HYDROGEN CLOUDS BEFORE REIONIZATION: A LOGNORMAL MODEL APPROACH

Hongguang Bi, Li-Zhi Fang, Longlong Feng, and Yipeng Jing

Received 2002 November 14; accepted 2003 July 31

ABSTRACT

We study the baryonic gas clouds (the intergalactic medium) in the universe before reionization with the lognormal (LN) model, which has been shown to be dynamically legitimate in describing the fluctuation evolution in quasilinear as well as nonlinear regimes in recent years. The probability distribution function of the mass field in the LN model is long-tailed and so plays an important role in rare events, such as the formation of the first generation of baryonic objects. Since in this model the nonlinear field is directly mapped from the corresponding linear field, we can calculate the density and velocity distributions of the intergalactic medium at very high spatial resolutions. We simulate the distributions at a resolution of 0.15 kpc from z = 7 to 15 in the low-density cold dark matter cosmological model. We analyze statistics on the hydrogen clouds at high redshifts, including column densities, clumping factors, sizes, masses, and spatial number densities. One of our goals is to identify which hydrogen clouds are going to collapse. By inspecting the mass density profiles and the velocity profiles of clouds, we find that the velocity outflow significantly postpones the collapsing process in less massive clouds, even when their masses are larger than the Jeans mass. This indicates that the formation of collapsed clouds with small mass at high redshift is substantially suppressed. Consequently, only massive (>10^5 M_☉) clouds can form objects at higher redshifts, and less massive (10^4–10^5 M_☉) collapsed objects are formed later. Although the mass fraction in clouds with sizes larger than the Jeans length is already larger than 1% at z = 15, there is only a tiny fraction of mass (10^-5) in collapsed clouds. If all the ionizing photons and the ~10^-2 metallicity observed at low redshift are produced by the first 1% of the mass of collapsed baryonic clouds, the majority of that first generation of objects would be occurring not much earlier than z = 10.

Subject headings: cosmology: theory — intergalactic medium — large-scale structure of universe

1. INTRODUCTION

The lognormal (LN) model of the clustering of the cosmic mass field was used as a phenomenological description of the density and velocity distributions of the intergalactic medium (IGM; baryonic gas) in the redshift range 2 ≤ z ≤ 5 in the study of the Lyα forest of QSOs’ absorption spectra (Bi 1993; Bi, Ge, & Fang 1995; Bi & Davidsen 1997). Since then, the LN model has gained substantial support from dynamical analysis and numerical simulation of the nonlinear evolution of the cosmic mass field. First, it has been found that in the nonlinear regime the dynamics of the growth (irrotational) mode of density perturbations can be sketched by the random-force–driven Burgers equation (Berera & Fang 1994; Buchert, Dominguez, & Peres-Mercader 1999). An IGM model based on the random-force–driven Burgers equation proposed by Jones (1999) yields intermittency of the IGM mass density field, and its probability distribution function (PDF) is found to be lognormal. Numerical simulations have directly shown that a lognormal PDF is a good approximation of the nonlinear density and velocity fields (e.g., Coles & Jones 1991; Hui, Kofman, & Shandarin 2000; Yang et al. 2001). The LN model can be used to describe the cosmic gravitational clustering not only in quasilinear regimes but also in highly nonlinear regimes. Recently, a detailed study of dynamical systems consisting of dark matter and the IGM also supports the LN model (Matteurese & Mohayee 2002).

In this paper, we use the LN model to study semianalytically the formation and evolution of the IGM clouds before reionization. A basic property of the LN model is that the PDF of the mass field is long-tailed and therefore plays an important role in rare events, such as the formation of the first-generation baryonic objects. A number of semianalytic models have been developed to describe structure formation in the early universe, such as those based on extrapolations of linear theory (e.g., Couchman & Rees 1986; Madau, Meikin, & Rees 1997; Tozzi et al. 2000), analytic models (e.g., Tegmark et al. 1997; Miralda-Escudé 1998; Valageas & Silk 1999; Miralda-Escudé 1998; Valageas & Silk 1999; Volonteri et al. 2000), the Press-Schechter (1974) formalism (e.g., Haiman & Loeb 1998), and numerical simulations (e.g., Abel et al. 1998; Bromm, Coppi, & Larson 1999; Gnedin 2000; Cen & Haiman 2000). See also reviews in Barkana & Loeb (2001) and Madau (2002). As a semianalytical approach, the nonlinear field in the LN model is mapped from the linear density field. Therefore, one can take advantage of this to simulate the spatial distributions of density and velocity, so as to calculate the density and velocity profiles of clumpy regions in the distributions, called “hydrogen clouds” hereafter.

The Jeans length of the IGM is below 1 kpc before reionization. To study hydrogen clouds on mass scales as small as the Jeans mass, ~10^4 M_☉, it is necessary to calculate the IGM distribution on scales as least as fine as 0.2 kpc. Using
the LN model, we are able to simulate the IGM distribution on such small scales. We focus on the abundances of clouds that are going to collapse. Since an IGM cloud that is going to collapse is a necessary condition of hosting star formation, the abundance of such a collapsing cloud can set effective constraints on the formation of the first-generation baryonic objects.

The paper is organized as follows: In § 2 we present an updated introduction of the LN model of the clustering of the cosmic field. The basic features of the IGM clustering around the epoch of reionization is discussed in § 3. In § 4 we analyze the statistical properties of the baryonic clumps (hydrogen clouds) based on the simulation of the LN model. In § 5 we identify the clouds that match the conditions of collapse. Finally, in § 6 we summarize the results and our conclusions.

2. LOGNORMAL MODEL UPDATED

The basic assumption of the LN model is that the PDF of the density field \( \rho(x) \) is lognormal, i.e.,

\[
\rho(x) = \rho_0 e^{X(x)},
\]

where \( X(x) \) is a Gaussian field with mean \( \bar{X} \) and variance

\[
\left[(X - \bar{X})^2\right]^{1/2} = \sigma_0.
\]

From equation (1), the mean of \( \rho(x) \) is

\[
]\bar{\rho} = \rho_0 e^{\bar{X} + \frac{1}{2}(X - \bar{X})^2}.\]

The overall average density \( \bar{\rho} \) should always satisfy \( \bar{\rho} = \bar{\rho}_0 \), which is required by mass conservation. Therefore, equation (2) requires

\[
\bar{X} = -(1/2) \left[(X - \bar{X})^2\right] = -\sigma_0^2/2.
\]

The density fluctuation or density contrast of the field is then \( \delta = (\rho - 1)/\bar{\rho} \).

