the usual manner. Since the two correlations are identical for scales up to $100\,h^{-1}$ Mpc, one can start from the LOS correlation, replace it with the transverse correlation, and obtain the power spectrum.

4.4 Column density distribution

In this section we study the results for the column density distribution. As we have discussed in Section 3, we shall consider two separate cases for the evolution of T_0 and $T_{\rm m}$, i.e. (i) (T_0 and $T_{\rm m}$) $\propto (1+z)^{3\gamma-3}$, and (ii) T_0 and $T_{\rm m}$ independent of z. The comparison between the two cases is shown in Fig. 7, where we plot the quantity $f(N_{\rm H\,I})$ for the two different cases. $f(N_{\rm H\,I})$ is related to $({\rm d}N_{\rm pk}/{\rm d}z{\rm d}N_{\rm H\,I})$ through the relation (Bi & Davidsen 1997)

$$f(N_{\rm H\,I}) = ({\rm d}N_{\rm pk}/{\rm d}z\,{\rm d}N_{\rm H\,I})/(1+z). \tag{71}$$

As we can see from Fig. 7, the difference between the two cases becomes more significant at higher column densities. We have checked and found that $T_{\rm m}$ has very little effect on the column density distribution. The difference between the two cases at $z=3.5$ is a result of the fact that the value of T_0 is different for the two cases at that redshift. However, we have already mentioned that any uncertainty in the knowledge of T_0 can be compensated to some extent by changing ϵ. Hence, for study of the column density distribution we shall consider only the case where both $T_{\rm m}$ and T_0 evolve as $(1+z)^{3\gamma-3}$.

We shall now discuss the dependence of the column density distribution on the following parameters: (i) cosmological models, (ii) ϵ, and (iii) γ.

In what follows we try to get constraints on our model parameters using the observed column density distribution obtained from Hu et al. (1995) and Kim et al. (1997) at three mean redshifts, 2.31, 2.85 and 3.35. In Fig. 8 we have plotted the quantity $f(N_{\rm H\,I})$ for these redshifts. The observational data points are the points with error bars in the figure.

For all the plots in Fig. 8, we have fixed $T_{\rm m}(z=2.5) = 40\,000$ K and $T_0(z=2.5) = 20\,000$ K. In the leftmost panel of each row in the figure we plot the predicted column density