The PDF of the variable \( \rho \) defined by equations (1) and (2) is

\[
\rho(\rho/\bar{\rho}) = \frac{1}{(\rho/\bar{\rho})\sigma_0 \sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left[ \ln(\rho/\bar{\rho}) + \sigma_0^2/2 \right]^2 \right\}, \quad \rho \geq 0.
\]

Obviously, it is normalized as \( \int_0^\infty \rho(x) \, dx = 1 \). The variance of \( \rho \) is given by (Vanmarcke 1983)

\[
\sigma = (e^{\sigma_0^2} - 1)^{1/2}.
\]

This means the second moment is \( \langle \rho^2 \rangle = \exp(\sigma_0^2) \). The higher order moments are

\[
\langle \rho^n \rangle = \exp\left[ n^2 - n \sigma_0^2/2 \right].
\]

Equation (5) shows

\[
\left\langle \frac{\rho^n}{\bar{\rho}} \right\rangle^{1/n} > \left\langle \frac{\rho^2}{\bar{\rho}} \right\rangle^{1/2};
\]

i.e., the moment is divergent when \( n \to \infty \). This indicates that the PDF equation (3) is long-tailed. The probability of long-tailed events \( \rho \gg 1 \) given by equation (3) is much larger than that of a Gaussian PDF.

Now we should identify the Gaussian field \( X(x) \). When the density perturbation is small, \( \delta(x) \ll 1 \), the mass field should be in the linear regime \( \delta_0(x) \); i.e., we need \( \delta(x) \simeq \delta_0(x) \). This yields \( X = \delta_0(x) - \sigma_0^2/2 \), or

\[
\rho(x) = \rho_0 \exp\left[ \delta_0(x) - \sigma_0^2/2 \right].
\]

where \( \sigma_0 = \langle \delta_0^2 \rangle^{1/2} \) is the variance of the linear Gaussian field of \( \delta_0(x) \) on the scale \( R \) considered. When \( \delta_0 \) is small, equation (4) gives \( \sigma = \sigma_0 \). Equation (6) is a basic assumption of the LN model. With equation (6), the nonlinear mass field \( \rho(x) \) is given by an exponential mapping from the corresponding linear density field \( \rho_0(x) \) or \( \delta_0(x) \).

The first successful application of the LN model mapping equation (6) was modeling the IGM distribution observed in QSO Ly \( \alpha \) forests (Bi 1993; Bi & Davidsen 1997). However, the LN model is useful not only for the weakly nonlinear regime but also for highly nonlinear evolution. The dynamical study of the lognormal PDF can be traced back to the adhesion approach, which sketches the nonlinear evolution of the growth mode of cosmic gravitational clustering with the Burgers equation (Gurbatov, Saichev, & Shandarin 1989). This equation yields a reasonable mass function of clumps (Vergassola et al. 1994). Considering the stochastic nature of field variables, cosmic clustering should actually be described by the random-force-driven Burgers equation (Berera & Fang 1994; Buchert et al. 1999). On the other hand, the lognormal PDF is found to be a good approximation of the density field described by this equation. Therefore, the lognormal model is dynamically legitimate for describing cosmic clustering in quasilinear as well as nonlinear regimes. This is also supported by numerical simulation. The one-point distribution of the cosmic mass and velocity fields in nonlinear regimes, such as the number density of galaxies, pairwise velocity, and angular momentum, from observed data and simulation samples is in good agreement with the lognormal distribution (e.g., Kofman et al. 1994; Hui et al. 2000; Yang et al. 2001; Pando et al. 2002).

3. IGM DISTRIBUTION WITH LOGNORMAL MODEL

3.1. Jeans Length of Baryonic Matter

In a homogeneous universe with the mean mass density \( \bar{\rho} \), the Jeans length of gaseous baryonic matter or the IGM is defined by \( \lambda_J \equiv v_s(\pi/G\bar{\rho})^{1/2} \), where \( v_s \) is the sound speed of the gas. For the current study, it is convenient to use a comoving scale \( x_b \) given by

\[
x_b \equiv \frac{1}{H_0} \left[ \frac{2\gamma k_B T_m}{3 \bar{\rho} \Omega(1 + z)} \right]^{1/2},
\]

which is \( 2\pi \) times smaller than \( \lambda_J \). In equation (7), \( T_m \) and \( \mu \) are the mean temperature and molecular weight of the gas, \( \Omega \) is the cosmological density parameter of total mass, and \( \gamma \) is the ratio of specific heats. The corresponding Jeans mass is \( m_J = (4\pi/3)\lambda_J^3 \bar{\rho} \).

Primordial baryons, created at the time of nucleosynthesis, recombines with electrons to become neutral gas at \( z \sim 1000 \), which corresponds to a Jeans mass of about \( 10^6 M_\odot \). Before \( z = 137 \), the residual ionization of the cosmic gas keeps its temperature locked to the cosmic microwave background temperature. After that, the gas cools down adiabatically during the expanding of the universe.
Assuming $\gamma = 5/3$, i.e., hydrogen temperature $T \propto \rho_b^{2/3}$, where $\rho_b$ is the mean mass density of baryonic matter, the evolution of the comoving Jeans length will depend approximately on $(1+z)^{1/2}$. At $z = 10$, the hydrogen temperature is $\sim 1.8$ K (Medvigy & Loeb 2001), the Jeans length $x_b \sim 1$ kpc, and the Jeans mass drops to about $10^4 M_\odot$. Note that the Jeans mass is 6 mag smaller than that of the galaxies we observed. Figure 1 plots the comoving Jeans length $x_b$ as a function of the cosmic scale factor $1/(1+z)$.

During reionization, the baryonic gas is heated by the UV ionizing background with temperatures from 1.8 to $\sim 1.3 \times 10^4$ K, and there is a 2 mag increase in $x_b$ and a 6 mag increase in the Jeans mass (e.g., Ostriker & Gnedin 1996). In Figure 1 the reionization is assumed to happen at $z = 7$ instantly; i.e., the IGM is assumed to be almost completely neutral and ionized before and after the reionization, respectively. It should be emphasized that the assumption of reionization redshift $z = 7$ does not affect the result of clustering before reionization, as discussed below. That is, the calculation of clustering at $z = 12$ is independent of whether the reionization redshift is at $z = 7$ or 10, for instance.

After reionization, the IGM temperature is maintained at about $\sim 10^4$ K by the UV background photons, and therefore $x_b$ will gradually increase with the decrease of $z$ because of the factor $1+z$ in equation (1). Note that the He II ionization occurs at redshift 3.3, which leads to an increase of the IGM temperature from $1.3 \times 10^4$ K before $z = 3.3$ to $2.5 \times 10^4$ K after $z = 3.3$ (Theuns et al. 2002). This is shown in Figure 1 by the small jump in $x_b$ at redshift $\sim 3.3$.

3.2. IGM Mass Density Field in Lognormal Model

The linear evolution of the IGM mass field $\rho_0(x)$ driven by the gravity of the dark matter mass field $\rho_{dm}(x)$ has been studied by using various assumptions of the thermal property of the IGM (e.g., Nusser 2000; Matarrese & Mohaye 2002). A common conclusion is

$$\delta_0(k, t) = (1 + \text{decaying terms})\delta_{dm}(k, t) + \text{decaying terms}$$

if $k \ll k_3$, (8)

where $\delta_0(k, t)$ and $\delta_{dm}(k, t)$ are, respectively, the Fourier transform of density fluctuations of the IGM, $\delta_0(x) = [\rho_0(x)/\rho_0] - 1$, and the dark matter, $\delta_{dm}(x) = [\rho_{dm}(x)/\rho_{dm}] - 1$. The decaying terms in equation (10) depend on the initial conditions. That is, regardless of specific assumptions of thermal processes and initial condition, the linear fluctuations of the IGM mass field on scales larger than the Jeans length $x_b$ always follows the dark matter mass field (Fang et al. 1993):

$$\delta_0(x) = \frac{1}{2\pi x_b^2} \int \delta_{dm}(x_1) \frac{1}{|x - x_1|} e^{-|x-x_1|/x_b^2} dx_1.$$ (9)

Obviously, $\langle \delta_0 \rangle = \langle \delta_{dm} \rangle = 0$.

With the LN model of §2, the nonlinear mass field of the IGM $\rho_0(x)$ has to be given by an exponential mapping from the corresponding linear density field of dark matter $\rho_0(x)$ as

$$\rho_0(x) = \tilde{\rho}_0 \exp \left[ \delta_0(x) - \frac{\sigma_0^2}{2} \right],$$ (10)

where $\sigma_0$ is the variance of the linear Gaussian field of $\delta_0(x)$ on the scale of the Jeans length. To simplify equation (12), we use the normalized mean baryonic matter density $\bar{\rho}_0 = 1$ below.

The number of $\sigma_0$ is shown in Figure 2, in which we use the low-density flat cold dark matter model (LCDM) with the density parameter $\Omega_0 = 0.3$, the cosmological constant

FIG. 1.—Jeans length of baryonic gas (IGM) as a function of the cosmic factor $1/(1+z)$.

FIG. 2.—Variances of the density fluctuations at the Jeans length as a function of the cosmic scale factor. The dotted line is for the linear fluctuation, and the solid line is for the nonlinear evolution given by the LN model, i.e., $\sigma^2 = \exp(\sigma_0^2) - 1$. 

64 32 16 8 4 2 1 0.5

Redshift $z$

0.1 0.01 0.001

$\chi$ (h $^{\text{ cm}}$/s$^2$)

0.1 0.01 0.001

0.01 0.05 0.1 0.5

$\sigma_0$

0.01 0.05 0.1 0.5

$\sigma_0$

0.001 0.01 0.1 1

Scale Factor 1/(1+z)
\[ \Omega = 0.7, \text{ and the Hubble constant } h = 0.7. \] The linear
power spectrum \( P(k) \) is given by the fitting formula given by
Eisenstein & Hu (1998). The linearly increasing of \( \sigma_0 \) with
the cosmic factor \( 1/(1 + z) \) is due to the linear increasing of the
density perturbations. The variance \( \sigma \) of the lognormal
PDF (eq. [3]) is also plotted in Figure 2. We see from Figure
2 that the first time that the variance \( \sigma_0 \) reaches the order of
1 is in the redshift period \( 15 > z > 7 \). This should be the
epoch of the formation of the first generation of collapsed
objects.

When the fluctuations \( \delta(x) \) are small, \( \delta_0(x) \ll 1 \), equation
(10) yields \( \delta(x) \approx \delta_0(x) \). This is the linear solution equation
(8). On the other hand, on scales \( x \ll x_s \), equation (10) gives
the nonlinear relation between the IGM density distribution
and the dark matter gravitational potential as

\[ \rho_b(x) \propto \exp \left[ -\frac{\mu m_p}{\gamma kT} \psi_{dm}(x) \right]. \] (11)

Equation (11) is the well-known isothermal hydrostatic
solution, which describes highly clumped structures such as
intracluster gas (Sarazin & Bahcall 1977). Therefore, the
density distribution \( \rho_b(x) \) in equation (10) is consistent with
highly nonlinear distribution of the baryonic mass in the
dark matter gravitational potential wells. Note that the primordial
potential \( \psi_{dm} \) remains linear and Gaussian much longer than the density (Brainerd, Scherrer, & Villumsen
1983; Bagla & Padmanabhan 1994). Equation (11) directly
shows that the IGM field \( \rho_b(x) \) is lognormal. Thus, if we
require that the mapping between the IGM field \( \rho_b(x) \) and
the linear dark matter field \( \psi_{dm}(x) \) satisfy (1) the linear relation
equation (8) in the linear regime and (2) the exponential
relation equation (11) in the highly nonlinear regime, then
the mapping of equation (10) probably is the most reasonable
one. Therefore, the lognormal mapping equation (10)
can uniformly describe the IGM distribution from linear to
weakly nonlinear to highly nonlinear regimes, corresponding
to density contrasts of \(<1, \approx 1, \text{ and } \gg 1 \), respectively,
without introducing extra parameters before the dynamics
of the clouds are dominated by hydrodynamic processes,
such as cooling and heating during star formation.