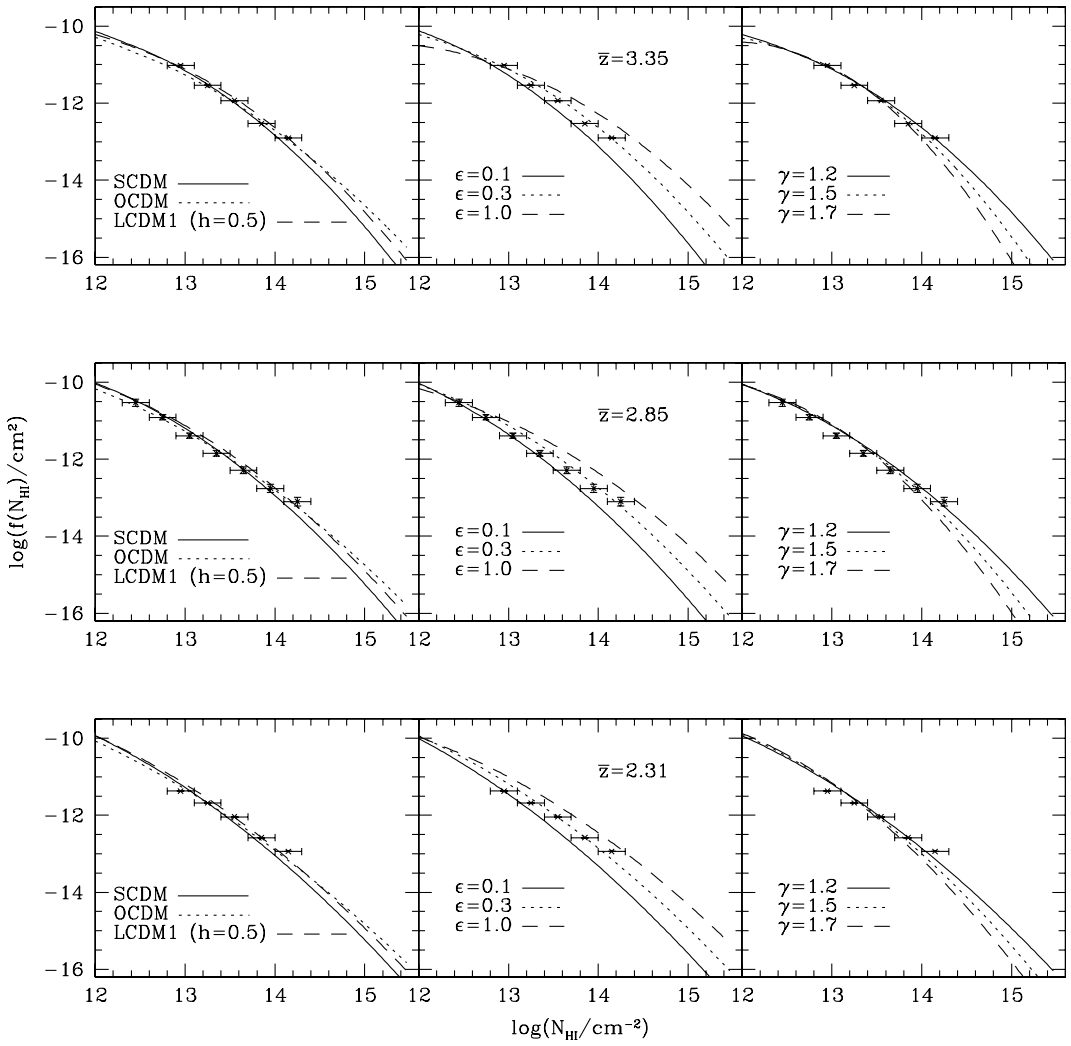

Figure 8. $f(N_{\rm H\,I})$ as a function of $N_{\rm H\,I}$ for three redshifts. We show the dependence of $f(N_{\rm H\,I})$ on cosmology (left), ϵ (centre) and equation of state γ (right). For all the plots we have taken $T_{\rm m}(z=2.5) = 40\,000$ K and $T_0(z=2.5) = 20\,000$ K. For the left-hand plots, we fixed $\gamma = 1.2$, $\epsilon = 0.3$. We have not plotted the curve for the LCDM2 model because it completely overlaps with that for the LCDM1 model. The centre plots are for the LCDM1 model with $\gamma = 1.2$. The right-hand plots are also for LCDM1 with $\epsilon = 0.3$.

distribution for various cosmological models for a given set of IGM parameters ($\gamma = 1.2$) and $\epsilon = 0.3$. We did not plot the LCDM2 model because it completely overlaps with the LCDM1 model. In all the redshift bins it is clear that the SCDM curves fall steeply at the higher column density end compared with other models. This is consistent with the results noted by Gnedin & Hui (1996). However, in our method of obtaining $f(N_{\rm H\,I})$, the SCDM model can be made to fit the data by slightly increasing the value of ϵ. The OCDM and LCDM1 curves fit the observed distribution up to $\log(N_{\rm H\,I}/{\rm cm}^{-2}) = 14.2$ for $\epsilon = 0.3$.

In the middle panel of each row in Fig. 8 we plot the predicted distribution for the three assumed values of ϵ for LCDM1 with $\gamma = 1.2$. Increasing the value of ϵ enhances the column density of a cloud (with fixed peak density), and consequently the value of $f(N_{\rm H\,I})$ increases in the high column density region. It is clear that the observed distribution is consistent with $\epsilon = 0.3$. This means that, in the case of LCDM, the effective length of the overdense region is about (1/3) of the coherence scale R^*. The value of R^* depends on the baryonic power spectrum, its typical value being of the order of a few hundred kpc; for the LCDM model with $\gamma = 1.2$ and $T_{\rm m}(z = 2.5) = 40\,000\,{\rm K}$, we get $R^* \simeq 60\,h^{-1}\,{\rm kpc}$. Hence the typical length of an individual overdense region is $\sim 20\,h^{-1}\,{\rm kpc}$. Zhang et al. (1998) infer that the typical LOS scalelength of an absorbing cloud is $\sim$ 15–35 h^{-1} kpc through their hydrodynamical simulations. Our values are consistent with theirs.