The LN model equations (9)–(10) are successful in fitting
the transmitted flux of the QSO Ly\( \alpha \) absorption spectrum
(Bi & Davidsen 1997; Feng & Fang 2000). The lognormal
PDF of the IGM field also gives better fitting to the recent
detected intermittency of the transmitted flux with high-
resolution, high signal-to-noise ratio samples of QSO Ly\( \alpha \)
absorption spectra (Jamkhedkar, Zhan, & Fang 2000;
Zhan, Jamkhedkar, & Fang 2001; Feng, Pando, & Fang
2001, 2003; Jamkhedkar et al. 2003).

3.3. Volume Filling Factor and Cumulative Mass Fraction
of IGM Clouds

We now study the formation and collapse of overdense
hydrogen clouds in the IGM field with lognormal mapping.
First, we demonstrate the lognormal PDF with the volume
filling factor \( V(\rho) \), which is the fraction of volume with
density larger than a given \( \rho \). From equation (3) or (10) we have

\[ V(\rho) = \int_{\rho}^{\infty} p(\rho) d\rho = \frac{1}{2} \text{erfc} \left( \frac{\sigma_0}{\sqrt{2}} + \frac{\ln \rho}{\sqrt{2} \sigma_0} \right). \] (12)

Figure 3 shows \( V(\rho) \) for \( \sigma_0 \) on the Jeans length scales at
redshifts \( z = 30, 20, 10, \text{ and } 7 \). For a uniform Gaussian random
field, we have roughly half the space volume with a
density lower than the mean \( \rho = 1 \) and half with larger than
\( \rho = 1 \). However, Figure 3 shows that \( V(>1) < \frac{1}{2} \) even when
\( z = 30 \). That is, the mass field of the IGM has already deviated
from a Gaussian field at very early time. Figure 3 shows
also that most of the IGM has fallen into clumps and most
of the space is occupied by very low density gas. This is why
the LN model is able to fit intermittent fields (Pando et al.
2002; Feng et al. 2003). In these fields, the mass is concentra-
ted in peaks or spikes, which are randomly and widely
scattered in space and surrounded by low mass density.

The long tail of the LN model can be more easily seen
with the cumulative mass fraction \( M(>\rho) \), which is the fraction
of mass in regions having mass density larger than a given \( \rho \), given by

\[ M(>\rho) = \int_{\rho}^{\infty} x \rho(x) dx = \frac{1}{2} \text{erfc} \left[ \frac{\ln(\rho/\bar{\rho})}{\sqrt{2}\sigma_0} - \frac{\sigma_0}{2\sqrt{2}} \right]. \] (13)

Figure 3 shows \( M(>\rho) \) for \( \sigma_0 \) on the Jeans length scales at
redshifts \( z = 20, 12, 10, \text{ and } 7 \). The curves of \( M(>\rho) \) at high \( \rho \)
show clearly the long tail. For instance, for the \( z = 7 \) curve,
\( M(>100) \) is less than \( M(>10) \) by only a factor of about 10.
That is, the mass fraction of large-mass events (\( \rho = 100 \)) in
in R can be 10% that of small-mass events (\( \rho = 10 \)). This is
because that the variance \( \sigma_0 \) at \( z = 7 \) is about 1.8. Objects of
\( \rho = 10 \) correspond to a \( \ln \rho/1.8 \approx 1.3 \sigma_0 \) event, and \( \rho = 100 \)
to a \( \ln \rho/1.8 \approx 2.6 \sigma_0 \) event. Therefore, the probabilities of
\( \rho = 10 \) and \( \rho = 100 \) are different by only a factor of about
10. On the other hand, if the PDF is Gaussian, \( \rho = 10 \) objects corresponds to a 1.3 \( \sigma_0 \) event, while the high-density \( \rho = 100 \) corresponds to a 7.7 \( \sigma_0 \) event. In this case, the number of \( \rho = 100 \) objects is completely negligible with respect to \( \rho = 10 \) objects.

Figure 3 shows that \( M(>\rho) \) is significant for \( \rho = 10 \) and \( z > 10 \). On the other hand, we always have \( V(>\rho) \sim 0 \) with \( \rho = 10 \), regardless of redshift. This indicates that a significant part of mass concentrates in a small volume. That is, we have dense objects at high redshifts. The long-tailed events may not be important at low redshifts, since at those times \( \sigma_0 \) on large scales is close to 1 and the formation of large-mass objects is no longer rare.

4. STATISTICS OF HYDROGEN CLOUDS

4.1. Simulations of the IGM Distribution

To study the clustering and collapse of hydrogen clouds, we produce simulation samples of spatial distribution of gas \( \rho(x) \) with the LN model developed in last section. In order to quickly grasp the features of these distributions, we simulate one-dimensional distributions. The details of the simulation procedure are given in Bi & Davidsen (1997). A brief description follows.

We first simulate the one-dimensional density and velocity distributions in Fourier space, \( \delta_0(k) \) and \( v(k) \), which are two Gaussian random fields. Both \( \delta_0(k) \) and \( v(k) \) are given by the power spectrum \( P_0(k) \) as follows (Bi 1993; Bi et al. 1995):

\[
\delta_0(k, z) = D(z)[u(k) + w(k)] , \tag{14}
\]

\[
v(k, z) = F(z) \frac{H_0}{c} \frac{\alpha(k)}{k} w(k) , \tag{15}
\]

where \( D(z) \) and \( F(z) \) are the linear growth factors for fields \( \delta_0(x) \) and \( v(x) \) at redshift \( z \). The fields \( w(k) \) and \( u(k) \) are Gaussian with power spectra given by

\[
P_u(k) = \alpha^{-1} \int_k^\infty P_0(q) 2\pi q^{-1} dq , \tag{16}
\]

\[
P_v(k) = \int_k^\infty P_0(q) 2\pi q dq - P_u(k) , \tag{17}
\]

where \( P_0(k) \) is the power spectrum of the three-dimensional field \( \delta_0(x) \). Functions \( \alpha(k) \) in equation (16) are defined by

\[
\alpha(k) = \frac{\int_k^\infty P_0(q) q^{-3} dq}{\int_k^\infty P_0(q) q^{-1} dq} . \tag{18}
\]

From equation (9), we have

\[
P_0(k) = \frac{P_{dm}(k)}{(1 + x_h^2 k^2)^2} , \tag{19}
\]

where \( P_{dm}(k) \) is the dark matter power spectrum in three dimensions. Thus, for a given \( P_{dm}(k) \) and \( x_h \), one can produce the distributions \( \delta_0(k) \) at grid points \( k_i, i = 1, 2, \ldots, N \), in Fourier space. The spatial distributions in the real line-of-sight space \( \delta_0(x) \) can be obtained by using a fast Fourier transform. Since velocity follows linear evolution longer than density, we can use the linear \( v(k) \) and its Fourier counterpart as a velocity field.

The Jeans length before reionization is about 1 \( h^{-1} \) kpc. This requires the resolution of the simulation to be less than 0.2 kpc. Our simulation range is 40 Mpc in comoving space. The total number of pixels is 262,144, so the pixel size is 0.152588 kpc. Three random samples of the density field with a comoving size of 1 Mpc are plotted for \( z = 15, 10, \) and \( 7 \) in Figure 4. The prominent spikes correspond to the dense hydrogen clouds that are collapsing-object candidates. Such hydrogen clouds occur not only at redshift \( z = 7 \) but also at \( z = 15 \). Although the number and height of the spikes decrease with higher redshifts, this \( z \)-dependence
is actually not very sharp. The typical size of the clouds shown in Figure 4 is of the Jeans length, but they have different heights. This indicates that the probability of spike events of the same size is not sharply dependent on the their height; i.e., the probability does not sharply depend on the mass density $\rho$ of the spikes but only on $\ln \rho$. This is an effect of the lognormal long tail.

### 4.2. Basic Properties of Hydrogen Clouds

To identify hydrogen clouds, we first smooth the simulated density field with a Gaussian filter of dispersion $x_h$. With the smoothed field, baryonic or hydrogen clouds are identified as the regions between two successive minima in the field. For each clouds, we have a mass density profile between the two minima. The maximum between the two successive minima is the central position of the cloud. The width $D$ is defined to be the FWHM of the top density. The column density $N_{\text{HI}}$ of an absorber is obtained by summing gas densities in each pixel from the first minimum to the next minimum in the smoothed field. For each cloud, we assign a peculiar velocity to be the velocity at the top density, and the internal velocity profile is the peculiar velocity of each pixel relative to the center velocity.

Since column density depends on both hydrogen number density and the size of clouds, it may not be a good indicator of the density contrast of hydrogen clouds. For instance, two clouds with the same column density may have a 10-fold difference in their density contrast due to a 10-fold difference in their sizes. If such cases are common, one cannot measure mass density with the column density. However, statistically we have good reason to use column density $N_{\text{HI}}$, to characterize the mass density of baryonic clouds. Figure 5 plots the relation between the column density $N_{\text{HI}}$, and density $\rho$ of clouds identified from one realization of the one-dimensional field. Figure 5 shows a tight correlation between the column density $N_{\text{HI}}$ and the mass density $\rho$ of clouds. For a given $N_{\text{HI}}$, the dispersion of $\rho$ is no greater than 20%. That is, the column density is mainly determined by the clouds’ mass density, not their size. A similar correlation has also been found among the precollapsed halos identified from N-body simulation samples (Xu, Fang, & Wu 2000). Thus, the IGM clustering in the dark age can be approximately described by the statistics of the hydrogen clouds with the number $N_{\text{HI}}$.

We first calculate $N_i(>N_{\text{HI}}, z) h \text{Mpc}^{-1}$, which is the one-dimensional comoving number density of clouds with column densities greater than a given $N_{\text{HI}}$. The redshifts are taken to be $z = 7, 10$, and 15. The differential number density is $dN_i(>N_{\text{HI}}, z)/d\ln(N_{\text{HI}})$, which is plotted in the top left panel of Figure 6. We see here that clouds with $N_{\text{HI}} \geq 10^{20}$ at $z = 7$ are much more numerous than at $z = 15$. This indicates that most clouds with $N_{\text{HI}} \geq 10^{20}$ formed at redshfits of less than 15.

The top right panel of Figure 6 presents the cumulative mass fraction of clouds with column density greater than a given $N_{\text{HI}}$. It also shows that the mass fraction of clouds with $N_{\text{HI}} > 10^{20.5}$ underwent a significant evolution from $z = 15$ to 7. The mass fraction of $N_{\text{HI}} > 10^{20}$ clouds at $z = 15$ is about 1%, but it is about 10% at $z = 7$. For clouds with column densities $10^{19.0} < N_{\text{HI}} < 10^{19.75}$, the number density and mass fraction at redshifts of 7–15 are comparable.

The bottom left and right panels of Figure 6 give respectively the mean top density and the mean comoving width $D$ of $N_{\text{HI}}$ clouds. These panels show that the clouds with $N_{\text{HI}} \geq 10^{20}$ at $z = 7$ have smaller size (width) and higher central density than at $z = 15$ and 10. This shows that hydrogen clouds are collapsing in the period from $z = 15$ to $z = 7$. The mass density of $N_{\text{HI}} \geq 10^{19.25}$ clouds has overdensity $\rho > 5$ at all redshifts. This means that the $N_{\text{HI}} \geq 10^{19.25}$ clouds were already going to collapse as early as $z = 15$.