The effect of γ on the column density distribution can be seen from the rightmost panels in Fig. 8 where we plot the results of the LCDM1 model for three values of γ. In the models with higher values of γ, the $f(N_{\rm H\,I})$ curve falls sharply at the higher column density end at all redshifts, which is consistent with the conclusions of Hui et al. (1997). The reason for this sharp fall is that the neutral hydrogen density is lower (for a given baryonic density) when the value of γ is larger, and hence there are fewer clouds with high column densities. At low redshifts, it is clear that lower values of γ are preferred for $\epsilon = 0.3$. However, one can get a consistent fit for higher values of γ by increasing the value of ϵ. For example, at $\bar{z} = 2.31$, one can fit the data reasonably well with ϵ in the range 0.3 to 0.5 and γ in the range 1.2 to 1.6, but it is impossible to fit the data with $\gamma > 1.6$, for any choice of ϵ at $\bar{z} = 2.31$. This implies that we cannot accommodate γ larger than 1.6 at this redshift even by changing the values of T_0 or other IGM parameters. At $\bar{z} = 2.85$, one can fit the observations for $\gamma = 1.7$ (the highest value of γ that we are considering in this paper), by taking $\epsilon = 0.4$. For the case where $\bar{z} = 3.35$, we can see that we can marginally fit the observations for the whole range of γ considered in this paper with $\epsilon = 0.3$.

Finally, we comment on how our results compare with some of the hydrodynamical simulations. It has been observed through numerical simulations that the peak baryonic over-density and the neutral hydrogen column density are strongly correlated (Zhang et al. 1998; Davé et al. 1999). We tested equation (57) against the scatter plots showing the correlations. The parameters were taken to be the LCDM1 cosmological model with $\gamma = 1.2$ and $\epsilon = 0.3$ at $\bar{z} = 3$. These are the parameters that fit the observations for the column density distribution reasonably well (see Fig. 8). It turns out that the relation matches well with the median of the scatter plot in fig. 8 of Zhang et al. (1998). Davé et al. (1999) give an analytical formula which relates $n_{\rm B}({\rm peak})/n_0$ and $N_{\rm H\,I}$ by a power law (see equation 7). We find that one needs a much higher value of $\gamma (\sim 1.82)$ in our model to match the power-law index at redshift 3. Also, one needs $\epsilon \sim 1$ to match the overall scaling factor. If we take such values of ϵ and γ, then our results for the column density distribution will not be able to match the observations. We believe that such a discrepancy arises because of the difference in the methods used for assigning a column density to a cloud.