To study the three-dimensional statistics of the clouds, we assume that most clouds are spherical in three-dimensional space with scale $D$. This is to approximate the nonlinear dynamical evolution of clouds as a spherical collapsing process. This approximation is probably poor for clouds with low $N_{\text{HI}}$ but is better for higher $N_{\text{HI}}$. Thus, we can estimate the baryonic mass of the cloud with $m_{\text{HI}}D^3$. The result is plotted in the top panel of Figure 7. We see that the $M$-$N_{\text{HI}}$ relation is insensitive to redshift. The bottom panel of Figure 7 shows the cumulative mass fraction of three-dimensional clouds with mass larger than a given $M$. It is interesting to see that the mass fraction of clouds with mass in the range $M < 10^{4.5} M_\odot$ is weakly dependent on redshift from $z = 7$ to 15. The mass fraction for $M \geq 10^{4.5} M_\odot$ is significantly dependent on redshift.

The top panel of Figure 8 gives the three-dimensional cumulative number density of clouds with column density larger than a given $N_{\text{HI}}$. The bottom panel of Figure 8 is similar to the top, but with respect to the mass $M$ of clouds. We see again that the cumulative number densities of clouds at $N_{\text{HI}} \approx 10^{19.25}$ or $M \approx 10^{4.5} M_\odot$ are not strong functions of redshifts from 7 to 15. The cumulative number densities of massive clouds $N_{\text{HI}} > 10^{20}$ or $M > 10^5 M_\odot$ decreases with higher redshift quickly. From these statistics, we can conclude that in the LN model, the baryonic clouds with mass $M > 10^5 M_\odot$ underwent strong evolution in the epoch $15 > z > 7$, but that for clouds with mass $M < 10^{4.5} M_\odot$, the evolution is moderate.

![Fig. 5.—Relation between $\rho$ and column density $N_{\text{HI}}$ of hydrogen clouds identified from one realization.](image-url)
5. CLOUDS THAT ARE GOING TO COLLAPSE

5.1. Density Profile and Collapse

We now identify which clouds are going to collapse. Obviously, collapsed IGM clouds are a necessary condition for hosting star formation. To study the details of the collapse, we calculate the density profile within each cloud. The mean density profiles of clouds along the line-of-sight are plotted in Figure 9, in which the horizontal axis is the comoving distance from the center of the clouds and redshifts are taken to be 15, 12, 10, and 7. For each redshift, the successive curves from lower to higher correspond to column density \( N_{\text{HI}} = 10^{16.25} \pm 0.5 m \), with \( m = 0, 1, \ldots, 8 \).

Figure 9 shows that all density profiles are approximately exponential, i.e., \( \rho \propto \exp(-x/2x_0) \), where \( x_0 \) is of the order of \( x_b \). The exponential profile comes from the exponential factor in equation (11). In Figure 9, we can see that the comoving density profile of clouds with \( N_{\text{HI}} = 10^{17.75} \) is almost independent of redshift, and therefore they are expanding with the Hubble flow. These clouds are comoving with the Hubble expansion, which is expected. From Figure 5, we see that \( N_{\text{HI}} = 10^{17.75} \) corresponds to about \( \rho = 1 \); i.e., these “clouds” have the same density as the background universe. This should follow the Hubble expansion.

For clouds with \( N_{\text{HI}} < 10^{17.75} \), or \( \rho < 1 \), the comoving size is greater with smaller redshift. In other words, these “clouds” actually are in voids. They are a part of the voids. Therefore, their expansion is faster than Hubble streaming.

For clouds with \( N_{\text{HI}} \geq 10^{18.25} \), the comoving size is smaller for smaller redshifts. That is, they are in the turnaround phase. The mean density of these clouds at \( z = 7 \) is \( \rho \sim 6 \) (Fig. 6). These clouds have decoupled from the Hubble expansion. Their expansion is slower than the Hubble expansion. However, we should not simply identify all clouds with \( N_{\text{HI}} \geq 10^{18.25} \) as collapsing, because they may still be physically expanding, but only with expanding velocity less than the Hubble velocity.
5.2. Velocity Profile and Collapse

Whether a cloud is collapsing should also be investigated through its velocity profile. We calculate the peculiar velocity profile for all clouds identified from 3000 realizations of a one-dimensional IGM field. The mean one-dimensional velocity profiles for clouds of various $N_{\text{HI}}$ are plotted in Figure 10, in which the horizontal axis is the comoving distance from the centers of the clouds, and the velocity is measured with respect to the centers of the clouds. The redshifts are taken to be $z = 7$ (solid lines), 10 (dotted lines), and 15 (dashed lines). Figure 10 also shows the Hubble flow $v_H$ with respect to the centers of the clouds. The dotted lines of Figure 10 actually show $-v_H$. The physical velocity profiles are then given by $v + v_H = v + (-v_H)$. Figure 11 is the same as Figure 10, but showing a zoom-in profile of the central parts of the clouds.

In Figures 10 and 11, we can classify baryonic clouds into three types: (1) The velocity profile $v$ is negative, $v < 0$, when separation is less than 0 and positive, $v > 0$, when separation is greater than 0. These clouds are expanding away from their centers; i.e., their physical expansion velocity is faster than the Hubble flow. (2) The velocity profile $v$ is positive, $v > 0$, when separation is less than 0 and negative,
when separation is greater than 0, but the physical velocity profile is negative, $v + v_{\text{H}} < 0$, when separation is less than 0 and positive, $v + v_{\text{H}} > 0$, when separation is greater than 0. These clouds are decoupled from the Hubble flow, but they still remain in the expanding phase and have not yet reached their maximum expansion. (3) The physical velocity profile is positive, $v + v_{\text{H}} > 0$, when separation is less than 0 and negative, $v + v_{\text{H}} \leq 0$, when separation is greater than 0. These clouds are in the phase of maximum expansion, or starting to collapse. In other words, their infall velocity is greater than the Hubble velocity in physical space.