We next test the column density distributions against some of the recent simulation results. In the column density range $10^{12.8}$–$10^{14.3}\,{\rm cm}^{-2}$, one can fit a power law of the form $f(N_{\rm H\,I}) \propto N_{\rm H\,I}^{-\beta_{\rm H\,I}}$. The values of $\beta_{\rm H\,I}$ for our LCDM1 model with $\gamma = 1.2$ and $\epsilon = 0.3$ are $\beta_{\rm H\,I} = 1.70 \pm 0.02$ for $\bar{z} = 2.31$, $\beta_{\rm H\,I} = 1.65 \pm 0.04$ for $\bar{z} = 2.85$ and $\beta_{\rm H\,I} = 1.56 \pm 0.04$ for $\bar{z} = 3.35$. When compared with fits to the observational points, the corresponding power-law indices are 1.35 ± 0.03, 1.39 ± 0.26 and 1.59 ± 0.13, respectively (Kim et al. 1997). All errors given here are 2σ. The simulations of Zhang et al. (1997) using the SCDM model produce $\beta_{\rm H\,I} = 1.39 \pm 0.06$ in the range $2 \times 10^{12} < N_{\rm H\,I}/{\rm cm}^{-2} < 10^{14}$ for $z = 3$. Our curves are also in quite good agreement with the P^{3}M–SPH simulations (using the SCDM model) by Theuns et al. (1998a,b) in the same column density and redshift ranges. More recently, Machacek et al. (2000) have performed hydrodynamical simulations for various cosmological models. The LCDM model ($\Omega_{\rm m} = 0.4$, $\Omega_\Lambda = 0.6$, $h = 0.65$, $\sigma_8 = 1.0$) with $\Omega_{\rm B}h^2 = 0.015$ gives $\beta_{\rm H\,I} = 1.61 \pm 0.04$ for $z = 2$ and $\beta_{\rm H\,I} = 1.48 \pm 0.04$ for $z = 3$ in the range $10^{12.8} < N_{\rm H\,I}/{\rm cm}^{-2} < 10^{14.3}$. They note that the power-law index is higher for the SCDM model, and lower for the OCDM model. This is also consistent with what we get from our model.

5 CONCLUSIONS

We have presented a simple analytic expression for the correlation function and the column density distribution for the low H I column density systems seen in the spectra of high-redshift QSOs. We have used our results to get constraints on various cosmological and IGM parameters. We summarize our main results below.

(1) One cannot rigorously define a power spectrum from the LOS correlation function. However, since the LOS and transverse correlations are identical below scales of $\sim 100\,h^{-1}\,{\rm Mpc}$, it is possible to obtain the power spectrum by a Fourier transform of the LOS correlation function (provided that the IGM parameters are known).

Previous studies have attempted to recover the power spectrum of density fluctuations from observations of the IGM (Croft et al. 1998, 1999; Hui 1999). We show that it is difficult to recover a unique power spectrum from the H I correlation function without constraining the IGM parameters, especially γ. We have shown that the shape of the LOS correlation function at a particular mean redshift becomes less and less sensitive to the DM power spectrum as the universe evolves after reionization. We feel that the correct approach in studying this issue is to constrain the cosmological models using cosmic microwave background or supernova data, and apply these constraints to study the IGM parameters using H I correlation functions.

(2) The LOS correlation function and its evolution are much more sensitive to γ than to $T_{\rm m}$. Using observations that give the shape of the LOS correlation at a particular redshift, one can constrain the value of γ, even with ill-constrained values of $T_{\rm m}$, provided that the background cosmology is known. However, the more uncertain is the value of $T_{\rm m}$, the less constrained is γ. Carrying out such an exercise for different redshifts, it will be possible to constrain the evolution of γ. However, for this study,

one needs accurate observational data at different redshift bins that are not affected by evolutionary effects. *Thus one can use the study of the correlation function as an independent method to constrain the reionization history of the Universe.*

(3) The analytic column density distribution for H I, like $\xi_{\rm LOS}$, is less sensitive to $T_{\rm m}$ than to γ. The distribution, when compared with observations, favours a lower value of γ, although at redshifts >2.5 one can marginally fit the observations with higher values of γ. Our model clearly rules out $\gamma > 1.6$ at redshift 2.31.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the Indo-French Centre for Promotion of Advanced Research under contract No. 1710-1. TRC is supported by the University Grants Commission, India. We thank the anonymous referee for suggestions which helped us to improve the clarity of the paper. We also thank Patrick Petitjean for useful comments.

REFERENCES

Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Bechtold J., Yee H. K. C., 1994, AJ, 110, 1984
Bechtold J., Crotts A. P. S., Duncan R. C., Fang Y., 1994, ApJ, 437, L83
Bergeron J., Boisse P., 1991, A&A, 243, 344
Bi H., 1993, ApJ, 405, 479
Bi H., Davidsen A. F., 1997, ApJ, 479, 523
Bi H. G., Börner G., Chu Y., 1992, A&A, 266, 1
Bi H., Ge J., Fang L., 1995, ApJ, 452, 90
Black J. H., 1981, MNRAS, 197, 553
Bond J. R., Szalay A. S., Silk J., 1988, ApJ, 324, 627
Cen R. Y., Miralda-Escudé J., Ostriker J. P., Rauch M. R., 1994, ApJ, 437, L9
Coles P., Jones B., 1991, MNRAS, 248, 1
Cristiani S., D'Odorico S., Fontana A., Giallongo E., Savaglio S., 1995, MNRAS, 273, 1016
Cristiani S., D'Odorico S., D'Odorico V., Fontana A., Giallongo E., Moscardini L., Savaglio S., 1997, in Petitjean P., Charlot S., eds, Proc. 13th IAP Astrophys. Colloq., Structure and Evolution of the Intergalactic Medium from QSO Absorption Line Systems. Editions Frontières, Paris, p. 165
Croft R. A. C., Weinberg D. H., Katz N., Hernquist L., 1998, ApJ, 495, 44
Croft R. A. C., Weinberg D. H., Pettini M., Hernquist L., Katz N., 1999, ApJ, 520, 1
Crotts A. P. S., Bechtold J., Fang Y., Duncan R. C., 1994, ApJ, 437, 79
Davé R., Hernquist L., Katz N., Weinberg D. H., 1999, ApJ, 511, 521
Dinshaw N., Impey C. D., Foltz C. B., Weymann R. J., Chaffee F. H., Jr, 1994, ApJ, 437, L87
D'Odorico S., Cristiani S., D'Odorico V., Fontana A., Giallongo E., Shaver P., 1997, in Petitjean P., Charlot S., eds, Proc. 13th IAP Astrophys. Colloq., Structure and Evolution of the Intergalactic Medium from QSO Absorption Line Systems. Editions Frontières, Paris, p. 392
Doroshkevich A. G., Shandarin S. F., 1977, MNRAS, 179, 95
Efstathiou G., Bond J. R., White S. D. M., 1992, MNRAS, 258, 1–P
Eke V. R., Cole S., Frenk C. S., 1996, MNRAS, 282, 263
Fang L. Z., Bi H., Xiang S., Börner G., 1993, ApJ, 413, 477
Gnedin Y. G., Hui L., 1996, ApJ, 472, L73
Hernquist L., Katz N., Weinberg D. H., Miralda-Escudé J., 1996, ApJ, 457, L51
Hu E. M., Kim T., Cowie L. L., Songaila A., 1995, AJ, 110, 1526
Hui L., 1999, ApJ, 516, 519
Hui L., Gnedin Y. G., 1997, MNRAS, 292, 27
Hui L., Gnedin Y. G., Zhang Y., 1997, ApJ, 486, 599
Khare P., Srianand R., York D. G., Green R., Welty D., Huang K., Bechtold J., 1997, MNRAS, 285, 167
Kim T., Hu E. M., Cowie L. L., Songaila A., 1997, AJ, 114, 1
Machacek E. M., Bryan G. L., Meiksin A., Anninos P., Thayer D., Norman M., 2000, ApJ, 532, 118
McDonald P., Miralda-Escudé J., Rauch M., Sargent W. L. W., Barlow T. A., Cen R., Ostriker J. P., 2000, ApJ, 543, 1
McGill C., 1990, MNRAS, 242, 544
Miralda-Escudé J., Cen R., Ostriker J. P., Rauch M., 1996, ApJ, 471, 582
Petitjean P., Surdej J., Smette A., Shaver P., Mücket J., Remy M., 1998, A&A, 334, L45
Rauch M. et al., 1997, ApJ, 489, 7
Riediger R., Petitjean P., Mücket J. P., 1998, A&A, 329, 30
Sarazin C. L., Bahcall J. N., 1977, ApJS, 34, 451
Schaye J., Theuns T., Leonard A., Efstathiou G., 1999, MNRAS, 310, 57
Schaye J., Theuns T., Rauch M., Efstathiou G., Sargent W. L. W., 2000, MNRAS, 378, 817
Shaver P. A., Robertson J. G., 1983, ApJ, 268, L57
Shaver P. A., Boksenberg A., Robertson J. G., 1982, ApJ, 261, L7
Smette A., Surdej J., Shaver P. A., Foltz C. B., Chaffee F. H., Weymann R. J., Williams R. E., Magain P., 1992, A&A, 389, 39
Smette A., Robertson J. G., Shaver P. A., Reimers D., Wisotzki L., Koehler T., 1995, A&AS, 113, 199
Srianand R., 1997, ApJ, 478, 511
Steidel C. C., 1993, in Shull J. M., Thronson H. A., eds, Proc. 3rd Teton Astronomy Conference, The Environment and Evolution of Galaxies. Kluwer, Dordrecht, p. 263
Theuns T., Leonard A., Efstathiou G., 1998a, MNRAS, 297, L49
Theuns T., Leonard A., Efstathiou G., Pearce F. R., Thomas P. A., 1998b, MNRAS, 301, 478
Zhang Y., Anninos P., Norman M. L., 1995, ApJ, 453, L57
Zhang Y., Anninos P., Norman M. L., Meiksin A., 1997, ApJ, 485, 496
Zhang Y., Meiksin A., Anninos P., Norman M. L., 1998, ApJ, 495, 63