Obviously, only clouds of type 3 can be identified as collapsing or collapsed clouds. That is, in Figures 10 and 11, only the $N_{\text{H}_1}$ clouds having velocity profiles $v$ (solid lines) equal to or greater than the negative Hubble flow $-v_{\text{H}}$ (dotted lines) are possible spots of star formation. Therefore, not all clouds with a size greater than the Jeans length are in the collapsing phase. Although clouds with small $N_{\text{H}_1}$ have a size greater than the Jeans length, the velocity outflow of the clouds significantly postpones their collapse. This leads to a suppression of the formation of low-mass collapsed clouds at high redshifts. This result is consistent with the simulation result (Gnedin 2000), which shows that the formation of low-mass objects in the redshift range $12 > z > 7$ is significantly suppressed with respect to the Jeans length prediction.

5.3. Constraints on First-Star Formation

Besides the Jeans length of the IGM in equation (9), we have not introduced any other parameters related to the hydrodynamic processes of star formation in this paper. Even so, we can already set some constraints on the formation of first-generation stars by considering that collapsed clouds are the necessary environment for hosting star formation.

As mentioned in the last subsection, only clouds of type 3 are collapsed or collapsing and can play the role of host of star formation. Thus, first-generation IGM clustering can be seen from the properties of the type 3 clouds and their
redshift dependence. Table 1 lists the basic properties of these clouds, including the minimum column densities \((\log N_{\text{HI}})_{\text{min}}\), the minimum mass \(M_{\text{min}} (M_\odot)\), and the minimum density \(\rho_{\text{min}}\). All the minimum densities \(\rho_{\text{min}}\) of the collapsing clouds are greater than 10 or even \(\gg 10\). Note that \(\sigma_0\) is of the order of 1; events of \(\rho_{\text{min}} > 10\) are rare, or very rare. Therefore, the first generation of star formation relies on the long tail of the PDF.

| Value | \(z = 15\) | \(z = 12\) | \(z = 10\) | \(z = 7\) |
|-------|-------------|-------------|-------------|-------------|
| \((\log N_{\text{HI}})_{\text{min}}\) | 22.25 | 20.25 | 19.5 | 18.75 |
| \(M_{\text{min}} (M_\odot)\) | \(10^{6.5}\) | \(10^{5.3}\) | \(10^{4.4}\) | \(10^{3.5}\) |
| \(\rho_{\text{min}}\) | \(3.1 \times 10^3\) | \(1.9 \times 10^2\) | 35 | 19 |

Figure 12 shows the redshift dependence of the cumulative mass fraction \(M(>M_J)\) of clouds larger than the Jeans mass and \(M(>\rho_{\text{min}})\) of clouds in collapsing and collapsed baryonic objects. One can clearly see from Figure 12 that the collapse of hydrogen clouds is significantly suppressed when \(z > 12\). The evolution of \(M(>M_J)\) is moderate in the redshift range \(7 < z < 15\), while the redshift dependence of \(M(>\rho_{\text{min}})\) is dramatic when \(z > 12\). This is because the suppression of collapse is stronger at higher redshift and weaker at lower redshift. At early-universe \(z = 15\), about 1% of the mass is already in clouds with sizes larger than the Jeans length. However, there is only a tiny fraction of mass \((\sim 10^{-8})\) in clouds that are collapsed. This is because at redshift 15 only clouds with masses greater than \(10^{6.5} M_\odot\) can collapse and host star formation. The mass fraction in the collapsed clouds is higher for smaller redshift. Therefore, the formation history of the first-generation baryonic objects lasts the entire epoch from \(z = 15\) to 10.

It has been estimated that if all the ionizing photons and the \(\sim 10^{-2}\) metallicity observed at low redshift are produced...
by the first-generation stars, the mass fraction of the IGM falling into the collapsed clouds should not be much less than 1% (Haiman & Loeb 1998; Ostriker & Gnedin 1996) before the reionization. Using the 1% mass to quantify the first generation of baryonic objects, we can conclude from Table 1 and Figure 12 that the generation of sources responsible for reionization is likely to occur not much earlier than \( z = 10 \), and therefore the reionization itself should take place not much earlier than that era.

6. DISCUSSION AND CONCLUSIONS

We have outlined an evolutionary picture of IGM clustering with the LN model, which is capable of sketching cosmic clustering of both dark matter and the IGM in weakly nonlinear, as well as nonlinear, regimes. With fine numerical simulations, we can make a number of important statistics of the IGM clouds before reionization. We focus, in particular, on the abundances of hydrogen clouds that are going to collapse, which would be candidates for harbors of first-generation stars. While the details of first-star formation depend on many chemical and hydrodynamic factors, the formation of collapse candidates is a necessary condition for the formation of stars or galaxies. Therefore, we can set an effective constraint on the history of the formation of the first objects.

A remarkable feature of the LN model is that the PDF of the mass field is long-tailed. The long-tailed events, such as the formation of massive clouds at high redshifts, have a much greater probability than a Gaussian field. The rare events of the first-generation collapsed clouds are sensitive to the long tail. For a Gaussian PDF, there are very few events of mass perturbation much greater than \( \sigma_0 \), and therefore there are no massive objects formed at early time for a Gaussian PDF. However, the LN model predicts a much higher probability of forming massive clouds at high redshift.

The LN model provides the density as well as the velocity distributions of the IGM. Comparing the density profile with the velocity profile of clouds, it is revealed that clouds with the Jeans mass are not collapsing immediately at high redshift. Typically, the IGM cloud around a group of dark...
matter potential wells stays in the phase of slowly expanding in proper space for a while. The smaller the mass of hydrogen cloud, the longer the time it remains in this phase. Only clouds with high enough mass can enter the phase of collapsing in proper space. Thus, as shown in Table 1, the minimal mass of hydrogen clouds in collapse is larger for higher redshifts. That is, the clustering in the LN model is not of the bottom-up hierarchy at scales less than $10^5 - 10^6 M_{\odot}$. Rather, massive collapsed clouds formed first, and less massive clouds later. It should be pointed out that this feature cannot simply be explained by the redshift dependence of the Jeans mass. Although the Jeans mass is smaller at smaller redshift, it is not enough to explain the significant evolution of $M_{\text{min}}$ from $10^5$ to $10^4 M_{\odot}$ in the redshift range $15 < z < 10$. This feature is again due to the suppression of the formation of collapsed clouds with small mass on high redshifts. The lognormal mapping makes the evolution of $M_{\text{min}}$ much stronger than that of the Jeans mass $M_J$.