APPENDIX A: DETAILED CALCULATIONS FOR THE COLUMN DENSITY DISTRIBUTION

Suppose that we are looking at the IGM along any one direction, at some redshift z. Then the linear density field $\delta_{\rm B}^{(1{\rm D})}(x,z)$ along that axis will be described by a one-dimensional Gaussian random field, with a power spectrum

$$P_{\rm B}^{(1{\rm D})}(k,z) = \frac{1}{2\pi} D^2(z) \int_k^\infty {\rm d}k' k' \frac{P_{\rm DM}(k)}{[1+x_{\rm b}^2(z)k^2]^2}. \tag{A1}$$

From now on we shall derive all the expressions at a particular redshift z, and we shall not write the explicit z-dependence on the quantities.

To define the column density, we associate each local maximum or peak in the linear density field with a Lyα cloud. The column density corresponding to such a cloud is given by equation (56). We have expressed $\delta_{\rm B}^{(1{\rm D})}$(peak) in terms of $N_{\rm H\,I}$ in equation (58).

Using the properties of a Gaussian random field, we can derive the joint probability distribution for the three Gaussian random fields $\delta_{\rm B}^{(1{\rm D})}$, $\delta_{\rm B}^{(1{\rm D})''}$, $\delta_{\rm B}^{(1{\rm D})'}$. The probability that the field and its second derivative have values $\delta_{\rm B}^{(1{\rm D})}$ and $\delta_{\rm B}^{(1{\rm D})''}$, respectively, at the peak $\delta_{\rm B}^{(1{\rm D})'} = 0$ is

$$\mathcal{P}[\delta_{\rm B}^{(1{\rm D})}, \delta_{\rm B}^{(1{\rm D})''}, \delta_{\rm B}^{(1{\rm D})'} = 0]\, {\rm d}\delta_{\rm B}^{(1{\rm D})}\, {\rm d}\delta_{\rm B}^{(1{\rm D})''} |\delta_{\rm B}^{(1{\rm D})''}|\, {\rm d}x = \frac{1}{(2\pi)^{3/2}\sigma_1\Sigma}$$
$$\times \exp\left[\frac{1}{2\Sigma^2}\left(\sigma_2^2 \delta_{\rm B}^{(1{\rm D})^2} + 2\sigma_1^2 \delta_{\rm B}^{(1{\rm D})} \delta_{\rm B}^{(1{\rm D})''} + \sigma_0^2 \delta_{\rm B}^{(1{\rm D})''^2}\right)\right]$$
$$\times {\rm d}\delta_{\rm B}^{(1{\rm D})}\, {\rm d}\delta_{\rm B}^{(1{\rm D})''} |\delta_{\rm B}^{(1{\rm D})''}|\, {\rm d}x, \tag{A2}$$

where

$$\sigma_{\rm m}^2 \equiv \sigma_{\rm m}^2(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} {\rm d}k k^{2m} P_{\rm B}^{(1{\rm D})}(k,z), \tag{A3}$$

and

$$\Sigma^2 = \sigma_0^2\sigma_2^2 - \sigma_1^4. \tag{A4}$$

Note that $\sigma_0 = \Delta$, defined in equation (8).

For convenience, let us define some dimensionless quantities

$$\nu \equiv \frac{\delta_{\rm B}^{(1{\rm D})}}{\sigma_0}, \quad \lambda \equiv -\frac{\delta_{\rm B}^{(1{\rm D})''}}{\sigma_2}, \quad \kappa \equiv \frac{\sigma_1^2}{\sigma_0\sigma_2}. \tag{A5}$$

ν and λ measure the field and its second derivative, respectively; κ is a measure of the width of the power spectrum. One can use these quantities to obtain the number of peaks (clouds) per unit length:

$$\frac{{\rm d}N_{\rm pk}}{{\rm d}x} = \mathcal{P}[\delta_{\rm B}^{(1{\rm D})}, \delta_{\rm B}^{(1{\rm D})''}, \delta_{\rm B}^{(1{\rm D})'} = 0]\,{\rm d}\delta_{\rm B}^{(1{\rm D})}\,{\rm d}\delta_{\rm B}^{(1{\rm D})''}|\delta_{\rm B}^{(1{\rm D})''}|. \tag{A6}$$

Using equation (39), one can convert the above expression to the number of clouds per unit redshift interval. After simplification, the relation becomes

$$\frac{{\rm d}N_{\rm pk}}{{\rm d}z} = \frac{d_H(z)}{(2\pi)^{3/2}\sqrt{1-\kappa^2}R*} \times \exp\left\{-\frac{1}{2}\left[\frac{(\nu-\kappa\lambda)^2}{1-\kappa^2} + \lambda^2\right]\right\}\lambda\,{\rm d}\lambda\,{\rm d}\nu. \tag{A7}$$

The λ-integration can be carried out to obtain

$$\frac{{\rm d}N_{\rm pk}}{{\rm d}z\,{\rm d}\nu} = \frac{d_H(z)}{(2\pi)^{3/2}R*}\left\{\sqrt{1-\kappa^2}\exp\left[-\frac{\nu^2}{2(1-\kappa^2)}\right] + \kappa\nu\sqrt{2\pi}\,{\rm e}^{-\nu^2/2} - \sqrt{\frac{\pi}{2}}\kappa\nu\,{\rm erfc}\left[\frac{\kappa\nu}{\sqrt{2(1-\kappa^2)}}\right]{\rm e}^{-\nu^2/2}\right\}. \tag{A8}$$

We are interested in the quantity

$$\frac{{\rm d}N_{\rm pk}}{{\rm d}z\,{\rm d}N_{\rm H\,I}} = \frac{{\rm d}N_{\rm pk}}{{\rm d}z\,{\rm d}\nu}\frac{{\rm d}\nu}{{\rm d}N_{\rm H\,I}}, \tag{A9}$$

which is straightforward to obtain from equation (A8), provided that we know ν as a function of $N_{\rm H\,I}$. Equations (58) and (A5) give ν in terms of $N_{\rm H\,I}$, and they can be used to calculate ${\rm d}\nu/{\rm d}N_{\rm H\,I}$ (for $n_{\rm H\,I} \ll n_{\rm B}$):

$$\frac{{\rm d}\nu}{{\rm d}N_{\rm H\,I}} = \frac{1}{\beta\Delta N_{\rm H\,I}}. \tag{A10}$$

Thus we get an analytic expression for the number of clouds per unit redshift interval per unit column density range $({\rm d}N_{\rm pk}/{\rm d}z\,{\rm d}N_{\rm H\,I})$ as a function of $N_{\rm H\,I}$.

This paper has been typeset from a TEX/LATEX file prepared by the author.