It should be pointed out that there are factors leading us to overestimate the number of collapsed hydrogen clouds. First, we count all type 3 clouds (§5.2) as being hosts of star formation. Obviously, collapsing clouds of type 3 are not yet hosts. Second, the total mass of baryons in the collapsed objects is estimated by the mass of collapsed clouds. This may also be a source of overestimation. Last but not least, the LN model seems to overestimate the abundance of saturated absorption lines of Ly$\alpha$ (Matarrese & Mohayaee 2002); it may indicate an excess of dense clouds. Nevertheless, all these factors actually strengthen the results on the upper limit of the mass fraction in first-generation objects at a given redshift and consolidate the conclusion that the generation of sources responsible for reionization and the reionization itself is likely to occur not much earlier than $z = 10$.

We may improve the overestimation problem by considering the hydrodynamics of the IGM. However, we would rather trade the overestimation for the uncertainty of IGM hydrodynamics. That is, in the current approach, the conclusions are not dependent on parameters related to the hydrodynamic processes of star formation, such as cooling timescale, rate of star formation, and efficiency of producing reionization photons by stars. Our conclusions are, however, sensitively dependent on the power spectrum of initial density perturbations of a few $h^{-1}$ kpc. Therefore, it would be useful to obtain information on the initial power spectrum at such small scales.

We thank Ping He and Hu Zhan for help during the preparation of this paper. We also thank the referee and editor for their helpful comments.

REFERENCES

Abel, T., Anninos, P., Norman, M. L., & Zhang, Y. 1998, ApJ, 508, 518
Bagla, J. S., & Padmanabhan, T. 1994, MNRAS, 266, 227
Barkana, R., & Loeb, A. 2001, preprint (astro-ph/0010468)
Berger, A., & Fang, L. Z. 1994, Phys. Rev. Lett., 72, 458
Bi, H. G. 1993, ApJ, 405, 479
Bi, H. G., & Davidsen, A. F. 1997, ApJ, 479, 523
Bi, H. G., Ge, J., & Fang, L. Z. 1995, ApJ, 452, 90
Brainerd, T. G., Scherrer, R. J., & Villumsen, J. V. 1993, ApJ, 418, 570
Bromm, V., Coppi, P. S., & Larson, R. B. 1999, ApJ, 527, L5
Buchert, T., Domínguez, A., & Perez-Mercader, J. 1999, A&A, 349, 343
Cen, R., & Haiman, Z. 2000, ApJ, 542, L75
Chiu, W. A., & Ostriker, J. 2000, ApJ, 534, 507
Coles, P., & Jones, B. 1991, MNRAS, 248, 1
Couchman, H. M. P., & Rees, M. J. 1986, MNRAS, 221, 53
Eisenstein, D., & Hu, W. 1998, ApJ, 496, 605
Fang, L. Z., Bi, H. G., Xiang, S. P., & Börner, G. 1993, ApJ, 413, 477
Feng, L. L., & Fang, L. Z. 2000, ApJ, 535, 519
Feng, L. L., Pando, J., & Fang, L. Z. 2001, ApJ, 555, 74
———. 2003, ApJ, 587, 487
Gnedin, N. Y. 2000, ApJ, 542, 535
Gurbatov, S. N., Saitchev, A. I., & Shandarin, S. F. 1989, MNRAS, 236, 385
Haiman, Z., & Loeb, A. 1998, ApJ, 503, 505
Hui, L., Kofman, L., & Shandarin, S. F. 2000, ApJ, 537, 12
Jamkhedkar, P., Feng, L. L., Zheng, W., Kirkman, D., Tytler, D., & Fang, L. Z. 2003, MNRAS, in press
Jamkhedkar, P., Zhan, H., & Fang, L. Z. 2000, ApJ, 543, L1
Jones, B. T. 1999, MNRAS, 307, 376
Kofman, L., Bertschinger, E., Gelb, J. M., Nusser, A., & Dekel, A. 1994, ApJ, 420, 44
Madau, P. 2002, preprint (astro-ph/0212555)
Madau, P., Meikin, A., & Rees, M. J. 1997, ApJ, 475, 429
Matarrese, S., & Mohayaee, R. 2002, MNRAS, 329, 37
Miralda-Escudé, J. 1998, ApJ, 501, 15
Miralda-Escudé, J., Haeblult, M., & Rees, M. J. 2000, ApJ, 530, 1
Nusser, A. 2000, MNRAS, 317, 902
Ostriker, J. P., & Gnedin, N. Y. 1996, ApJ, 472, L63
Pando, J., Feng, L. L., Jamkhedkar, P., Zheng, W., Kirkman, D., Tytler, D., & Fang, L. Z. 2002, ApJ, 574, 575
Press, W. H., & Schechter, P. 1974, ApJ, 187, 425
Sarazin, C. L., & Bahcall, J. N. 1977, ApJS, 34, 451
Tegmark, M., Silk, J., Rees, M. J., Blanchard, A., Abel, T., & Palla, F. 1997, ApJ, 474, 1
Theuns, T., Schaye, J., Zaroubi, S., Kim, T.-S., Tzanavaris, P., & Carswell, B. 2002, ApJ, 567, L103
Tozzi, P., Madau, P., Meikin, A., & Rees, M. J. 2000, ApJ, 528, 597
Valageas, P., & Silk, J. 1999, A&A, 347, 1
Vervegassola, M., Dubrulle, B., Frisch, U., & Noullez, A. 1994, A&A, 289, 325
Xu, W., Fang, L. Z., & Wu, X. P. 2000, ApJ, 532, 728
Yang, X. H., Feng, L. L., Chui, Y. Q., & Fang, L. Z. 2001, ApJ, 560, 549
Zhan, H., Jamkhedkar, P., & Fang, L. Z. 2001, ApJ, 555, 58